

CHAPTER 4

Collaborating with the Institutional Review Board (IRB)

Kathleen Murphy (Northwestern University, ret.)

4.1 Introduction

This chapter is focused on the institutional review board (IRB),¹ an administrative body created at a university or other organization to review research to ensure ethical protection of participants involved. This chapter describes what the IRB does and does not do and what researchers, data providers, and related stakeholders can expect from IRB review of research that involves humans. While all research uses information in various formats that is “data,” for the purpose of this chapter, the focus will be on research that accesses and uses administrative data in different forms, formats, and contexts. This may include research activity where administrative data are the central feature or

Copyright © Kathleen Murphy.

Cite as: Murphy, Kathleen. “Collaborating with the Institutional Review Board (IRB).” In: Cole, Shawn, Iqbal Dhaliwal, Anja Sautmann, and Lars Vilhuber (eds.), *Handbook on Using Administrative Data for Research and Evidence-based Policy*. Cambridge, MA: Abdul Latif Jameel Poverty Action Lab. 2020.

¹There are various names for similar boards such as Research Review Board (RRB) (Chicago Public Schools, 2020), Research Ethics Committee (REC) (NHS Health Research Authority, 2020) or some similar naming convention for boards established to conduct ethical and regulatory review of human research.

where the data are part of a larger project. There may be different contexts such as international research or collaborative research (or both) where there are different regulatory requirements, as well as different review processes. Some data-driven projects will include only existing administrative data, while others include retrospective or prospective data alone or in conjunction with other research methods, such as experimental interventions, surveys, interviews, or observation. Some projects only involve the analysis of data, while others can include multiple iterations of experimental and comparison interventions as well as innovative analysis of multiple data sets, which are linked by a subset of identifiers. In the United States, the IRB review of such projects takes all of these design factors into consideration in the context of a well-established ethical and regulatory process as described in section 4.5.

The goal of this chapter is to provide researchers, data providers, data stewards, and other stakeholders with the tools they need to understand the IRB process. The chapter provides a practical understanding of what an IRB considers and how an IRB processes human research including data driven proposals. This includes how an IRB considers data acquisition, data management, data storage, and data retention in the conduct of research. The chapter references the ethical principles as well as the application of the federal, state, local, and institutional guidelines for research in as much as the IRB has oversight of these principles and guidelines in the United States. The text includes discussion of related international considerations, which may inform ethical and regulatory deliberation. Finally, the chapter provides practical strategies for collaborating with the IRB, which has oversight of the research.

There are a number of resources in the literature that identify the advantages of big data and administrative data for conducting research. That is not reiterated here except to endorse that the ease of use, reduced burden on participants and researchers and the long-term availability of administrative data makes this approach a logical way to contribute to the knowledge base. For additional information see Feeney et al. (2015); Connelly et al. (2016); Collmann and Matei (2016). For

more detailed descriptions of the use of administrative data for public policy and the public good see for example Card et al. (2011); Collmann and Matei (2016); Figlio, Karbownik and Salvanes (2016).

4.2 What is the IRB?

The ethical guidance and regulatory requirements for IRB review of all human research includes the ethical principles of the Belmont Report (United States 1978) and the US Department of Health and Human Services (HHS) Office for Human Research Protections (OHRP) regulations found in the part of the Code of Federal Regulations (CFR) referred to as “Title 45: Public Welfare, Part 46—Protection of human subjects, Subpart A—Basic HHS Policy for Protection of Human Research Subjects” or 45 CFR 46 (Code of Federal Regulations, 2017a; Office for Human Research Protections, 2016b). Throughout this chapter, regulatory citations are in reference to this section of the CFR.

The IRB is an administrative body that reviews human research (defined by 45 CFR 46.102 (e)(1)) to ensure the ethical protection of participants from the reasonably foreseeable risks of harm caused by research. The harms the IRB considers include physical, psychological, social, legal, and economic risks as well as community or group harms. For example, an inadvertent disclosure of sensitive or identifiable information is a common risk in social and behavioral research because the disclosure can result in social, psychological, or legal harm. All IRBs include the risks that need to be considered in the conduct of research in the protocol and consent templates, as well as in reviewer guides and on their websites. See for example, University of California, Irvine² and the Northwestern University³ protocol templates.

An IRB or ethics review process may be part of an academic institution; a medical facility; a federal, state, or local agency; or any other organization or commercial entity that chooses to conduct human research. Entities that receive federal funds for any reason and conduct human

²<https://research.uci.edu/compliance/human-research-protections/irb-members/assessing-risks-and-benefits.html> (accessed 2020-12-15).

³<https://www.irb.northwestern.edu/templates-forms-sops> (2020-12-15).

research are required by federal mandate in the United States to have an IRB.

IRB membership and the organization and function of an IRB is defined in the regulations 45 CFR 46: 107, 108. An IRB will consist of a minimum of five members of diverse backgrounds and expertise, including scientists and non-scientists, in order to provide complete and adequate review of human research. In IRBs with a large volume of projects, minimal risk research activity is generally reviewed by full-time employed IRB office staff who are also board members and qualified to review. Greater than minimal risk studies must always be reviewed at a convened meeting referred to as Full Board review.

Table 4.1: Categories of review conducted by an IRB

Review Type	Regulatory Authority	Risk	Description
Exempt	Ethical principles of Belmont (respect for persons, beneficence, and justice)	Minimal risk (often anonymous or deidentified data)	Briefer application and typically reviewed in the IRB office
Expedited	Belmont and 45 CFR 46.111	Minimal risk (identifiable, personal or sensitive information)	Reviewed in the office by one or more IRB members. If expedited reviewer does not approve, the study may go to the full board
Full Board	Belmont and 45 CFR 46.111	Greater than minimal risk (could include minimal risk research that does not fit in exempt or expedited review categories)	All studies involving prisoners and certain research with vulnerable populations regardless of risk such as children, fetuses, and neonates. Projects can only be disapproved at a convened meeting

In addition to the internal organization or agency-based IRB, organizations and independent researchers that do not have their own IRB

can contract with independent IRBs which can be both commercial or non-profit. Independent IRBs also can serve in the role as a central IRB where multiple (academic or clinical) institutions are conducting the same research and either want to contract with an independent IRB or are required by regulation to rely on one IRB for oversight of the whole project. The reliance agreement process, where one IRB agrees to rely on another IRB for oversight, can be with a commercial IRB or with an IRB that is, for example, located in an academic institution where that IRB has agreed to serve as the IRB of record for a multisite project. For the regulatory guidance on the reliance process see 45 CFR 46.114.

Independent IRBs also may be an option for a data provider who would like to submit research projects for ethical oversight when there is no federal requirement to do so. This chapter is not focused on independent or central IRBs but for more information about central IRBs and institutional IRBs see Wandile (2018).

At the center of the ethics review process is the Belmont Report (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979), which summarizes the ethical principles and guidelines IRBs use when reviewing research involving human subjects. Three core principles are identified:

1. *Respect for persons* allows individuals to be self-directed and make informed, voluntary decisions about whether they wish to participate in research and is the fundamental ethical rationale for the consent process and the elements of the consent document.
2. *Beneficence* assesses the risks and benefits of participating in research, recognizing the obligation of the researcher to minimize risks while maximizing the benefits of participation.
3. *Justice* directs investigators to recruit and enroll those who would benefit from the outcome of the research and to not impose undue risks on those who would not otherwise be helped by the research.

The principles of the Belmont Report are codified in federal regulations 45 CFR 46 to protect the rights and welfare of humans recruited to participate in federally funded research activities. Although the federal regulations specifically apply to non-exempt research projects in

organizations that receive federal funds, academic institutions have routinely applied these same regulatory guidelines to federally and non-federally funded or even unfunded projects, simply because the regulatory standards are ethically reasonable.

It is in the context of these ethical principles and regulatory requirements that IRBs are charged with the responsibility of reviewing research involving human participants. The definition of human research is discussed in section 4.3 in more detail, but it is in this context that the IRB has the authority to approve, monitor, modify, and disapprove all research activities that fall within its jurisdiction. These regulations apply to research conducted in the United States or by US-based researchers conducting research in another country.

4.3 IRBs and International Research

Human research can take place anywhere in the world and there are over 1,000 laws, regulations, and guidelines on human research protections in 133 countries (Office for Human Research Protections, 2020). OHRP annually compiles the most relevant regulations and agencies⁴ that regulate research in each country. Some, though not all countries, have regulations and guidance regarding social and behavioral research activities. Countries that do have such guidance tend to have more restrictive data protection rules and regulations than those in the United States. For example, in the European Union, the General Data Protection Regulation (GDPR)⁵ (European Parliament and Council of the European Union, 2016) covers the protection of all personal data of which research data are but a subset. GDPR special category data include race and ethnic origin; religious or philosophical beliefs; political opinions; trade union memberships; biometric data used to identify an individual; genetic data; health

⁴<https://www.hhs.gov/ohrp/sites/default/files/2020-international-compilation-of-human-research-standards.pdf> (accessed 2020-12-15).

⁵GDPR is legislation in the European Economic Area that protects persons with regard to the processing of personal data and on the free movement or sharing of those data. GDPR is comprehensive, encompassing all personal data not just research data.

data; and data related to sexual preferences, sex life, and/or sexual orientation. Similarly, the consent documents in the countries of the European Economic Area (EEA) have more prescriptive and restrictive requirements than in the US (Office for Human Research Protections, 2018). Whatever the country, researchers need to be cognizant of the local country regulations that may apply. For example, *respect for persons* as articulated in the Belmont Report applies in other countries, it just may be defined differently.

In addition, when research is taking place in a country where the regulations are different, researchers in the United States will be held to the standard of what is referred to as equivalent protections (45 CFR 46.101(h)); additional guidance can be found in Office for Human Research Protections (2016a). This means the researcher based in the US (who is subject to review by an IRB) and conducting research internationally is responsible for utilizing strategies to mitigate risk and protect participants at the level that would be required if the research was conducted in the United States. One example is the age of majority and consent to participate in research. In most US states the age of majority and consent is 18, while in some countries, such as Germany, Italy, Paraguay, and Ecuador, the age of consent is 14. A US researcher conducting research in Paraguay will be expected to use 18 as the age of consent to participate by the IRB. Another example, the Family Educational Rights and Privacy Act (FERPA) is a US law (20 U.S. Code § 1232g; 34 CFR part 99) and not applicable in other countries; however, if using education data from another country where education data does not have privacy and confidentiality protections, the IRB will expect that the research will apply equivalent protections as would exist under FERPA. In this example, data providers, data stewards, and researchers would need to address the use and collection of data in relation to minors when requesting IRB review.

4.4 What an IRB Does Not Do

Just as important as what the IRB does do, is what it does not do. As stated earlier, the mission of an IRB is the protection of participants in

research from risks associated with the research. To do this, an IRB must contribute to the development of training, policies, and practices that facilitate this purpose. However, there are a number of related oversight and regulatory activities required for some research activities that are **not** the purview of the IRB, though they contribute to the IRB process.

The IRB does not manage the grants or mechanisms for funding the research and is not involved in developing conflict of interest management plans. Additionally, while the IRB in some institutions may serve as the privacy board, as is the case for biomedical research, this is not a regular IRB function. The IRB typically does **not** have the responsibility to create or finalize data sharing agreements such as data use agreements (DUAs) and data transfer agreements or other contracts such as non-disclosure agreements (NDAs). Finally, data safety plans for sensitive restricted data are most often developed outside of the IRB. However, non-disclosure agreements and data safety plans have implications for the IRB review of the data management plan in the protocol (the specific and detailed design for how a research study will be conducted, which is submitted to the IRB for review).

The IRB will conduct an administrative review of these agreements and plans and, when applicable, hold the researcher accountable. For example, if there is a reported conflict of interest as part of the COI management plan where the principal investigator (PI) is prohibited from conducting data analysis because of a vested interest in the outcome, the IRB will make sure that is written into the protocol and reflected in any consents that are in use. Similarly, when applicable, the IRB will require that the DUA be uploaded into the IRB record and that the data protections outlined in the data sharing agreement are written into the IRB protocol. However, the IRB is not a signatory or even an intermediary in these agreements. The designated official on the institution side is the responsible party for signing the DUA or NDA, and for processing the funding, evaluating conflict of interest, or establishing the appropriate data security mechanism. While data providers can rely on the IRB monitoring and enforcing any of these activities as they relate to data protection and protection of participants, the IRB is

not the responsible party for initiating them.

In addition, researchers need to know their own institutional policies and practices as to where each of these related activities fit with IRB review. For example, in some institutions, the IRB review may not proceed until the DUA is in place. In other institutions, the finalizing of the DUA is contingent on the IRB approval. While both the IRB and the data sharing agreement processes can typically be started at the same time, the researcher and data provider need to know what the sequence is for final approval. A key point is in all research requiring approval, the data security evaluation and compliance with FERPA or the Health Insurance Portability and Accountability Act (HIPAA) regulations must be in place before IRB approval can be processed.

4.5 What the IRB Will Do to Ensure the Protection of Participants

The first order of ethical challenge in all research is the risk of harm. When it comes to the use of administrative data in research, the risk of harm stems from the potential for violations of privacy, confidentiality, or informed consent (even if the research project as a whole may expose participants to additional risks). All of the stakeholders in data-driven, human research that are subject to IRB need to start with the federal regulations that govern the IRB review of research. The criteria for IRB review are articulated in 45 CFR 46.111 (Code of Federal Regulations, 2017b). This part of the regulation outlines seven specific elements that **must** be in every non-exempt research project protocol, which all IRBs use to determine whether research can be approved. The following have been abbreviated from the regulations for the purpose of this handbook; **all of the following** must be met:

- (1) “Risks to subjects are minimized by using procedures that are consistent with sound research design and that do not unnecessarily expose subjects to risk.” (45 CFR 46.111(a)(1)(i))

To evaluate sound research design in a data driven project, the IRB will consider whether the variables of the data set, the sample size,

and the proposed analysis are consistent with the intended purpose of the study. There must be scientific merit to the study and there must be consistency between the purpose and the data being used.

- (2) “Risks to subjects are reasonable in relation to anticipated benefits, if any, to subjects, and the importance of the knowledge that may reasonably be expected to result.” (45 CFR 46.111(a)(2))

A primary risk to the subjects directly related to the use of administrative data or linkage of such data with survey data is the re-identification of participants, either by an external party or by one of the stakeholders in the project. This is in addition to any other risks associated with the project unrelated to the use of administrative data, such as the risks to participants due to the intervention itself. The IRB will work with researchers to anticipate risks to individual participants and to ensure there are adequate mechanisms in place to protect participants from harm, such as loss of income, retaliation, or punishment. Risk mitigation with administrative data is often focused on levels of access and security with regard to the collection, transfer, storage, and access management of data. In addition to protecting subjects from the risks of disclosure to outside parties, projects may also need to mitigate the risks of reidentification by the data provider; the researcher and data provider may consider an arms-length agreement, which prevents the data provider from accessing the identified data and provides another measure of protecting subjects. There are multiple ways to protect individuals and their related information through technology and by de-identifying that data. The researcher will work with the IRB, in addition to their institution’s general counsel and IT where appropriate, to manage the risks and security procedures for working with administrative data.

For example, in a study where a researcher collaborates with a bank to evaluate a microfinance program, it is possible for researchers to uncover fraud or deception by individual participants in the course of the project. Logically the bank will want to know that information, but that places the participant at risk of harm by having participated in the research. In this example, it would not be unusual for an IRB

to require a research team to state in the protocol that the DUA must prevent access to, or sharing of, identifiable information with the bank or must otherwise restrict the bank's use of linked administrative data to protect participants from retaliation or punishment.

(3) "Selection of subjects is equitable." (45 CFR 46.111(a)(3))

This means that for all research, the data being used or collected are a logical reflection of the purpose of the study and representative of the population most likely to benefit from the study. For data-driven projects that analyze a set of existing data, this would not generally be an issue. The primary concern in this case is that the data used must be logically connected to the purpose of the research project. However, some projects may use an existing administrative data set to select a study sample as in the case of randomized controlled trials that use administrative data as a census to select participants. This selection process should be free of biases; any biases could lead to the benefits and burdens of the research being unequally distributed. This can be an issue if there are biases within the administrative data. The IRB will consider the usage of administrative data for sample selection as it relates to the Belmont Report principle of justice: the people selected to be recruited to participate in the research are those most likely to be affected by the problem being studied and to benefit from the research.

(4) "Informed consent will be sought from each prospective subject or the subject's legally authorized representative, in accordance with, and to the extent required by, 45 CFR 46.116" (45 CFR 46.111(a)(4))

The typical standard for research with human subjects is that there is signed written consent. With projects where the data were originally collected for purposes other than research, consent for the data to be used for future research is rarely part of the original agreement between those subjects and the data collector. If consent is present, oftentimes the agreement that the data can be used for research is buried in the details at the end of the Terms of Service as to belie the concept of "informed" consent. Similarly, governments rarely use "consent" in the IRB sense of the term when collecting administrative data, as they

do not obtain data for research purposes. Instead, in the US, the government may use terms like Privacy Impact Assessments (PIAs), System of Records Notices (SORNs), and Computer Matching Agreements (CMAs) to alert the public to additional uses of data. These protocols do establish a legal floor for the use of the data, but they do not reflect the ethical intent of informed consent as articulated in the federal regulations. For projects that only use retrospective administrative data, an IRB will typically look for an explanation in the research protocol for why it is not possible or reasonable to obtain written consent. In research projects that combine administrative data with survey data or other direct subject contact, the informed consent procedure for the new data collection can also include consent to the use of the administrative data. To that end, the researcher needs to decide whether individuals who meet the criteria for the ongoing research activities are free to decline the use of the administrative data and still participate in the rest of the study. If use of the data is a mandatory requirement for participation, that needs to be stated in the consent. If it is optional, then it needs to be added to the consent form as an “optional element” to make it clear that it is not a requirement of participation.

- (5) “Informed consent will be appropriately documented or appropriately waived in accordance with 45 CFR 46.117.” (45 CFR 46.111(a)(5))

This is referred to by IRBs as documentation of consent and the rationale is consistent with element 4 that the standard practice is signed written consent. However, there are many circumstances in which a waiver of documentation of consent is appropriate either because it is not practical, such as with a phone interview or an online survey, or for safety reasons in which written consent would endanger the person to have their name attached to a study. This is most likely to occur with participants who are vulnerable. For example, interviews with sex workers in countries where it is illegal or with individuals in domestic violence shelters could be at heightened risk if their names were on a document.

- (6) “When appropriate, the research plan makes adequate provision for monitoring the data collected to ensure the safety of subjects.”
(45 CFR 46.111(a)(6))

Monitoring data collection is not an issue for projects using existing data in isolation or data that will be collected anonymously, especially if the data are used retrospectively. However, this may apply to a study that uses administrative data to observe participants over time during their participation in a project. For example, consider a randomized controlled trial that uses administrative data to study the implementation of a new social policy. As part of the assessment, the study uses unemployment records, medical records, or other sources to assess measures related to socio-economic status, employability, and markers of depression. In such a scenario, the IRB will typically require real time monitoring of those data so that researchers can intervene in outstanding circumstances. Some examples where intervention is warranted include the instance of a participant reporting suicidal ideation, lack of ready access to food, clean water, or health care, or any increased risk of harm caused by a change in the policy being studied. In situations where it is unclear that the benefits to society outweigh the harm to participants, the research may need to be stopped to protect the participants. The only way to recognize the harm is to monitor the data as they are generated. The IRB expects researchers to recognize the probability and the magnitude of the harm and to address it in the protocol. While monitoring data may not be an issue, the protocol needs to address why that is the case.

- (7) “When appropriate, there are adequate provisions to protect the privacy of subjects and to maintain the confidentiality of data.” (45 CFR 46.111(a)(7))

Confidentiality is a key factor for IRB deliberation of all research including projects using administrative data. Unintended disclosure of sensitive, private information is one of the primary risks of participation in research, and appropriate measures to manage the risk must be in place to protect participants and their related data. The more sensitive the data being used or collected, the more robust the data protec-

tion plan must be. Several of the chapters in this handbook discuss in detail the different strategies available to protect subject privacy and confidential data; those details will not be reiterated, but this chapter emphasizes that appropriate strategies must be elements provided in the protocol for IRB review.

The above seven elements are required for IRB approval of a research project. There is far more detail about the specifics of what is required with informed consent including when it can be altered or waived (Code of Federal Regulations, 2017d) and how it must be documented in the actual regulations. It is important to note that while all IRBs are using the same federal regulations, there may be different interpretations of the application of the regulations, especially around the requirement of consent and when it can be altered or be waived. Data providers can rely on the IRB review process to address each of the seven elements required for IRB approval and to approve only those projects that have adequate protections in place. Researchers, on the other hand, need to understand the basic regulatory requirements and to work with their own IRB to understand how the principles and regulations are being applied to their specific study. Similarly, researchers can go a long way in helping themselves navigate the IRB process by addressing each of the specific regulatory requirements in their protocol and related documents submitted to the IRB. The rest of this chapter is focused on the practical concerns for IRBs regarding specific research projects, the IRB related questions that must be asked and answered, and the manner in which IRBs think about the answers.

4.6 Considerations of the IRB

Being able to understand how and what the IRB considers when reading over a new project will inform the researcher what to include when submitting a new project proposal to the IRB. If the project proposal is framed how an IRB considers projects, the review process will likely be more collaborative and quicker, with far fewer changes requested.

4.6.1 Is the Study Human Research or Not Human Research (nHR)?

The first consideration is whether IRB review is needed and involves two questions to come to a conclusion. To decide whether a project is human research the following questions are considered in sequence by an IRB. If the answer to **any** of these questions is no, the study is not human research (nHR) and it does not require IRB review. For additional guidance, the OHRP provides decision charts⁶ (Office for Human Research Protections, 2020) to help map the process of how to think about the question, “Is an Activity Human Subjects Research Covered by 45 CFR Part 46?”

1. **Is it research?** In this context research is defined as a systematic investigation designed to contribute to generalizable knowledge (45 CFR 46.102(l)). There are two concepts to consider: systematic collection of information and generalizable knowledge. If a project does not meet both requirements then it does not require IRB review as it is not a research activity and is therefore not human research. It should be noted that generalizability can be a nuanced concept that is more multifaceted than just statistical generalizability, although data driven projects tend to be most closely linked to statistical generalizability (Lee and Baskerville, 2003). Nonetheless, when there is a systematic investigation (secondary analysis) of existing data and the investigation is intended to contribute to generalizable knowledge, the activity is research.
2. **Does the research involve human subjects?** It is possible to have a systematic collection of data that are routinely collected about people such as birth, death, taxes, participation in programs, insurance cost, medical care, etc. This collection of data is not for research purposes so while it is systematic, it is not research at the outset, because it is not intended to contribute to generalizable knowledge. Managing the data does not change that assessment. In the course of working with one (or many) administrative data

⁶<https://www.hhs.gov/ohrp/regulations-and-policy/decision-charts/index.html> (accessed 2020-12-15).

sets over time, the data provider or researcher may also use these data for activities that do not constitute research. For example, if a researcher assists a government data provider in managing their administrative data both for a research project and to improve the government's internal processes, the latter usage is not a research activity. Managing and organizing data to make data more accessible is still not intended to contribute to generalizable knowledge, so this would not meet the definition of research.

For research to be considered human subjects research, the investigator must be conducting research about a living individual. The federal definition of "human subject" includes that the researcher "(i) obtains information . . . through intervention or interaction with the individual, and uses, studies, or analyzes the information . . . ; or (ii) obtains, uses, studies, analyzes, or generates identifiable private information" (45 CFR 46.102(e)(i-ii)).

There is a regulatory "or" so if either factor is true (intervention/ interaction *or* identifiable private information) then the study is considered to involve human subjects. However, the timing of when the interaction or identifiable information occurs matters. If data were collected for non-research purposes and the data source removed the identifiers from the data before providing it to the researcher, it is research but without human identifiers, so there are no people for the IRB to protect. On the other hand, if the researcher receives identifiable data and is the one to remove the identifiers, then the human subjects have come into contact with the research and the study would require IRB review. The details of the lifecycle of the data matter for IRB review. For additional guidance, the OHRP has produced decision charts⁷ to help IRBs, institutions, and researchers.

While an activity might not meet the federal definition of human research, some institutions may still require researchers to undergo the IRB process; researchers must be aware of their local IRB policies and practices. In addition, many journals, conferences, and workshops require documentation of IRB review; in response, most IRBs have de-

⁷<https://www.hhs.gov/ohrp/regulations-and-policy/decision-charts/index.html> (accessed 2020-12-15).

veloped an abbreviated process for submitting a description of nHR and the IRB will verify whether additional IRB review is necessary, and provide documentation of this process for the researcher

If a study is determined to be human research, there are additional questions to be considered regarding IRB review.

4.6.2 Is the Study Federally Funded?

In addition to the Department of Health and Human Services, there are 19 other federal agencies that are signatories to 45 CFR 46 and include the OHRP regulations for the protection of humans in research in their own regulations. The issue of federal versus non-federal funding (including no funding) is important for two reasons. The first is that most non-exempt federally funded projects are under the purview of 45 CFR 46 and therefore require IRB review. In addition, even if a project is not federally funded, institutional policy may require IRB review. In particular, this is the case if the institution where the research is occurring has a Federalwide Assurance under which there is an agreement that all research will be subject to 45 CFR 46 (Office for Human Research Protections, 2017). Data providers may also require an IRB review, even absent federal funding, as a condition for supplying data for research projects. While most academic institutions have an IRB, private organizations and private individuals are not compelled to use IRB review if their research is not federally funded. For example, private corporations like Amazon, Facebook, and Google can conduct research without IRB review, as they are not constrained in the same way by the federal regulations.

4.6.3 Is the Researcher an Agent Such That the Institution is Engaged in HR?

The follow up to the funding question is the question of engagement in the research. It is possible to be a collaborator on a research project and not be engaged in the IRB sense of the term. If an institution is

not engaged, then IRB review is also not needed. Engagement centers around the question of agency and whether the researcher is an agent of the institution or organization for which the local IRB has oversight. The definition of “agent” will be defined by the institution or organization, not by the individual. The guidance from OHRP about engagement states, “In general, an institution is considered *engaged* in a particular non-exempt human subjects research project when its employees or agents for the purposes of the research project obtain: (1) data about the subjects of the research through intervention or interaction with them; (2) identifiable private information about the subjects of the research; or (3) the informed consent of human subjects for the research.” (Code of Federal Regulations, 2017c) There are nuances to *engaged* and OHRP has detailed guidance regarding what it means to be engaged and examples of *not engaged* in research. The examples in the guidance are helpful to researchers, data providers, and IRBs to consider.

In addition, where there are multiple researchers collaborating on the same research study, some of the researchers and their institutions may not be engaged in HR if their role does not involve access to actual people or identifying information. In multi-site projects, determining who is an agent and what institutions are engaged can get complicated. Engagement is ultimately a decision that is up to the IRB of each institution. Neither can an outside IRB or other external party decide whether another IRB should be involved. Data providers, data stewards, and researchers need to be clear that it is never the place of one institution’s IRB to decide for another that they are not engaged. Data providers, data administrators, and any relevant stakeholders, including researchers, need to know that individual researchers will always be held accountable by their own IRB for verification of engagement. Note that this is distinct from determining the IRB of record for a multi-site research project.

4.6.4 Is the Project Exempt From the Regulations or Non-Exempt (Expedited or Full Board Review)?

The final question is directly related to the level of review. There are three primary distinctions between projects that are eligible for exempt review and those eligible for non-exempt review: risk of harm as it relates to identifiability of the data, vulnerability of the participants, and matters of research consent and waiver of consent.

Identifiability of the Data and Retention of the Identifiers

The most common difference between exempt and non-exempt research is related to the level of risk of harm to participants. *Minimal risk* and *greater than minimal risk* are the two levels of risk that IRBs consider. Minimal risk is defined in the regulations⁸ as “... the probability and magnitude of harm or discomfort anticipated in the research are not greater in and of themselves than those ordinarily encountered in daily life or during the performance of routine physical or psychological examinations or tests.” Anything else is considered as greater than minimal risk.

The probability and magnitude harm are the important concepts related to an assessment of the difference between minimal risk and greater than minimal risk. The magnitude of harm relates to the nature of the harm and the vulnerability of the participants in the research and is somewhat more concrete than assessing the probability of harm. For the IRB, magnitude of harm starts with what could possibly go wrong and then what would be the actual harm to the participant. For projects using administrative data, a common risk of harm is the possibility of linking research information directly to an individual. This can be further exacerbated when combining administrative data with primary data collection. If there is a loss of privacy and confidentiality, the IRB always considers the types of harm that may be related to psychological, legal, social, economic, group, or community harms

⁸<https://www.law.cornell.edu/cfr/text/45/46.10245> CFR 46.102 (j) (accessed 2020-12-15).

with regard to the actual content of the information. Even if reidentification occurs, the level of harm that may result can vary depending on the information in the data. In addition, even if the data collected in a study have been de-identified, there needs to be an assessment of the probability of the re-identification. De-identification is a first line of defense against many harms, but it is not infallible. As technology, software, and algorithms improve, it is increasingly possible to reidentify people based on just a few concrete data points (see chapter 5 for more details).

With personally identifiable or sensitive information, the researcher will be required to provide the IRB with a rigorous data protection and data management plan minimizing the risk of identification or re-identification of participants. The relevant margin that the IRB needs to consider is the additional risk of harm that occurs due to the use of the data for the proposed research project. While collecting and storing the original data may entail risks, these would be incurred with or without the research. From this perspective, the use of an isolated data set under an appropriate data management plan typically does not appreciably change the risk of individuals in the data. Probability and magnitude of harm become more challenging for IRBs, data providers, and researchers when the research is combining multiple data sets. This applies both to combining different sources of administrative data as well as when combining administrative data with primary data collection. The researcher needs to specifically communicate to the IRB not only the risk of each data set in use but also the probability and magnitude of harm of any combined data set. It is important that data providers, data stewards, researchers, and IRBs are informed, informative, and realistic about the probability and magnitude of harm in a study that is engaging in secondary analysis of one or more data sets. That discussion must include the reality of the protection afforded by de-identification as well as the robustness of the overall data protection plan if identifiers are retained. In that regard, it is always a good strategy to include a statement in the research protocol: even if re-identification could be possible, the principal investigator commits to ensuring that the study team will not re-identify participants.

It should be noted that anonymous and de-identified data are not subject to the GDPR of the European Union provided that the research team had no role in the collection of the data with identifiers and has no access to the identifiers going forward. If identifiers are collected by the research team, the definition of “special categories” of data require a more robust data protection plan.

De-Identified Data, Risk, and IRB Review De-identified data once contained identifiers, but by the time of the new use they no longer contain sufficient identifiers to link information to specific individuals with any degree of certainty. The level of IRB review for de-identified data is contingent on who originally collected the data and whether the data are coded or whether a key exists. IRBs need to know when, where, and how the data were de-identified in the life cycle of the research. The IRB will take note of whether the producer of the data (Institution A) is removing the identifiers or whether the recipient of the data (Institution B) is removing identifiers. If Institution B is receiving de-identified data from Institution A, with no access to a code or key and no one on the study team had anything to do with the original collection of the data, it is probable that such a study would not meet the definition of human research. If the study personnel from Institution B were involved with the original collection, will have access to the key of identifiers, or will be removing the identifiers, the study could be exempt. Such a study could be reviewed by expedited procedure if, for example, the PI from Institution B is listed on the original grant proposal as a Co-PI.

Identifiable Private Information and Restricted Data The regulatory code defines identifiable private information as follows: “*Private 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)” And

“Identifiable private information is private information for which the identity of the subject is or may readily be ascertained by the investigator or associated with the information” (45 CFR 46.102 (e)(4)(5)).

Restricted data is a distinction that is at the discretion of the holder of the data. Restricted data are typically described as both private and identifiable by source of the data or data steward. This means there is a process that the researcher must go through in order to obtain access to and use the data. The definition of “restricted” is made by the data source, not by the IRB; the IRB will respect the designation and the level of review required by the source.

The study protocol submitted to the IRB must specify the type of data, the source of the data, and whether the identifiers (if any) will be removed or retained. If there are identifiers or if there is a plan to retain identifiers long term, there must be a data protection plan that specifies where the data will be stored, for how long, and who will have access. The greater the risk to participants of inadvertent disclosure of identifiable private information, the more robust the data protection plan must be.

Vulnerability of the Participants

The second consideration for IRBs in determining whether a project is exempt or non-exempt is regarding the perceived vulnerability of the study population. Vulnerable populations⁹ are defined in the regulations (45 CFR 46 Subpart B, C, and D), including children, prisoners, and other groups of people who are considered to need additional protections due to social or economic conditions. Most human research with vulnerable populations is likely to be non-exempt and subject to regulatory review, although it can depend on the purpose of the study and whether any of the information is already publicly available.

⁹For vulnerable populations under Federal protection see 45 CFR 46 Subpart B regarding pregnant women, human fetuses, and neonates, Subpart C regarding prisoners, and Subpart D regarding minors. Other vulnerable populations identified by IRBs might include situations in which there might be a power differential such as student and instructor, employee and employer; a cognitive or physical disability; or difference that requires additional protections such as literacy, SES, language, or other social status.

Consent and Waiver of Consent

The third consideration that distinguishes exempt from non-exempt studies is the issue of informed consent. Exempt projects have a consent process but are not required to meet the documentation or other requirements for consent as detailed in 45 CFR 46.116 and 45 CFR 46.117 criteria. A waiver of consent or waiver of documentation of consent is not necessary; instead, participant consent may be achieved through distribution of information sheets.

If a study contains personally identifiable information, there is an increased risk of harm, so the study will likely be considered non-exempt. Non-exempt review includes a regulatory requirement for the IRB to review consent. For example, with a non-exempt study that proposes to use administrative data that was obtained without consent, the IRB has to determine whether consent is needed at the point of the research or whether it can be waived. The standard regulatory requirement for all HR is for there to be an informed consent process and a signed written document. For the IRB to waive consent, there are specific regulatory criteria—all of which must be met. Researchers must address in their study protocol the following criteria as part of a rationale for the request for the IRB to waive consent 45 CFR 46.116(f)(3):

- i. “The research involves no more than minimal risk.” Researchers should use the regulatory definition of minimal risk (see section 4.6) in a study specific way in the rationale for the request for a waiver. There is no option for waiver of consent for studies determined to be greater than minimal risk.
- ii. “The research could not practicably be carried out without the requested waiver or alteration.” This refers to the research design. Often there is not a reasonable or feasible way to ask hundreds of people for the consent to use their administrative or other pre-existing data that have been or will be collected over time, because the current research team does not have access to the individuals from which the data were originally collected. The more distant the researcher is from the initial collection of the data, the more likely an IRB will grant a waiver of consent based on this criterion,

- provided there is a robust data protection plan.
- iii. "If the research involves using identifiable private information... , the research could not practicably be carried out without using such information... in an identifiable format." While similar to practicability, this criterion relates more directly to the retention and use of identifiers. From an IRB perspective, this is usually the key part of the ethical deliberation to waive consent for identifiable private information. While the IRB will typically respond more favorably if the researcher plans not to retain identifiers, sometimes the identifiers are needed to connect different data files and data collected over time. For instance, randomized controlled trials often remove but store the identifying information separately from the rest of the data so that subjects can be reidentified in the future as needed, such as in the case of adverse events that need to be remedied. If the identifiers need to be retained, the IRB simply requires the researcher to provide the rationale for why the identifiers are needed and the plan for how the identifiers, or the key to the identifiers, will be kept separate from the actual data. It is also useful to provide a plan for the end-of-study removal of identifiers.
 - iv. "The waiver or alteration will not adversely affect the rights and welfare of the subjects." People have a fundamental right to consent to participate in research. In order to provide a rationale for why a waiver of consent does not affect rights and welfare, the protocol needs to address the issue of protection of privacy of the individual and confidentiality of the data. For example, depending on the original source of the contact information, it could be ethically feasible to justify a waiver consent for using retrospective data to identify potential participants for recruitment to research. Similarly, if all the other elements are addressed and the raw data are to be de-identified (if the benefit of the study is greater than the risk to participants of using their information without consent) this could be a circumstance when rights and welfare would not be placed at risk. Alternatively, when a waiver could adversely affect rights and welfare, it is unlikely to be granted. For example, in a situation where the research poses greater than a minimal

risk to the subjects and the researchers are performing a direct intervention or otherwise interacting with the subjects, the subjects are available and there are no logistical hurdles to obtaining the waiver of consent. Similarly, it would be unusual for an IRB to waive a parent's right to consent (give permission) for their minor child to participate in research because parents, as guardians for their children, have a fundamental right to determine consent for the child to participate in research. Although waiving parent permission is not a welfare issue per se, waiving parent permission could be considered to negatively affect the parent's rights.

- v. "Whenever appropriate, the subjects or legally authorized representatives will be provided with additional pertinent information after participation." Access to participants after the study depends entirely on the research project. In a study that performs primary data collection, notifying participants through some sort of report or posting on a website may be feasible. However, it can also be true that providing a summary to participants is not possible due to passage of time, or is not appropriate due to the relevance of the findings to the individual. This can happen with a project where the researchers do not have direct access to the subjects in the data, as can be the case for studies only using administrative data. Feasibility and appropriateness are considered by the IRB when determining whether researchers need to provide additional information to the subjects of a study.

All of these criteria for granting a waiver of consent use the regulatory "and," meaning that all criteria must be addressed. Researchers who are requesting a waiver of consent need to be proactive about addressing all five of the criteria.

4.7 Strategies for Communicating with the IRB

Working with the IRB should be a collaborative process. While the IRB's authority to approve or reject proposed research projects may frustrate researchers, it is important to emphasize that the purpose

of the IRB is to protect participants and ensure that human research meets the requisite ethical and regulatory criteria.

At any given time, IRB staff are reviewing potentially hundreds of projects from different disciplines, with differing funding sources, and with different regulatory requirements. A project protocol that clearly and directly addresses the criteria from the perspective of the IRB will undergo a more efficient and effective review process.

Communicating effectively and constructively with the IRB is key to getting studies reviewed in a productive and timely manner. The following are some strategies for communicating with an IRB:

1. The protocol templates required by IRBs are constructed to address the ethical and regulatory considerations that must be present for IRB approval. Although protocol templates may vary between IRBs in terms of format and the order of the elements, they are all designed to collect the information required to consider any project in light of the 45 CFR 46.111 criteria.
2. Because IRBs must consider whether a project is exempt or non-exempt, it is important to focus particular attention on the specific interactions with participants and/or their identifying information. The IRB is less concerned about the theory underlying the purpose of the project and more focused on the risks to participants. This includes needing specific detail of the how, when, why, and where of interactions with participants or their identifying information.
3. The protocol should indicate whether current study staff are related or unrelated to the original collection of the data. The protocol should be specific about who is doing what on the study.
4. The IRB needs to know the details of the data collection, access, storage, and management of any retrospective or prospective data used by the research project. There should be data collection instruments or a data dictionary, or both, included with the other study documents. If the information collected is identifiable and sensitive, there needs to be commensurate plan for mitigating risk of harm to the participants.
5. The protocol should address what identifiers will be collected, received, or accessed by the study team. In addition, the retention of

identifiers over the life of the project must be addressed. The IRB will focus on the risk associated with retaining identifiers as well as the risk associated with re-identification of de-identified data. The IRB will also want to know about the risk to participants associated with combining multiple data sets.

6. If the study is collaborative or multi-site, there needs to be a description of what each collaborator and site is doing on the project and a specific articulation of what each collaborator is doing in terms of IRB review. Questions that should be addressed include: what part of the research is happening at what institution, organization, or country, and by whom? If all institutions or organizations are doing the same thing, who is conceptually in charge of the research? For studies subject to the Revised Common Rule's Cooperative Research Provision (45 CFR 46.114), which institution will be the IRB of record?
7. Identify the type of data sharing agreement and the process for establishing it. The process will vary by institution or organization, so researchers should know what policies and procedures apply. The data sharing agreement is not an IRB function, but it can affect the IRB process.
8. Every protocol submitted to the IRB for review stands on its own merit and every IRB has their own way of applying the regulations. Just because one IRB found a project to be exempt, does not mean that another IRB will find the same. Similarly, even within the same IRB, just because one reviewer determined that a project did not need IRB review, that does not mean that another reviewer would come to the same conclusion. Consistency within and between IRBs is a challenge, especially with complicated research: the collaborative process is therefore an important feature. The more information the IRB has to work with, the more consistent the results of the review.

The part of a protocol that relates to the use of administrative data is often easy to write and fast to review if it contains all the relevant information. Researchers facing pushback from an IRB should be able to have a dialogue with the reviewers where the IRB can explain its

decisions and why it is making certain recommendations or requesting specific protections.

The goal of this chapter has been to provide a practical guide to researchers and other stakeholders on managing IRB procedures. It is important to emphasize that while this chapter addresses a wide variety of potential problems and concerns, in practice almost every university where research takes place has a well-functioning IRB, which performs the critical, but typically routine, work of providing oversight of research. Nearly all research proposals are able to satisfy IRB concerns, though they may sometimes require some adjustment to satisfy the principals laid out above.

About the Author

Kathleen Murphy, PhD, MSW, MLIS, is a Certified IRB Professional, retired from Northwestern University IRB in 2019 after serving as a Board member, Vice Chair, and Manager of the Social and Behavioral IRB. She came to Northwestern as the first Social Science Data Librarian in 2006. Prior to Northwestern, Kathleen was a clinical social worker in private practice specializing in work with children. She has taught clinical practice and quantitative and qualitative methods for many years. Kathleen has served on different IRBs over the years and the opinions and perspectives included here are based on her experience and are not presented as the perspective of any given IRB.

References in Chapter 4

- Card, David E., Raj Chetty, Martin S. Feldstein, and Emmanuel Saez.** 2011. “Expanding Access to Administrative Data for Research in the United States.” American Economic Association Report January. <https://www.aeaweb.org/content/file?id=1319>.
- Chicago Public Schools.** 2020. “The Research Review Board.” <https://www.cps.edu/research/Pages/Research.aspx> (accessed 2020-07-03).
- Code of Federal Regulations.** 2017a. “45 CFR 46 Subpart A—Basic HHS Policy for Protection of Human Research Subjects.” <https://www.ecfr.gov/cgi-bin/retrieveECFR?gp=&SID=83cd09e1c0f5c6937cd9d7513160fc3f&pitd=20180719&n=pt45.1.46&r=PART&ty=HTML#sp45.1.46.a> (accessed 2020-09-22).
- Code of Federal Regulations.** 2017b. “45 CFR §46.111 Criteria for IRB approval of research.” https://www.ecfr.gov/cgi-bin/retrieveECFR?gp=&SID=83cd09e1c0f5c6937cd9d7513160fc3f&pitd=20180719&n=pt45.1.46&r=PART&ty=HTML.#se45.1.46_1111 (accessed 2020-09-15).
- Code of Federal Regulations.** 2017c. “45 CFR §46.116 General requirements for informed consent.” https://www.ecfr.gov/cgi-bin/retrieveECFR?gp=&SID=83cd09e1c0f5c6937cd9d7513160fc3f&pitd=20180719&n=pt45.1.46&r=PART&ty=HTML.#se45.1.46_1116 (accessed 2020-09-15).
- Code of Federal Regulations.** 2017d. “45 CFR §46.117 Documentation of informed consent.” https://www.ecfr.gov/cgi-bin/retrieveECFR?gp=&SID=83cd09e1c0f5c6937cd9d7513160fc3f&pitd=20180719&n=pt45.1.46&r=PART&ty=HTML.#se45.1.46_1117 (accessed 2020-09-15).
- Collmann, Jeff, and Sorin Adam Matei.** 2016. *Ethical Reasoning in Big Data: An Exploratory Analysis*. Computational Social Sciences, Switzerland:Springer.
- Connelly, Roxanne, Christopher J Playford, Vernon Gayle, and Chris Dibben.** 2016. “The role of administrative data in the big data revolution in social science research.” *Social Science Research*, 59: 1–12. <https://doi.org/10.1016/j.ssresearch.2016.04.015>.
- European Parliament and Council of the European Union.** 2016. “Regulation (EU) 2016/679.” <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN> (accessed 2020-09-15).
- Feeney, Laura, Jason Bauman, Julia Chabrier, Geeti Mehra, and Michelle Woodford.** 2015. “Administrative Data for Randomized Evaluations.” J-PAL North America. <https://www.povertyactionlab.org/resource/using-administrative-data-randomized-evaluations> (accessed 2020-01-30).
- Figlio, David N., K. Karbownik, and K.G. Salvanes.** 2016. “Education Research and Administrative Data.” In *Handbook of the Economics of Education*. Vol. 5 of *Chapter 2*, 75–138. Elsevier. <https://doi.org/10.1016/B978-0-444-63459-7.00002-6>.
- Lee, Allen S., and Richard L. Baskerville.** 2003. “Generalizing Generalizability in Information Systems Research.” *Information Systems Research*, 14(3): 221–243. <https://doi.org/10.1287/isre.14.3.221.16560>.

- National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research.** 1979. “The Belmont report: ethical principles and guidelines for the protection of human subjects of research.” Department of Health, Education, and Welfare. <https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/read-the-belmont-report/index.html>.
- NHS Health Research Authority.** 2020. “Research Ethics Committees in the UK.” <https://www.hra.nhs.uk/about-us/committees-and-services/res-and-recs/research-ethics-committees-overview/> (accessed 2020-07-03).
- Office for Human Research Protections.** 2016a. “Equivalent Protections.” <https://www.hhs.gov/ohrp/international/equivalent-protections/index.html> (accessed 2020-09-15).
- Office for Human Research Protections.** 2016b. “Federal Policy for the Protection of Human Subjects (‘Common Rule’).” <https://www.hhs.gov/ohrp/regulations-and-policy/regulations/common-rule/index.html> (accessed 2020-09-15).
- Office for Human Research Protections.** 2017. “Federalwide Assurance (FWA) for the Protection of Human Subjects.” <https://www.hhs.gov/ohrp/register-irbs-and-obtain-fwas/fwas/fwa-protection-of-human-subjecct/index.html> (accessed 2020-08-21).
- Office for Human Research Protections.** 2018. “Compilation of European GDPR Guidances.” <https://www.hhs.gov/ohrp/international/GDPR/index.html> (accessed 2020-09-15).
- Office for Human Research Protections.** 2020. “Human Subject Regulations Decision Charts.” <https://www.hhs.gov/ohrp/regulations-and-policy/decision-charts/index.html> (accessed 2020-09-15).
- Wandile, Pranali M.** 2018. “Central IRB vs. Institutional IRB—Advantages and Disadvantages for Multicenter Trials.” *Clinical Researcher*, 32(4): 28–38. <https://doi.org/10.14524/CR-17-0009>.

Appendix

A data-only protocol template can be found in the Online Appendix at admindatahandbook.mit.edu/book/v1.0/irb.html#irb-appendix.