

Healthy Marriage and Relationship Education Program Evaluation Toolkit

August 2024



Introduction

The Administration for Children and Families (ACF) provides grants to fund healthy marriage and relationship education (HMRE) programs to strengthen and improve the quality of relationships. The programs offer a range of services from relationship education for high school students to marriage and relationship skills building for adult couples. Grant recipients may be funded to also conduct descriptive or impact evaluations of their funded programs. Independent local evaluators support grant recipients in conducting their local evaluations. ACF is committed to ensuring these program evaluations provide information that is useful and meaningful for HMRE programs, ACF, and the field at large. To accomplish this, ACF provides standards and guidance to help local evaluators develop and implement rigorous evaluations that produce quality evidence. These standards and guidance are developed to help grant recipients and evaluators conduct high-quality descriptive and impact evaluations to answer a range of research questions.

Mathematica developed this toolkit to help HMRE local evaluators understand key program evaluation concepts, common evaluation challenges, and strategies to prevent or overcome challenges. The toolkit consists of eight briefs. The briefs are standalone documents that can be read in any order. It is important to note that the briefs provide general information and tips and are not intended to be a comprehensive guide to program evaluation or the topic. The briefs use examples that are tailored to the populations served by HMRE programs—namely, youth, adult individuals, and adult couples. The briefs are focused on the following topics:

1. Enhancing the informed consent process and enrolling participants in a program evaluation.

This brief describes the importance of obtaining informed consent from participants and contains five tips that HMRE program staff and evaluators can use to improve procedures for obtaining informed consent and enrolling participants in evaluations.

2. Improving survey response rates. This brief describes the importance of maximizing the number of participants that respond to surveys and includes six tips that HMRE evaluators can use to improve their survey response rates.

3. Monitoring program dosage. This brief introduces the concept of dosage and contains four tips that HMRE program staff and evaluators can use to collect information about dosage and address program retention challenges that might affect dosage.

4. Understanding and mitigating attrition. This brief describes (1) what attrition is and why it can be a problem; (2) how to calculate and assess it; and (3) steps evaluators can take to mitigate it.

5. Using mixed methods. This brief explains why and when programs and evaluators might use mixed methods research in an HMRE evaluation, followed by a description of three commonly implemented designs. The brief closes with questions HMRE programs and evaluators can consider when choosing an approach.

6. Cleaning and preparing data for analysis. This brief describes how data errors can occur throughout an HMRE evaluation and provides four tips for how to avoid them.

7. Creating equivalent research groups. This brief contains information on: (1) how study design choice affects group equivalence; (2) how to regularly monitor equivalence throughout the course of the study; and (3) based on attrition or design, how to establish baseline equivalence for the final analytic sample before conducting analysis.

8. Powering an evaluation to detect effects. This brief first describes common errors in statistical testing, then recommends five steps to help evaluators conduct a power analysis and calculate effect sizes they can use throughout an evaluation.

Table of Contents

Introduction.....2

1. Enhancing the informed consent process and enrolling participants
in a program evaluation.....5

2. Improving survey response rates 9

3. Monitoring program dosage 15

4. Understanding and mitigating attrition..... 21

5. Using mixed methods..... 26

6. Cleaning and preparing data for analysis 31

7. Creating equivalent research groups..... 36

8. Powering an evaluation to detect effects.....41

Suggested citation

Friend, Daniel, Rebecca Piatt, Avery Hennigar, Armando Yañez, Brandon Hollie, Sonia Alves, and Angela Valdovinos D'Angelo. (2024). "Healthy Marriage and Relationship Education Program Evaluation Toolkit." OPRE Report #2024-138. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Acknowledgements

Many people contributed to this toolkit. First, we acknowledge staff at the OPRE in the Administration for Children and Families at the U.S. Department of Health and Human Services. We are particularly grateful for the direction and feedback from Samantha Illangasekare, Rebecca Hjelm, and Kathleen McCoy. We want to extend our gratitude to Sarah Avellar and Angela D'Angelo for reviewing drafts of the briefs. We also extend our appreciation to Effie Metropoulos and Bridget Gutierrez for editing and Yvonne Marki-Korosec and Gwyneth Olson for designing the graphics in this toolkit.

Acronyms

ACF	Administration for Children and Families
ETAP	Evaluation Technical Assistance Partner
FPS	Family Assistance Program Specialist
HMRE	Healthy Marriage and Relationship Education
HMRF	Healthy Marriage and Responsible Fatherhood
IRB	Institutional Review Board
nFORM	Information, Family Outcomes, Reporting, and Management system
OFA	Office of Family Assistance
QED	Quasi-experimental design
RCT	Randomized controlled trial
TA	Technical assistance
WWC	What Works Clearinghouse

Enhancing the informed consent process and enrolling participants in a program evaluation

Rebecca Piatt, Avery Hennigar, Sonia Alves, and Daniel Friend

The success of a Healthy Marriage and Relationship Education (HMRE) program evaluation begins with clearly describing the evaluation to potential participants, obtaining their informed consent, and enrolling them in the study (Box 1.1). Describing the study accurately—including its purpose, procedures, risks, and benefits—and obtaining informed consent is both a legal and ethical requirement.¹ However, research and evaluation concepts and terminology may be unfamiliar to potential participants, participants may have had negative experiences with research, and/or participants may not fully understand what they are signing up for. HMRE program staff and evaluators' early interactions with potential participants are important for addressing these issues and fully informing and encouraging them to enroll in the study. This brief contains five tips that HMRE program staff and local evaluators can use to improve procedures for obtaining informed consent and enrolling participants in evaluations.

Box 1.1. Key terms

Informed consent. The process in which staff explain to potential participants the purpose, risks, benefits, and voluntary nature of the evaluation. Staff must properly inform participants of all possible risks and benefits so that they can make a voluntary decision about whether to participate in the evaluation. If the evaluation involves youth, a parent's or guardian's consent needs to be obtained first, followed by the youth's assent, or agreement, to participate (Shah et al. 2023).

Enrollment. Enrolled participants are those who are eligible to participate (they meet the study's inclusion criteria) and who consent to participate. Enrollment in an evaluation means the participant plans to participate in the program and the evaluation activities for their study group (for example, by attending workshops or participating in data collection activities), as appropriate (Bruursema 2015).



Tip 1: Develop a consent form that is concise and easy to understand

A consent form is a document that briefly describes the evaluation's purpose and procedures as well as the risks and benefits of participation. An effective consent form presents the potential participant with enough information to understand the evaluation and to make an informed decision about whether to participate.

Although consent forms must be thorough (Box 1.2), they should also be brief and easy to understand (Grant 2021; Kadam 2017). HMRE evaluators should distill the form to the essential elements that a potential participant would need to make an informed decision. Keep sentences and paragraphs on the form concise—for example, less than seven lines per paragraph and less than 12 words per sentence.

¹ HMRE evaluators must obtain approval from their local Institutional Review Board (IRB) on all consent procedures and forms.

Box 1.2. Common elements of a consent form

- Purpose of the evaluation
- Eligibility criteria
- Description of the evaluation procedures and expected duration of participation
- Benefits and risk of participating
- How confidentiality will be protected and limits to those protections, if any
- Data protection procedures
- Compensation, if any
- A statement that participation is voluntary
- Evaluation team contact information
- Any specific IRB requirements

Source: U.S. Department of Health and Human Services (2016).

Consider using bulleted lists rather than paragraphs to keep information brief and consumable. Providing succinct information and using accessible language will help potential participants understand the evaluation and their role in it.

Developing the right language for participant audiences may be challenging when describing more complicated research and evaluation processes and procedures. It may be hard to avoid using certain evaluation terms that participants might not be familiar with—for example, randomization, intervention, or comparison group. In general, consent forms should avoid technical terminology when possible or provide definitions and examples when introducing new or complicated concepts. After crafting the consent form, consider obtaining feedback on its overall readability and comprehension from program participants not enrolled in the evaluation or from other community members. Including input from community members can increase enrollment of underrepresented groups in research studies (Markman et al. 2023).

In addition, HMRE programs and local evaluators should think carefully about the participants they want to enroll in the evaluation and tailor the form to them. Identify key

characteristics that might influence a participant's ability to understand the consent form, such as education level and primary language. For example, if participants generally have lower educational attainment, make sure the consent form is tailored to a sixth- to eighth-grade reading level using Microsoft Word's accessibility functions. Or, if the HMRE program is primarily recruiting from a community with high rates of Spanish speakers, make sure the Spanish-language version of the consent form is translated to fit the words and phrases used by the local population by having multiple team members or other native Spanish speakers review it.



Tip 2: Assess intake procedures to incorporate evaluation consent and enrollment

Because HMRE programs typically have a centralized intake process for enrolling participants, the processes of engaging participants in the program and in the evaluation can be linked. Typically, programs recruit people who are eligible for and interested in program services, then they obtain consent for participating in an evaluation of the program (Avellar et al. 2017). HMRE program staff and local evaluators should work together to identify the best point in the intake process for evaluation consent and enrollment to occur (Box 1.3).

Box 1.3. Methods for assessing existing procedures

- Use customer journey mapping (also called [experience diagramming](https://www.luma-institute.com/experience-diagramming/))² to map a participant's journey through the intake and enrollment processes. Illustrating their experiences may help reveal existing processes or steps to simplify or eliminate.
- Interview past participants about their experiences with consent and enrollment to determine areas that could be streamlined and forms that could be combined.

² Luma Institute. "Experience Diagramming." <https://www.luma-institute.com/experience-diagramming/>

Ideally, program and evaluation consent and enrollment processes should take place simultaneously for all potential participants (Feeny 2021). Integrating program and evaluation consent processes will allow potential participants to hear about the HMRE program; learn about the evaluation; provide consent (if they choose); and complete program intake forms and evaluation data collection (for example, a baseline survey) at the same time. Combining program and evaluation enrollment into a single step also lessens the need to follow up with participants to obtain consent and complete baseline data collection after program enrollment.



Tip 3: Identify who will consent and enroll participants

Although the consent form is an essential aspect of the process, obtaining informed consent is more than just completing a written form. Informed consent is a process that involves conversing with potential participants and answering their questions about the evaluation (Markman et al. 2023). Staff in charge of obtaining informed consent and enrolling participants in the study should feel comfortable explaining all aspects of the evaluation using plain language and should be able to answer any questions (Kadam 2017). Staff who will consent and enroll participants should also be able to connect with participants on a personal level. For example, they should demonstrate sensitivity and familiarity working with the identified community (Feeny 2021). Developing rapport with participants who have had adverse experiences with evaluations is especially important (Box 1.4).

Box 1.4. Understand participants' experiences with research

Understanding the research experiences of potential participants, especially those from traditionally excluded groups such as undocumented immigrants, people of color, and families with low incomes, is critical to tailoring the consent and enrollment procedures to their needs (Doucet 2021). Introducing the study is an excellent opportunity to build rapport with participants and to discuss their past experiences with research. For example, while describing details of the evaluation, ask whether participants have had previous experiences with research and whether they were positive or negative. If they were negative, ask open-ended questions to understand what contributed to the negative experiences and describe how this evaluation would differ. If they were positive, describe how the current evaluation would replicate those positive experiences, as appropriate.

When selecting who will serve in this role, it is best practice to have the same person both consent and enroll a participant in an evaluation.³ Involving multiple staff in different parts of this process can introduce systematic differences between participants and could affect who consents to the evaluation, which could potentially affect the external validity of the study (Zief et al. 2011). HMRE evaluators should monitor the consent rates for each staff person to identify those who are particularly successful or unsuccessful at consenting and enrolling participants and adjust accordingly.



Tip 4: Equip staff with proper training and tools

Staff members who will consent and enroll participants will need the right training and tools. Training could focus on how to communicate important information about the evaluation accurately and empathetically. By the end of training, these staff should feel confident describing the purpose, procedures, risks, and benefits of the evaluation to potential participants (Feeny 2021). Training should also prepare staff to answer potential participants' questions and address their concerns, for example, by role-playing an informed consent process where a participant is concerned about the study maintaining their privacy.

³ For larger evaluations, multiple staff may serve this role, or one staff person may consent and enroll multiple participants in a group setting.

HMRE local evaluators may consider including community members, particularly from underrepresented populations, in trainings to provide context about the community and insight into local preferences and standards. For example, if the HMRE program is working with fathers of color, staff could invite fathers from the community, community leaders, or staff from other community organizations that serve fathers to discuss how potential participants might perceive participating in research and the group's historical experience with research, which might influence their perceptions of the evaluation. Similarly, if an HMRE program is working in a school, evaluators may want to seek input from members of the parent-teacher association or from the principal on how to best communicate the evaluation to students, parents, and teachers in their school district.



Tip 5: Keep track of consent rates to adjust strategies as needed.

HMRE local evaluators should track who is and is not consenting to participate in the evaluation and use that information to improve their strategies for consenting and enrolling participants (Zief et al. 2011). Evaluators can track which staff have particularly high consent rates and observe what they are doing to help other staff improve. If the HMRE program has previously conducted evaluations of the program it should use, use existing data to help inform the current approach and set realistic benchmarks for enrollment (Avellar et al. 2017). Furthermore, HMRE local evaluators can analyze demographic information early in the evaluation to understand which populations they are unable to reach, and refine their consenting approach accordingly. If youth participants aren't returning parental consent forms, for instance, staff can consider providing incentives or reminders through multiple formats (text, email, and so on).

References

- Avellar, S., K. Borradaile, and D. Strong. "Tips for Enrolling and Retaining Evaluation Participants." Evaluation Technical Assistance Brief No. 4. Washington, DC: Mathematica, November 2017. <https://www.mathematica.org/publications/tips-for-enrolling-and-retaining-evaluation-participants-evaluation-technical-assistance-brief>.
- Bruursema, T. "Recommendations for Successfully Recruiting and Retaining School Participation in a Teen Pregnancy Prevention Impact Evaluation." Evaluation Technical Assistance Brief No. 9. U.S. Department of Health and Human Services, Office of Adolescent Health, June 2015. <https://opa.hhs.gov/sites/default/files/2020-07/tppeval-tabrief9.pdf>. Accessed July 7, 2023.
- Doucet, F. "Identifying and Testing Strategies to Improve the Use of Antiracist Research Evidence Through Critical Race Lenses." New York, NY: William T. Grant Foundation, 2021. https://wtgrantfoundation.org/wp-content/uploads/2021/01/Doucet_Digest_Issue-6.pdf. Accessed July 7, 2023.
- Feeny, L. "Define Intake and Enrollment Process." Cambridge, MA: Abdul Latif Jameel Poverty Action Lab, March 2021. <https://www.povertyactionlab.org/resource/define-intake-and-consent-process>.
- Grant, S.C. "Informed Consent—We Can and Should Do Better." *JAMA Network Open*, vol. 4, no. 4, April 2021, pp. 1–3.
- Kadam, R.A. "Informed Consent Process: A Step Further Towards Making It Meaningful!" *Perspectives in Clinical Research*, vol. 8, no. 3, July–September 2017, pp. 107–112.
- Markman, K.M., N.P. Weicker, A.K. Klein, and R. Sege. "Community-Engaged Training in Informed Consent." *Journal of Clinical and Translational Science*, vol. 7, no. 1, April 2023. doi:10.1017/cts.2023.534.
- Shah, P., I. Thornton, D. Turrin, and J.E. Hipkind. "Informed Consent." Treasure Island, FL: StatPearls Publishing, June 2023. <https://www.ncbi.nlm.nih.gov/books/NBK430827/>.
- U.S. Department of Health and Human Services. "Informed Consent Checklist (1998)." March 2016. <https://www.hhs.gov/ohrp/regulations-and-policy/guidance/checklists/index.html>. Accessed July 7, 2023.
- Zief, S., J. Henke, J. Deke, R. Cole, and M. Thomas. "Frequently Asked Questions: Evaluation Start-Up." Evaluation Technical Assistance Update No. 2. U.S. Department of Health and Human Services, Office of Adolescent Health, July 2011. <https://opa.hhs.gov/sites/default/files/2020-07/tpp-update2-july2011.pdf>. Accessed July 7, 2023.

Improving survey response rates

Rebecca Piatt, Daniel Friend, Avery Hennigar, and Angela Valdovinos D'Angelo

A survey is a useful way to hear directly from participants in Healthy Marriage and Relationship Education (HMRE) programs. Thus, surveys are a common method to collect data in HMRE evaluations. A survey response rate is the percentage of participants in a sample (out of all participants in the sample) who take a particular survey. Because missing survey responses affect what HMRE programs can learn from an evaluation (Box 2.1), it is important to try to maximize the number of participants who respond to each survey. This brief includes six tips that HMRE local evaluators can use to improve their survey response rates.

Box 2.1. Downsides of low response rates

Missing survey responses mean missing out on learning about participants' characteristics, experiences, and outcomes. Low response rates can also lead to misleading results, known as nonresponse bias. Nonresponse bias can happen when people who are unable or unwilling to fill out surveys systematically differ from those who complete them. When nonresponse bias occurs, results might be higher or lower than they should be.

Source: American Association for Public Opinion Research (2023); Czajka and Beyler (2016).



Tip 1: Assess the root causes of low survey response rates and develop tailored strategies to overcome them

Participants may not respond to surveys for many reasons, including their perceptions of surveys (for example, not understanding or not being adequately informed of the value of research, not being interested in a survey completion incentive, or concerns about privacy) and the survey's characteristics (such as mode or length). Programs can also experience challenges with tracking, locating, and contacting participants (Dillman et al. 2014; Sheehan 2001; Singer and Ye 2013; de Leeuw et al. 2008).

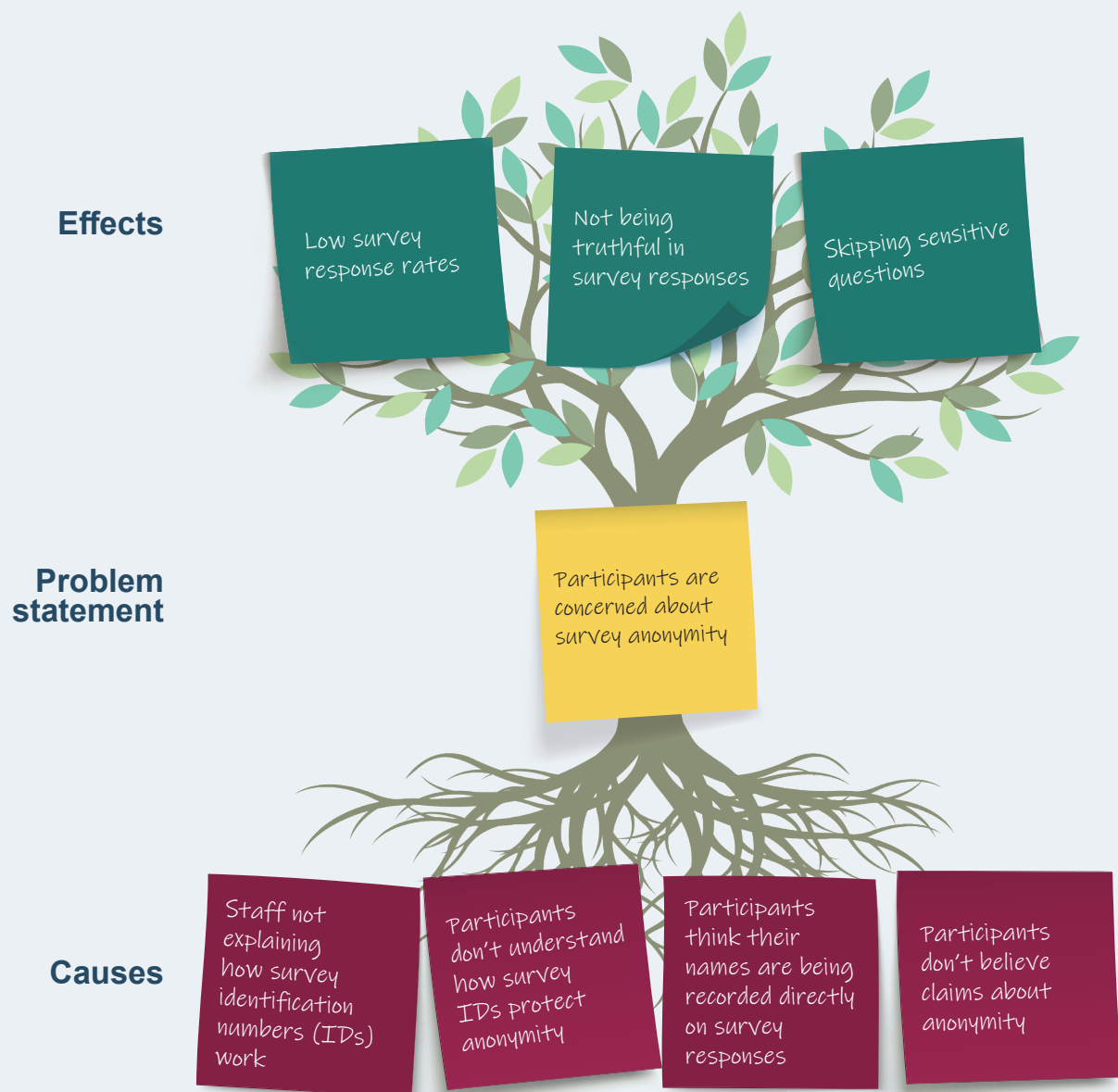
Examining why participants are not completing surveys can help evaluators select strategies to boost response rates. To do so, evaluators can use a combination of staff knowledge, participant feedback, and data to figure out why response rates are low. For instance, evaluators can involve program staff and participants in brainstorming about the potential root cause or causes of the issue (Box 2.2). In addition, evaluators could consider conducting the following exploratory activities:

- Asking program staff and participants if they understand the purpose and importance of completing surveys
- Asking participants if they are hesitant to complete surveys because of privacy concerns
- Using baseline data to explore whether certain groups or types of participants are completing surveys at different rates
- Using existing data to determine the percentage of participants who don't respond to surveys and cannot be located by program or evaluation staff

Box 2.2. Unpack why response rates are low

HMRE local evaluators can consider setting aside time in a regular staff meeting to reflect on the challenge of low response rates. One method evaluators may try is using a [problem tree analysis](#)¹ to unpack why response rates are low and to understand the causes and effects of the problem in a systematic way. During the activity, staff use a tree template to map causes (on the roots) and effects (on the branches) of a problem to better understand it. A problem tree analysis helps untangle complex problems, reveals their various causes and effects, builds a shared understanding of the problem, and provides direction for solving it (Luma Institute n.d.).

For example, consider that staff hear from participants that their main concern with completing surveys is that their name will be tied to their responses. When asked more about this challenge, participants voice that they are hesitant to complete surveys because they are concerned about their personal data use being reported to the federal government. To unpack this challenge, staff complete the problem tree activity below.



¹ ODI. "Planning tools: Problem Tree Analysis." 2009. <https://odi.org/en/publications/planning-tools-problem-tree-analysis/>

After determining some reasons why response rates may be low, program staff and evaluators can rank them and brainstorm potential solutions to test. To generate solutions, evaluators can ask program staff and participants for their ideas about how to solve the top response rate challenge. For instance, if the main challenge is locating participants for a follow-up survey, staff might suggest sending out an address reminder card for participants to indicate if their address changes during the evaluation. Or, if the main challenge is that participants are not completing surveys due to privacy concerns, staff might explain measures to protect participant anonymity. Evaluators and program staff can participate in group brainstorming activities, such as a [creative matrix](#)², to invite everyone in the group to identify new and innovative ideas for addressing the challenge(s) (Box 2.3). After selecting one promising strategy, evaluators can work with program staff to run a small pilot—for example, with one cohort of participants—to see whether employing the selected strategy increases response rates.

Box 2.3. Brainstorm ideas for addressing the challenge

After completing the problem tree analysis activity, evaluators and program staff brainstormed solutions to the identified problem using a creative matrix. To unpack solutions to the challenge, staff came up with the following ideas (written in the sticky notes) for how to help participants understand that their survey responses are anonymous. After brainstorming five potential ideas, staff decide to first try creating a script to explain to participants how survey identification numbers protect their anonymity.

Solution categories	How might we help participants understand that their survey responses are anonymous?	
...with technology?	Add text at the top of the survey explaining how surveys are anonymous	Create a lecture slide explaining privacy protection in surveys
...with people?	Create a script for staff to use to explain how survey IDs protect anonymity	Train staff to field privacy questions
...with partnerships?	Set up meetings for evaluators and program staff to review progress after trying solution(s)	



Tip 2: Clearly explain the importance of surveys to increase willingness to respond

People are more likely to complete a survey when they understand the importance of their participation (Fowler 1995). Before administering a survey, staff should clearly describe why they are asking participants to complete it (Centers for Disease Control and Prevention [CDC] 2018). HMRE local evaluators can develop a brief script for survey administrators to use when distributing surveys to participants. This script might include a description of the rationale behind the evaluation including the overarching goals and purpose—why it is critical to hear participants' perspectives, and how responses may improve the program for others in the future.

² Maad Labs. "Creative Matrix." <https://www.maadlabs.io/strategic-design-kit/creative-matrix>.

Evaluators should consider tailoring language in the script to the people and communities served by the program. For example, evaluators can consider providing examples of how they adapted or expanded the program based on past participant feedback. If the program experiences high rates of absences or dropouts, staff administering the survey can remind participants that the HMRE program plans to collect survey data from participants even if they leave the program. If the program is serving youth in school, staff can differentiate the current evaluation surveys from other tests, exams, or assessments that students are used to.



Tip 3: Consider when and how to administer surveys to increase responses

Evaluators should think about administering surveys when it's most convenient for participants to complete them (for example, during workshop sessions) (de Leeuw et al. 2008). Responses to follow-up surveys (for example, a six-month follow-up) are typically harder to obtain than surveys administered at enrollment because of participant absences, dropouts, or difficulty locating participants once the program has ended. However, it's important to obtain survey responses from all participants—even those who don't complete the program—to make sure their experiences are captured in the data. Evaluators might work with program staff to host an in-person event for participants who have not completed their follow-up surveys where evaluators could gather responses—such as a dinner; additional workshop or service; or a family night, date night, or program alumni night. Administering follow-up surveys when many participants are together eliminates the need to follow up and locate them outside of the program.

Also, evaluators should think about ways to bundle surveys together so that participants can complete them in one sitting—for example, offer the required exit and local evaluation surveys at the same time.³ The brief on consent and enrollment discusses methods for assessing existing procedures to determine how to bundle and streamline surveys.



Tip 4: Obtain multiple modes of contact information to track and contact participants

At the beginning of the program, evaluators should request participants' contact information and permission to use different methods to contact them (Box 2.4). Having multiple ways of contacting participants increases HMRE local evaluators' chances of reaching them (CDC 2018). Staff obtaining contact information for participants should ask their preference for outreach—for example, text, direct message on social media, phone call, and so on. A participant might not respond to emails but may reply to other forms of outreach such as texts. Evaluators should tailor their outreach strategy to the people and communities the HMRE program serves.

In addition to participants' contact information, evaluators should collect information for an alternative contact—for example, a friend, mother, or neighbor. Staff can ask for permission to contact a participant's alternative contact in case they cannot reach the participant. Also, evaluators may consider assessing the stability of a participant's primary and secondary contact information by asking if they plan to move or switch any of their contact information in the next year.

Box 2.4. Examples of contact information to collect

- Mailing address
- Multiple phone numbers (cell, work, and so on)
- Email address
- Social media accounts and permission to contact (for example, Facebook, Instagram, Snapchat)
- Alternative contact information (for example, mother's or father's cell phone number)

³ HMRE local evaluators should be mindful of the length of time participants spend completing surveys—especially when bundling them. Consider including page breaks in surveys to help participants pace themselves while taking the survey. Make sure to select a survey platform that autosaves responses, in case participants exit the survey early.



Tip 5: Offer participants several ways to take surveys

Using multiple modes for local evaluation surveys can increase response rates, reduce bias, and improve the quality of the data gathered (Bethlehem 2010; Dillman et al. 2014). When reaching out to participants who haven't completed a local evaluation survey, evaluators should offer different modes for them to do it—for example, over the phone, in person, via the web, or by accessing the survey through a link in a text message. Providing multiple ways for participants to complete surveys recognizes that people have preferences for or access to different modes of communication (Box 2.5).

Box 2.5. Offer flexibility in local evaluation survey modes for a variety of populations

Below are some considerations for tailoring survey mode to the different populations served by HMRE programs:

- **People that live in rural areas.** Completing a web-based survey may be challenging for participants in rural areas due to a lack of internet connectivity. Offer a paper- or phone-based survey instead.
- **People with low incomes.** Completing surveys over the phone may be challenging if participants have limited call minutes or if their phone is temporarily out of service. Offer a paper- or web-based survey to these participants.
- **Couples.** Privacy while discussing sensitive topics over the phone may be a concern for people who live with others—for instance, couples that participate in an HMRE program together may not want to answer questions about their relationship together. Consider giving these participants the option to complete a web- or paper-based survey separately.
- **Youth in juvenile detention facilities, foster homes, or homeless shelters.** Participants in these situations have unique needs and may not have access to technology. Consider visiting these places to administer surveys in person, working closely with facility staff to administer surveys, or meeting participants at a neighborhood restaurant or library.
- **People with special needs, disabilities, or visual impairments.** Assess which format participants prefer based on their individual needs, such as a phone survey rather than a paper survey. Tailor the survey mode accordingly.

Note: Grant recipients that use nFORM surveys in their local evaluations should follow guidance given on administering that surveys. Evaluators should check with their FPS before making any survey mode changes.

ACF's [data collection and logistics manual](#)⁴ provides best practices for collecting performance measure data in the Information, Family Outcomes, Reporting, and Management system (nFORM 2.0). If a program's local evaluation uses nFORM surveys (for example, the Applicant Characteristics, Entrance, and Exit Surveys), reference the manual to create or revise the administration of these surveys. Evaluators can also consult [the manual for using nFORM for local evaluations](#),⁵ which includes information on how to grant local evaluators access to data. ACF also recommends that HMRE programs and evaluators consult their Family Assistance Program Specialist (FPS) for all data collection scenarios.



Tip 6: Use motivating and well-timed survey completion incentives

To be effective, the incentives for completing a survey should be enticing to participants (Singer and Ye 2013). HMRE programs and local evaluators should tailor incentives to something that will motivate participants and that they will be excited to receive. Evaluators can gather data on participants' preferences—for example, allow them to vote on which type of gift card they find most appealing. Incentives might differ based on cultural preferences, popular stores in the area, or age groups of the populations served. For example, using a gift card to a local store may be enticing for one population (like adults) but not for another (like youth), who may prefer a cash gift card.

⁴ Healthy Marriage and Responsible Fatherhood Resource Site for 2020 Grantees. "nFORM Performance Measures and Data Collection Logistics Manual." Administration for Children and Families, March 2023. https://hmrfggrantresources.info/sites/default/files/2023-03/PMDC%20Logistics_Mar2023.pdf.

⁵ 2020 Healthy Marriage and Responsible Fatherhood Resource Site for 2020 Grantees. "Using nFORM 2.0 for Local Evaluations." Administration for Children and Families, May 2021. https://www.hmrfggrantresources.info/sites/default/files/2022-04/nFORM_2.0_Evaluation_Guidance.pdf.

Evaluators can consider using higher incentive amounts for follow-up surveys than for entrance or exit surveys. Lower incentive amounts are usually fine for baseline surveys or surveys administered during enrollment because participants are more motivated at the beginning of services. Higher amounts may be needed to encourage responses six to 12 months after programs end because participants have likely ended their involvement with the program. Evaluators can work with program leaders to tie incentive to evaluation milestones that align with the HMRE program—for example, tying incentives to attending all workshop sessions *plus* completing the local evaluation survey.⁶

Incentives don't only need to be for participants. Evaluators might also offer incentives or verbal congratulations to staff who consistently obtain high response rates. If evaluators cannot link response rates to individual staff, they might consider offering group incentives, such as a staff lunch, if the group achieves an overall high response rate in a particular month or quarter. Acknowledging program and evaluation staff reinforces how important they are to achieving high response rates and gaining valuable feedback from participants.

References

- American Association for Public Opinion Research. "Best Practices for Survey Research." 2023. <https://aapor.org/standards-and-ethics/best-practices/#166811232459-8b1678f0-d862>. Accessed June 20, 2023.
- Bethlehem, J. (2010). "Selection Bias in Web Surveys." *International Statistical Review*, vol. 78, no. 2, pp. 161–188.
- Centers for Disease Control and Prevention (CDC). "Data Collection Methods for Program Evaluation: Questionnaires." Evaluation Brief No. 14. U.S. Department of Health and Human Services, November 19, 2018. <https://www.cdc.gov/healthyyouth/evaluation/pdf/brief14.pdf>. Accessed June 20, 2023.
- Czajka, J., and A. Beyler. "Background Paper: Declining Response Rates in Federal Surveys, Trends and Implications." Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services, June 15, 2016.
- de Leeuw, Edith D., Joop J. Hox, and Don A. Dillman. *International Handbook of Survey Methodology*. Routledge, 2008, pp. 221–238.
- Dillman, Don A., Jolene D. Smyth, and Leah Melani Christian. *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. John Wiley & Sons, 2014.
- Fowler, F.J. "Improving Survey Questions: Design and Evaluation." Sage, 1995.
- Healthy Marriage and Responsible Fatherhood. "Healthy Marriage and Responsible Fatherhood Grants; Information, Family Outcomes, Reporting, and Management (nFORM 2.0) System User Manual." March 2023. https://hmrfgresources.info/sites/default/files/2023-03/nFORM_UserManual_March2023.pdf. Accessed July 7th, 2023.
- Luma Institute. "Problem Tree Analysis." n.d. <https://www.luma-institute.com/problem-tree-analysis/>.
- Sheehan, K.B. "E-Mail Survey Response Rates: A Review." *Journal of Computer-Mediated Communication*, vol. 6, no. 2, 2001, <https://doi.org/10.1111/j.1083-6101.2001.tb00117.x>.
- Singer, E., and C. Ye. "The Use and Effects of Incentives in Surveys." *The ANNALS of the American Academy of Political and Social Science*, vol. 645, no. 1, 2013, pp. 112–141.

⁶ HMRE programs must stay within OFA's guidelines on maximum allowable incentives. Programs can contact their FPS for more information.

Monitoring program dosage

Avery Hennigar, Armando Yañez, Daniel Friend, and Rebecca Piatt

Like the comparable definition in medicine, a participant's "dosage" of a program refers to how much of a program they receive or participate in. Dosage helps Healthy Marriage and Relationship Education (HMRE) program staff and local evaluators understand and track how much of an intervention is delivered and to whom. Examining dosage can also reveal how well an intervention is being implemented and can help uncover key challenges programs faced while implementing it (Bronte-Tinkew et al. 2007; James Bell Associates 2008; James Bell Associates 2018; Loeb et al. 2017). Assessing dosage is important for both descriptive and impact evaluations, because participant outcomes can differ based on the amount or type of programming or services they receive (Zhai et al. 2010).

Dosage and program retention (that is, sustained participation in program services) are closely linked. For federal HMRE programs, dosage is typically captured by recording (1) the number of workshop sessions participants attend or (2) their level of participation in supplemental program components like financial literacy workshops or case management (Box 3.1). Although HMRE programs work to maximize the amount of the program content participants receive, which is one important factor in maintaining fidelity to the program curriculum, there is often a range in the amount of programming different participants receive. This brief contains four tips that HMRE program staff and evaluators can use to collect information about dosage and address program retention challenges that might affect dosage.

Box 3.1. Measuring dosage

The goal of measuring dosage is to create a quantitative metric that shows the amount of programming participants receive. Dosage can be measured in different ways. Common metrics include:

- The number of hours of a service participants received
- The percentage of workshop sessions participants attended
- A binary indicator of whether or not a participant "completed" the program

Note: Program completion can be defined in various ways. For HMRE programs funded by the Administration for Children and Families, participants who attended at least 90 percent of the program's primary workshops are considered as having completed the program.



Tip 1: Collect and monitor data on dosage

HMRE programs are required to track clients' participation in workshops in the Information, Family Outcomes, Reporting, and Management system (nFORM 2.0). Evaluators can use nFORM 2.0 data to learn about program retention by identifying participants who complete the program and participants who exited the program early. These data can also help evaluators identify when in the program participants are most likely to exit. These retention metrics can give important insights into program dosage. For example, examining how much of the program participants are receiving throughout implementation can help program staff and evaluators discover and address retention challenges—such as when only one member of a couple attends most workshop sessions—and signal whether participants are regularly receiving enough of the intervention for the evaluation findings to show changes in participant outcomes.

Table 3.1 illustrates how a program might track retention using a table or spreadsheet. In this example, the table is set up to track retention in the program's primary workshop—a key performance measure for ACF-funded HMRE programs—and it distinguishes between two study groups (those who receive the program in person and those who receive it virtually). The last row in the table shows that participants assigned to virtual program delivery attend workshop sessions at a higher rate than participants assigned to in person program delivery (73 versus 66 percent, respectively, attended at least 90 percent of workshop hours). It is possible that the difference in dosage might lead to higher post-test scores in the virtual group because they received a greater dose of program content. Collecting and analyzing these data can help evaluators identify when it might be necessary to work on boosting attendance in one of the study groups to ensure they limit bias in their effects estimates.

Table 3.1. Example of a table to track program retention

Program retention categories	Study group 1 (In-person delivery)		Study group 2 (Virtual delivery)	
	Number of participants	Percent	Number of participants	Percent
Workshop enrollment	200	100%	400	100%
Did not attend any workshop hours	25	13%	30	8%
Attended less than 50 percent of workshop hours	18	9%	27	7%
Attended more than 50 percent but less than 90 percent of workshop hours	25	12%	50	12%
Attended more than 90 percent of workshop hours	132	66%	293	73%

Source: Adapted from a hypothetical example in McInerney and Roemer (2017).



Tip 2: Engage participants and staff to explore underlying retention challenges and work together to solve them

Evaluators can consider asking participants and program staff questions to deepen their understanding of the program's retention challenges (Box 3.2). For example, during enrollment, program staff or local evaluators can ask participants if they expect anything to prevent them from completing the program and clearly describe everything the program does to support attendance (for example, making child care or travel vouchers available, providing options for participating virtually, or attending a make-up class for missed workshop sessions). Throughout implementation, program staff and evaluators can use both quantitative and qualitative data to deepen their understanding of retention challenges. For example, quantitative data can show things such as at which session participants are most likely to stop attending or the average number of sessions a participant completes. Other data sources, like interviews with participants who are often absent from workshop sessions, can help programs and evaluators identify common barriers to participation.

Box 3.2. Ask questions to understand why participants leave the program early

Consider asking participants—both those who actively participate and those who do not—the following questions to further understand what keeps some participants from attending sessions.

1. Are workshops or services offered at times that are convenient for you? How about make-up options? If not, what times would work better?
2. What, if anything (such as a job, transportation, or child care), is getting in the way of your attending workshops or participating in evaluation activities? If nothing gets in your way, what helps you attend workshops regularly?
3. Do the workshops cover topics that interest you? Why or why not?
4. How is the workshop content relevant to your cultural practices, beliefs, and values? If it isn't, how could it be changed?
5. How do facilitators or other staff actively encourage or motivate you to participate during workshop sessions? How could they improve?
6. What do you think of the incentives the program offers for attending workshops? How would you change them to encourage more people to attend?

Source: Ooms and Wilson 2004; Baumgartner et al. 2022.

In addition to engaging program staff and participants to understand challenges, evaluators can converse with these groups to find and develop promising strategies. For example, evaluators can engage participants and staff in interactive activities designed to help them brainstorm solutions (Baumgartner et al. 2022). If an HMRE program realizes attendance is particularly challenging for adult couples who have different work schedules or do not have child care during workshop sessions, the program could bring staff and participants together to ask if offering longer classes on the weekends instead of several shorter classes throughout the week would be helpful, and/or if they would utilize on-site child care during workshops.¹

There is no one-size-fits-all strategy that will solve every dosage-related challenge faced by HMRE programs. To devise a strategy that could boost attendance, in addition to talking with participants and program staff, evaluators can consult the research literature. ACF has published several resources to help HMRE programs address common implementation challenges (for example, see this [guide](#)²). After conducting background research, evaluators can go back to staff and participants to see which solutions might best promote workshop attendance within the particular community or program. Addressing retention-related challenges early may help increase dosage for future participants who will not face the same barriers to participation.



Tip 3: Offer flexibility to encourage attendance

A common challenge for HMRE programs is that participants cannot attend every scheduled session (Baumgartner et al. 2022). For adults, work schedules can get in the way; for programs serving youth in schools, disruptions like assemblies or drills, changing class schedules, or youth transferring schools can lead to missed workshop sessions. HMRE programs may consider the following strategies to overcome scheduling conflicts. Box 3.3 offers additional strategies, specifically for programs offering workshops to youth in schools.

¹ Note that HMRE programs should discuss any programmatic changes with their FPS to make sure they comply with the federal guidelines before implementing changes.

² Friend, Daniel, Lauren Mattox, Annie Buonaspina, Avery Hennigar, Scott Baumgartner, and Angela Valdovinos D'Angelo (2022). Strategies for addressing common implementation challenges in healthy marriage and relationship education programs: A guide for supporting design and improvement efforts, OPRE Report #2022-314, Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

- Offer a make-up session (Bodenlos et al. 2021)
- Provide virtual workshops or services to participants who face significant barriers, such as having no transportation or child care (Friend et al. 2023)
- Offer programming in multiple locations close to the population served (Friend et al. 2023)
- Consider hiring a mix of full-time and hourly staff to increase the program’s ability to offer services on a flexible schedule outside of regular business hours (Friend et al. 2023)

■ **Box 3.3. Strategies to address dosage-related challenges for programs serving in-school youth**

For HMRE programs serving youth in schools, a lack of strong partnerships with schools or districts and parental buy-in can make it harder to keep youth participating in a program (Baumgartner et al. 2022; Friend et al. 2022). The following strategies may help address participation challenges in school settings:

- Strengthen partnerships with schools and parents to encourage youth to participate in services (McCallum and Adamek 2020).
- Engage parents in services by sharing program content before the program starts (for example, by sending supplemental materials home with youth or sharing content via text messages to parents) (McCallum and Adamek 2020).
- Ensure the class setting is comfortable, and classes give students enough opportunities to engage (for example, by allowing them to choose where they sit, and allowing youth to opt out of certain activities or discussions if they are uncomfortable) (Logan et al. 2022).
- Every youth in the program might not currently be in a romantic relationship. Focus on making the content relevant for all youth by emphasizing that the skills being taught can apply to non-romantic relationships, need to be practiced, and can help make a future romantic relationship live up to their expectations (Goesling and Alamillo 2018).



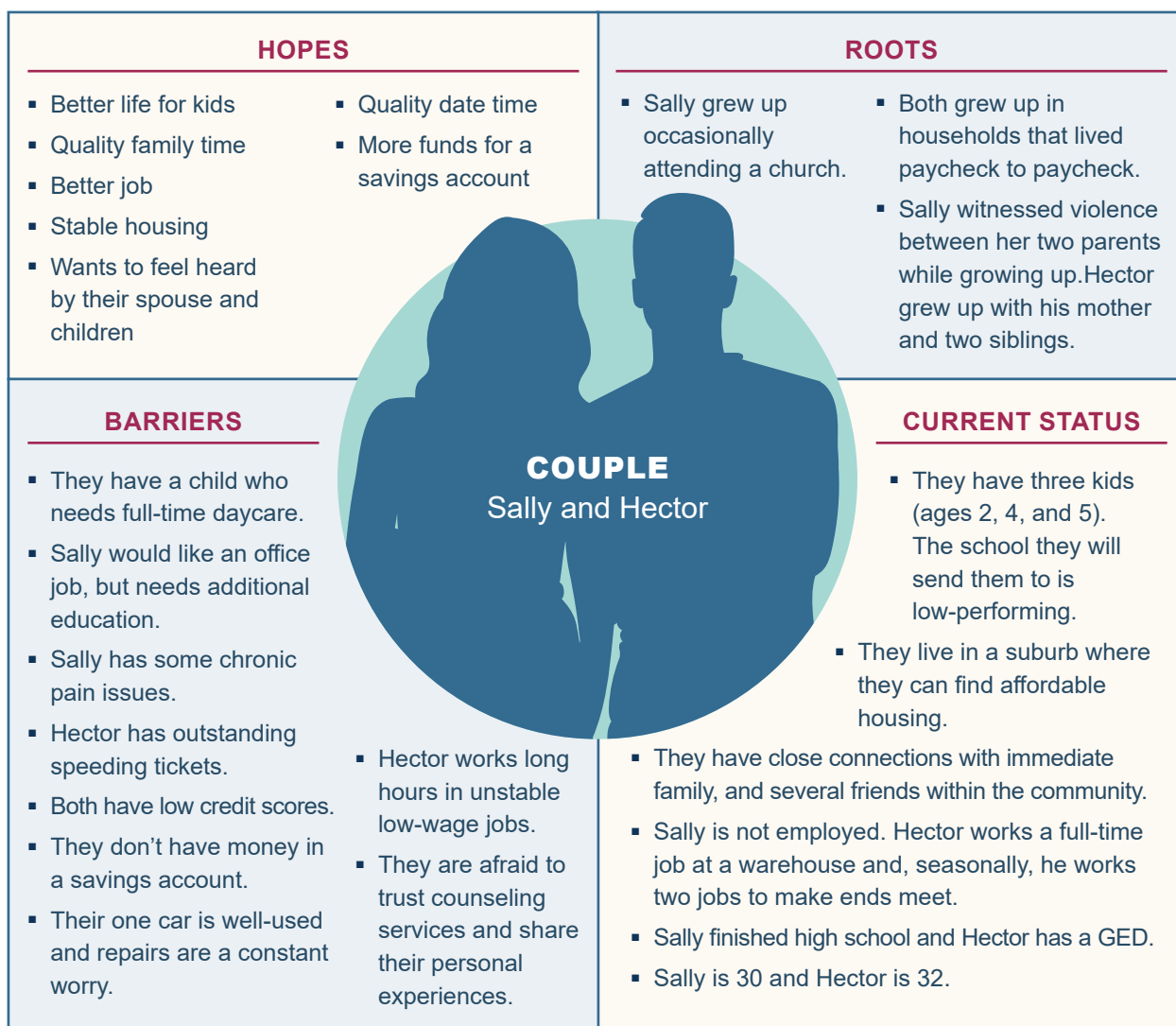
Tip 4: Offer incentives that are relevant to participants

Another strategy for encouraging attendance is to offer incentives for participating in workshops (Friend et al. 2022). Effective incentives include cash or gift cards, food (for example, providing a meal at workshop sessions), recognition (for example, holding a program graduation), or goods and services (for example, providing free child care). Offering incentives at various points throughout the program can help maintain high attendance, but only if participants are interested in or motivated by them. For example, gas cards would not be motivating if most participants take the bus.

To identify incentives that align with participants’ interests, programs should consider asking participants which incentives would be motivating to them or using a [persona profile](#)³ (that is, a summary of a typical participant’s characteristics, to understand their key characteristics, behaviors, goals, or needs). Persona profiles help programs think through the characteristics of their target population (Lemon and Verhoef 2016). Creating a profile of a typical participant may help programs to develop tailored incentive strategies. Figure 3.1 is an example of a persona profile of an HMRE program that serves couples. In this example, this activity reveals that couples with children may enjoy incentives for activities they can do as a family or during a date night—as opposed to incentives they can use individually—since the couple noted that they hoped for more quality time.

³ Luma Institute. “Persona Profile.” n.d. <https://www.luma-institute.com/persona-profile/>.

Figure 3.1. Example persona activity for an HMRE program that serves couples



Program staff or local evaluators should keep track of the incentives offered to see if they are working.⁴ Tracking the types of incentives used, their value, and participants' engagement in workshops can help programs determine whether the incentives offered—and when they were offered—are effective (Greene et al. 2012).

⁴ Note that HMRE programs should discuss any changes to incentives with their FPS to make sure they comply with the federal guidance.

References

- Baumgartner, Scott, Daniel Friend, Robert G. Wood, Annie Buonasina, and Hannah McInerney. "Developing Strategies to Address Implementation Challenges Facing Healthy Marriage and Relationship Education Grantees." OPRE Report #2022-36. Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, 2022.
- Bronte-Tinkew, Jacinta, Allison Horowitz, Zakia Redd, Kristin A. Moore, and Sherylls Valladares. "A Glossary of Research Terms for Out-of-School Time Program Practitioners." Research-to-Results Fact Sheet, publication no. 2007-02, Child Trends, 2007.
- Friend, Daniel, Lauren Mattox, Annie Buonasina, Avery Hennigar, Scott Baumgartner, and Angela Valdovinos D'Angelo. "Strategies for Addressing Common Implementation Challenges in Healthy Marriage and Relationship Education Programs: A Guide for Supporting Design and Improvement Efforts." OPRE Report #2022-314, Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, 2022.
- Friend, Daniel, and Diane Paulsell. "Research to Practice Brief: Developing Strong Recruitment Practices for Healthy Marriage and Relationship Education (HMRE) Programs." OPRE Report #2020-78. Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, 2020.
- Goesling, Brian, and Julia Alamillo. "Five Tips for Teaching Healthy Marriage and Relationship Education in Schools." Research to Practice brief. OPRE Report #2018-101. Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, 2018.
- Greene, Kaylin M., Bora Lee, Nicole Constance, and Kathryn Hynes. "Examining Youth and Program Predictors of Engagement in Out-of-School Time Programs." *Journal of Youth and Adolescence*, vol. 42, 2013, pp. 1557–1572.
- James Bell Associates. "Evaluation Brief: Conducting a Process Evaluation." James Bell Associates, August 2008. <https://www.jbassoc.com/wp-content/uploads/2018/03/Conducting-Process-Evaluation.pdf>. Accessed June 15, 2023.
- James Bell Associates. "Formative Evaluation Toolkit." December 2018. <https://www.acf.hhs.gov/cb/resource/formative-evaluation-toolkit>. Accessed June 15, 2023.
- Lemon, K.N., and P.C. Verhoef. "Understanding Customer Experience Throughout the Customer Journey." *Journal of Marketing*, vol. 80, no. 6, 2016, pp. 69–96.
- Loeb, S., S. Dynarski, D. McFarland, P. Morris, S. Reardon, and S. Reber. "Descriptive Analysis in Education: A Guide for Researchers." (NCEE 2017–4023). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, 2017.
- Logan, D., A. Wilson, H. Netschytailo, and V. Martinez. "Students Offer Lessons on Enhancing Engagement in Youth Programs." Child Trends, 2022. <https://doi.org/10.56417/5422o9735x>.
- McCallum, D., and K. Adamek. "Strategies for Engaging Parents and Caregivers." Tip sheet. Office of Population Affairs, Office of the Assistant Secretary for Health, U.S. Department of Health and Human Services.
- McInerney, Hannah, and Grace Roemer. "Reviewing Workshop Participation in the QPR/PPR." Tip sheet. Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, 2023.
- Ooms, Theodora, and Pamela Wilson. "The Challenges of Offering Relationship and Marriage Education to Low to Income Populations." *Family Relations*, vol. 53, no. 5, 2004, pp. 440–447.
- U.S. Department of Health and Human Services. "Frequently Asked Questions: Reporting Implementation Findings." U.S. Department of Health and Human Services, Office of Adolescent Health, 2011. Accessed June 15, 2023.
- Williamson, Hannah C., Angela V. D'Angelo, and Stephanie Warren. "Recruitment and Retention of Diverse Couples in Relationship Education with Integrated Economic Services." *Family Process*, 2023.
- Zhai, Fuhua, C. Cybele Raver, Stephanie M. Jones, Christine P. Li-Grining, Emily Pressler, and Qin Gao. "Dosage Effects on School Readiness: Evidence From a Randomized Classroom-Based Intervention." *Social Service Review*, vol. 84, no. 4, 2010, pp. 615–655.

Understanding and mitigating attrition

Daniel Friend, Avery Hennigar, Brandon Hollie, and Rebecca Piatt

Even the most well-designed and well-implemented evaluation is likely to lose sample members over the course of the study because at least a few of the participants will not complete all data collection (for example, will not provide exit or follow-up data). (See, for example, Crutzen et al. 2015.) This sample loss is called attrition. Attrition can weaken a Healthy Marriage and Responsible Education (HMRE) evaluation by introducing bias. That bias, known as attrition bias, in turn affects the evaluator's ability to form rigorous conclusions. This brief describes (1) what attrition is and why it can be a problem; (2) how to calculate and assess it using guidance from the [What Works Clearinghouse](#) (WWC)¹ and (3) steps evaluators can take to mitigate it.

? What is attrition bias?

Although attrition is a problem for any evaluation, it is particularly problematic for evaluations using experimental designs like randomized control trials (RCTs) (Box 4.1). In an RCT design, participants are randomly assigned into one of two or more groups or conditions—typically a treatment group and a control group. The process of random assignment balances any differences between the groups in their observed and unobserved characteristics. Therefore, in an RCT, study groups are considered to be equivalent at the time of study enrollment (Box 4.2). Because the two groups are equal on other factors at baseline (that is, study enrollment), evaluators can attribute any observed differences between the outcomes of the treatment group and the outcomes of the control group to the HMRE program (Hariton and Locasio 2018; Shadish and Galindo 2010). However, when an RCT has a high rate of attrition, there is a risk that the study will no longer have equivalent groups and may be subject to attrition bias. Attrition bias occurs when the loss of sample through attrition creates an imbalance or inequivalence between the research groups (Gustavson et al. 2012; Miller and Hollist 2007; WWC 2022).

■ Box 4.1. Should attrition only be a concern for RCTs?

No. Attrition is concerning for any type of study. Attrition threatens an RCT's main strength—creating equivalent groups at baseline through random assignment—but it can also threaten other research designs (such as descriptive or quasi-experimental). For example, attrition can reduce the statistical power to detect effects or outcomes because it reduces the sample size for analysis (Button et al. 2013; Shadish and Luellen 2014). Additionally, nonresponse bias—which happens when people who did respond systematically differ from those who did not—can negatively affect any study design (Dong et al. 2011; Miller and Hollist 2007).

■ Box 4.2. Key terms

Attrition. Occurs when outcome data for a participant or participants are not available.

Baseline equivalence. Refers to the assumption that research groups are equal on observed and unobserved characteristics before they enroll in the evaluation and before they receive the HMRE program.

Bias. Any systematic error in a study.

Source: WWC 2022.

¹ The WWC is operated by the U.S. Department of Education, Institute of Education Sciences. The WWC guides program evaluators using causal designs to provide rigorous evidence of program efficacy. The WWC contains information on when attrition is problematic for RCTs and on appropriate next steps. The Administration for Children and Families adapts clearinghouse standards, like those from the WWC, to provide evaluators with guidance on analysis and reporting of HMRE local evaluations.

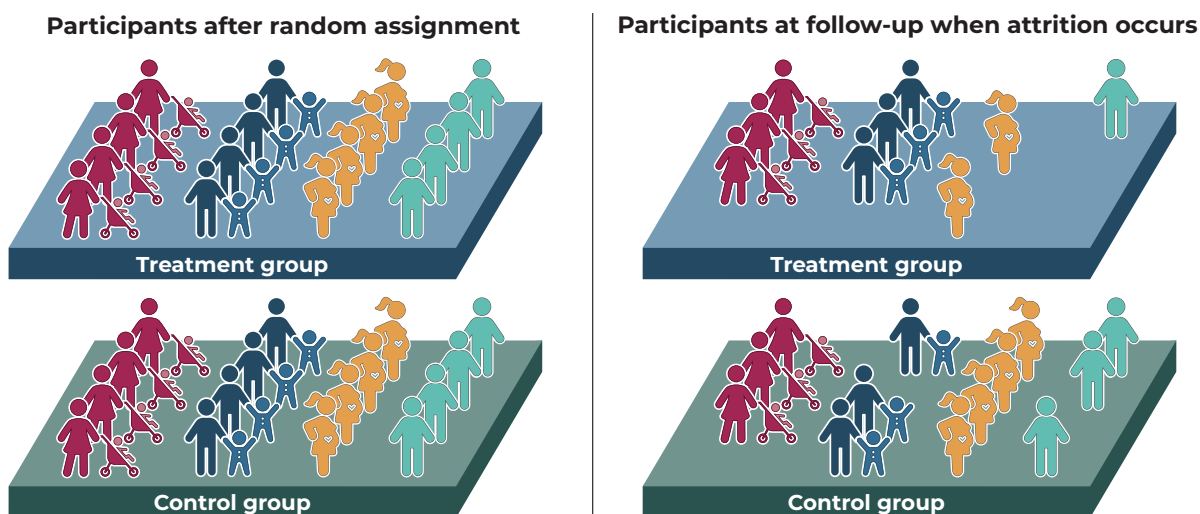
Attrition can happen for a variety of reasons. For example, a participant could:

- Not be located at the time of follow-up data collection
- Be unavailable at the time of data collection
- Decide to drop out of the study
- Refuse to participate in evaluation activities or withdraw their consent from the study

Regardless of the reason, it is important to closely monitor and mitigate attrition in the study to reduce the risk of bias.

Figure 4.1 is a visual example of attrition in an HMRE program using an RCT evaluation design.² The figure shows the composition of the research groups immediately following randomization at the start of the evaluation before they receive the HMRE program (the left column) and again six months later after a follow-up survey (the right column). After random assignment, the intervention and comparison groups are equivalent in terms of demographic and other characteristics at baseline. At the six-month follow-up, only a subset of participants completed the survey. As a result, the intervention and control group now have different characteristics. For example, the treatment group only has two pregnant individuals at follow-up, whereas the control group has four. Additionally, the treatment group only has one single person at follow-up, whereas the control group has three. In other words, because the groups are no longer similar, there is no certainty that any differences in outcomes at the six-month follow-up were caused by the HMRE program.

Figure 4.1. Example of attrition in an HMRE program



Key: The different icons signify participants with different demographic characteristics. The red icons in the first row are female participants with children. The blue icons in the second row are male participants with children. The yellow icons in the third row are pregnant women. The teal icons in the fourth row are single men.

² This figure shows an HMRE program that serves adult individuals. However, many HMRE programs served paired partners simultaneously. For more information on challenges that evaluators face in analyzing data from paired partners, and strategies they can consider to address these challenges, see this brief: <https://www.acf.hhs.gov/opre/report/analyzing-data-paired-partners-program-evaluation-strategies-overcome-common-challenges>.

Attrition is outcome and time specific, meaning it is possible to have low attrition for some outcomes or time points and high attrition on others (WWC 2022). As a result, it is important to calculate attrition for *each* time point and for *each* outcome (Box 4.3).

Box 4.3. Examples of attrition, varied across multiple outcomes and time points

Multiple time points. An HMRE local evaluator noticed high attrition at the first follow-up after baseline. After investigating the issue, they learned participants preferred to complete surveys online. In later follow-up periods, the HMRE program offered participants the chance to complete the survey by phone or online, which led to improved response rates at subsequent follow-ups. As a result, the first follow-up had higher attrition than the later follow-ups.

Multiple outcomes. An HMRE evaluation examined the program's impact on several outcomes: communication, relationship satisfaction, and intimate partner violence. However, some participants skipped the section on intimate partner violence because they thought the questions were too sensitive. In this instance, attrition for this outcome was higher than it was for the others, making the outcome variable on intimate partner violence challenging to analyze.



How to calculate and assess attrition

Evaluators need to calculate two types of attrition: overall and differential (Box 4.4). Overall attrition is the total number of participants across both research groups with missing data—in other words, the total number of participants that did not participate in data collection divided by the total number of participants enrolled in the evaluation or randomized (Box 4.5).

Differential attrition compares the differences in the attrition rates of the research groups. To calculate differential attrition, compute the attrition rate for both research groups, and take the absolute value of the difference between the two rates (Box 4.5). In Figure 4.1, the treatment group had more attrition than the control group, which can affect study quality (Gustavson et al. 2012; Miller and Hollist 2007). For example, participants who did not find the HMRE program relevant and stopped attending after a few sessions might not have responded to the follow-up survey. This would lead to the treatment group consisting of only people

Box 4.4. Key terms, continued

Overall attrition. Refers to attrition for the whole sample (treatment and control groups) calculated as the percentage of randomly assigned units for which the evaluators do not observe outcome data.

Differential attrition. Calculated by taking the absolute value of the percentage point difference between attrition rates for the treatment group and the comparison group.

Source: WWC 2022.

Box 4.5. How to calculate overall and differential attrition

Overall attrition calculation

$$\frac{\text{\# of participants without observed outcome data in the analysis}}{\text{\# of participants randomized}}$$

Differential attrition calculation

$$\frac{\text{\# of treatment participants without observed outcome data in the analysis}}{\text{\# of treatment participants randomized}} - \frac{\text{\# of control participants without observed outcome data in the analysis}}{\text{\# of control participants randomized}}$$

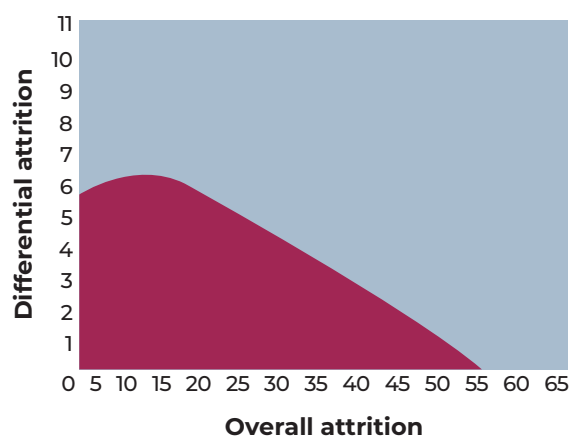
who completed the program, and the evaluation would therefore find favorable outcomes when none exist simply because only participants who found the program relevant to them were included in the outcome measures (Miller and Hollist 2007).

When evaluating an HMRE program, evaluators can follow the guidance published by the WWC to ensure they are conducting analysis in a rigorous and systematic way. WWC standards for causal design studies require that researchers conducting an RCT calculate both overall and differential attrition and use a ratio of the two to determine if the attrition is within an acceptable threshold range (WWC 2022). The WWC developed two threshold ranges—the cautious and optimistic threshold. The Administration for Children and Families (ACF) recommends that most HMRE studies use the cautious threshold, which is a more conservative approach to assessing the potential influence of attrition bias on study results. Under the cautious threshold, higher rates of attrition are considered unacceptable.

According to the WWC attrition standard, evaluations are classified as having either “high” or “low” attrition based on both overall and differential attrition. Figure 4.2 shows this threshold in different colors: low attrition rates fall within the maroon threshold, and high attrition rates fall within the blue threshold.

The WWC standards also account for an important trade-off between overall and differential attrition—namely, that a study can have a higher overall rate of attrition if it has a low rate of differential attrition and still be deemed acceptable. As shown in Figure 4.2, overall attrition could be approximately 50 percent and still be considered low if differential is minimal (1 percent or less). If an evaluation has low attrition, attrition bias is not a concern, and the evaluator can proceed with their planned analyses. If an evaluation has high attrition, the evaluator is required to demonstrate baseline equivalence of participants in the final analytic sample. In other words, the evaluator should show that the research groups are still similar for those who completed data collection (WWC 2022) ([See the seventh brief in this toolkit on creating equivalent research groups](#)).

Figure 4.2. WWC attrition threshold for HMRE local evaluations



Source: WWC 2022.



Mitigating attrition

The best way to mitigate the effects of attrition bias in an evaluation is to ensure high response rates. See the second brief in this toolkit on improving survey response rates, which focuses on response rates and discusses strategies for maximizing response rates, including clearly explaining the importance of surveys to participants, obtaining several modes of contact information, offering participants different ways to complete surveys, and using motivating and well-timed incentives.

Evaluators should also consider how to minimize differential attrition, as that is often more important than minimizing overall attrition. For example, as shown in Figure 4.2, having low overall attrition but high differential attrition can put an evaluation in the “high” attrition category and require the evaluator to demonstrate baseline equivalence. Evaluators can minimize differential attrition by striving to collect data at equal rates from all participants, regardless of the treatment condition or baseline characteristic. Evaluators should consider that reaching people in one group could require more effort than reaching people in another group. For example, it may be easier to contact participants in the treatment group because they are continuing to receive services from the HMRE program.

Regularly monitoring attrition by examining response rates at each follow-up period during and after data collection and examining missing data across outcomes can help evaluators and HMRE program staff proactively address it. For example, identifying particular subgroups of participants who are not responding to follow-up data collection could inform a strategy to improve response rates. Regularly calculating overall and differential attrition throughout data collection can help evaluators monitor whether the study is on track to meet WWC attrition thresholds. Early monitoring of attrition rates can also help evaluators be proactive about demonstrating baseline equivalence if the study is at risk of attrition that falls outside the acceptable threshold.

References

- Abshire, Martha, Victor D. Dinglas, Maan Isabella A. Cajita, Michelle N. Eakin, Dale M. Needham, and Cheryl Dennison Himmelfarb. "Participant Retention Practices in Longitudinal Clinical Research Studies with High Retention Rates." *BMC Medical Research Methodology*, vol. 17, no. 1, 2017, pp. 1–10.
- Button, Katherine S., John P.A. Ioannidis, Claire Mokrysz, Brian A. Nosek, Jonathan Flint, Emma S.J. Robinson, and Marcus R. Munafò. "Power Failure: Why Small Sample Size Undermines the Reliability of Neuroscience." *Nature Reviews Neuroscience*, vol. 14, no. 5, 2013, pp. 365–376.
- Cole, Russell, and Seth Chizeck. "Sample Attrition in Teen Pregnancy Prevention Impact Evaluations." Mathematica Policy Research, 2014.
- Crutzen, Rik, Wolfgang Viechtbauer, Mark Spigt, and Daniel Kotz. "Differential Attrition in Health Behaviour Change Trials: A Systematic Review and Meta-Analysis." *Psychology & Health*, vol. 30, no. 1, 2015, pp. 122–134, DOI: [10.1080/08870446.2014.953526](https://doi.org/10.1080/08870446.2014.953526).
- Dong, Nianbo, and Mark W. Lipsey. "Biases in Estimating Treatment Effects Due to Attrition in Randomized Controlled Trials and Cluster Randomized Controlled Trials: A Simulation Study." *Society for Research on Educational Effectiveness*, 2011.
- Dorey, Frederick J. "In Brief: Statistics in Brief: Statistical Power: What Is It and When Should It Be Used?" *Clinical Orthopedics and Related Research*, vol. 469, no. 2, 2011, pp. 619–620.
- Fleiss, J.L. *Statistical Methods for Rates and Proportions*. 2nd ed. John Wiley & Sons, 1981.
- Gustavson, Kristin, Tilmann von Soest, Evalill Karevold, and Espen Røysamb. "Attrition and Generalizability in Longitudinal Studies: Findings From a 15-Year Population-Based Study and a Monte Carlo Simulation Study." *BMC Public Health* vol. 12, 2012, p. 1–11.
- Hariton, Eduardo, and Joseph J. Locascio. "Randomised Controlled Trials—The Gold Standard For Effectiveness Research." *BJOG: An International Journal of Obstetrics and Gynaecology*, vol. 125, no. 13, 2018, p. 1716.
- Kost, Rhonda G., and Joel Correa da Rosa. "Impact of Survey Length and Compensation on Validity, Reliability, and Sample Characteristics for Ultrashort-, Short-, and Long-Research Participant Perception Surveys." *Journal of Clinical and Translational Science*, vol. 2, no. 1 2018, pp. 31–37.
- Miller, Richard B., and Cody S. Hollist. "Attrition Bias." Paper 45, Faculty Publications, Department of Child, Youth, and Family Studies, University of Nebraska - Lincoln, 2007.
- Shadish, W.R., and R. Galindo. "Randomized Control Trials." In *The Corsini Encyclopedia of Psychology*, edited by I.B. Weiner and W.E. Craighead, 2010. <https://doi.org/10.1002/9780470479216.corpsy0770>.
- Shadish, W.R., and J.K. Luellen. "Attrition." In *Wiley StatsRef: Statistics Reference Online*, edited by N. Balakrishnan, T. Colton, B. Everitt, W. Piegorisch, F. Ruggeri, and J.L. Teugels. 2014. <https://doi.org/10.1002/9781118445112.stat06672>.
- Yan, Ting, and Richard Curtin. "The Relation Between Unit Nonresponse and Item Nonresponse: A Response Continuum Perspective." *International Journal of Public Opinion Research*, vol. 22, no. 4, 2010, pp. 535–551.
- What Works Clearinghouse. *Procedures and Standards Handbook, Version 5.0*. U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, 2022. <https://ies.ed.gov/ncee/wwc/Handbooks>.

Using mixed methods

Daniel Friend, Rebecca Piatt, and Brandon Hollie

Mixed methods research, or research that involves collecting both quantitative and qualitative data, can help Healthy Marriage and Relationship Education (HMRE) programs and evaluators answer a wide array of research questions and deepen their insights into the questions they are investigating (Box 5.1; Creswell and Creswell 2018; Johnson and Onwuegbuzie 2004). For example, quantitative data could help HMRE programs and evaluators determine if their program achieved its intended goals, and qualitative data could help answer how and why it did or did not.

This brief explains why and when programs and evaluators might use mixed methods research in an HMRE evaluation, followed by a description of three commonly implemented designs. The brief closes with questions HMRE programs and evaluators can consider when choosing an approach.

Box 5.1. Key terms

Quantitative research. A research method for testing objective theories by examining the relationship among variables. Quantitative data are often close-ended, numerical data (such as numerical responses on a survey measure).

Qualitative research. A research method for exploring and understanding the meaning individuals or groups associate with a social or human problem. Qualitative data are often open-ended, non-numerical, or descriptive data (for example, responses to an in-depth interview or focus group).

Source: Creswell and Creswell 2018.

? Why use mixed methods research?

There are many advantages to using a mixed methods design. A mixed methods research design is particularly useful when attempting to answer research questions that neither quantitative nor qualitative methods can answer alone. Mixed methods research can allow evaluators to make more substantive conclusions because the strengths of the qualitative methods can offset the weaknesses of the quantitative methods, and vice versa (Creswell and Creswell 2018; Johnson and Onwuegbuzie 2004). For instance, quantitative data can provide numerical information—such as percentages, rates, or trends—while qualitative data can provide detailed insights into participants’ perspectives and experiences. In HMRE contexts, quantitative research can help demonstrate that an HMRE program improves participants’ relationship skills, but evaluators might need qualitative data to understand *how and why* relationship skills improved. Similarly, if an HMRE grant recipient is interested in testing the impact of the program or generalizing findings to broader populations or settings, they will need to collect quantitative data. Conducting mixed methods research can help HMRE grant recipients answer complex research questions that are generalizable while simultaneously providing a deeper understanding of the issue being investigated (Dawadi 2021).

Types of mixed methods research

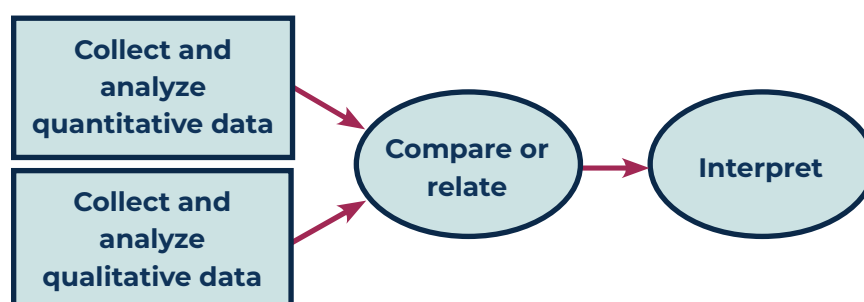
There are several types of mixed method research designs. This section focuses on three common designs—*convergent parallel* or *triangulation*, *explanatory sequential*, and *exploratory sequential*. This section explains each type of design, discusses when to consider using it, and gives an example scenario of how to use it.



Convergent parallel or triangulation design

In a convergent parallel design (Figure 5.1), researchers simultaneously (but separately) collect quantitative and qualitative data, analyze them separately, and triangulate the results (Creswell and Creswell 2018). Evaluators can use a triangulation design to collect different but complementary data about an outcome of interest.

Figure 5.1. Convergent parallel design



Source: Creswell and Creswell 2018.

A central purpose of this design is to give researchers the ability to directly compare and contrast quantitative or statistical results with qualitative data to reveal similarities and discrepancies. Accordingly, it is important to measure the same or similar variables, constructs, or concepts for each data collection method (Creswell and Creswell 2018). For example, a survey can have a question about how satisfied participants were with the HMRE program, and a focus group can explore the reasons why they were or weren't satisfied with the program. Box 5.2 is a scenario of how an HMRE program can use a triangulation design to evaluate their program.

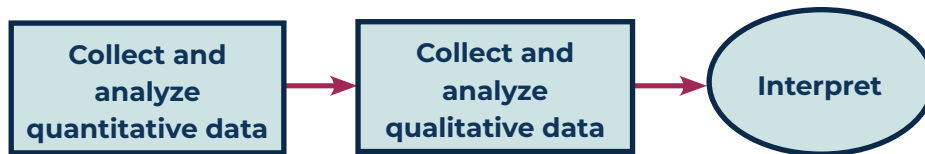
Box 5.2. A triangulation design for evaluating an HMRE program

An HMRE program that serves couples is interested in learning more about how their program influences couples' level of satisfaction with their relationships. The evaluation seeks to answer two research questions: (1) Did participants' satisfaction with their relationship increase in the period from enrollment to six months after program services ended? and (2) What aspects of the program contributed to changes in participants' relationship satisfaction? To answer these research questions, the evaluator decides to triangulate quantitative and qualitative data and compare the results. Specifically, they administer a series of quantitative survey measures in which participants rate their communication skills and feelings of closeness and affection toward their partner. They also conduct a series of focus groups with couples at the six-month follow-up to collect qualitative data on their program experiences and how their relationship with their partner changed since participating in the program.

Explanatory sequential design

The intent of an explanatory sequential design is to have qualitative data explain quantitative findings in detail. It involves two phases of data collection: (1) collecting and analyzing quantitative data, followed by (2) designing a qualitative data collection approach based on the quantitative findings (Figure 5.2).

Figure 5.2. Explanatory sequential design



Source: Creswell and Creswell 2018.

Evaluators should use this design when they want a better explanation of the results of their quantitative analysis—particularly when the results might be innovative or different from what the literature suggests (for example, when demonstrating the efficacy of a novel approach). Evaluators might also use this design when they wish to create meaningful subgroups based on their quantitative findings. For example, an evaluator may choose to purposively sample focus group participants that reported high and low scores on a survey measure on cooperative co-parenting. This type of mixed method design can also be useful in program improvement (Box 5.3).

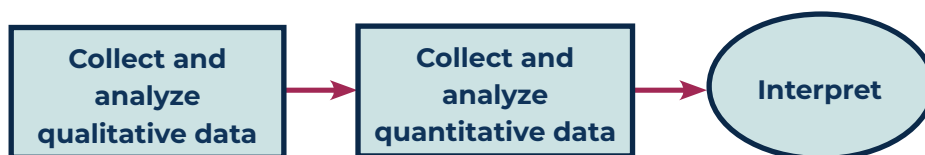
Box 5.3. Explanatory sequential design for program improvement

An HMRE program for parents learns from a quantitative exit survey that 60 percent of participants were satisfied with the program. The program and evaluator decide to purposefully select a group of participants who reported being satisfied with the program and a group of participants who reported being dissatisfied with the program and conduct qualitative, one-on-one interviews to understand what contributed to their experiences. After analyzing the data, the evaluator finds common themes among the dissatisfied group—namely, that they think the program is too long, and that it should provide meals before workshops. Meanwhile, themes from the satisfied group include appreciation for on-site child care, engaging material, and the transportation assistance. The program uses these results to continue to support aspects of the program that participants appreciated, and tweak aspects of the program that could be improved (for example, keep offering child care, and start offering meals before workshops).

Exploratory sequential design

In an exploratory sequential design, evaluators use qualitative data to develop a survey or other quantitative measure (Creswell and Creswell 2018). This design involves two phases of data collection: (1) collecting and analyzing qualitative data, followed by (2) designing and administering a survey based on the qualitative data (Figure 5.3).

Figure 5.3. Exploratory sequential design



Source: Creswell and Creswell 2018.

Evaluators should choose this design when they need to explore a topic in more depth before collecting more data. Exploratory sequential designs can be useful when appropriate measures or instruments are not available; there is no guiding framework or theory; or the evaluator wants to generalize established findings to new groups. For example, an exploratory sequential design can help programs and evaluators tailor program content or measures used in evaluations to make them more culturally appropriate for specific populations. Box 5.4 gives a scenario of when an HMRE program might use an exploratory sequential design.

Evaluators may also use this design when conducting an implementation study or when undertaking a program improvement effort. For example, if an HMRE program decides to implement a new content delivery approach (such as offering pre-recorded, asynchronous workshop sessions), their evaluator may choose to collect qualitative feedback from staff and participants on the new mode, then design a survey to collect more information on how to enhance delivery of the asynchronous content.

Box 5.4. An exploratory sequential design to inform program delivery

An HMRE program is interested in developing a program to improve the co-parenting relationships of recently divorced parents. To understand the needs and experiences of divorced parents, the evaluator starts by holding a focus group with providers serving divorced families. After collecting, coding, and analyzing these data, the evaluator finds common challenges among this group in forming new romantic relationships, poor communication about parenting roles and expectations, and difficulty accessing co-parenting services in the neighborhood. Based on these findings, the evaluator designs a community survey to understand if these themes are common to a larger group of people and to assess community interest in co-parenting services. The evaluator also designs a survey for legal professionals working with divorced families to assess their interest in and ability to refer participants to the HMRE program.



Choosing a mixed methods design

Although this brief explores three common mixed methods designs, they are not the only designs to choose from. There are many others—for example, embedded, transformative, and multiphase designs (Creswell 2008; Creswell and Creswell 2018). Incorporating these types of designs into HMRE evaluations is beyond the scope of this brief, but evaluators can use the reference section to find more resources on these types of designs.

Given the many options evaluators have for mixed methods approaches, they might wonder how to go about selecting a specific design type. Selecting a suitable design is based on many factors, including the evaluation's research questions, resources, timing, and capabilities (Creswell and Creswell 2018; Creswell et al. 2011). For example, if the evaluation budget is limited, and an evaluator's research questions can be answered primarily with quantitative data, they might consider the explanatory sequential design. Table 5.1 provides guidance evaluators can use to help determine their approach based on suggestions from prominent experts on mixed methods.

Table 5.1. Considerations when choosing a mixed method design approach

Areas to consider	Questions	Guidance
<i>Research questions</i>	<ul style="list-style-type: none"> What type of data are required to answer the question? Will both types of data be given equal weight in terms of answering the research questions? 	<ul style="list-style-type: none"> Evaluators might first consider the theoretical framework guiding the HMRE program and the evaluation. Some theories may lend themselves to collecting certain types of data. Evaluators should be able to explain the rationale behind their decisions about the data types they need to answer the questions, and their prioritization of the data types. This rationale should align with the analytic approach (for example, how the evaluators are merging data).
<i>Resources</i>	<ul style="list-style-type: none"> Does the evaluation have enough resources to support a mixed methods design? 	<ul style="list-style-type: none"> Evaluators should assess all the resources available to them—that is, funding, staff, and time. For example, the evaluator should have funds available to collect both types of data, ample staff to implement the approach, and enough time to analyze and interpret the data.
<i>Timing</i>	<ul style="list-style-type: none"> Should the data be collected concurrently or sequentially? At what stage of the evaluation will the data be integrated? 	<ul style="list-style-type: none"> Evaluators should consider the sequence of the data collection. For example, if the evaluator wants to select focus groups based on survey responses, a sequential approach is warranted. Evaluators should also consider when the data will be merged. Each of the design types discussed in this brief merge different data types at different stages. For example, a triangulation design merges data at the interpretation stage.
<i>Capabilities</i>	<ul style="list-style-type: none"> Do evaluator staff have the required skills to collect and analyze the data? 	<ul style="list-style-type: none"> Lead evaluators should consider their own abilities and those of their staff. They will need the right skill sets to collect, analyze, merge, and interpret each data type.

Source: Based on Creswell and Creswell 2018; Creswell et al. 2011.

References

- Creswell, J. W. *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research*. Third edition. Pearson, 2008.
- Creswell, John W., and J. David Creswell. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. Fifth ed. SAGE Publications, Inc., 2018.
- Creswell, John W., and Vicki L. Plano Clark. *Designing and Conducting Mixed Methods Research*. Sage, 2018.
- Creswell, J.W., A.C. Klassen, V.L. Plano Clark, and K.C. Smith. *Best Practices for Mixed Methods Research in the Health Sciences*. National Institutes of Health, Office of Behavioral and Social Sciences Research, 2011.
- Dawadi, Saraswati, Sagun Shrestha, and Ram A. Giri. "Mixed-Methods Research: A Discussion On Its Types, Challenges, and Criticisms." *Journal of Practical Studies in Education*, vol. 2, no. 2, 2021, pp. 25–36.
- Johnson, R. B., and A.J. Onwuegbuzie. "Mixed Methods Research: A Research Paradigm Whose Time Has Come." *Educational Researcher*, vol. 33, no. 7, 2004, pp. 14–26.

Cleaning and preparing data for analysis

Avery Hennigar, Rebecca Piatt, Sonia Alves, and Daniel Friend

Consistent data monitoring and regular data cleaning help evaluation teams quickly address errors and improve data quality for analysis and reporting (Osborne 2010). Without data monitoring and cleaning, staff and participants can unknowingly introduce errors that can lead to incorrect or biased conclusions. This brief describes how data errors can occur throughout a Healthy Marriage and Relationship Education (HMRE) evaluation and provides tips for how to avoid them.

Common sources of errors

Data errors can be introduced at various stages of an evaluation, including during data collection, entry and cleaning, and analysis. Regardless of the source, data errors are problematic if they are not identified and fixed. Such errors can introduce issues in the analysis and lead evaluators to incorrect conclusions.

- **Errors during data collection.** Data collection errors can occur when participants inadvertently select the wrong responses, misunderstand questions that may be confusingly worded, provide responses outside of the expected ranges, or enter values in the incorrect field. Logistical issues with survey administration also can cause errors, such as incorrectly programming a survey's skip logic. HMRE local evaluators and program staff may also introduce errors during data collection by administering surveys inconsistently.
- **Errors during data entry and cleaning.** Several mistakes can occur during data entry and data cleaning—including incorrectly transferring data from the survey, entering values in the wrong field, entering values incorrectly, or accidentally deleting or duplicating entries.
- **Errors during data analysis.** Evaluators can create data errors by incorrectly extracting data from a database or assigning the wrong values to a variable during data cleaning.

Box 6.1. Three examples of errors

1. Consider an HMRE program that uses an online survey for its evaluation. During survey development, the evaluator incorrectly programmed a skip pattern, which resulted in participants missing questions about their parenting skills. Because the data set includes incomplete data, the evaluator cannot fully answer the research questions at the end of the study.
2. An evaluator correctly entered the survey responses into the database for a question with a 5-point rating scale by coding responses 1 (strongly disagree) through 5 (strongly agree) according to the codebook's guidelines but mistakenly coded "I don't know" values as 6, instead of as a missing data value. This is particularly problematic if these data are later summed or averaged. The values will be inaccurate and not comparable to other published studies because they will include an additional value that potentially increases the sum or mean.
3. To create a summary score, some items might be coded as they are and others might be reverse-coded (with high values switched to low values and vice versa), so higher values always represent a favorable (or unfavorable) response. If an evaluator does not reverse-code items correctly during data cleaning, or the values are flipped when they should not be, the analysis will be incorrect.

Using strong data preparation and cleaning methods throughout the evaluation can reduce the risk of such errors. Preparing and checking data for analysis should occur on an ongoing basis—rather than after data collection is complete. HMRE local evaluators can use the following four tips to ensure that their data are accurate, free of errors, and ready for analysis.



Tip 1: Create a comprehensive plan for data collection and entry

An important first step when launching an evaluation is to create a comprehensive data collection plan. A thorough plan will help HMRE programs and evaluators achieve high-quality data collection that is consistent, complete, timely, and secure. The data collection plan should be a living document. Evaluators should update it and any associated materials regularly (HMRF Resource Site for 2020 Grantees 2022). The plan should include the following details (Box 6.2, Institute of Education Sciences 2021):

- Evaluation design and research questions
- Description of sample
- Data sources
- Data collection and analysis methods
- Person or persons responsible for data collection
- Security and storage method¹
- Evaluation timeline

Box 6.2. Data collection plan template

Evaluators can use the [Data Collection Plan Template](#)² on the Information, Family Outcomes, Reporting, and Management (nFORM) resources page to compile these details.

HMRE local evaluators should provide training to the program and evaluation staff who will be involved in data collection and data entry (Avellar et al. 2017). The training should provide guidance to staff on administering the surveys correctly and consistently across participants, which will reduce the risk of errors and bias. The training approach and mode should be based on the staff's evaluation experience—for example, less experienced staff may need more intensive in-person training, while more experienced staff may benefit from shorter virtual or prerecorded trainings to refresh their knowledge. Numerous nFORM resources, such as the user manual, training videos, and monthly office hours, are available to use when developing staff trainings (HMRF Resource Site for 2020 Grantees 2023).

Tip 2: Develop procedures for regularly monitoring, cleaning, and checking data for errors



Data cleaning is the process of detecting (for example, identifying strange or unexpected patterns), diagnosing (for example, identifying true extreme values vs. true normal values), and editing (for example, correcting, deleting, or leaving data unchanged) faulty data. The process of data cleaning can help create a final data set that is as accurate, complete, and ready for analysis as possible (Institute of Education Sciences 2021; Van den Broeck et al. 2005).

Box 6.3. Remember to preserve original data files

Maintain an original copy of the data. Never alter, trim, or recode data in the original file. Before data cleaning begins, create a copy of the data set and then a clean data file from the copy.

Evaluators often create a codebook to outline how variables will be cleaned (but not replaced) (Box 6.3). A codebook is a useful document for establishing clear coding procedures for survey instruments (Institute of Education Sciences n.d.). The nFORM resources include a [data dictionary](#),³ which can be a useful starting place for creating a codebook. HMRE evaluators may consider incorporating additional

¹ It is important to include proper security protections in an evaluation's application to an institutional review board and to follow them throughout the study. For example, evaluators should create guidelines for storing hard-copy files of survey instruments in locked filing cabinet drawers and for using password protection for spreadsheets that include participant data.

² HMRF Resource Site for 2020 Grantees, nFORM 2.0 Team. "Data Collection Plan Template." Administration for Children and Families, 2022. <https://www.hmrfgrantresources.info/resource/data-collection-plan-template>.

³ HMRF Resource Site for 2020 Grantees. "Data Dictionary." Administration for Children and Families, 2022. <https://www.hmrfgrantresources.info/resource/data-dictionary>.

variables from their local evaluation into a copy of the nFORM data dictionary to ensure consistency in participant IDs and demographic information. Relevant staff should be trained on how to navigate the codebook so that they feel comfortable using it.

HMRE local evaluators can use a variety of data cleaning techniques to screen, diagnose, and edit identified data errors. In addition to checking for errors, data cleaning can include standardizing variable types, such as changing all dates to follow the same format, and running a spell-check to correct misspellings that could cause coding or frequency count errors. Cleaning data regularly can identify patterns in the data errors so that they can be fixed early in the data collection process. In addition, developing a clean, organized data file can make the data archiving process more efficient. Many data archiving repositories require files to be accessible and interpretable, so data cleaning is an important step in preparing files for submissions.

Table 6.1 provides examples of techniques for checking errors during the data cleaning process and how to remedy the errors during or after data cleaning.

Table 6.1. How to catch and fix common data errors

Technique	Description	Example error	Example remedy
<i>Descriptive analysis</i>	Evaluators can calculate descriptive statistics—such as mean, range, or the distribution of variables—to make sure they seem reasonable and theoretically possible.	An evaluator of an HMRE program that serves youth in high school checks the age variable on the local evaluation survey and notices that there are several maximum values of 34 years old. However, the program only serves youth up to age 24, so these values are implausible.	The evaluator decides to restrict the range in the age variable on the survey to a maximum value of 24 years old.
<i>Double entry</i>	When conducting data entry, evaluators can ask two or more staff to enter the same survey, then check for discrepancies. Double entering each survey is the gold standard. However, if resources are limited, consider selecting a small subset of surveys to double-enter.	After comparing double-entered surveys, the lead evaluator of an HMRE program that serves Hispanic couples notices that one of the data entry staff members entered the ethnicity variable for all couples as a value of 1 in the data file, while another staff member entered the ethnicity variable for the same couples as 2 in the data file.	The lead evaluator checks the codebook for how data should be entered and verifies that the ethnicity variable for Hispanic couples should be coded as 1. The lead evaluator instructs the staff member who entered erroneous data to correct past data entered and use 1 moving forward.
<i>Logic checks</i>	Evaluators can review whether answers to related questions make logical sense. With online surveys, logical skip patterns or checks can be built into the survey to expedite data cleaning.	When reviewing survey data, an evaluator staff member notices that participants who have not had past experiences with intimate partner violence (IPV) are being asked to complete the section on IPV.	The staff member notifies their supervisor to correct the survey logic.

Technique	Description	Example error	Example remedy
<i>Data visualizations</i>	Evaluators can use graphs, such as bar charts or scatter plots, to identify improbable values or outliers.	After conducting a cohort of workshops, an evaluator creates a scatter plot of the relationship between workshop attendance and demographic characteristics, a key research question in the program's evaluation. The evaluator notices that attendance recorded in nFORM is low among mothers, even though written attendance records show that most mothers participated in make-up sessions and completed 90 percent of the workshop content.	The evaluation team convenes a meeting with HMRE program staff to compare nFORM and written attendance records and make sure that nFORM data entry is timely and accurate for this and future cohorts.

Source: Adapted from Institute of Education Sciences (2021).

Tip 3: Interpret unexpected findings alongside the program team

Collecting feedback from people with a variety of experiences - including program staff, evaluators and participants - can promote stronger research and program improvement.

HMRE local evaluators can share preliminary analyses through reports, presentations, or data visualizations on a regular cadence, such as monthly or quarterly, or during technical assistance (TA) meetings with family assistance program specialists (FPSs) and evaluation technical assistance partners (ETAPs). During these calls, evaluators can share preliminary findings and ask program staff to provide additional context for them.

For example, consider an evaluation that finds that more youth are reporting relationship violence at program exit compared to baseline—even after gaining skills to recognize unhealthy relationship patterns or abusive partnerships through the program. Discussing such findings with program staff may suggest ways to change the data collection procedures or instruments to improve data quality (Box 6.4). If evaluators make tweaks to existing processes or procedures, they should clearly document their decisions and report them in the study's final report.

Box 6.4. Reflect on the data collected with program and evaluation staff

Program staff may offer important insights into program operations that could explain unexpected findings and inform tweaks to data collection. Suppose the evaluation team notices findings trending in the opposite direction than expected, such as decreases in relationship satisfaction after program completion. Evaluators can engage staff in discussions about why this is happening. These conversations might reveal that variables are not being reverse-coded correctly or that the program served a greater number of couples in unstable relationships—who break up after receiving more information about healthy relationships. The evaluator can use this context to (1) correct the way data are coded; or (2) describe the study's unexpected findings in their final report.



Tip 4: Cautiously correct data errors

If evaluators find data errors, they can address them by removing duplicates, correcting entry errors, and removing outliers when appropriate. More advanced techniques include imputing data (that is, replacing missing values or incomplete data with estimated values) or transforming variables (that is, changing the scale, format, or distribution of data to make it more consistent or suitable for a particular analysis). Determining how to correct data errors is a nuanced process. Evaluators should carefully select an approach, as it can influence the analyses and types of conclusions that can be drawn (Box 6.5).

It is important to document how each data error is handled so that it can be explained in final reports and other publications. This will increase transparency and allow other researchers to accurately reproduce results. Decisions on how to handle data errors should not be made in isolation, but rather discussed as a program and evaluator team with input from TA providers, FPSs, and other staff or experts who are familiar with the program and data.

Box 6.5. Not all outliers should be removed

Finding an outlier in the data does not necessarily mean it is incorrect and should be removed. Outlier values should only be removed if they are determined to be an error after careful scrutiny.

References

- Avellar, Sarah, Kelley Borradaile, and Debra Strong. "Evaluation Technical Assistance Brief: Tips for Enrolling and Retaining Evaluation Participants." Washington, DC: Mathematica Policy Research, November 2017.
- Healthy Marriage and Responsible Fatherhood. "nFORM Performance Measures and Data Collection Logistics Manual." March 2023. https://hmrfggrantresources.info/sites/default/files/2023-03/PMDC%20Logistics_Mar2023.pdf. Accessed July 7, 2023.
- Healthy Marriage & Responsible Fatherhood (HMRF) Resource Site for 2020 Grantees. "Propelling Forward to Grant Year 3." September 2022. <https://www.hmrfggrantresources.info/resource/propelling-forward-grant-year-3>. Accessed October 27, 2023.
- Institute of Education Sciences. "Guidelines for a Codebook." n.d. <https://ies.ed.gov/ncee/edlabs/regions/central/resources/pemtoolkit/pdf/module-7/CE5.3.2-Guidelines-for-a-Codebook.pdf>. Accessed October 30, 2023.
- Institute of Education Sciences. "Program Evaluation Toolkit." October 2021. <https://ies.ed.gov/ncee/rel/Products/Region/central/Resource/100644>. Accessed July 7, 2023.
- Osborne, Jason W. "Data Cleaning Basics: Best Practices in Dealing with Extreme Scores." *Newborn and Infant Nursing Reviews*, vol. 10, no. 1, March 2010, pp. 37–43. <https://www.sciencedirect.com/science/article/pii/S1527336909001779?via%3Dihub>.
- Van den Broeck, J., S. Argeseanu Cunningham, R. Eeckels, and K. Herbst. "Data Cleaning: Detecting, Diagnosing, and Editing Data Abnormalities." *PLoS Medicine*, vol. 2, no. 10, 2005, pp. 966–970. <https://doi.org/10.1371/journal.pmed.0020267>.

Creating equivalent research groups

Daniel Friend, Angela Valdovinos D'Angelo, Avery Hennigar, Armando Yanez, and Rebecca Piatt

The goal in causal evaluation is to ensure that the differences between two or more study groups are due to the intervention and not to initial (or baseline) variations between groups. To accomplish this, evaluators need to make sure their research groups demonstrate baseline equivalence—meaning the groups are “equivalent” on measured baseline characteristics (for example, age and sex). Evaluators need to monitor and ensure baseline equivalence during the ongoing enrollment into the evaluation and at the end of the study during the analysis stage. Equivalent research groups are essential to accurately estimate the impact of a program. If groups are not equivalent, then they initially differed on characteristics that could be related to outcomes of interest, and the evaluator cannot draw causal conclusions due to the high risk of bias. This brief contains information on: (1) how study design and implementation affect group equivalence, (2) why regular monitoring of equivalence is needed throughout the course of a study, and (3) when demonstrating baseline equivalence for the analytic sample is necessary before conducting analysis. The content of this brief is primarily based on guidance from the [What Works Clearinghouse](#) (WWC).¹



Select and implement an appropriate evaluation design to help develop equivalent groups

Using a randomized controlled trial (RCT) or a quasi-experimental design (QED) enables evaluators to determine whether a program caused a particular outcome—as long as the groups are equivalent at baseline. Below, we briefly describe how each design promotes the creation of equivalent groups:

- **RCT.** RCTs use random assignment to create groups that are equal on observed and unobserved characteristics at the time that participants are randomly assigned—which typically occurs at enrollment into the evaluation (Shadish et al. 2002). For example, a Healthy Marriage and Relationship Education (HMRE) evaluation could randomly assign people to receive the HMRE program or not after they complete a baseline survey at program intake. Randomly assigning participants balances out any initial, potentially systematic differences between the groups, since assigning people by chance alone helps to ensure baseline characteristics of both groups will be inherently balanced, on average. If done correctly, randomization enables evaluators to assume that any observed differences between the two groups are due to the program itself.

¹ The WWC is operated by the U.S. Department of Education, Institute of Education Sciences. The WWC guides program evaluators using causal designs to provide rigorous evidence of program efficacy. The WWC contains information on demonstrating baseline equivalence. The Administration for Children and Families adapts clearinghouse standards, like those from the WWC, to provide evaluators with guidance on analysis and reporting of HMRE local evaluations.

- **QED.** QEDs do not use random assignment and instead evaluators select a nonrandom comparison group. When conducting a QED, evaluators should strive to select a comparison group that is as similar to the treatment group as possible, apart from the offer of services (Box 7.1). For example, an HMRE evaluator who is studying a program offered in a school district may select a comparison group from another school district within the same county that is not offering HMRE services. Although selecting an appropriate comparison group can help create equivalent research groups, it is not the same as an RCT. As participants are not randomly assigned to the treatment condition, the HMRE program, the treatment and comparison groups in the QED might not be equal on the observed *and* unobserved characteristics at baseline. As a result, evaluators must show that the groups are equivalent in their final analytic sample (WWC 2022).

Box 7.1. Options for selecting an appropriate comparison group for a QED

- Select a comparison group from an area (such as a nearby county) where HMRE services are not offered or the services are substantively different from what is being offered to the treatment group.
- Regardless of the area selected, characteristics of the area and its inhabitants should be similar to those where HMRE services are being offered. Before beginning data collection, use existing data sources (for example, Census data) to help determine the comparison group's characteristics before collecting baseline data from them. Consider using proxy measures if it is not possible to directly assess an outcome from the secondary source. For example, federally sponsored, publicly available data sources, like the Census, likely do not include measures of relationship satisfaction and coparenting. Marriage and divorce rates, child support claims, child living arrangements, and so on could potentially serve as proxy measures for these outcomes.
- Construct a baseline or pre-test survey to identify demographics and outcomes of interest that includes key covariates related to the outcomes of interest. For example, covariates of mental health might influence outcomes related to relationship satisfaction. The same could be true for covariates related to poverty and socioeconomic status. Note that it is necessary to apply any eligibility criteria for the treatment group to the comparison group.

Sources: Handley et al. (2018); Shadish et al. (2002).

Regardless of design choice, evaluators will need appropriate procedures to implement the design as intended to achieve equivalence. In an RCT, group equivalence can be compromised if the assignment procedure was not actually random (Office of Adolescent Health 2014). For example, an HMRE local evaluator or other staff member responsible for randomization might move randomly assigned participants to a different group. Or a staff member conducting random assignment might switch a couple originally assigned to the control group into the treatment group because the staff member perceives that the couple needs services—which would compromise the design by creating an imbalance between the two groups. To properly maintain the study's design, evaluators should clearly communicate key information about the study design to all staff, including how it could be compromised, and regularly monitor that random assignment is being implemented as expected. For example, if half of the randomly assigned group received HMRE services and the other half did not, then a periodic check of random assignment status should show a roughly 50/50 split in terms of sample size.

In a QED, nonequivalent groups can occur when the comparison group is selected from a different population than the treatment group (Office of Adolescent Health 2014). For example, groups would likely be nonequivalent if an HMRE local evaluator using a QED to evaluate a program for single Hispanic mothers selects comparison group members from another area (for example, county or community) that consists of married, non-Hispanic women. To help prevent this, evaluators should use the information in Box 7.1 to help select an appropriate group.



Regularly monitor equivalence to determine whether adjustments are needed

Regular monitoring throughout the evaluation can help evaluators assess when groups may be at risk of nonequivalence. As an evaluator sees groups becoming nonequivalent, they can modify their approach. For example, take a scenario where an evaluator is monitoring attrition and finds that, if the current rate of attrition persists, they will be at risk of having a high-attrition RCT. The evaluator then examines several key baseline characteristics and finds they are different between the experimental and control groups—again noting that, if the pattern continues, the groups will not be equivalent. Based on these conditions, the evaluator can work with staff to improve data collection efforts (for example, for participants with the nonequivalent characteristics) to improve response rates and make the treatment and comparison groups more similar at enrollment.

Sometimes, the evaluation procedures (or lack thereof), rather than attrition, contribute to nonequivalence. Early monitoring of equivalence could help an evaluator identify when staff are not adhering to evaluation procedures. For example, staff could incorrectly be randomly assigning participants. In this instance, early monitoring could help the evaluator revise the procedures or retrain staff to address the issue and correct any imbalances between the groups.

Equivalence between the groups is typically demonstrated by calculating effect size estimates (that is, standardized mean differences; Box 7.2) between the treatment and comparison groups on key baseline characteristics that are expected to influence the outcomes of interest (WWC 2022). To construct a monitoring tool for baseline equivalence, the evaluator will need to compute effect sizes to demonstrate baseline equivalence between the research groups. To do this, they will need each research group's means, standard deviations, and sample sizes for each baseline variable they wish to examine (Table 7.1). Using this information, evaluators should select the appropriate effect size estimate calculation and continue to monitor how these estimates fluctuate throughout data collection.

Box 7.2. Using effect sizes to demonstrate baseline equivalence

The suggested calculation of effect sizes depends on the type of variable.

- **Hedges' g.** An effect size index used with continuous variables such as number of children or ratings of relationship satisfaction.
- **Cox's Index.** An effect size index used with dichotomous variables like whether a participant was married at baseline (yes/no).

The [WWC](#) provides more information on how to calculate these effect sizes. Additionally, several effect size calculators are available online or come standard with statistical software.

Source: WWC 2022.

Table 7.1. Illustrative baseline equivalence monitoring tool

Baseline measure	Intervention group			Comparison group			Effect size of difference
	Mean (or %)	Standard deviation	Sample size	Mean (or %)	Standard deviation	Sample size	
<i>Knowledge of healthy relationships</i>	5.2	4.7	100	8.4	6.9	150	0.53
<i>Relationship satisfaction</i>	4	1	102	4.2	1	150	0.20
<i>Coparenting skills</i>	4	1.75	102	3.75	1.25	150	0.17



Demonstrate baseline equivalence before conducting final analyses

Demonstrating baseline equivalence means that an evaluator shows that the participants in the research groups composing their final analytic sample *remain* similar in terms of their baseline characteristics. Design choices and procedures undoubtedly influence how equivalent research groups are formed, but the need to demonstrate baseline equivalence will differ depending on whether the study is an RCT or a QED. Evaluators using a QED *must always* demonstrate baseline equivalence—as assignment to the treatment or comparison groups is nonrandom (WWC 2022). As a result, an evaluator employing this design needs to demonstrate baseline equivalence to ensure the research groups are similar, to attribute differences in observed outcomes at the end of the study to the program.

An RCT must demonstrate baseline equivalence if it has high rates of attrition. If an RCT loses too many participants over time, the groups may no longer be equivalent in terms of their initial characteristics, particularly when those differences are related to outcomes of interest (Shadish et al. 2002). For example, attrition-related bias would be introduced if participants who did not participate in follow-up data collection were, on average at baseline, less satisfied with their relationship than those who did respond to follow-up surveys. RCTs with high rates of attrition must demonstrate that the study groups are still equivalent at baseline based on who completed each round of data collection and for each outcome the study seeks to examine (WWC 2022). See the fourth brief in this toolkit on understanding and mitigating attrition for more information on calculating and determining when attrition is “high.”

Demonstrating baseline equivalence should be done *before* conducting any analyses. This is because, if the groups are not equivalent, the evaluator may be able to adjust their analysis to account for this. Evaluators can determine whether the two groups are similar by comparing whether the effect size estimates for each characteristic fall within an acceptable range.

Below is guidance for interpreting differences in effect sizes to determine whether groups are equivalent (WWC 2022):

- **Effect size is less than 0.05.** This indicates the groups are equivalent because baseline differences are relatively small and thus not a concern. In this case, the evaluator can proceed with their planned analysis. In Table 7.1, the coparenting variable falls within this range.
- **Effect size is greater than 0.05, but less than or equal to 0.25.** This indicates the groups are not equivalent. As a result, a statistical adjustment is required if the effect size is within this range. For example, the evaluator should use regression covariate adjustments or an analysis of covariance—where the covariate is any baseline variable that falls within this effect size range. In Table 7.1, the relationship satisfaction variable falls within this range.
- **Effect size is greater than 0.25.** This indicates the groups are not equivalent, and the differences are too large for statistical adjustment. In this instance, evaluators should consult their evaluation technical assistance providers and federal project specialists for more guidance on this issue. There may be additional analytic options for the evaluator to try that are too nuanced to detail in this brief. Typically, though, this means that the study cannot demonstrate baseline equivalence and can no longer be considered a causal evaluation, as the threat of bias is too great to make causal claims. In Table 7.1, the knowledge of healthy relationships variable falls within this range.

References

Handley, M.A., C.R. Lyles, C. McCulloch, and A. Cattamanchi. "Selecting and Improving Quasi-Experimental Designs in Effectiveness and Implementation Research." *Annual Review of Public Health*, vol. 39, 2018, pp. 5–25. <https://doi.org/10.1146/annurev-publhealth-040617-014128>.

Office of Adolescent Health. "Baseline Inequivalence and Matching." U.S. Department of Health and Human Services, 2014. <https://opa.hhs.gov/sites/default/files/2020-07/baselineinequivalence-tabrief.pdf>. Accessed June 15, 2023.

Shadish, W.R., T.D. Cook, and D.T. Campbell. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Houghton Mifflin, 2002.

What Works Clearinghouse. *What Works Clearinghouse Procedures and Standards Handbook, Version 5.0*. U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance (NCEE), 2022. <https://ies.ed.gov/ncee/wwc/Handbooks>.

Powering an evaluation to detect effects

Avery Hennigar, Daniel Friend, Rebecca Piatt, and Angela Valdovinos D'Angelo

An important underlying principle when conducting statistical analyses for any evaluation is that no matter how much data an evaluator collects, they can never prove or disprove a hypothesis with absolute certainty. Instead, evaluations can present findings that can reflect the outcomes or other characteristics of a population or populations based on what was observed in the sample. It is always possible that an evaluator's conclusion could be wrong because sometimes, just by chance, a sample does not represent the population. Evaluators should carefully design and execute an evaluation that minimizes the risk of any errors that could interfere with their ability to draw accurate conclusions.

This brief first describes common errors in statistical testing, then recommends five steps to help evaluators conduct a power analysis and calculate effect sizes they can use throughout an evaluation.



Understanding Type I and Type II errors

In statistical testing, there are two types of errors that could lead to incorrect conclusions:

- **Type I error.** A false positive, which happens when an evaluator concludes a relationship exists when it does not. For example, an evaluator finds that a Healthy Marriage and Relationship Education (HMRE) program improves relationship skills for youth, but in reality, it does not.
- **Type II error.** A false negative, which happens when an evaluator concludes a relationship does not exist when it does. For example, an evaluator finds that an HMRE program does not improve relationship skills for youth, but in reality, it does.

An evaluator can aim to minimize a Type I error by choosing a statistical significance level. Significance level or p -value represents the chance of committing a Type I error. For example, if the p -value an evaluator selects is 0.05, this means there is a 5 percent chance of committing a Type I error.

To avoid a Type II error, an evaluator needs to ensure the evaluation is adequately powered to detect effects. Power is driven by two factors: (1) the estimated size or magnitude of the effect (also called effect size); and (2) the sample size (Box 8.1). Before beginning an evaluation, evaluators need to minimize the occurrence of a Type II error by determining what effect size to expect and computing the sample size that will give them enough power to detect that effect size.

Box 8.1. Key terms

Effect size. A quantitative indicator that measures how strongly the HMRE program is related to an outcome. For example, in a causal evaluation, it is the magnitude of difference between two groups. In a descriptive evaluation, this could be the magnitude of difference between two time points.

Minimum detectable effect (MDE). The smallest effect that is likely to produce an impact estimate at a given statistical significance level.

Power analysis. An assessment of a study's chance of detecting meaningful impacts or effects if they exist.

Sources: Bloom 1995; IES 2021.

Effect sizes help put statistically significant findings in context and make sense of them. Although a p -value can provide information about whether a difference exists above and beyond chance, it does not reveal the size of this difference (Sullivan and Feinn 2012). In addition, it is possible that even if a difference is significant, it might not be meaningful. Evaluators should consult previous literature in the field, program staff, and participants' feedback to determine if significant differences are practically or substantively meaningful.

Accurately estimating the effect size is a process that starts before any data are collected. Designing and executing a study with adequate power to detect a statistically significant effect (if one exists) take careful planning and monitoring.



Step 1. Identify outcomes of interest that closely align with the program's theory of change

The first step in a power analysis is to determine the expected effect size. To do this, the evaluator must identify the outcome measures of interest. For an evaluation to be successful, there should be a clear, well-supported path between how the intervention is expected to achieve its results and in which aspects of a participant's life changes are likely to be evident. Selected outcome measures should reflect what the intervention is expected to influence (Coster 2013). A logic model or theory of change can illustrate these linkages, and the evaluator should consult it when selecting outcome measures.

To identify specific outcomes, evaluators can consult the literature as well as speak with program and evaluator staff to answer the following:

- What is the program trying to change?
- Does this measure align with the program's activities?
- How can this outcome be measured?
- Would assessing this outcome be meaningful to participants in the program?



Step 2. Determine the minimum detectable effect for each outcome of interest

After identifying an evaluation's outcomes, evaluators should next identify the minimum detectable effect (MDE) for each outcome (Box 8.1). To do this, evaluators can consult existing research to determine what change is reasonable to expect for a given outcome. The MDE should align with what other HMRE or similar programs have found in previous studies with similar participants at similar time points. For example, if an evaluator is examining an HMRE program serving primarily Spanish-speaking couples with a curriculum that integrates HMRE and content on economic stability, the evaluator should identify research on a similar program serving Spanish-speaking couples to inform MDE estimates.

Evaluators may consider reviewing findings from earlier large HMRE evaluations, such as [Building Strong Families](#),¹ [Supporting Healthy Marriage](#),² [Parents and Children Together](#),³ or [Strengthening Relationship Education and Marriage Services](#).^{4,5} Evaluators can use the effect sizes these evaluations detected to determine if their own estimated MDE is realistic. Note, however, that these studies have large sample sizes, and therefore have the power to detect smaller effects than studies with smaller sample sizes.

¹ Office of Planning, Research, and Evaluation. "Building Strong Families, 2002-2013." Administration for Children and Families. <https://www.acf.hhs.gov/opre/project/building-strong-families-2002-2013>.

² Office of Planning, Research, and Evaluation. "Supporting Healthy Marriages, 2003-2014." Administration for Children and Families. <https://www.acf.hhs.gov/opre/project/supporting-healthy-marriages-2003-2014>.

³ Office of Planning, Research, and Evaluation. "Parents and Children Together, PACT Evaluation, 2011-2020." Administration for Children and Families. <https://www.acf.hhs.gov/opre/project/parents-and-children-together-pact-evaluation>.

⁴ Office of Planning, Research, and Evaluation. "Strengthening Relationship Education and Marriage Services (STREAMS), 2015-2022." Administration for Children and Families. <https://www.acf.hhs.gov/opre/project/opre/research/strengthening-relationship-education-and-marriage-services-streams>.

⁵ Evaluators can also review the local evaluation final reports from the 2015–2020 cohort of Healthy Marriage and Responsible Fatherhood programs. <https://www.acf.hhs.gov/ofa/programs/healthy-marriage-responsible-fatherhood/data-reports>.



Step 3. Run a power analysis to determine the sample size necessary to detect the MDE

The next step is to conduct a power analysis for each of the selected outcomes (Box 8.1). A power analysis uses MDEs along with other factors to estimate the likely sample size needed to statistically detect a difference of the specified effect size. There are several key pieces of information evaluators need to conduct most power analyses:

- 1. Power and significance level.** Power level is similar to the significance level or p -value discussed earlier. The power level represents the chances of *not* committing a Type II error. The minimum power level threshold is typically 80 percent, indicating that an evaluation has an 80 percent chance of not committing a Type II error. Power and significance level have an inverse relationship—increasing the power level increases the probability of a Type I error. The same is true for significance level; setting a lower significance level decreases the probability of a Type I error, but increases the probability of a Type II error. As a result, there is always a trade-off in these decisions. An evaluator must weigh the severity of each error type to determine where to set each threshold.
- 2. MDE.** This is the smallest effect the evaluation is powered to detect. In social science research, Cohen's d , a standard effect size metric, is typically used for MDEs, but other metrics exist given the type of outcome (Table 8.1). MDEs are often classified as small ($d = 0.2$), medium ($d = 0.5$), or large ($d \geq 0.8$) (Carson 2012). These broad categories can be used as a general guide, but MDEs may be smaller or larger depending on the context of the evaluation, such as the design of the study, the timing of follow-up measures, the characteristics of participants, and the quality of the outcome measures (Ferguson 2009; IES 2021; Sullivan and Feinn 2012). For example, an evaluator who seeks to measure outcomes at two years post-program completion might expect smaller effects, as treatment effects tend to diminish over time. Or an evaluator may be interested in an outcome that is hard to measure or has no reliability information. In this instance, the evaluator might also expect smaller effects due to the measurement quality.
- 3. Type of hypothesis test.** The type of statistical hypothesis testing can influence power. For example, a one-sided test will require a smaller sample to adequately power an evaluation than a two-sided test does. Evaluators typically use a two-sided test because they are interested in whether the program has an effect regardless of whether the average difference between the research groups is higher or lower.

Table 8.1. Description of effect size estimates

Index	Description	General guidance for interpreting an effect size	Notes
Between groups			
Cohen's d	$d = M_1 - M_2 / s$ where $M_1 - M_2$ is the difference between the group means, and s is the standard deviation of either group	Small: 0.2 Medium: 0.5 Large: 0.8 Very large: 1.3	A useful measure during the planning stage when conducting a power analysis
Odds ratio (OR)	$\frac{(\text{Group 1 odds of outcome})}{(\text{Group 2 odds of outcome})}$ If OR = 1, the odds of the outcome are equally likely in both groups	Small: 1.5 Medium: 2 Large: 3	Used for binary variables to compare the odds of an outcome occurring as a result of one intervention compared with another

Index	Description	General guidance for interpreting an effect size	Notes
Relative risk or risk ratio (RR)	$\frac{(\text{Group 1 probability of outcome})}{(\text{Group 2 probability of outcome})}$ If RR = 1, the outcome is equally probable in both groups	Small: 2 Medium: 3 Large: 4	Compares the probabilities of outcomes occurring from one intervention to another
Measures of association			
Pearson's <i>r</i> correlation	Range from -1 to 1	Small: 0.2 Medium: 0.5 Large: 0.8	Measures the degree of linear relationship between two quantitative variables
<i>r</i> ² coefficient of determination	Range from 0 to 1 Typically expressed as a percentage	Small: 0.04 Medium: 0.25 Large: 0.64	Represents the proportion of variance in one variable explained by the other

Note: Adapted from Sullivan and Feinn (2012).

There are many statistical programs that evaluators can use to conduct their power calculations, including free or open-source ones such as [G*Power](#). Some of these calculators have user-friendly interfaces that produce estimates of sample size based on the details the user enters about the evaluation, such as MDEs, type of statistical test, and desired significance level.

Power analyses are outcome specific, so evaluators should choose a sample size large enough to give the study power to detect all outcomes. For example, an evaluator might pick the largest sample size specified across all the outcomes to make sure the study is powered to detect all the outcomes' effects. It can also be a good idea to consider different scenarios and run a few calculations to understand what the study will be powered for if the sample size is not achieved (for example, assuming the program was only able to enroll enough people to achieve 75 percent of the intended sample size). This allows evaluators to identify which outcomes are well powered or underpowered if the desired sample size is lower than expected.



Step 4. Review power and sample size estimates with program staff to determine monthly enrollment targets

Once evaluators determine the sample size the evaluation needs, they can discuss these results with program staff and ask if these enrollment goals seem feasible. It can help to break down the desired sample size into monthly enrollment goals. When calculating monthly targets, evaluators should be sure to consider times of year when enrollment might be slow (for example, in November and December around the holidays) and devise a plan to over-enroll in busier months to make up for it. If the monthly enrollment goals are not feasible as the evaluation progresses, evaluators may need to go back to Step 1 and revise their outcome measures to select an outcome that could be adequately powered if there were a more realistic estimate of sample size.

Running initial power calculations when designing a study is an important first step, but evaluators should remember they might have to adjust power calculations as the evaluation progresses. Evaluators should consider running these calculations on a regular schedule (at least annually) to determine how the pace of enrollment is aligning with power estimates.



Step 5. Calculate and report the effect size estimate and compare it to the broader field

After completing follow-up data collection and conducting the planned evaluation analyses, a best practice is to calculate the actual effect size for the outcomes of interest. Table 8.1 provides an overview of common effect size estimates that evaluators can calculate. Evaluators should report the effect size estimates in their final report and other study publications, such as in the abstract and results sections. Effect size estimates are important to the field—for example, in conducting meta-analyses and to inform MDEs in future studies.

References

- Bloom, H. S. "Minimum Detectable Effects: A Simple Way to Report the Statistical Power of Experimental Designs." *Evaluation Review*, vol. 19, no. 5, 1995, pp. 547–556.
- Coster, W. J. "Making the Best Match: Selecting Outcome Measures for Clinical Trials and Outcome Studies." *American Journal of Occupational Therapy*, vol. 67, no. 2, pp. 162–170.
- Ferguson, C. J. "An Effect Size Primer: A Guide for Clinicians and Researchers." *Professional Psychology: Research and Practice*, vol. 40, no. 5, 2009, pp. 532–538.
- Institute for Education Sciences. "Effect Size Basics: Understanding the Strength of a Program's Impact." 2021. https://ies.ed.gov/ncee/edlabs/regions/west/relwestFiles/pdf/4-2-3-14_Effect_Size_Infographic_Final_508c.pdf.
- Office of Evaluation Sciences. "Effect Size and Evaluation: The Basics." <https://oes.gsa.gov/assets/files/effect-size-evaluation-basics.pdf>.
- Sullivan, G. M., and R. Feinn. "Using Effect Size – Or Why the P Value Is Not Enough." *Journal of Graduate Medical Education*, vol. 4, no. 3, 2012, pp. 279–282.