

Chapter 6. Gather Credible Evidence

What's Inside?



What this chapter contains

- A discussion of identifying data sources and selecting measures
- Recommendations for developing data collection instruments and data collection procedures

Who can use this chapter

- Evaluation team members completing an evaluation plan
- All staff engaged in or responsible for data collection

Click the links below to view the relevant section

[Practice Culturally Responsive and Equitable Evaluation Methods when Gathering Data](#)



[Monitor Data Collection Activities](#)



[Protect Study Participants](#)



[Create Data Collection Procedures](#)



[Introduction](#)



[Identify the Best Data Sources](#)



[Select or Develop Data Measures](#)



[Develop Instruments](#)



Introduction

This chapter serves as a companion to [chapter 5](#), delving deeper into data collection, beginning with planning activities related to evidence gathering. The following are data collection decisions you will need to make:

- Which sources you will use to obtain data
- What data elements you need and what measures you will use to collect your data
- How you will structure your data collection instruments
- What procedures you will use to collect data
- How you will continue to monitor data collection and protect study participants

While these decisions are presented in sequential order, you will likely go back and forth developing answers to these questions according to data and data source availability, budget, and effort until you have a complete and feasible plan.

Identify the Best Data Sources

While methodologies presented in chapter 5 describe techniques for how to collect data, you also need to determine where you will collect data and from whom or what. Two types of data sources are available: primary¹ and secondary.² An example of a primary data source is a survey of individuals participating in the program being evaluated for the specific purpose of collecting information to be used in an evaluation. An example of a secondary data source is program administrative data. Program administrative data, such as household composition, income, and program attendance, may be collected to inform program management, service eligibility, service tracking, and reporting needs, but the data can also provide important information on variables of interest for the evaluation.

Making use of administrative data

Administrative data are often underused when building knowledge to inform social service policy and program design. Using such data can be a cost-effective way to answer policy-relevant evaluation questions by eliminating the need for gathering primary data. However, consider the logistics and time required to access administrative data and assess quality or usability. Potential considerations related to administrative data follow:

- Data quality, including amount of missing data and consistency in coding and timing of data collection
- Time lag between waves of data collection and data availability
- Approvals for using administrative data for evaluation purposes
- Developing data use scope and rules for a data sharing agreement
- Data transfer from data storage entity to evaluator
- Creating or clarifying information for a data dictionary or codebook (OPRE, 2019)

For additional guidance on the use of administrative data, see OPRE (n.d.) for sample reports that demonstrate and discuss the use of administrative data for evaluation supporting social services.

¹ A primary data source is an original data source; that is, the data are collected firsthand by the researcher for a specific research purpose or project (Salkind, 2010).

² Secondary data refer to data that have already been collected for some other purpose (Allen, 2017).

One strategy to identify data sources is to brainstorm all the possible data sources that could inform your evaluation questions and identify the benefits and limitations of each.

Example: You are interested in knowing whether children in foster care feel safe and comfortable during visits from birth parents at your family visitation center. Potential primary data sources include the participating children themselves, family visitation center staff, children's case managers, children's foster parents, and children's birth parents. Potential secondary data sources include visitors' logs tracking the frequency and setting of visitations from birth parents, court reports written by social workers, and documentation of the birth parents' adherence to court-ordered case plans (e.g., parenting classes, drug treatment services).

Each data source provides unique information about the children's experiences. While the evaluation team might first consider collecting data directly from the children, this might not be the best choice. Some children may be too young to engage in the data collection effort, or your IRB may determine such questions would be too emotionally difficult for children to answer. You may determine birth and foster parents are too busy to respond to a data collection request, and collecting data from children's case managers is best given your constraints.

Assess how well each potential source will provide accurate and high-quality data. Consider how complete and thorough the data are likely to be. For example, individuals generally provide more accurate information about themselves than can be obtained from a secondary source. However, program participants may not have access to or be able to recall the specific data needed. Another consideration is survey response. If you plan to conduct a survey of program participants, will they be particularly difficult to reach or track (such as people with unstable housing)? If so, you may decide an administrative data source that captures information about all or most evaluation participants is a better data source than a survey. A survey could suffer from low response (and therefore would not be representative based on your ability to contact evaluation participants).

Metadata: data about your data

Did you know you can improve your evaluation skills and future evaluation efforts by learning about your current data collection efforts? Many online survey platforms offer access to survey metadata. Metadata is information about how your survey data collection effort unfolded. Depending on your survey platform, you may be able determine the average time it took people to complete your survey, the items they spent the most and least time on, which types of equipment survey respondents used, and information such as the time of day and location where they completed the survey. This information can help you identify items that may need to be revised, improve communications around expectations of time needed to complete survey, and help time reminders for survey completion.

While you need at least one data source for each data element you collect, you do not need to limit yourself to one. Using more than one data source to measure the same data element is called triangulation.³ Triangulation can increase the quality and rigor of your study. For example, you could measure substance use through both a standardized self-report measure on a study and through an evaluation participant blood test. You could measure child behavior by interviewing both a child's primary parent and primary

³ Triangulation refers to the use of multiple methods or data sources in qualitative research to develop a comprehensive understanding of a phenomenon. Triangulation also has been viewed as a qualitative research strategy to test validity through a convergence of information from different sources (Carter et al., 2014).

teacher. This would let you balance the relative advantages and weaknesses of each source and check for consistency across sources.

In deciding the best sources for information, the goal is to obtain the most accurate and complete data available within the cost, time, and burden constraints of your evaluation. Consider the following questions when evaluating data sources:

- Are useful secondary data sources available to address the evaluation question?
- Is the data source accurate (i.e., does it provide relevant and correct information about the concepts you intend to measure)?
- Is the data source reliable (i.e., does it yield consistent results)?
- Is the data source timely (i.e., is it available for analysis and interpretation when needed; does it cover the time period needed) and within the budget of an evaluation?
- Is the data source comprehensive and complete (i.e., does it provide sufficient detail or contextual information to be meaningfully interpreted; do you anticipate high levels of missing data)?
- Will collecting information from a particular data source pose an excessive burden (i.e., will it take much evaluation team time or budget to secure, or much participant time)?

Identifying the appropriate data source often involves considering tradeoffs. For example, you may obtain more in-depth information about services from interviews with program staff, but because of time and budget constraints, you choose to rely on case records or program logs. In this case, you must document the limitations of using a secondary data source when obtaining this information and discuss the implications when sharing results.

The rest of this chapter focuses on primary data collection. Many factors should be considered when using secondary data. For example, you must understand the limitations of a specific dataset, such as data from the American Community Survey or the need to develop an extraction tool for systematic data extraction from documents. Such elements are outside the scope of this Guide. See the “To learn more” textbox at the end of this chapter for secondary data resources.

Identifying and prioritizing information needs

When making decisions about your information needs, you and your evaluation team may consider collecting certain types of interesting data. However, if the data are not directly related to at least one of your evaluation questions, you should resist the urge. Limiting data collection to “must know” is more time-efficient, cost-effective, and respectful of data providers’ time. The greater the time commitment required, the harder it will be to recruit participants. It is preferable to achieve a high response to your survey using a concise data collection instrument.

Select or Develop Data Measures

Data Elements

You will need a measure for each of your data elements. Data elements are ideas you want to capture information about. They are more than just outcome⁴ domains. They include every concept you need to answer all your evaluation questions (implementation and outcome) and any that can provide context for your findings (such as gender, age, geographic location, or educational attainment). For each evaluation question, develop a list of all the data elements you need to measure, then develop measures to capture or operationalize the data. Below are two examples.

You are evaluating your mentoring program targeting adolescents who would be first-generation college students, and you want to know who is participating in the program. Your implementation evaluation question is, “What are the characteristics of adolescents who enrolled in the mentoring program?” As a first step, you will identify all the characteristics you want data about. Those characteristics might include age, gender, race/ethnicity, family/household structure, career/education aspirations, and relationships with peers. You will also define enrollment (e.g., completed all the registration paperwork). For each data element, you will identify exactly how you measure it. Some constructs, such as gender, might seem simple on the surface, but you will need to think through how you will define them. For example, will you ask for gender assigned at birth or current gender identity? Will you offer noncisgender options, such as nonbinary, transwoman, or transman? Concepts such as career aspirations are even more complicated to convert to measures. Luckily, many other evaluators and researchers have grappled with developing strong, valid measures for many concepts, as discussed below under data measures.

Measurement terminology

Measurement can take numerous forms. A measure is all the information you will use to operationalize a data element. Sometimes you will use a single item (e.g., how old are you) as a measure. Other times you may use multiple items to collectively measure a data element. This typically happens when you are measuring a more complex data element. For example, you may develop several interview questions (items) to capture staff experience with program recruitment or use a scale to measure child development that has 20 or more items, such as the Ages and Stages Questionnaires (Squires & Bricker, 2009).

In another example, you are evaluating your local Head Start program and want to know the impact of the program on socioemotional learning. Your outcome evaluation question is, “To what extent do children at Alexandria’s Head Start demonstrate improvements in socioemotional learning after we institute a mindfulness program?” You need to determine what elements of socioemotional learning you expect will change as a result of the mindfulness program. The CASEL framework (CASEL, 2020) identifies five elements of socioemotional learning: self-awareness, self-management, responsible decision-making, relationship skills, and social awareness. You will need to visit your logic model and program materials to select specific types of socioemotional learning you will measure. You will then need to determine how you will measure self-awareness, for example, among preschool children.

⁴ Outcomes are variables that are monitored during a study to document the impact of a given intervention or exposure (Ferreira & Patino, 2017).

Data Measures

Measures are the tools you will use to assess each of your data elements precisely. All data elements must be operationalized. Operationalization means turning your more abstract concepts into measurable observations. If you want to measure earnings, for example, you could operationalize this element as participant-reported wages per hour, weekly take-home pay as documented on a pay stub, average salary for participants' work titles as documented by local labor market information, or the amount indicated on a participant's W-2. Much like data sources, you will need to assess the options for measuring each of your data elements based on quality, accessibility, feasibility, and precision.

Quality measures typically have the following characteristics (Blocklin et al., 2019):

- **They are reliable.** Good measures have demonstrated capabilities to collect information consistently. This means the same result can be achieved repeatedly using the same methods under the same circumstances. Typical reliability measures include internal consistency, test-retest reliability, or interrater reliability.
- **They are valid.** Good measures are those that truly capture the concept they intend to measure. Measurement validity has several different aspects, such as face, content, and criterion (Price et al., 2017).

Outcome measures should also exhibit the following qualities:

- **They are sensitive to change.** Good outcome measures can be expected to capture change over time and within the timeframe of data collection and expected sample size. For example, a measure such as Adverse Childhood Experiences (Felitti et al., 1998) is static; you cannot experience a reduction in the number of childhood experiences. Some measures may not be sensitive enough to capture change during your data collection timeframe. For example, a drug or alcohol use measure that asks about use over the previous 6 months would not be appropriate for a short-term program. If the time between pre- and posttest measurement is less than 6 months, the posttest measurement is still capturing behavior before the program started. Finally, some measures may not be sensitive enough to capture your expected level of change. If you expect your program to make small improvements in quality, a single item (e.g., that asks "is your marriage good" with yes or no as response options) may not capture enough nuance or detail. A more complex measure that captures multiple aspects of a marriage would be better suited to your needs.
- **They are not overaligned.** Good outcome measures are not too closely aligned with or tailored to the intervention being tested. Overalignment can occur when a measure is developed for a study of a specific program or when the intervention's creator develops the measure. Overaligned measures are a particular problem in impact analyses because they give the treatment group an advantage in appearing to have improved in that data element. For example, if you are providing a parenting course based on the book "1-2-3 Magic," you shouldn't measure parenting techniques by asking a question based directly on the book such as, "Do you follow the no talking, no emotion rule?" This kind of measure will skew your posttest comparisons. Only people who have completed the program will understand this measure; people who have enrolled but not yet started the program and people in your comparison group will be unable to answer this question accurately.

Selecting measures for each data element. Continuing with the three measurement examples above (gender, career aspirations, and self-awareness), you may find your program already captures some elements in an administrative database (such as gender). As discussed in section A, secondary data sources can provide a low-cost source of measures. However, ensure the data contained in those secondary sources truly reflect the concept you wish to measure. Review how the data are collected and coded to ensure you are meeting your information needs.

If you do not already have the data, you could develop your own measure to capture the data, or you could use a measure developed by another researcher. Available measures already in use typically offer many advantages.

Developing qualitative measures

While some evaluations use existing measures for some or all their quantitative data, almost all evaluations develop at least some new measures to support their qualitative data collection. For example, you may write your own questions or items for interviews and focus groups. You should consult question development resources to provide advice and principles related to good question development, such as Jacob and Furgerson (2012) and Krueger (2002). Sources such as the Office of Management and Budget provide copies of interview and focus group protocols used in federal research and evaluation. You may be able to copy or adapt questions used in similar data collection efforts.

Researchers often invest heavily in developing measures to ensure they will produce high-quality data. They are typically composed of scales, meaning they have many components or questions, which help to measure a complicated concept such as self-awareness. They have often undergone testing to confirm they are reliable and valid. Some measures also have corresponding instructions for use, coding, scoring, and even interpreting scores (such as values that indicate thresholds for high, medium, and low values).

However, no measure is suitable for every concept for every target population and type of data collection effort. For example, you may find self-awareness measures, but they are designed for elementary aged-children to self-report, and you need a measure a teacher can complete about a 3-year-old. Or, you may find available measures of career aspirations that have been validated with White, urban males, but they will likely not resonate with the Vietnamese American girls in rural Louisiana participating in your program. Table 6.1 shows considerations to help you compare the advantages and disadvantages of available versus new measures. You may end up with a combination of available and new measures.

Table 6.1. A Comparison of Existing Versus New Data Collection Measures

Type of Measure Used	Advantages	Disadvantages
Using an available measure	<ul style="list-style-type: none"> ■ Often standardized ■ Usually established as valid and reliable ■ Can offer comparisons or benchmarks from other studies or surveys 	<ul style="list-style-type: none"> ■ Not always appropriate for all cultural or ethnic populations ■ May not be useful for specific program ■ May have use implications (e.g., cost, administration restrictions)
Developing a new measure	<ul style="list-style-type: none"> ■ Can align more closely with program objectives ■ Offers more flexibility to increase cultural sensitivity and relevance of measurement content and language 	<ul style="list-style-type: none"> ■ Might not be seen as valid and reliable ■ Can be less psychometrically sound (i.e., less likely to measure what it is meant to with reliability and validity) ■ Good measurement development is difficult (time and resource investment)

Criteria for selecting data measures

Use the following questions to assess your measurement options:

- Does the measure address a program domain (e.g., for a parenting training, choosing a measure that captures change in knowledge or skills)?
- Does it have acceptable values of validity and reliability (e.g., consistency measures, test-retest statistics, face validity)?
- Is the measure appropriate for participants with regard to age or developmental level, language, and ease of use?
- Does the measure respect and reflect participants' cultural backgrounds (e.g., are definitions, concepts, and items in the measure relevant to the participants' community and experience)?
- Have you pilot tested the measure to uncover any difficulties and ensure it can be completed in a reasonable timeframe?

Develop Instruments

After you've identified all your measures, data sources, and methodologies, you will likely need to develop one or more instruments to collect primary data. The most common instruments are interview protocols, focus group protocols, and surveys (which can be self-administered electronically, on paper, or through a data collector asking questions and recording answers). Important elements of instruments follow:

- **Introductions and consent.** Respondents should understand the purpose of the evaluation they are participating in, why their participation is being requested, what their participation entails, and how their data will be used (see [Protect Study Participants](#) below for more information).
- **Ground rules.** You will likely develop instructions or expectations for each type of data collection. Focus groups need to operate using common understanding of participation (developed by either the evaluator or by group consensus), such as confidentiality, respect, and ensuring all participants have space to speak. Interviewees should receive information about how their names will be recorded (if not in consent form), how you will handle recordings, and how interviewees can indicate they do not want to answer a question. Survey respondents need to know how to mark their responses, where to pose questions if they have problems, and how to save and submit their responses.
- **Item flow.** You will need to order all your measures (single or multiple item) in a logical fashion. You might start with easier, more factual questions, and save more personal questions for later in the survey or interview. You may not want to leave vital measures until the end of any instrument in case you run out of interview time or the survey respondent quits the survey before it is over. You will need to carefully plan any skip logic patterns (where, depending on the answer to one question, your respondent is sent to a different subsequent question). Skip logic is often used to make a survey easier for respondents because they skip questions that do not apply to them (e.g., there is no need to ask a respondent the ages of their children if they indicate they have no children). Additional attention and time might be needed to ensure surveys are coded correctly and to clean and prepare a dataset with skipped items.

- **Thank-you, closing, next steps.** Evaluation participants give their precious time to the evaluation, and you should thank them for that. Before you end data collection, you may also consider giving evaluation team contact information in case participants have questions later. This offers a way for participants to follow the evaluation if they want to read public reports, and if appropriate, for the evaluation team to provide referrals or hotlines to address potential traumas or issues that sensitive evaluation questions may have revealed.

After your instruments are compiled, test them before data collection. Your testing should ensure your instrument is visually correct (on screens or on paper), automatic skip patterns work or paper instructions are clear, response options are coded correctly (e.g., an item that directs respondents to “choose all” doesn’t limit them to one response), paper instructions are clear, and electronic data are saved in your system after a survey is submitted. For all instruments, you should practice fielding—or pilot testing—them to see how long it takes to collect the data, make sure the questions flow, and if appropriate, the data collector is comfortable with the instrument.

To the extent possible, data collection staff should conduct the pilot testing. Ask them to take notes and make comments on the process of administering or using each instrument. Then, review these notes and comments to determine whether changes to the instruments or data collection procedures are needed. As part of pilot testing, instruments should be reviewed to assess the number of incomplete answers, unlikely answers, comments on items that may be included in the margins, or other indicators that suggest revisions are necessary.

Be mindful

Understand you may be asking personal, uncomfortable, and potentially traumatizing questions of your evaluation participants. Data collectors should be trained to understand how evaluation participants may experience data collection and understand how to respond to participants’ concerns or reactions. Determine whether you should develop a resource list to share with evaluation participants at the end of data collection. A resource list should provide local or national organizations and hotlines that can respond to issues that surfaced during data collection. For example, a federally funded evaluation on romantic relationships probed on intimate partner violence, and all instruments closed with information about the National Domestic Violence Hotline.

Create Data Collection Procedures

Important components of evaluation rigor are consistency and accuracy in data collection. Ideally, you will develop written data collection procedures and use those procedures to train data collection staff, monitor data collection efforts, and describe your methodology in final reports and articles. Note that some evaluation funders may have specific requirements for data security which programs will need to follow.

Data collection procedures should consider the following:

- **Who is in charge of and participates in each data collection effort.** You may name staff roles (e.g., case manager) or staff names and their role on each data collection effort (e.g., Dr. Hernandez will lead 10 focus groups). You may also identify the skills, knowledge, training, and experience individuals engaged with data collection should have. For example, you may require all data collectors to receive training about trauma and retraumatization and all evaluation team members to agree to evaluation participant protections such as signing nondisclosure agreements or holding human subjects research training certificates.
- **When each data collection effort occurs.** This includes dates (e.g., staff interviews occur annually in May) and time in relation to evaluation participation (e.g., baseline data are collected within 2 weeks of study enrollment, and posttests take place within 2 weeks after the 10-week program ends).
- **Where and how each data collection effort occurs.** Data collectors need checklists to ensure data collection occurs as intended and consistently across data collectors. Procedures can include instructions, for example: Intake forms are completed on a tablet by evaluation participants alone in a private room. Evaluation staff wait outside to answer questions if needed. You should also address the use of any incentives such as gift cards and when you provide them (e.g., after consent, after completion of all forms) and any other requirements, such as signed acknowledgement forms.
- **How online data collection sites are accessed.** You need to keep your online data collection and storage sites secure. Research the security features of any online survey or data collection system you might use. Ask about protections they have in place, certifications of their systems, and their other clients. Employ strong access controls, such as two-factor authentications, strong and frequently changed passwords, and access limited to only necessary evaluation team members. Consider downloading the data at the end of data collection to analyze on a physical computer and deleting the copies online.
- **How data are handled and protected.** A clear chain of control for transfer of data is important. Strong data security procedures are also needed to reduce the likelihood of breaches or identification of evaluation participants.
- **How data are stored.** Both paper instruments and electronic files need to be kept as secure as possible. Security related to paper forms could include storage in a locked hotel room safe, secure shipping to the office, and ultimately storage in a locked file folder. Electronic files should be stored on restricted access folders that only necessary evaluation team members can access. Develop procedures to keep original and working copies of electronic files in case of accidental deletions or corruptions.
- **How evaluation participant privacy is maintained.** Many evaluations ask sensitive or personal questions related to income, involvement with child welfare, parent-child relationships, or experience with violence. Evaluation designs may require two or more data collection waves, with information collected from the same participants. Evaluations need to connect data across waves to the same person but keep the participant's identity separate from their sensitive data.

Most evaluations address this problem by using unique identifiers. Each evaluation participant is assigned an identification (ID) number. Only the ID number is included on files containing responses and other data. Never include information that can be used to identify a person in the same file that contains

the data. A separate crosswalk file should be maintained with contact and personal identification information so you can link a participant to an ID number as needed. The datafile and crosswalk should be kept in separate locations on a cloud or computer drive with different access restrictions. Individuals involved in the data collection must respect participant privacy by using this information only to track participation (not to review an individual's data). Data collectors must not discuss anything they have learned about an individual during the data collection.

- **What to do if things go wrong.** It is almost inevitable your evaluation will encounter a problem at some point. Good data collection procedures plan for common challenges (e.g., completed intake form goes missing, a data collector quits) and provide guidance to the evaluation team on how to handle problems and whom to contact (e.g., evaluation team leadership). You may also want to discuss how to handle rarer but challenging problems; for example, a situation where a data collector feels unsafe or is threatened.

Data security principles

Several frameworks define data security, including the CIA Triad (Samonas & Coss, 2014), the General Data Protection Regulation (Tamburri, 2020), and the Precision Medicine Initiative (PMI) Data Security Policy Framework (All of Us Research Program, n.d.). The PMI framework is based on work conducted by the National Institute for Standards and Technology, offering five principles:

- **Identify:** Develop a data security plan, use risk management approaches to develop protection decisions, have your plans reviewed by independent parties, be transparent in your approaches and plans.
- **Protect:** Employ access control measures, conduct awareness and training efforts, execute data security plans, maintain data infrastructures.
- **Detect:** Audit events and logs, employ a detection and alert system, share information threats with similar organizations, report anomalies to organizational leadership for future prevention planning.
- **Respond:** Develop and employ an incident response plan, test the incidence response system regularly, notify individuals when their data were part of a breach, assign a staff member to be the accountability point of contact for breaches.
- **Recover:** Establish and implement incident and breach recovery plans, communicate when the data platform is resecured, document lessons learned.

Everyone engaged in collecting evaluation data must be trained in data collection procedures. Training should include the following:

- An item-by-item review of each of the instruments to be used in data collection, including a discussion of the meaning of each item, why it was included in the instrument, and how it is to be completed
- A review of all instructions on administering or using the instruments, including instructions to the respondents
- A discussion of potential problems that may arise in administering the instrument, including procedures for resolving the problems

- A practice session when data collection staff administer the instrument to one another, use it to extract information from existing case records or program logs, or complete it themselves if it is a written questionnaire
- A discussion of respondent confidentiality, including administering an informed consent form, answering respondents' questions about confidentiality, keeping completed instruments in a safe place, and procedures for submitting instruments to the appropriate person
- A discussion of the need for frequent reviews and checks of the data and for meetings of data collectors to ensure consistent data collection

Ideally, you will develop a training guide or manual for data collection staff to read. You would then conduct a training or a series of trainings with data collection staff with ample time for questions and discussions. To anticipate staff turnover, record the trainings so new staff can be onboarded quickly. If your data collection period is lengthy (e.g., over 6 months), you should schedule regular refresher trainings. You should also update the manual and hold additional trainings if data collection procedures change, or if you move into a new phase of data collection (e.g., begin a follow-up data collection phase that involves a new mode of data collection, such as a phone survey).

It is useful to develop a manual that describes precisely what is expected in the information collection process. This will be a handy reference for data collection staff and useful for new staff hired after the initial evaluation training occurred.

Protect Study Participants

[Chapter 5](#) provides an overview of the IRB process and discusses the importance of obtaining IRB approval prior to starting any data collection. All research and evaluation efforts must take adequate steps to protect the privacy and confidentiality of all human data. This includes working with available or secondary data and any primary data collection efforts.

IRB protections for “vulnerable populations”

The Code of Federal Regulations dictates the way IRBs work. The code requires that the IRB pay special attention to evaluations conducted with certain kinds of people: pregnant women, fetuses, and newborns; children; and incarcerated people. While you should not avoid conducting an evaluation with these populations, you will need to spend additional time thinking through how to ensure their safety in your study. For example, a study of a certain medication or therapy may have different (negative) effects on fetuses, children may not be able to understand the full ramifications of participating in an evaluation, and people who are incarcerated may be vulnerable to coercion based on their imprisonment. Your IRB may extend the spirit of protections for vulnerable populations to other groups such as non-Native English speakers or individuals with cognitive impairments. If your evaluation will collect data from any of these populations, you must communicate closely with your IRB to ensure your evaluation adequately protects all participants. For more information, see Shivayogi (2013).

Ensure all your data collection efforts align with the procedures your IRB approved. Make sure any deviations, including adverse effects or data breaches, are reported to your IRB within the approved timeframe. Adverse effects include a negative reaction an evaluation participant has to the evaluation, such as crying during an interview. A data breach occurs whenever an unauthorized individual gets or could have gotten access to data. Common data breaches occur when someone uses a nonsecure method of data transfer or leaves data unattended, or a computer with data stored on it is lost.

Monitor Data Collection Activities

Throughout the data collection timeframe, monitor and examine data collection efforts. Data collection needs to proceed according to your plans and be consistent and accurate over time; across data collectors; and across people, sites, and treatment and comparison groups. Failure to administer data collection instruments correctly or consistently can significantly degrade the quality of your data and in some cases make it unusable.

Data collection monitoring activities can include the following:

Establish a routine and timeframe for submitting completed instruments. This information may be included in your data collection manual. Data collectors should submit their completed instruments to the appropriate receiver as soon as possible (e.g., upload completed observation tools to the required folder each night of a site visit). A process should be in place to quickly check completed instruments for accuracy and completeness. Ideally, errors should be identified in time to resolve them (e.g., call the evaluation participant back and complete missing questions). Many data collection efforts use software to create automatic checks, such as pausing a survey if a required item is not answered or flagging responses that seem likely to be incorrect (e.g., a birthday in the 1800s). You may need to retrain some or all data collectors if some mistakes occur frequently.

Conduct random observations of the data collection process. A member of the evaluation team may be assigned to observe the data collection process at various times during the evaluation. This person, for example, may sit in on an interview session to ensure all procedures are being conducted correctly.

Conduct checks of data coding and privacy procedures. If you use paper-and-pen data collection forms, someone will need to enter the data into a datafile. Develop a data coding quality check plan where another individual spot checks the datafiles against the original forms to ensure all data are entered and entered correctly. Monitor procedures to assign evaluation participants unique identifiers and the construction of separate contact and evaluation datafiles.

Conduct quality checks with respondents. As an additional quality control measure, it can be helpful to assign someone on the evaluation team to routinely check with a sample of respondents to determine whether the instruments were administered in the expected manner. This individual may ask respondents if they were given the informed consent form to sign and if it was explained to them, where they were interviewed, whether their questions about the interview were answered, and whether they felt the attitude or demeanor of the interviewer was appropriate.

Sampling error: A threat to the evaluation

Chapter 5 discusses internal and external validity and how they affect the credibility of an evaluation. Sampling errors are one way an evaluation can have reduced internal validity.

Encourage staff to view the evaluation as an important part of the

program. If program staff are given responsibility for data collection, they will need your support for this activity. Their priority is providing services or training to participants, and collecting evaluation information may be a secondary goal. You will need to emphasize to your staff that the evaluation is part of the program, and evaluation information can help them improve their services or training to participants. The best way to demonstrate the value of evaluation data is to provide concrete examples, such as a mock data report (at project beginning) along with regular, complete reports and periodic findings presentations.

Monitor completion rates. Aim to collect data from as many members of your sample as possible. Missing data can affect the accuracy of your findings across the whole population, or eligible units of observation, you are studying (see textbox). Monitor overall completion rates compared with your goal and assess completion rates by subgroups. Relevant analyses could compare completed intake forms by intake coordinators, survey completion rates by community, and interview rates by treatment and comparison groups. If you find differences in completion rates, work to address them. For example, retrain intake coordinators or spend more time and effort to obtain responses from underresponding community members.

Monitor for missing data

Track your response rates and missing data rates. Overall, the higher the rate of missing information or nonresponse, the less confident you can be that the data you collect represents responses you would have gotten from all the eligible entities. For example, if people who didn't find employment after completing your training program were less likely to complete the postprogram survey, your evaluation may find artificially high estimates of program outcomes on employment. In this way, missing data affects the validity and generalizability of your evaluation. See [chapter 7's section](#) on assessing data quality for more information.

After evaluation information is collected, you can begin to analyze it. To increase the benefits of the evaluation to you, program staff, and program participants, this process should be ongoing or occur at specified intervals during the evaluation. Information on the procedures for analyzing and interpreting evaluation information are discussed in the following [chapter](#).

Practice Culturally Responsive and Equitable Evaluation When Gathering Data

Sources for data, either primary or secondary, can have inherent bias. Whether you are selecting or developing instruments to use with program participants or compiling already available administrative data, use a CREE approach to think through how bias might be built into the data. One type of bias is the selection of those who will have the opportunity to provide data. Think about those you might reach with a mailed survey versus by telephone or online. How does the time of day and day of the week of a focus group exclude some and give advantage to others? For secondary data, it might be difficult to identify biases because the instruments and data collection methods were decided in the past by others. Depending on the data source, a description of evaluation methods might be available for you to review.

Community members, members of your program's target population, and other advocates for human service program recipients can help improve your data and data collection efforts. Community member input can benefit data instrument decisions and identify concerns respondents may have about participating in data collection. Consider the following recommendations:

- Select measures that are culturally appropriate and credible. [Chapter 5](#) discussed ensuring you select appropriate outcomes. Measures selected should also appropriately operationalize those outcomes. Community members should vet measurement options and provide advice related to their selection. For example, you may learn that one positive parenting measurement uses more culturally relevant language than another.

Community engagement in Tribal evaluation initiatives

In addition to following CREE best practices, some evaluations may be required to work alongside community members throughout the data collection process. Evaluations subject to multijurisdictional oversight, such as Tribal governments, need to abide by all Tribal regulations and recommendations. Tribal members are protected by their own sovereign nation's oversight. Evaluators should recognize that Native American and Indigenous populations have a history of being harmed by research and evaluation. The National Congress of American Indians recommends that evaluations that include tribal members adhere to the following principles:

1. Indigenous knowledge is valid and should be valued.
2. Research is not culturally neutral.
3. Responsible stewardship includes the task of learning how to interpret and understand data and research.
4. Tribes must exercise sovereignty when conducting research and managing data.
5. Research must benefit Native people.

For more information, see National Congress of American Indians Policy Research Center (2009).

- Select measures that have been “normed” and tested with members of your program’s target population. Measures developed for and tested with people like your evaluation participants are much more likely to accurately capture the concept they intend to measure.
- Work closely with many members of your community and target population if you decide to develop your own new measures or adapt current measures. Consider consulting with a subject matter expert who is also a member of the target population to help develop and test the new measure.
- Ask members of your target population to review and help pilot test your data collection instruments. Ensure the measures and response options are relevant, the order of items is sensible, the instructions are clear, and the language is accessible and at the appropriate reading level. If the instrument is online, make sure it works on mobile devices, tablets, and computers. Talk with community members to get a sense of accessibility issues such as Wi-Fi, public computers, and how much data a respondent might need to complete the instrument. Test sending the instrument to various email platforms to determine if the transmission is flagged as spam and ask community members to help develop email language more likely to be seen as authentic and not a scam.

Culturally and linguistically responsive data collection

To understand the effectiveness of federally funded nutrition assistance programs in Puerto Rico, a study team developed a household survey on food security, economic well-being, and coping strategies following natural disasters such as hurricanes. The study team used a standard measure of food security, validated in Spanish, that a Puerto Rican member of the study’s technical expert panel vetted. The study team needed to develop new items or revise others that had been used with other populations to capture data on other topics such as shopping habits. To ensure these items were clear and response choices were appropriate, the team pretested the survey instrument in Puerto Rico. A local, trusted business helped recruit pretest participants and advised the study team on the incentive amount. Two Puerto Rican bilingual evaluators translated the survey into Spanish and back into English (back translation is a method to assess the quality of the translation). During the pretest, participants read each question aloud and recited their thought process so the interviewer could discern any confusion or hesitation. Interviewers also asked pretest participants about likely reasons a sample member might not respond. Based on the responses, the survey materials were revised to emphasize that all responses would be kept private and would not affect the respondent’s benefits in any way (Wilson, 2021).

- Consider including community members in developing your data collection training. Community members can help elevate concerns evaluation participants may have, highlight important cultural practices to build trust and show respect, and give other relevant advice.
- Hire members of your evaluation’s target community to be data collectors. Evaluation participants may be more likely to consent to evaluation data collection and provide more thorough or accurate information if they feel they are speaking with someone who understands and values them. Community data collectors are building important evaluation skills and can use your employment as job and education references.

- Consider compensating community members for their valuable time spent participating in evaluation planning and the evaluation participants' effort in providing data. If you can't provide financial remuneration, think about covering transportation costs, providing food and meals, and asking community members what other contributions they would find valuable.

To learn more ...

- [Best Practices in Creating and Adapting Quality Rating and Improvement System \(QRIS\) Rating Scales](#) (Burchanai, Tarullo, & Zaslow, 2016)
- [Enhancing Rigor, Relevance, and Equity in Research and Evaluation Through Community Engagement](#) (OPRE, 2021)
- [Supporting the Use of Administrative Data in Early Care and Education Research: Resource Series](#) (OPRE, 2019)
- [Types of Data Used for Impact Evaluation](#) (Courtney, 2021)

References

- All of Us Research Program. (n.d.). *Achieving the principles through a precision medicine initiative data security policy framework*. U.S. Department of Health and Human Services, National Institutes of Health. <https://allofus.nih.gov/protecting-data-and-privacy/precision-medicine-initiative-data-security-policy-principles-and-framework-overview/achieving-principles-through-precision-medicine-initiative-data-security-policy-framework>
- Allen, M. (2017). *Secondary data*. SAGE Encyclopedia of Communication Research Methods. <https://dx.doi.org/10.4135/9781483381411.n557>
- Blocklin, M., Hyra, A., Kean, E., & Porowski, A. (2019). *Building capacity to evaluate child welfare community collaborations to strengthen and preserve families (CWCC) grantee local evaluation plan and implementation plan templates*. <https://omb.report/icr/201906-0970-001/doc/98252801.pdf>
- Burchinal, M., Tarullo, L. & Zaslow, M. (2016). *Best practices in creating and adapting Quality Rating and Improvement System (QRIS) rating scales* (OPRE Research Brief 2016-25). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation. https://www.acf.hhs.gov/sites/default/files/documents/opre/ccepra_qris_531_508compliant_66_b508.pdf
- Carter, N., Bryant-Lukosius, D., DiCenso, A., Blythe, J., & Neville, A. J. (2014). The use of triangulation in qualitative research. *Oncology Nursing Forum*, 41(5), 545–547. <https://pubmed.ncbi.nlm.nih.gov/25158659/#:~:text=Triangulation%20refers%20to%20the%20use,of%20information%20from%20different%20sources>
- CASEL. (2020). *CASEL's SEL framework*. <https://casel.org/casel-sel-framework-11-2020/>
- Courtney, M. (2021). *Types of data used for impact evaluation*. U.S. Department of Health and Human Services, Children's Bureau. <https://www.acf.hhs.gov/sites/default/files/documents/opre/Types%20of%20Data%20Used%20for%20Impact%20Evaluation-oct-2021.pdf>
- Felitti, V. J., Anda, R. F., Nordenberg, D., Edwards, V., Koss, M., & Marks, J. S. (1998). Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults. *American Journal of Preventive Medicine*, 14(4), P245–258. [https://www.ajpmonline.org/article/S0749-3797\(98\)00017-8/fulltext](https://www.ajpmonline.org/article/S0749-3797(98)00017-8/fulltext)
- Ferreira, J. C., & Patino, C. M. (2017). *Types of outcomes in clinical research*. National Library of Medicine. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5790671/>
- Jacob, S. A., & Furgerson, S. P. (2012). Writing interview protocols and conducting interviews: Tips for students new to the field of qualitative research. *The Qualitative Report*, 17, 1–10. <https://files.eric.ed.gov/fulltext/EJ990034.pdf>
- Krueger, R. (2002). *Designing and conducting focus group interviews*. <https://www.eiu.edu/ihec/Krueger-FocusGroupInterviews.pdf>

- National Congress of American Indians Policy Research Center. (2009). *Research that benefits Native People: A guide for Tribal Leaders*. <https://www.ncai.org/policy-research-center/research-data/NCAIModule1.pdf>
- OPRE (Office of Planning, Research, and Evaluation). (n.d.). *OPRE publications, administrative data, and research*. U.S. Department of Health and Human Services, Administration for Children and Families. <https://www.acf.hhs.gov/opre/project/opre-publications-administrative-data-and-research>
- OPRE. (2019). *Supporting the use of administrative data in early care and education research: Resource series*. U.S. Department of Health and Human Services, Administration for Children and Families. <https://www.acf.hhs.gov/opre/report/supporting-use-administrative-data-early-care-and-education-research-resource-series>
- OPRE. (2021). *Enhancing rigor, relevance, and equity in research and evaluation through community engagement*. U.S. Department of Health and Human Services, Administration for Children and Families. <https://opremethodsmeeting.org/meetings/2021/>
- Price, P. C., Jhangiani, R. S., Chiang, I-C. A., Leighton, D. C., & Cuttler, C. (2017). *Research methods in psychology* (3rd ed). <https://opentext.wsu.edu/carriecuttler/>
- Ryan, K., Gannon-Slater, N., & Culbertson, M. J. (2012). Improving survey methods with cognitive interviews in small- and medium-scale evaluations. *American Journal of Evaluation*, 33(3) 414–430. <https://journals.sagepub.com/doi/full/10.1177/1098214012441499>
- Salkind, N. (2010). *Primary data source*. Encyclopedia of Research Design. <https://dx.doi.org/10.4135/9781412961288.n333>
- Samonas, S., & Coss, D. (2014). The CIA strikes back: Redefining confidentiality, integrity and availability in security. *Journal of Information System Security*, 10(3). <http://www.proso.com/dl/Samonas.pdf>
- Shivayogi, P. (2013). Vulnerable population and methods for their safeguard. *National Library of Medicine*, 4(1), 53–57. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3601707/>
- Squires, J., & Bricker, D. (2009). *Ages & stages questionnaires* (3rd ed.) <https://products.brookespublishing.com/Ages-Stages-Questionnaires-Third-Edition-ASQ-3-P569.aspx>
- Tamburri, D. (2020). Design principles for the General Data Protection Regulation (GDPR): A formal concept analysis and its evaluation. *Information Systems*, 91. <https://doi.org/10.1016/j.is.2019.101469>
- Wilson, C. (2021). *Food security status and well-being of Nutrition Assistance Program (NAP) participants in Puerto Rico: Revised pretest findings and recommendations* [Insight Policy Research internal memo]. September 21.