

Roadmaps to Building Evidence in Child Welfare

RESEARCH REPORT

A Guide to Writing High-Quality Evaluation Reports in Child Welfare

Sarah Prendergast

Katrina Brewsaugh

OPRE Report #2022-43

August 2022

A Guide to Writing High-Quality Evaluation Reports in Child Welfare

Sarah Prendergast and Katrina Brewsaugh

OPRE REPORT #2022-43
AUGUST 2022

SUBMITTED TO

Kathleen Dwyer and Alysia Blandon, project officers
Office of Planning, Research, and Evaluation
Administration for Children and Families
US Department of Health and Human Services

Contract Number: HHS P233-2015-000641

SUBMITTED BY

Michael Pergamit, Principal Investigator
Urban Institute
500 L'Enfant Plaza SW
Washington, DC 20024

This report is in the public domain. Permission to reproduce is not necessary. Suggested citation: Prendergast, Sarah, and Katrina Brewsaugh. 2022. *A Guide to Writing High-Quality Evaluation Reports in Child Welfare*. OPRE Report #2022-43. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, US Department of Health and Human Services.

DISCLAIMER

The views expressed in this publication do not necessarily reflect the views or policies of the Office of Planning, Research, and Evaluation, the Administration for Children and Families, or the US Department of Health and Human Services. This report and others sponsored by the Office of Planning, Research, and Evaluation are available at <http://www.acf.hhs.gov/programs/opre>. Cover image by Ridofranz/iStock.



Sign-up for the OPRE newsletter



Follow OPRE on Twitter @OPRE_ACF



[facebook.com/OPRE.ACF](https://www.facebook.com/OPRE.ACF)



Connect on LinkedIn [company/opreacf](https://www.linkedin.com/company/opreacf)



Follow OPRE on Instagram @opre_acf



Contents

Acknowledgments	iv
Executive Summary	v
A Guide to Writing High-Quality Evaluation Reports in Child Welfare	1
Guide Overview	1
What to Include in the Background Section of Your Evaluation Report	3
What to Include in the Methods Section of Your Evaluation Report	8
How to Report Implementation Evaluation Results	18
How to Report Impact Evaluation Results	19
What to Include in the Discussion Section of Your Evaluation Report	23
Guide Summary	24
Appendix A. Evaluation Report Checklist	25
Appendix B. Summary of Clearinghouse Insights in This Guide	28
Appendix C. Additional Clearinghouse Resources	39
References	41
About the Authors	42
Statement of Independence	43

Acknowledgments

This report was funded by the Administration for Children and Families of the Department of Health and Human Services. We are grateful to them and to all our funders, who make it possible for Urban to advance its mission. We would also like to thank our project officers Kathleen Dwyer and Alysia Blandon for their guidance and input as well as Cara Kelly and other reviewers at OPRE and the Children’s Bureau for their valuable comments. The Supporting Evidence Building in Child Welfare project (<https://www.acf.hhs.gov/opre/project/supporting-evidence-building-child-welfare-2016-2025>) includes conducting rigorous evaluations of child welfare programs, practices, and policies as well as building evaluation capacity in the child welfare field. All the Roadmaps to Building Evidence in Child Welfare products can be found at <https://www.urban.org/projects/roadmaps-building-evidence-child-welfare>.

The views expressed are those of the authors and should not be attributed to the Urban Institute, its trustees, or its funders. Funders do not determine research findings or the insights and recommendations of Urban experts. Further information on the Urban Institute’s funding principles is available at [urban.org/fundingprinciples](https://www.urban.org/fundingprinciples).

We appreciate the representatives from six clearinghouses who participated in interviews to inform our project and reviewed this guide: Blueprints for Health Youth Development, California Evidence-Based Clearinghouse for Child Welfare, National Institute of Justice CrimeSolutions, Home Visiting Evidence of Effectiveness Review, Title IV-E Prevention Services Clearinghouse, and Social Programs That Work.

Executive Summary

Introduction

Although the body of evidence on the effectiveness of child welfare programs is growing, to date very few child welfare programs have received the highest ratings from clearinghouses. Clearinghouses summarize the available evidence on programs and rate the evidence of their effectiveness in a systematic, transparent way. In addition to the sound design and execution of an evaluation, a high-quality evaluation report is also necessary to document evidence on the effectiveness of your child welfare program. All evaluation reports should contain the same basic elements so that readers can understand the research and judge its quality. Different clearinghouses also set specific evaluation reporting standards for authors to follow. You can ensure your evaluation report is high quality by following best practices. You can increase your chances of meeting clearinghouse criteria by understanding the various clearinghouses' reporting standards.

This report is one of several activities to support evidence building in child welfare through a contract to the Urban Institute funded by the Administration for Children and Families of the Department of Health and Human Services. The Supporting Evidence Building in Child Welfare¹ project is conducting rigorous evaluations of child welfare programs, practices, and policies as well as helping building evaluation capacity in the child welfare field.

Purpose

The purpose of this guide is to support evaluators in writing high-quality reports that can contribute to the evidence base of child welfare programs and support children's and families' well-being. This guide walks you through the steps required to write a high-quality impact evaluation report. It also includes a checklist (appendix A) and a summary of clearinghouse insights (appendix B) that you can easily reference as you write.

¹ For more information, see "Supporting Evidence Building in Child Welfare," US Department of Health and Human Services (HHS), Administration of Children and Families (ACF), Office of Planning Research and Evaluation (OPRE), accessed September 1, 2021, <https://www.acf.hhs.gov/opre/project/supporting-evidence-building-child-welfare-2016-2025>.

Key Findings and Highlights

High-quality evaluation reports can help readers understand a program’s effectiveness, make decisions about whether to fund or adopt a program, and replicate an evaluation’s findings in another study.

To write a high-quality evaluation report,

- include background, methods, results, and discussion sections;
- include detailed information about the program and how it was implemented during the evaluation;
- describe the evaluation design and how it was carried out;
- describe the analysis plan and how analyses were conducted;
- describe the evaluation’s findings and limitations; and
- summarize the findings and explain how the evaluation contributes to the child welfare field.

Methods

The Urban Institute project team summarized their expertise in writing high-quality evaluation reports to develop this guide. In addition, we reviewed information available on the websites of six relevant clearinghouses about their evidence review requirements for child welfare program evaluations. We also conducted interviews with representatives from each of the six clearinghouses to better understand their requirements and gain insights into the types of information commonly missing in reports.

Recommendations

Authors should ensure evaluation reports include all the information needed to clearly communicate the evaluation’s design, execution, and findings. In general, clearinghouse reviewers noted that authors should err on the side of including more detail, rather than less, to ensure an accurate review. Evaluators seeking a review from a specific clearinghouse should consult that clearinghouse’s published standards when writing evaluation reports.

A Guide to Writing High-Quality Evaluation Reports in Child Welfare

Guide Overview

Developing an evidence-based program involves years of hard work and intensive resources spent on program development, implementation, and continued evaluation. You can share information on your program's effectiveness with stakeholders through evaluation reports. Over time, your program's evaluation reports become a body of evidence that tells a story about if and how the program works. When carefully prepared, reports contain information that help readers understand your program's effectiveness and whether to adopt or fund the program.

Stakeholders can review evaluation reports on their own or rely on systematic reviews carried out by evidence clearinghouses. Clearinghouses review available program materials and published evaluation reports to rate how well a program works based on standardized criteria. For this reason, many states and funders require the use of programs that have a positive evidence rating from a clearinghouse. Programs with high ratings may have a better chance of achieving more funding or program adoption. For example, the Family First Prevention Services Act allows states to receive federal Title IV-E funds for prevention services only for programs meeting specific evidence criteria as determined by an independent systematic review.²

Evaluation reports also provide critical information to program developers to help them determine whether their program works as intended or if they should make changes. Researchers may use an evaluation report as a guide to replicate the prior evaluation's findings. Though evaluation reports play a large role in the evidence-building process, their usefulness varies widely, depending on the quality of information included.

Throughout this guide, we discuss best practices for writing the major sections of an impact evaluation report: background, methods, results, and discussion. For each section, we list key information that should be included and discuss why including this information is important. We also include examples from a child welfare evaluation report. Appendix A provides a checklist that you can easily reference as you write. We note reporting practices that clearinghouse staff members mentioned as important for the evidence review

² For more information, see "Title IV-E Prevention Program," HHS, ACF, OPRE, updated September 3, 2021, <https://www.acf.hhs.gov/cb/title-iv-e-prevention-program>.

process. We summarize these practices in appendix B and include additional clearinghouse resources in appendix C.

Box 1 lists the six clearinghouses interviewed for this guide.

BOX 1

Interviews with Clearinghouse Representatives

The Urban Institute conducted interviews with the six clearinghouses listed below to understand better their evidence review requirements and gain insights into the types of information commonly missing from reports. Two clearinghouses, Prevention Services and HomVEE, have published their own reporting guidelines, and we encourage authors to consult them if seeking a rating from these clearinghouses.

- California Evidence-Based Clearinghouse for Child Welfare (CEBC)^a
- Blueprints for Healthy Youth Development (Blueprints)^b
- Title IV-E Prevention Services Clearinghouse (Prevention Services)^c
- Social Programs That Work (Social Programs)^d
- National Institute of Justice CrimeSolutions (CrimeSolutions)^e
- Home Visiting Evidence of Effectiveness (HomVEE)^f

^a CEBC, accessed September 1, 2021, <https://www.cebc4cw.org/>.

^b Blueprints, accessed September 1, 2021, <https://www.blueprintsprograms.org/>.

^c Prevention Services, accessed September 1, 2021, preventionservices.acf.hhs.gov.

^d Social Programs, accessed September 1, 2021, <https://evidencebasedprograms.org/>

^e CrimeSolutions, accessed September 1, 2021, <https://crimesolutions.ojp.gov/>.

^f HomVEE, accessed September 1, 2021, <https://homvee.acf.hhs.gov/>.

Two of the clearinghouses we interviewed, CEBC and HomVEE, may sometimes require programs to submit published journal articles for consideration.³ Journal articles often have page limits that make it challenging to include all the information noted in this report that should be included in an evaluation report. However, many journals allow additional information to be published through online supplements, and all clearinghouses accept such materials as part of their review.

³ If a home visiting model's rating is based solely on randomized controlled trials, then HomVEE requires at least one favorable outcome to be reported in a peer-reviewed journal. CEBC only includes peer-reviewed journal articles in evidence reviews.

A strong evaluation report is only one part of the evidence-building process.⁴ Equally important is following best practices in study design and execution. We encourage readers to review our companion guide, *Ten Key Design Elements for Rigorous Impact Evaluations in Child Welfare: A Desk Reference for Evaluators* (Brewsaugh and Prendergast 2022), for information on key steps for designing strong impact evaluations and clearinghouses' acceptance of certain designs.

What to Include in the Background Section of Your Evaluation Report

A high-quality evaluation report begins with a description of the program and relevant background information. Throughout this guide, we use *program* to describe any intervention that seeks to promote change (e.g., prevention programs, practice models, policies).

Provide an Overview of the Program

Describe the program's main components as well as its goals and desired outcomes. Clearly describe the main components to help others replicate the study. Discuss the program's conceptual framework, theory of change, or logic model to explain how the program's components are hypothesized to impact the outcomes of interest. Program descriptions help readers understand why the program was developed and how it is intended to work.

The following information should also be included:

- the target population (i.e., intended participants of the program)
- typical dosage (i.e., duration, frequency)
- the implementation setting where the program takes place (e.g., in the home, in a clinic, online)
- who delivers the program or service

⁴ For more information on concepts discussed in this guide, see "Child Welfare Evidence-Building Academy: A Training on Rigorous Evaluation Design and Implementation," HHS, ACF, OPRE, October 20, 2021, <https://www.acf.hhs.gov/opre/training-technical-assistance/child-welfare-evidence-building-academy-training-rigorous>.

This information about the program allows readers to compare your program with similar programs. An evaluation report that includes these details also helps stakeholders determine if the model aligns with what they seek to implement in their agency.

State whether the program has a book, manual, or other material that clearly and completely describes how it should be implemented. Tell the reader where the materials can be located (e.g., a website). Program manuals or similar materials allow readers to adopt the program with fidelity, or adherence, to the model.

Some programs require people who deliver the program to participants (i.e., program implementers) to pay per use of a tool, pay for training, or hire the developer or franchisee to provide the program. State whether the program (or parts of it) are proprietary. This information can help stakeholders determine whether the program fits within their budget or intended implementation method.

Clearinghouse insights: All clearinghouses require authors to describe the program. Prevention Services requires authors to submit a manual that details program components, implementation practices, and fidelity measures. If the program already has a manual or other material that includes the information listed above, it may be acceptable to cite the program materials instead of providing a full description in the report, so long as the materials are publicly available. Clearinghouse staff members said that evaluation reports commonly lack a citation to the program manual (Prevention Services, CEBC) and a thorough discussion of program components (CrimeSolutions). Two clearinghouses (CrimeSolutions and Blueprints) require authors to discuss the conceptual framework and theoretical base.

Provide an Overview of the Evaluation

Although the program overview describes the program in general, you should begin the evaluation overview with a description of the program as implemented during the evaluation period. Describe the timing of the intervention (start and end dates) and the implementation setting (i.e., home, office, clinic, online), indicating whether it was a usual care setting. Describe who delivered the program or services and provider qualifications. Also describe the geographic setting in which the program was implemented, including demographics of the population and whether the area was rural, urban, or suburban. This information can help readers understand which types of settings the findings apply to.

Next, describe the evaluation design. Impact evaluations test whether the program produces outcomes that are better than they would be without the program. State whether the impact evaluation used a randomized controlled trial (RCT) or quasi-experimental design (QED).

- RCT: Participants are assigned to receive the program (i.e., treatment group) or not (i.e., control group) based on a random process.
- QED: Participants are assigned to receive the program or not (i.e., comparison group) through a nonrandom process that usually involves matching, cutoff scores, or different time frames.

In RCTs and QEDs the control or comparison group condition of an evaluation still might receive services outside of the evaluation. For example, someone who is in the control/comparison condition of your evaluation could still participate in other services, like a universal preschool program, for which they are eligible. This situation is referred to as *services as usual*. When services as usual include minor enhancements for the purposes of the study, they are referred to as “enhanced” services as usual. For example, you might offer an informational parenting guide to participants in the control/comparison condition of your evaluation so that they receive something for participating. Describe the services that participants in the control/comparison condition of your evaluation may have received.

Box 2 describes additional considerations for different types of impact evaluation designs.

BOX 2

Considerations for Clustered, Multisite, and Multi-Arm Designs

Clustering (also referred to as nesting) refers to study designs in which study participants are related to some participants but not others (e.g., children within the same family or caseworkers within the same implementation site). When a program is implemented across more than one location, it is referred to as a multisite design. If the evaluation design was a clustered or multisite design, throughout this guide we indicate when you should mention whether any site-level differences in the program or any aspect of the study design occurred.

Some evaluation designs may include multiple treatment groups made up of different programs (program 1; program 2) or adaptations of a program (program; program plus adaptation). For these types of “multi-arm” evaluations, describe each arm of the study in the evaluation overview; clearly describe the evaluation design; and ensure that information covered in the subsequent sections addresses each arm of the study.^a

^a For more information on how to report multi-arm designs, see Edmund Juszczak, Douglas G. Altman, Sally Hopewell, and Kenneth Schulz, “Reporting of Multi-Arm Parallel-Group Randomized Trials Extension of the CONSORT 2010 Statement,” *JAMA* 321, no. 16 (2019): 16,010–1620.

Many impact evaluations also include an implementation evaluation. These evaluations collect information about how a program was implemented during the evaluation period and assess the extent to

which the program followed the program manual (i.e., fidelity). Sometimes program developers or implementers adapt a program, either intentionally or unintentionally. Adaptations could be made, for example, to the mode of delivery, number or frequency of intervention sessions, or type of service provider. Provide an overview of the implementation evaluation. Be sure to state whether any adaptations were made during the evaluation period and why. This information helps readers understand the extent to which the model was implemented with fidelity and is consistent with the original model described in the program manual or materials.

Discuss the services that the comparison group received (if any). If the comparison group received services, provide an overview of the services and who delivered them. Mention whether they were services as usual and whether services as usual included another manualized or evidenced-based program. If the comparison group received no services, indicate whether the group had the opportunity to participate at a later time (i.e., if there was a waiting list). As the use of evidenced-based programs has increased, it is common for services as usual in some locations to consist of one or more manualized programs. In such cases, clearly explain that those programs were considered services as usual for the study population (i.e., the sample of people selected to be in the evaluation at baseline). If this distinction is not explained clearly, the evaluation could be mistakenly classified by reviewers as a comparative study.⁵ The services that the control or comparison group received could impact the evaluation's results. Providing this information helps readers accurately interpret the findings.

Clearinghouse insights: Some clearinghouses examine the number and type of adaptations made to a program when determining whether the program is a “new” program. If the adaptations constitute a new program, the study will not be considered evidence for the effectiveness of the original model. CEBC, Social Programs, and Prevention Services consider whether a program was implemented in a usual care setting or other setting in their ratings. Prevention Services and CrimeSolutions only review evaluations in which the comparison group received either no services or services as usual. All clearinghouses said they allow an evidence-based program to be considered as services as usual if the evidence-based program truly reflects the services that participants would receive, regardless of the evaluation.

⁵ Comparative studies have more than one treatment group and often no control/comparison group. The goal is to see if one program works better than another program. A comparative study may test different programs (e.g., Alcoholics Anonymous and cognitive behavioral therapy) or different adaptations of a single program (e.g., in-home, in-office, or online).

List Research Questions and Hypotheses

List each research question and/or hypothesis. Clearly stated research questions and hypotheses are necessary for readers to understand the study design and what was tested.⁶ It is good practice to preregister⁷ the evaluation study, including research questions and/or hypotheses. If the study was preregistered, then the questions and hypotheses stated in this section should align with the preregistration plan. Preregistering the study can increase readers' confidence in the findings and overall study by demonstrating that you determined the research questions, analysis plans, and study design before you started analyzing the data. Preregistration guards against the temptation to keep running analyses until a favorable result is obtained, possibly by chance. Provide the reader with a link or other information on how to find the study's preregistration information.

If the design includes multiple outcomes, note which are primary and secondary outcomes.

- Designating an outcome as primary lets readers know that the outcome is closely aligned with the program's theory of change and that the change in that outcome was expected. Primary outcomes are sometimes referred to as confirmatory because they are tested to confirm whether your hypothesis was supported.
- Designating an outcome as secondary means that the outcome may provide context for the change in primary outcomes. Secondary outcomes should not be used to say whether the program was effective. Sometimes secondary outcomes are exploratory. This means that you did not have a hypothesis about the effect of the program on this outcome but explored it to determine whether to include it in future hypotheses.
- Effects on outcomes for subgroups may be considered primary or secondary.

If the study included an implementation evaluation, list each implementation research question.

Clearinghouse Insights: Clearinghouses vary in how they review studies that test for multiple outcomes. CEBC and CrimeSolutions look for a preponderance of evidence across all outcomes tested within a topic area. This may be difficult to achieve if many outcomes are specified for the evaluation.

⁶ You should develop your research questions and hypotheses when designing your evaluation. For more information on how to write clear research questions, see "Child Welfare Evidence-Building Academy: A Training on Rigorous Evaluation Design and Implementation," HHS, ACF, OPRE.

⁷ Evaluations can be preregistered by submitting a final evaluation plan to a public study repository (e.g., ClinicalTrials.gov, <https://www.clinicaltrials.gov>; Open Science, <https://osf.io/prereg/>; or American Economic Association, <https://www.socialsciceregistry.org>; all accessed September 1, 2021) before beginning an evaluation. Information submitted to the repository may include the evaluation's sampling procedures, sample size, group assignment process, hypotheses, outcome measures, data collection procedures, and analytic plan.

Prevention Services reviews outcomes individually. Blueprints only reviews outcomes that measure changes in behavior (e.g., drug use, risky activities, discipline techniques) and align with the program’s intent. CrimeSolutions only reviews studies with at least one behavioral outcome and will review nonbehavioral outcomes as secondary outcomes that would be weighted less than primary, behavioral outcomes.

Social Programs encourages researchers of new programs to specify only a few primary outcomes and designate the rest as secondary. It generally looks for favorable effects on primary outcomes that have clear policy importance. Blueprints examines whether a significant favorable change in a targeted outcome can be attributed to the program and that there are no harmful effects. HomVEE rates manuscripts about studies based on the highest evidence rating of any individual finding. Preregistration is not currently required by any of the six clearinghouses, but it does play a role in prioritizing a study for review. It may become a requirement in the future. Preregistration is particularly important for studies that include multiple outcomes or subgroup analyses because it gives reviewers confidence that reported findings were not simply the result of a “fishing expedition.”

What to Include in the Methods Section of Your Evaluation Report

The methods section of a report should include detailed information about how the evaluation was conducted.

Describe How the Sample Was Selected

In general, explain the eligibility, inclusion/exclusion criteria (noting whether anyone was excluded), identification of participants, recruitment procedures, and the sample selection process. Describe the population from which the sample was drawn and the geographic setting and time frames from which the treatment and comparison groups were drawn (if they differed between groups). If the evaluation obtained informed consent from participants, describe the informed consent process and the timing of consent (i.e., before or after group assignment). How you describe the sample will differ based on whether the evaluation is an RCT or QED.

FOR RCTS

Describe the randomization process by stating who performed the randomization and when, the randomization method, the unit of randomization, and whether participants were blind to group assignment

(i.e., whether participants knew they were in the treatment or control group). Common randomization methods include lottery draws, coin flips, algorithms, or a selection process such as every n th referral being assigned to treatment. Provide additional details if the method involves more complex randomization techniques such as stratification or blocking.⁸ For all RCTs, note treatment-control ratios and whether they differed by site/cluster or stratum/block.⁹

Randomization may involve two stages: first selecting clusters such as counties, schools, or offices, and then randomizing individuals within those clusters. For multisite or clustered RCTs, describe how sites or clusters were assigned to the treatment or control group. Describe who or what was randomized (the unit of randomization). Common randomization units include individual children or caregivers; families; schools, agencies, or clinics; and counties or regions.

Mention any deviations from the randomization protocol, the reason for deviation(s), and possible implications for internal validity. Internal validity is the extent to which the study can attribute differences detected between groups to participation in the program and not to other influences (known as confounding factors).

RCTs are designed to have high internal validity. When deviations from the protocol occur, the RCT's internal validity, and its ability to say whether the program caused any observed impacts, could be compromised. In such cases, the deviations would be referred to as threats to internal validity. For more information on internal validity, view our companion guide: *Ten Key Design Elements for Rigorous Impact Evaluations in Child Welfare: A Desk Reference for Evaluators* (Brewsaugh and Prendergast 2022).

FOR QEDS

Describe how the comparison group was created and if it was drawn from a different geographic setting or time frame from the treatment group. If the evaluation used a matching method, describe the method and the variables that were used for matching. Matching methods (e.g., propensity score matching, exact matching) are used in QEDs to match treatment and control group participants in an attempt to create

⁸ Stratification involves randomizing within subgroups and may be implemented if subgroup differences are expected. Block randomization is a sampling technique in which the numbers of treatment and control cases are determined by the size of a "block." Blocking can help avoid a situation in which randomization creates too many of the same group in a row. For more information on how to design RCTs that use stratification or blocking, "Child Welfare Evidence-Building Academy," HHS, ACF, OPRE; Brewsaugh and Prendergast (2022).

⁹ To learn more about what RCTs are, the value of conducting RCTs, the ethics of RCTs, what it takes to implement an RCT, and ways to customize an RCT to meet your needs, see Devlin Hanson and Michael Pergamit, *Conducting a Randomized Controlled Trial (RCT) in Child Welfare: A Guide to What, Why, and How for Child Welfare Agency Staff* (Washington, DC: Urban Institute, 2022), <https://www.urban.org/research/publication/conducting-randomized-controlled-trial-rct-child-welfare-guide-what-why-and>.

equivalent groups for comparison.¹⁰ For multisite or clustered QEDs, describe how sites were assigned to the treatment or comparison group. For instance, specify if sites were purposively selected based on specified criteria or if sites volunteered or were invited to be in the treatment or comparison group.

For all QEDs, it is important to mention the steps taken to limit selection bias. Selection bias may occur based on how participants are selected for treatment and comparison groups. Joiner bias (also called *volunteer bias*) is a type of selection bias that occurs when the people who agree to participate in a program differ from those who do not. Selection bias or joiner bias reflect threats to internal validity that could impact the results of the evaluation.

Clearinghouse insights: When rating the execution of a study, clearinghouses assess possible threats to internal validity and their potential impact. Clearinghouses vary in how much internal validity threats affect a study's rating. Social Programs and Blueprints expect to see no or only minimal threats to achieve a favorable rating. CEBC may use a study with serious internal validity threats in its rating process, though it could only be used to support their lowest rating of promising. CrimeSolutions expects authors to acknowledge threats and discuss what actions were taken to minimize their impact. HomVEE and Prevention Services rate a study as low, moderate, or high based on the presence of specific internal validity threats, such as the lack of baseline equivalence and problems with the integrity of random assignment. The handbooks for both HomeVEE and Prevention Services have flow charts detailing the effect specific threats will have on a study's rating (see appendix C).

Provide Sample Sizes for All Groups

Report sample sizes for all treatment and control or comparison groups. Include the following information:

- the number of participants invited to participate in the evaluation
- the number eligible
- the number randomized or assigned (for a QED)
- the timing of consent and number consented
- the number in each group at each measurement time point (i.e., response rates)

¹⁰ For more information on QEDs, their benefits and challenges, and an overview of four common QEDs, see Laura Packard Tucker, *Quasi-Experimental Designs in Child Welfare Evaluations: Opportunities for Generating Rigorous Evidence* (Washington, DC: Urban Institute, 2022), <https://www.urban.org/research/publication/quasi-experimental-designs-child-welfare-evaluations-opportunities-generating>.

Detailed sample counts at baseline, follow-up, and any other points included in analyses show study strengths or weaknesses. For example, sample size is related to statistical power, or the evaluation's ability to detect statistically significant group differences when differences actually exist. Providing the sample size helps readers determine whether the evaluation had enough statistical power to detect significant differences in outcomes of interest.

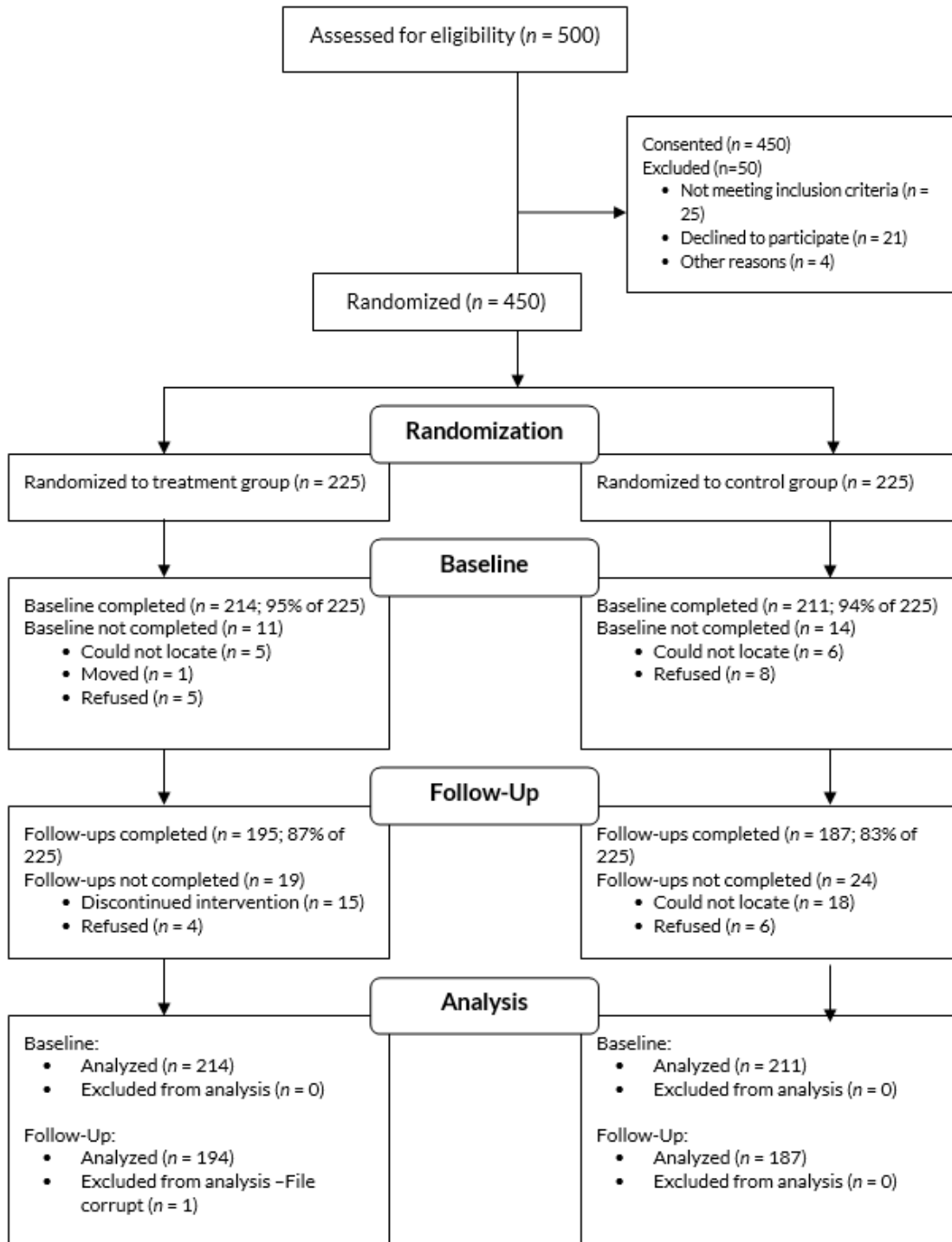
Clearly reported sample sizes at each time point noted above are also necessary for calculating attrition, or dropout, rates. Report the percentage of participants who remained at each time point relative to the sample of people selected to be in the evaluation at baseline. Note any differences in attrition rates separately for the treatment group(s) and the control/comparison group distinguishing between those who completed the program and those who did not. Differential attrition may threaten internal validity. In addition, the rate at which members of the treatment group drop out of a study or refuse to take it up might suggest you need to improve how you engage participants within the evaluation study or the program itself.

Ensure sample sizes are reported consistently throughout your evaluation report. Inconsistent sample sizes in the text or between the narrative and tables may cause readers to question the overall quality of the study and its methods. We recommend including a sampling flow chart to clearly illustrate sample sizes at all stages from recruitment through final measurement. The Consolidated Standards of Reporting Trials (CONSORT) developed a template that can be modified to produce a flow chart like the one shown in figure 1.¹¹

¹¹ For information about the Consolidated Standards of Reporting Trials (CONSORT), see <http://www.consort-statement.org/>, accessed September 1, 2021.

FIGURE 1

Example of an RCT Sampling Flow Chart That Clearly Illustrates Sample Sizes at All Stages from Recruitment through Final Measurement



Source: The authors developed this example for this guide, adapted from the Consolidated Standards of Reporting Trials (CONSORT).

Clearinghouse insights: Staff members at four clearinghouses (Blueprints, Social Programs, Prevention Services, and HomVEE) mentioned incomplete or unclear sample size information across all stages of the study as a common problem with evaluation reports.

Describe the Sample

Report descriptive statistics for the study population by treatment and control or comparison group. Provide descriptive statistics for the study setting and describe its social context. Include relevant demographic and geographic information, scores from pretest outcome measures, and child welfare (or other relevant system) involvement. Commonly reported characteristics are listed in table 1; however, the list is not exhaustive. Any other descriptive information that is relevant to the study should be included. Examples include involvement with other systems like Temporary Assistance for Needy Families, the Supplemental Nutrition Assistance Program, or school systems.

TABLE 1

Common Characteristics That Can Be Used to Describe the Study Population and Setting

Study context	Pretest outcome measures	Demographic characteristics	Child welfare involvement
<ul style="list-style-type: none"> ▪ geography ▪ urbanicity^a ▪ poverty rate ▪ racial/ethnic makeup ▪ homeownership rate 	<ul style="list-style-type: none"> ▪ all primary outcomes measured at baseline (if appropriate) 	<ul style="list-style-type: none"> ▪ age at study start ▪ race ▪ ethnicity ▪ gender ▪ education level ▪ household income ▪ household size ▪ marital status 	<ul style="list-style-type: none"> ▪ abuse and neglect referrals ▪ investigations ▪ removal history ▪ placement types ▪ age at entry/exit ▪ length of time in care ▪ number of placements
<p>Time frame</p> <ul style="list-style-type: none"> ▪ dates each group's data were collected 			

Source: The authors created this table to summarize commonly reported characteristics in the clearinghouses assessed.

Notes: Some items listed in the table may not apply to every evaluation. You should select items that are relevant to your specific study.

^a Urbanicity refers to whether a geographic area is considered urban rather than suburban, rural, or frontier. The federal "Rural-Urban Continuum Codes" (Economic Research Service, accessed September 1, 2021, <https://www.ers.usda.gov/data-products/rural-urban-continuum-codes.aspx>) from the US Department of Agriculture are commonly used to measure urbanicity.

A clear description of the evaluation sample and study setting allows readers to assess external validity—that is, to what populations or settings the study's findings may apply. Indicate whether the study population matches the target population described in the background; if it does not, describe why and how. This information helps readers understand whether any differences between the study population and the program's target population may have affected the study's findings.

Discuss Baseline Equivalence between the Control or Comparison and Treatment Groups

Baseline equivalence is the extent to which the treatment and control or comparison groups were similar to one another before the program began. You should show baseline equivalence at two points in the evaluation: preattrition (for the study population) and postattrition (for the analytic sample).

In discussing baseline or preattrition equivalence for the study population, include any baseline characteristics that may be related to the evaluation's outcomes (demographics, geographic setting, pretest outcome measures, or child welfare involvement as mentioned above). Sometimes the data required to measure and test baseline equivalence of pretest outcome measures are not available.¹² In such cases, best practice is to include measures correlated with the missing characteristics.

If a characteristic is statistically different between the groups at baseline, then the groups are not equivalent. Nonequivalence can threaten the evaluation's internal validity by introducing confounds. For an RCT, if randomization was successful, then the characteristics of the two groups should be statistically equivalent. Differences may reveal flaws in the randomization procedures, making the study a *de facto* QED. However, even well-designed and well-executed RCTs can yield nonequivalent groups, though the larger the sample size, the less likely this will happen. In such cases, differences should be addressed by using control variables in analyses. Because QEDs do not use random assignment as a tool to create equivalent groups, establishing baseline equivalence is especially crucial for increasing internal validity, identifying needed control variables, and ruling out alternative explanations for any significant findings.

The second time point at which you need to show equivalence is postattrition, using the same baseline characteristics mentioned above to examine the analysis sample. The analysis sample is the group of participants from the study population used to estimate the program's impact. Baseline characteristics could differ between groups in the analysis sample because of nonresponse at the time of baseline data collection and because of attrition between baseline and follow-up data collection. Examining equivalence across the study groups at this point shows the extent to which study groups may differ for reasons other than impacts of the program. Any differences in baseline characteristics measured at this point must be reported and controlled for in analyses.

Present baseline measures for each study group at each point in tables. Describe areas of nonequivalence and the analytic procedures used to control for nonequivalence in the narrative. Include the count, percentage, mean, standard deviation (if applicable), *p*-value, and effect sizes for group comparisons.

¹² One example comes from appendix B of the HomVEE handbook (Sama-Miller et al. 2020), which includes a table of outcomes that would not be assessable at baseline if participants in the sample were enrolled prenatally.

As noted above, sample size information should be reported clearly and consistently. Tables should be disaggregated by group and clearly labeled to indicate the sample being presented (i.e., study population or analysis sample). See figures 2 and 3 for examples of how to do this.

FIGURE 2
Example Baseline Equivalence Table for the Study Population of an RCT Testing a Housing Intervention's Impact on Child Welfare Involvement

TABLE C.1
Difference in Treatment and Control Group in Child Welfare Administrative Data Sample at Baseline

	N	Treatment mean	Control mean	P value
Household level				
Female	734	94%	92%	0.169
Age	734	30.4	30.5	0.879
<i>Race or ethnicity</i>				
	672			0.279
Hispanic		15%	17%	
Black		47%	40%	
Other		5%	6%	
White		34%	38%	
Prior criminal justice history	783	51%	47%	0.237
Prior domestic violence	784	35%	36%	0.807
Prior child protective services case	783	47%	41%	0.104
Prior child protective services case as a child	783	58%	53%	0.190
Homeless at randomization	785	48%	46%	0.518
Number of children	794	2.1	2.0	0.065
At least one child removed at randomization	794	40%	41%	0.896
All children removed at randomization	794	35%	36%	0.751
Child level				
Female	1,616	48%	49%	0.809
Age	1,603	5.3	5.4	0.639
Black	1,567	48%	49%	0.633
Hispanic	1,613	13%	13%	0.843
White	1,567	38%	38%	0.935
Other (includes Asian and Native American)	1,568	10%	10%	0.860
Case open at randomization	1,226	64%	60%	0.166
Removed at randomization	1,624	34%	35%	0.635
Months removed	548	4.0	3.3	0.121

Source: Referral data.

Source: Mike Pergamit, Mary K. Cunningham, Devlin Hanson, and Alexandra Stanczyk, *Does Supportive Housing Keep Families Together?* (Washington, DC: Urban Institute, 2019), <https://www.urban.org/research/publication/does-supportive-housing-keep-families-together>.

FIGURE 3

Example Baseline Equivalence Table for the Analysis Sample of an RCT Testing a Housing Intervention's Impact on Child Welfare Involvement

TABLE C.2

Difference in Treatment and Control Group Follow-Up Survey Respondents at Baseline

	N	Treatment mean	Control mean	Difference	P value
Female	531	94.6%	94.0%	0.6%	0.755
Age	531	30.1	30.2	-0.2	0.806
Race or ethnicity					0.596
Black	531	45.1%	47.4%	-2.3%	
White	531	36.4%	34.9%	1.4%	
Hispanic	531	14.0%	10.8%	3.3%	
Other	531	4.5%	6.9%	-2.4%	
Educational attainment					0.492
Less than high school or GED equivalent	528	36.3%	40.1%	-3.8%	
High school or GED	528	24.3%	27.9%	-3.6%	
Some college or higher	528	39.4%	32.0%	7.4%	
Lives with a partner	531	14.1%	14.0%	0.2%	0.961
Number of children in respondent's care	531	2.7	2.6	0.1	0.714
Reunification case at randomization	531	31.4%	31.5%	-0.1%	0.972
Family had prior child welfare involvement	528	46.8%	38.4%	8.3%*	0.069
Respondent had child welfare involvement as a child	528	57.5%	53.6%	3.9%	0.392
Family homeless at randomization (versus unstably housed)	530	47.4%	47.2%	0.3%	0.956
Respondent's history of...					
Domestic violence	529	33.9%	36.9%	-3.0%	0.494
Criminal justice involvement	528	52.6%	41.3%	11.3%**	0.015
Physical abuse in childhood	529	44.3%	44.3%	0.0%	0.992
Sexual abuse in childhood	528	39.5%	40.3%	-0.8%	0.861

Sources: Improving Family Services Survey baseline survey, referral data, and child welfare administrative data.

Notes: All variables are measured at randomization or at the time of the baseline survey. Means are weighted by site to adjust for site differences in treatment and control ratios. Imputed data are not used. Significant differences in race or ethnicity and educational attainment are measured with a chi-squared test.

* $p < 0.1$; ** $p < 0.05$.

Source: Mike Pergamit, Mary K. Cunningham, Devlin Hanson, and Alexandra Stanczyk, *Does Supportive Housing Keep Families Together?* (Washington, DC: Urban Institute, 2019), <https://www.urban.org/research/publication/does-supportive-housing-keep-families-together>.

Clearinghouse insights: All clearinghouses require evaluators to assess baseline equivalence. Information about baseline equivalence was mentioned by staff members at Blueprints, Social Programs, and Prevention Services as critical information that is commonly missing or incompletely reported. Prevention Services staff members mentioned that this information is also important for understanding the extent to which confounds exist. Both Prevention Services and HomVEE specify a set of demographic characteristics that are required for baseline equivalence of the study population and analytic sample, as

well as a list of acceptable methods and when they can be used to control for nonequivalence.¹³ In addition to the randomized/baseline sample, Blueprints looks for baseline equivalence on sociodemographic characteristics and pretest measures between participants who attrit versus those who remain in the program at follow-up (also known as the analysis sample). Not controlling for areas of nonequivalence was noted as a common reason for lower ratings.

Describe All Measures and Outcomes

Evaluation reports should include a thorough description of all primary and secondary data collection efforts for both implementation and outcome evaluations.

- Primary data are collected directly from a data source (e.g., a participant) for the purposes of an evaluation.
- Secondary data are collected for other purposes (e.g., child welfare agency operations, previous studies or surveys) and are accessed and used for the evaluation.

A detailed description helps readers understand how the data were collected and provides the information needed to interpret the findings and replicate the study.

FOR PRIMARY DATA COLLECTION

Provide an overview of each instrument¹⁴ (i.e., number of items, subscales) and cite its source, as applicable. Explain how and when the instrument was administered, how it was scored, and how scores can be interpreted. Include reliability and validity information from the literature for all instruments, and report reliability and validity information from the analysis sample. Reliability and validity information helps readers understand the extent to which the instruments measure what they were intended to measure (validity) and how consistently they measure it (reliability). It is especially important to report reliability and validity information if the instruments have not been previously validated or were developed specifically for the evaluation.

FOR SECONDARY DATA COLLECTION

State each data source, how and when the data were collected by the source, and the variables that were included in the study. If the administrative data include instruments, describe the instruments and their reliability and validity.

¹³ For more information, see WWC (2020).

¹⁴ Instruments may also be referred to as measures, measurement instruments, or measurement tools.

Clearinghouse insights: All six clearinghouses require information about reliability and validity and prefer instruments be well-established and standardized. A well-established instrument has been rigorously tested and shown to be reliable and valid or used in multiple published studies. Prevention Services, CEBC, CrimeSolutions, and HomVEE allow instruments that are developed for the evaluation as long as they meet reliability and validity reporting requirements.¹⁵ Clearinghouses also prefer reliability information (e.g., Cronbach's alpha) be calculated and reported with the analysis sample for all measurement instruments, even those that are well-established. Doing so is generally required if using a newer tool.

How to Report Implementation Evaluation Results

Implementation findings can help readers understand how the program was implemented during the evaluation period. You can connect this section to the background section by noting if implementation differed from the program description. Implementation findings also can provide context that helps explain outcome findings. For example, if an evaluation found that a parenting program had no effect on parenting outcomes, then an evaluator might explore implementation results to determine whether the program was implemented as planned. While implementation evaluations are not the focus of evidence reviews, such information can help others understand what it may take to successfully implement the program in a different location or context.

Report descriptive information on all implementation measures collected, as applicable. Be sure to note information on fidelity, or whether the program was implemented differently than it was supposed to be. For more information on developing fidelity measures, see our companion guide, *Ten Key Design Elements for Rigorous Impact Evaluations in Child Welfare: A Desk Reference for Evaluators* (Brewsaugh and Prendergast 2022). Commonly reported information includes the following:

- dosage (i.e., how much of the program participants received)
- frequency (i.e., how often participants received the program)
- duration (i.e., how long participants received the intervention)
- staffing (e.g., supervisor-to-staff ratios, turnover, or certifications held)

Provide a summary of qualitative implementation information collected during the evaluation, as applicable. This may be in the form of main themes from stakeholder interviews and focus groups or a

¹⁵ For more information about HomVEE's reliability and validity reporting requirements, see chapter III, section B.4 of Sama-Miller and colleagues (2020).

description of the infrastructure that supports the program (e.g., data systems, funding, program partnerships, community support). Discuss any implementation challenges stakeholders experienced and lessons learned. If applicable, be sure to discuss reasons why implementation may not have occurred as intended. For example, turnover may have caused high caseloads or a new statute may have allowed participants to stay in the program longer.

Clearinghouse insights: CrimeSolutions includes fidelity measures in its evidence rating. Blueprints, Prevention Services, and HomVEE recommend (but do not require) that authors describe any fidelity instruments that were included in the study. This information also provides context for interpreting outcomes.

How to Report Impact Evaluation Results

The results section typically includes a description of each step of the data analysis process and relevant statistical information for each analysis conducted.

Describe Analyses

Describe the overall analyses and subgroup analyses performed and state whether they included primary or secondary outcomes. This information helps readers understand the extent to which the findings should be used to assess the program's effectiveness.

For an RCT, note whether an intent-to-treat (ITT) or treatment-on-the-treated (TOT) analytical approach was used.

- The ITT estimate provides the impact of the offer of an intervention and includes all members of the initial sample in the group to which they were assigned. This includes all treatment group members regardless of whether they received the intervention. It also includes keeping people in the control group who somehow got the treatment. The ITT estimate helps policymakers know the full impact of offering the program to all people who are eligible.
- The TOT estimate uses analytic techniques to calculate the impact of the intervention for those people who received it. This includes people in the control group who somehow got the treatment. When using TOT, clearly define what *treated* entailed, such as how much of the program participants had to complete to be considered treated.
- For all analyses, explain the estimation method (e.g., regression, ANOVA) and list any control variables or weights used.

A detailed analysis description allows readers to accurately interpret the evaluation's findings and understand how to replicate them.

Some measures may have missing data; thus, it is important to report sample sizes for each variable or outcome throughout the report. Clearly describe how missing data were or were not handled (e.g., imputation, full information maximum likelihood, no approach). Because missing data may impact the results of the evaluation, it is important to be transparent about what approach (if any) was employed to address missingness.

Describe any other adjustments, such as clustering standard errors or corrections for multiple comparisons (e.g., adjusting p -values). A description of the methods used to adjust analyses for the clustered/nested (i.e., multilevel) nature of data helps readers know whether findings may be biased because of clustering. It is especially important to adjust for and report this information for multisite studies, as each site represents a cluster. Correcting for multiple comparisons reduces the chances of reporting false positives (i.e., findings that are significant by chance in the absence of a true effect). Describing any steps taken to adjust for multiple comparisons helps readers interpret significant findings.

Clearinghouse insights: Prevention Services and CEBC do not require ITT analyses. Blueprints and Social Programs require ITT analysis and consider TOT only as an exploratory analysis. Crime Solutions focuses on ITT if it is reported but will accept TOT when ITT analysis is not reported. HomVEE will focus on ITT if it is reported but will accept TOT if it meets *What Works Clearinghouse Version 4.1 Standards Handbook* standards.¹⁶ Two clearinghouses (Prevention Services and HomVEE) use standards from the *What Works Clearinghouse Version 4.1 Standards Handbook* for missing data.¹⁷ Most clearinghouses note that authors should be transparent about how they handled missing data.

Report Outcome Findings

Begin the section on outcome findings by reporting pooled findings (for the full sample) on primary and secondary outcomes, followed by any subgroup findings. For multisite or clustered designs, report pooled effects for the full analysis sample, not only findings by location. It is important to report pooled results so that readers can judge the overall impact of a program, even if there are differences by site or for certain subgroups.

¹⁶ For more information, see WWC (2020).

¹⁷ Missing data standards can also be found in Wilson and colleagues (2019), beginning on page 37, and in Sama-Miller and colleagues (2020), beginning on page 77.

When reporting effects, include all relevant information and statistics, including *p*-values and effect sizes (e.g., Cohen’s *d*, Hedge’s *g*). At minimum, report the information required for reviewers to compute effect sizes: the mean, standard deviation, and sample size for each group. Also state the measurement time frame (e.g., baseline, follow-up) for each outcome reported. Clearly listing the length of follow-up time frames helps readers understand whether effects were sustained and for how long.

Consider adding section headings to organize the findings (e.g., “Main Effects” and “Subgroup Effects”) and displaying the results in tables like the ones shown in figures 4 and 5. Please note that the study referenced in the example tables may not adhere to other clearinghouse design standards.

FIGURE 4

Example of a Table Reporting Outcome Evaluation Findings from the Full Sample

TABLE 33

Parenting

	<i>N</i>	Treatment mean	Control mean	Impact (difference)	<i>P</i> value
Average relationship quality across all children is excellent, very good, or good	511	17.7%	11.2%	6.5%**	0.029
Parenting practices of respondent in past six months (among those who lived with at least one child, at least some of the time, in the past six months)					
Average relationship quality across all children is excellent, very good, or good	416	16.7%	7.8%	8.9%**	0.004
Average number of times parent practiced corporal punishment or was physically aggressive to child	417	0.5	0.4	0.1	0.244
Average number of times parent was verbally aggressive to child	417	2.2	2.0	0.2	0.589
Average number of times parent exhibited neglectful behaviors toward child	417	0.2	0.1	0.1	0.185
Parenting warmth and supportiveness scale (0–4)	417	3.4	3.5	-0.1	0.316
Parenting stress scale (1–5)	416	2.4	2.4	0.0	0.950
Parenting skill scale (1–5)	416	4.4	4.4	0.0	0.637

Source: Improving Family Services Survey 12-month follow-up survey.

Notes: The weighted regression-adjusted models include the following control measures: time between randomization and follow-up survey; site; reunification case; respondent’s educational attainment, age, and race or ethnicity; number of children in respondent’s care; children’s ages; whether respondent lived with a partner; family homelessness at randomization; family history of child welfare involvement; respondent’s history of child welfare involvement as a child; respondent’s criminal history; respondent’s domestic violence history; respondent’s history of physical abuse as a child; and respondent’s history of sexual abuse as a child. All controls are measured at randomization or at the time of the baseline survey. Children for whom parental rights have been terminated are excluded from the analysis. On the parenting warmth and supportiveness scale, stress scale, and skill scale, higher values indicate greater warmth and supportiveness, stress, and skill, respectively.

** *p* < 0.05.

Source: Mike Pergamit, Mary K. Cunningham, Devlin Hanson, and Alexandra Stanczyk, *Does Supportive Housing Keep Families Together?* (Washington, DC: Urban Institute, 2019), <https://www.urban.org/research/publication/does-supportive-housing-keep-families-together>.

FIGURE 5

Example of a Table Reporting Outcome Evaluation Findings from a Subgroup of Preservation Families

TABLE E.9

Parenting, Preservation Families

	N	Treatment mean	Control mean	Impact (difference)	P value
Average relationship quality across all children is excellent, very good, or good	359	20.0%	9.6%	10.4%**	0.004
Respondent's parenting practices in the past six months (among those who lived with at least one child, at least some of the time, in the past six months)					
Average relationship quality across all children is excellent, very good, or good	320	21.1%	8.1%	13.0%**	0.001
Average number of times parent practiced corporal punishment or was physically aggressive to child	320	0.6	0.4	0.2*	0.069
Average number of times parent was verbally aggressive to child	320	2.3	1.9	0.4	0.361
Average number of times parent exhibited neglectful behaviors toward child	320	0.3	0.1	0.2*	0.097
Parenting warmth and supportiveness scale (0–4)	320	3.4	3.5	-0.1	0.204
Parenting stress scale (1–5)	320	2.5	2.5	0.0	0.956
Parenting skill scale (1–5)	320	4.4	4.4	0.0	0.539

Source: Improving Family Services Survey 12-month follow-up survey.

Notes: The weighted regression-adjusted models include the following control measures: time between randomization and follow-up survey; site; reunification case; respondent's educational attainment, age, and race or ethnicity; number of children in respondent's care; children's ages; whether respondent lived with a partner; family homelessness at randomization; family history of child welfare involvement; respondent's history of child welfare involvement as a child; respondent's criminal history; respondent's domestic violence history; respondent's history of physical abuse as a child; and respondent's history of sexual abuse as a child. All controls are measured at randomization or at the time of the baseline survey. Children for whom parental rights have been terminated are excluded. On the parenting warmth and supportiveness scale, stress scale, and skill scale, higher values indicate greater warmth and supportiveness, stress, and skill, respectively.

* $p < 0.1$; ** $p < 0.05$.

Source: Mike Pergamit, Mary K. Cunningham, Devlin Hanson, and Alexandra Stanczyk, *Does Supportive Housing Keep Families Together?* (Washington, DC: Urban Institute, 2019), <https://www.urban.org/research/publication/does-supportive-housing-keep-families-together>.

Clearinghouse insights: All clearinghouses except HomVEE base their reviews on results for the overall study population (i.e., pooled or main effects); subgroup analyses generally are not used to rate a program. Thus, authors should always state overall analyses before describing subgroup analyses. HomVEE is the only clearinghouse where a model can be classified as evidenced-based because of favorable findings from subgroup analyses, if the model otherwise meets HHS's criteria. You should refer to the HomVEE handbook¹⁸ for their specific subgroup analysis criteria.

Blueprints, Prevention Services, and Social Programs staff members stated that if there are theoretical reasons why a program may perform differently by setting (e.g., classroom, geographic location) or

¹⁸ For more information about HomVEE's subgroup analysis criteria, see chapter II, section B.2.b.ii of Sama-Miller and colleagues (2020).

subpopulation (e.g., gender, race/ethnicity), evaluators should specify subgroup analyses in the study's preregistration plan. As noted above, preregistering a study increases reviewers' confidence that the findings were based on planned hypotheses and that all planned analyses, not only those that yield significant impacts, are reported. If there are multiple comparison groups, Prevention Services will only review one comparison per outcome at each time point. None of the clearinghouses mentioned specific standards for adjusting for multiple comparisons, but in our interviews Blueprints, Social Programs, and HomVEE staff members said that such adjustments can be helpful to include. There is a general requirement to account for clustering if randomization occurred at the cluster level. Blueprints suggests doing so when randomization of individuals occurred at the site level as well. Clearinghouses generally use a p -value of at least 0.05 to represent statistical significance. Most (all but CEBC) incorporate effect sizes in their review process, though to varying degrees. CrimeSolutions and Prevention Services maintain a list of acceptable statistics for effect sizes.

What to Include in the Discussion Section of Your Evaluation Report

The discussion section provides an opportunity to summarize the findings reported in the results section and put them in context. Use plain language¹⁹ to briefly restate the findings of the evaluation. Indicate whether the findings were expected and align with previous evaluations of this program or related research. Discuss how the findings contribute to the broader literature. Provide a detailed discussion of the limitations of the study design or execution that could reduce the internal validity or generalizability of the results (e.g., sampling issues, reliability of measures). A thorough discussion of limitations can help readers understand the internal validity of the study (the confidence with which any findings can be attributed to participation in the program), and external validity (the extent to which the findings can be generalized to the intended population). Describe implications of the findings for the child welfare field in terms of research, programs, practice, or policies.

Clearinghouse insights: CrimeSolutions expects authors to acknowledge limitations and how they were addressed.

¹⁹ For more information on writing in plain language, see CDC (2018).

Guide Summary

Evaluation reports play an important role in the evidence-building process for the child welfare field:

- High-quality reports can support children and families by contributing to the growing evidence base of programs that help children grow up in safe and nurturing environments.
- A high-quality evaluation report may result in a positive clearinghouse rating, which could improve program funding opportunities (e.g., Prevention Services).
- Program details will help readers replicate the program in their jurisdictions.
- A strong evaluation report allows those not involved in the study to understand and replicate it.
- Perhaps most importantly, an evaluation report is often the only basis on which stakeholders, whether from clearinghouses or peers in the field, can assess the evidence about the effectiveness of a program.

Throughout this guide we highlight best practices that can be considered when writing high-quality evaluation reports. In general, strong evaluation reports include an overview of the program and current evaluation; a detailed description of the methods, analysis plan, and findings; and a discussion of limitations and policy or practice implications. Though we address components that are commonly included in evaluation reports, it is important to keep in mind that evaluation designs vary. We recommend including all information that could be relevant to a reviewer, even if it is not mentioned above.

Appendix A. Evaluation Report Checklist

What to include in the background section of your evaluation report

Provide an overview of the program

- Main components, goals, and desired outcomes
- Conceptual framework, theory of change, or logic model
- Target population, typical dosage, implementation setting, and who delivers the program or service
- Reference to book, manual, or other program materials
- Details on whether the program is proprietary

Provide an overview of the evaluation

- Implementation timing, setting (i.e., usual care, geographic setting), and who delivered the program or service
- Implementation evaluation details (as appropriate)
- Program adaptations
- Evaluation design
- Services received by comparison groups
- Differences by site (as appropriate)

List research questions and hypotheses

- Research questions and hypotheses
- Primary and secondary outcomes
- Implementation evaluation questions (as appropriate)

What to include in the methods section of your evaluation report

Describe how the sample was selected

- Eligibility
- Inclusion/exclusion criteria
- Identification of participants
- Recruitment procedures
- Sample selection
- Consent and timing of consent

Randomized Controlled Trials

- Who performed the randomization and when
- Whether participants were blind to group assignment
- How sites or clusters were assigned (as appropriate)
- Randomization method and unit of randomization
- Treatment-control ratios
- Deviations from protocol

Quasi-Experimental Designs

- Population from which sample was drawn
- Geographic setting and time frame of data collection
- Matching methods
- How sites were assigned (as appropriate)
- Steps taken to limit selection bias

Provide sample sizes for all groups

- Number of participants invited, number eligible, and number randomized or assigned
- Timing of consent and number consented
- Number in each group at each measurement period
- Participation rates for the overall sample
- Differences in attrition rates by group
- Sample sizes consistently reported

Describe the sample

- Descriptive statistics for the study population
- Descriptive statistics for the study setting
- Whether study population matched target population

Discuss baseline equivalence between groups

- Baseline equivalence for the study population
- Baseline equivalence for the analysis sample
- Describe nonequivalence and analytic procedures used to control for nonequivalence

Describe all measures and outcomes

Primary data collection

- Instruments
- How and when each instrument was administered
- How each instrument was scored and interpreted
- Reliability and validity information for each instrument

Secondary data collection

- Data sources
- How and when data were collected
- Variables obtained
- Measurement instruments
- Reliability and validity information for each instrument

How to report implementation evaluation results

Describe implementation evaluation results

- Descriptive information on all implementation measures
- Whether findings differed from the program model

How to report outcome or impact evaluation results

Describe analyses

- Overall analyses and subgroup analyses performed
- Estimation method
- Control variables or weights
- Sample sizes for each variable or outcome
- Missing data procedures
- Adjustments or corrections

Report outcome findings

- Main effects
- Subgroup effects, if applicable
- p -values and effect sizes
- Measurement time frame for each finding

What to include in the discussion section of your evaluation report

Discuss findings, limitations, and implications

- Restated findings
- Limitations of the study
- Contributions made to child welfare field

Appendix B. Summary of Clearinghouse Insights in This Guide

Table B.1 (see next page) summarizes clearinghouse requirements and insights mentioned during informational interviews. This is not a complete list of all clearinghouse requirements; we encourage readers to check clearinghouse websites, handbooks, and reporting guides (HomVEE and Prevention Services) to learn what is needed for a specific clearinghouse to review and assess a study.

TABLE B.1

Summary of Clearinghouse Insights Reviewed throughout This Guide to Writing High-Quality Evaluation Reports in Child Welfare

Evaluation component	Blueprints for Healthy Youth Development	The California Evidence-Based Clearinghouse for Child Welfare	National Institute of Justice CrimeSolutions	Home Visiting Evidence of Effectiveness Review	Title IV-E Prevention Services Clearinghouse	Social Programs That Work
Program components, implementation practices, and fidelity measures	<ul style="list-style-type: none"> Should provide a description of the program, measures of implementation fidelity, and any deviations from the protocol 	<ul style="list-style-type: none"> Must have a book, manual, or other written material that clearly and completely describes how the program was implemented Program developer completes questionnaire about the program and provides citation Does not expect authors to repeat information in a report that is covered by the questionnaire 	<ul style="list-style-type: none"> Must describe six specific elements of the program Acceptable to cite the program's manual Commonly missing: thorough discussion of program components 	<ul style="list-style-type: none"> Clearly describe the components of the model, state the model's name, and state how long it has been operating Acceptable to cite the program's manual Helpful to clearly describe how the intervention meets the clearinghouse's criteria 	<ul style="list-style-type: none"> Must have a book, manual, or other written material that clearly and completely describes how the program is implemented Acceptable to cite the program's manual Best practice to also describe comparison or control group services Commonly missing: citation to program manual 	<ul style="list-style-type: none"> Must provide a description of the program Acceptable to cite the program's manual
Conceptual framework and theoretical base	<ul style="list-style-type: none"> Discuss the conceptual framework or theory 	<ul style="list-style-type: none"> <i>Not mentioned</i> 	<ul style="list-style-type: none"> Discuss the conceptual framework or theory 	<ul style="list-style-type: none"> <i>Not mentioned</i> 	<ul style="list-style-type: none"> <i>Not mentioned</i> 	<ul style="list-style-type: none"> <i>Not mentioned</i>

Evaluation component	Blueprints for Healthy Youth Development	The California Evidence-Based Clearinghouse for Child Welfare	National Institute of Justice CrimeSolutions	Home Visiting Evidence of Effectiveness Review	Title IV-E Prevention Services Clearinghouse	Social Programs That Work
Program adaptations	<ul style="list-style-type: none"> ▪ Describe any changes made to the program ▪ Description of the intervention and fidelity is used to determine if the evaluation is a replication or of a new program 	<ul style="list-style-type: none"> ▪ Describe any changes made to the program ▪ Assesses whether the essential components were altered, consults with program developers, and makes decisions on a case-by-case basis ▪ If the adaptations constitute another manual or training, it would likely be considered a different program 	<ul style="list-style-type: none"> ▪ Describe any changes made to the program ▪ Assesses whether any of the six program elements were altered, consults with program developers, and decides on a case-by-case basis 	<ul style="list-style-type: none"> ▪ Describe any changes made to the program ▪ Currently does not have specific standards to assess program adaptations 	<ul style="list-style-type: none"> ▪ Describe any changes made to the program ▪ Assesses adaptations for significant changes in frequency or duration ▪ Adding, subtracting, or substantially modifying content ▪ Substantial content changes for different cultural or other groups ▪ Changing modality (e.g., individual to group therapy) ▪ Delivery by substantially different providers (e.g., paraprofessionals vs. nurses) 	<ul style="list-style-type: none"> ▪ Describe any changes made to the program ▪ Assesses whether the essential components were altered and makes decisions on a case-by-case basis

Evaluation component	Blueprints for Healthy Youth Development	The California Evidence-Based Clearinghouse for Child Welfare	National Institute of Justice CrimeSolutions	Home Visiting Evidence of Effectiveness Review	Title IV-E Prevention Services Clearinghouse	Social Programs That Work
Implementation setting	<ul style="list-style-type: none"> ▪ <i>Not mentioned</i> 	<ul style="list-style-type: none"> ▪ Describe implementation setting ▪ Considers whether a program was implemented in a usual care setting or other setting in its ratings 	<ul style="list-style-type: none"> ▪ Describe implementation setting 	<ul style="list-style-type: none"> ▪ Describe implementation setting ▪ Model must be designed or adapted and tested for delivery in parents'/care-givers' homes ▪ State whether participation in the program was voluntary 	<ul style="list-style-type: none"> ▪ Describe implementation setting ▪ Considers whether a program was implemented in a usual care setting or other setting in its ratings 	<ul style="list-style-type: none"> ▪ Describe implementation setting ▪ Considers whether a program was implemented in a usual care setting or other setting in its ratings
Services for control or comparison group	<ul style="list-style-type: none"> ▪ Describe services received by comparison or control groups ▪ Comparison group can receive an intervention beyond services as usual 	<ul style="list-style-type: none"> ▪ Describe services received by comparison or control groups ▪ Comparison group can receive an intervention beyond services as usual 	<ul style="list-style-type: none"> ▪ Describe services received by comparison or control groups ▪ Only reviews evaluations in which the comparison group received either no services or services as usual 	<ul style="list-style-type: none"> ▪ Describe services received by comparison or control groups ▪ Comparison group can receive an intervention beyond services as usual 	<ul style="list-style-type: none"> ▪ Describe services received by comparison or control groups ▪ Only reviews evaluations in which the comparison group received either no or minimal services or services as usual 	<ul style="list-style-type: none"> ▪ Describe services received by comparison or control groups ▪ Comparison group should receive services as usual or enhanced services as usual but will review evaluations for which comparison group received an intervention beyond services as usual

Evaluation component	Blueprints for Healthy Youth Development	The California Evidence-Based Clearinghouse for Child Welfare	National Institute of Justice CrimeSolutions	Home Visiting Evidence of Effectiveness Review	Title IV-E Prevention Services Clearinghouse	Social Programs That Work
Outcomes	<ul style="list-style-type: none"> ▪ Examines whether a preponderance of evidence indicates a significant favorable change in intended outcomes that can be attributed to the intervention and that there is no evidence of harmful effects ▪ Must be one of Blueprints’s targeted youth behavioral outcomes (problem behavior, education, emotional well-being, physical health, positive relationships) 	<ul style="list-style-type: none"> ▪ Looks for a preponderance of evidence across all outcomes tested in a topic area ▪ Required outcomes are defined for each topic area 	<ul style="list-style-type: none"> ▪ Studies are rated on whether the program demonstrated a change in behavior, rather than only attitudes, knowledge, or beliefs ▪ Studies with nonbehavioral outcomes can be reviewed, but they will likely be scored as secondary outcomes ▪ Primary outcomes must match the core intent of the program and relate to the clearinghouse’s target outcomes 	<ul style="list-style-type: none"> ▪ Manuscripts about studies are rated based on the highest evidence rating of any individual finding ▪ Must be in one of the target outcome domains listed in authorizing statute 	<ul style="list-style-type: none"> ▪ Studies must measure and report program or service impacts on at least one eligible target outcome. Eligible target outcomes differ by program or service area. 	<ul style="list-style-type: none"> ▪ Generally looks for a favorable effect on a primary outcome or outcomes ▪ Outcomes must have “clear policy importance” rather than intermediate outcomes that may or may not lead to important outcomes ▪ Encourages authors to clearly label which outcomes were primary ▪ Encourages evaluators of new programs to designate only one or a few primary outcomes and the rest as secondary

Evaluation component	Blueprints for Healthy Youth Development	The California Evidence-Based Clearinghouse for Child Welfare	National Institute of Justice CrimeSolutions	Home Visiting Evidence of Effectiveness Review	Title IV-E Prevention Services Clearinghouse	Social Programs That Work
Threats to internal validity	<ul style="list-style-type: none"> Expect no or minimal threats; e.g., assesses baseline equivalence, attrition, and issues with randomization; studies with bias are rated as insufficient or inconclusive 	<ul style="list-style-type: none"> Relies on peer review standard and suggests threats be discussed in limitations section 	<ul style="list-style-type: none"> Has detailed review criteria for internal validity threats; expects authors to acknowledge threats and discuss what actions were taken to minimize impact 	<ul style="list-style-type: none"> Handbook details impact of specific threats; e.g., RCTs and QEDs with confounding factors are given low ratings 	<ul style="list-style-type: none"> Handbook details impact of specific threats; e.g., all studies assessed for baseline equivalence; studies without integrity of randomization are reviewed using QED standards 	<ul style="list-style-type: none"> Expect no or minimal threats; e.g., assesses baseline equivalence, attrition and other issues that affect internal validity; studies with meaningful bias are considered inconclusive
Reporting sample sizes across all stages of evaluation	<ul style="list-style-type: none"> Commonly missing: incomplete or unclear sample size information across all stages of the study 	<ul style="list-style-type: none"> <i>Not mentioned</i> 	<ul style="list-style-type: none"> Commonly missing: incomplete or unclear sample size information across all stages of the study 	<ul style="list-style-type: none"> Commonly missing: incomplete or unclear sample size information 	<ul style="list-style-type: none"> Commonly missing: incomplete or unclear sample size information across all stages of the study 	<ul style="list-style-type: none"> Commonly missing: incomplete or unclear sample size information across all stages of the study

Evaluation component	Blueprints for Healthy Youth Development	The California Evidence-Based Clearinghouse for Child Welfare	National Institute of Justice CrimeSolutions	Home Visiting Evidence of Effectiveness Review	Title IV-E Prevention Services Clearinghouse	Social Programs That Work
Baseline Equivalence	<ul style="list-style-type: none"> Baseline equivalence must be assessed and reported for both RCT and QED Baseline equivalence is reviewed for both the assigned sample and the analysis sample Nonequivalence must be accounted for; methods not specified. Commonly missing or incompletely reported: baseline equivalence and differential attrition 	<ul style="list-style-type: none"> Report baseline equivalence and how it was controlled for No specific standards for how to correct for issues; discuss in limitations section 	<ul style="list-style-type: none"> Baseline equivalence must be assessed and reported for both RCT and QED Method for correcting issues depends on severity of nonequivalence and element on which treatment and control are nonequivalent Not accounting for nonequivalence will lower Statistical Adjustment score under Design Quality 	<ul style="list-style-type: none"> Baseline equivalence must be assessed and reported for the analysis sample for QED and high-attrition or compromised RCTs Uses WWC effect size thresholds to determine equivalence and whether adjustments are necessary Maintains a list of acceptable methods for controlling for nonequivalence 	<ul style="list-style-type: none"> Baseline equivalence must be assessed and reported for both RCT and QED Uses WWC effect size thresholds to determine equivalence and whether adjustments are necessary Maintains a list of acceptable methods for controlling for nonequivalence Commonly missing or incompletely reported: baseline equivalence 	<ul style="list-style-type: none"> Baseline equivalence must be assessed and reported Ideally adjust for any baseline differences using covariates Commonly missing or incompletely reported: baseline equivalence

Evaluation component	Blueprints for Healthy Youth Development	The California Evidence-Based Clearinghouse for Child Welfare	National Institute of Justice CrimeSolutions	Home Visiting Evidence of Effectiveness Review	Title IV-E Prevention Services Clearinghouse	Social Programs That Work
Measurement reliability and validity	<ul style="list-style-type: none"> ▪ Must include information about reliability and validity for each measure; no specific standards ▪ Administrative data generally considered reliable and valid 	<ul style="list-style-type: none"> ▪ Must include information about reliability and validity for each measure; no specific standards ▪ Administrative data generally considered reliable and valid ▪ Instruments can be developed for the evaluation, but only if reliability and validity information is reported ▪ Encourages researchers to think carefully about whether an existing measure can be used before developing a new measure 	<ul style="list-style-type: none"> ▪ Must include information about reliability and validity for each measure; no specific standards ▪ Administrative data generally considered reliable and valid ▪ Instruments can be developed for the evaluation, but only if reliability and validity information is reported and coefficients are high ▪ Prefers instruments that are well-established 	<ul style="list-style-type: none"> ▪ Must include information about reliability and validity for each measure ▪ Has standards for face validity and for reliability ▪ Include citation for well-established measures; prefers reliability and validity to be reported with study sample ▪ Administrative data generally considered reliable and valid ▪ Instruments can be developed for the evaluation, but only if reliability and validity information is reported 	<ul style="list-style-type: none"> ▪ Must include information about reliability and validity for each measure ▪ Has standards for reliability ▪ Administrative data generally considered reliable and valid; also demographics and medical or physical tests ▪ Instruments can be developed for evaluation as long as they meet criteria above and the information is reported 	<ul style="list-style-type: none"> ▪ Must include information about reliability and validity for each measure; no specific standards ▪ Administrative data generally considered reliable and valid ▪ Prefers instruments that are well-established; generally, does not accept instruments specifically developed for the evaluation

Evaluation component	Blueprints for Healthy Youth Development	The California Evidence-Based Clearinghouse for Child Welfare	National Institute of Justice CrimeSolutions	Home Visiting Evidence of Effectiveness Review	Title IV-E Prevention Services Clearinghouse	Social Programs That Work
Fidelity measures	<ul style="list-style-type: none"> ▪ <i>Recommended practice:</i> describe fidelity measures 	<ul style="list-style-type: none"> ▪ <i>Not mentioned</i> 	<ul style="list-style-type: none"> ▪ Includes fidelity measures in evidence ratings 	<ul style="list-style-type: none"> ▪ <i>Recommended practice:</i> describe fidelity measures 	<ul style="list-style-type: none"> ▪ <i>Recommended practice:</i> describe fidelity measures 	<ul style="list-style-type: none"> ▪ <i>Not mentioned</i>
Description of Analyses	<ul style="list-style-type: none"> ▪ Requires ITT analysis and considers TOT only as an exploratory analysis 	<ul style="list-style-type: none"> ▪ ITT is not required 	<ul style="list-style-type: none"> ▪ Focuses on ITT but will accept TOT if ITT is not reported 	<ul style="list-style-type: none"> ▪ Focuses on ITT if it is reported but will accept TOT if it meets the WWC Version 4.1 Standards 	<ul style="list-style-type: none"> ▪ ITT is not required 	<ul style="list-style-type: none"> ▪ Requires ITT analysis and will consider TOT as a secondary analysis
Missing data	<ul style="list-style-type: none"> ▪ Reviews are based on results for overall study population 	<ul style="list-style-type: none"> ▪ Reviews are based on results for overall study population 	<ul style="list-style-type: none"> ▪ Reviews are based on results for overall study population ▪ May review subgroup differences under specific circumstances 	<ul style="list-style-type: none"> ▪ Reviews are based on results for the overall study population ▪ Reviews can also be based on results from subgroup analyses that meet specific criteria^a ▪ Commonly missing: standard deviations reported for each outcome 	<ul style="list-style-type: none"> ▪ Reviews are based on results for overall study population ▪ If there are theoretical reasons for subgroup differences, then they should be specified in preregistration plan 	<ul style="list-style-type: none"> ▪ Reviews are based on results for overall study population ▪ If there are theoretical reasons for subgroup differences, then they should be specified in preregistration plan

Evaluation component	Blueprints for Healthy Youth Development	The California Evidence-Based Clearinghouse for Child Welfare	National Institute of Justice CrimeSolutions	Home Visiting Evidence of Effectiveness Review	Title IV-E Prevention Services Clearinghouse	Social Programs That Work
Multiple comparisons and adjustments	<ul style="list-style-type: none"> No standard for multiple comparisons but notes that they can be useful tests when examining multiple outcomes 	<ul style="list-style-type: none"> No adjustments required for multiple comparisons 	<ul style="list-style-type: none"> No adjustments required for multiple comparisons Reviewers apply greater weight to primary versus secondary outcomes when scoring studies 	<ul style="list-style-type: none"> No adjustments required for multiple comparisons 	<ul style="list-style-type: none"> If there are multiple comparison groups, will only review one comparison per outcome at each time point No adjustments required for multiple comparisons 	<ul style="list-style-type: none"> If there are more than two primary outcomes, then recommends that authors look to see if the effects would remain significant after adjusting for multiple comparisons; will do this on its own if authors do not Expects that multiple comparison adjustment is made across all primary outcomes, regardless of domain
Accounting for clustering in analysis	<ul style="list-style-type: none"> If a cluster design is used, then expects analysis accounts for clustering Recommends adjusting for clustering even if randomization occurred at the individual level 	<ul style="list-style-type: none"> If a cluster design is used, then expects analysis accounts for clustering 	<ul style="list-style-type: none"> If a cluster design is used, then expects analysis accounts for clustering Provides list of acceptable modeling approaches to accounts for clustering 	<ul style="list-style-type: none"> If a cluster design is used, then expects analysis accounts for clustering 	<ul style="list-style-type: none"> <i>Not mentioned</i> 	<ul style="list-style-type: none"> Assesses the appropriateness of clustering methods

Evaluation component	Blueprints for Healthy Youth Development	The California Evidence-Based Clearinghouse for Child Welfare	National Institute of Justice CrimeSolutions	Home Visiting Evidence of Effectiveness Review	Title IV-E Prevention Services Clearinghouse	Social Programs That Work
p-values	<ul style="list-style-type: none"> Not mentioned 	<ul style="list-style-type: none"> Not mentioned 	<ul style="list-style-type: none"> Uses a <i>p</i>-value of 0.05 	<ul style="list-style-type: none"> Uses a <i>p</i>-value of 0.05 	<ul style="list-style-type: none"> Not mentioned 	<ul style="list-style-type: none"> Generally uses a <i>p</i>-value of 0.05
Effect sizes	<ul style="list-style-type: none"> Should report effect sizes, differences in proportion, and significance levels of those differences or include information necessary for calculation Mentions that effect size is especially important when comparing two interventions 	<ul style="list-style-type: none"> Would not compute an effect size itself, but if the effect size is shared in the report, it is considered 	<ul style="list-style-type: none"> Reviewers screen to ensure information is reported for calculating effect size Maintains list of acceptable effect size statistics 	<ul style="list-style-type: none"> Strongly encouraged to include effect sizes but not required; will calculate effect size on its own if not provided 	<ul style="list-style-type: none"> Include author-computed adjusted mean or effect size and how it was computed Reviewers may compute effect size in Hedge's <i>g</i>, correcting as necessary for clustering Maintains list of acceptable effect size statistics 	<ul style="list-style-type: none"> Looks for effects in natural, commonly understood units and whether they are large enough to constitute meaningful improvements in someone's life
Limitations	<ul style="list-style-type: none"> Not mentioned 	<ul style="list-style-type: none"> Not mentioned 	<ul style="list-style-type: none"> Acknowledge limitations of the study and how they were addressed 	<ul style="list-style-type: none"> Not mentioned 	<ul style="list-style-type: none"> Not mentioned 	<ul style="list-style-type: none"> Not mentioned

Notes: Not all clearinghouses mentioned explicit requirements for each dimension listed in this table (denoted *Not mentioned*).

ITT = intent-to-treat; TOT = treatment-on-the-treated; WWC = What Works Clearinghouse.

^a See HomVEE's *Handbook of Procedures and Evidence Standards: Version 2* beginning on page 31 (December 2020, https://homvee.acf.hhs.gov/sites/default/files/2020-12/HomVEE_Final_V2_Handbook-508.pdf).

Appendix C. Additional Clearinghouse Resources

The California Evidence-Based Clearinghouse for Child Welfare (<https://www.cebc4cw.org>)

- CEBC Rating Policy and Procedures Manual (December 10, 2020, https://www.cebc4cw.org/files/CEBC_Rating_Policy_and_Procedures_Manual_12102020.pdf)
- Overview of the CEBC Scientific Rating Scale (May 13, 2019, <https://www.cebc4cw.org/files/OverviewOfTheCEBCScientificRatingScale.pdf>)

Blueprints for Healthy Youth Development (<https://www.blueprintsprograms.org>)

- Christine M. Steeger, Pamela R. Buckley, Fred C. Pampel, Charleen J. Gust, and Karl G. Hill, “Common Methodological Problems in Randomized Controlled Trials of Preventive Interventions,” *Prevention Science* 22 (2021): 1,159–72, <https://www.doi.org/10.1007/s11121-021-01263-2>.

Title IV-E Prevention Services Clearinghouse (preventionservices.acf.hhs.gov)

- *Handbook of Standards and Procedures, Version 1.0* (OPRE report 2019-56, April 2019, https://www.acf.hhs.gov/sites/default/files/documents/opre/psc_handbook_v1_final_508_compliant.pdf)
- *Reporting Guide for Study Authors* (OPRE report 2021-27, April 2021, <https://www.acf.hhs.gov/opre/report/title-iv-e-prevention-services-clearinghouse-reporting-guide-study-authors>)

Social Programs That Work (<https://evidencebasedprograms.org>)

- “Checklist for Reviewing a Randomized Controlled Trial of a Social Program or Project, To Assess Whether It Produced Valid Evidence” (January 2010, <https://evidencebasedprograms.org/related-resources/>)
- “Key Items to Get Right When Conducting Randomized Controlled Trials of Social Programs” (February 2016, <https://craftmediabucket.s3.amazonaws.com/uploads/PDFs/Key-Items-to-Get-Right-When-Conducting-Randomized-Controlled-Trials-of-Social-Programs.pdf>)

National Institute of Justice CrimeSolutions (<https://crimesolutions.ojp.gov>)

- “Practices Scoring Instrument” (June 21, 2013, <https://crimesolutions.ojp.gov/sites/g/files/xyckuh246/files/media/document/PracticeScoringInstrument.pdf>)

Home Visiting Evidence of Effectiveness (<https://homvee.acf.hhs.gov>)

- “Reporting Guide for Authors” (OPRE report 2020-181, May 28, 2021, <https://www.acf.hhs.gov/opre/report/home-visiting-evidence-effectiveness-reporting-guide-authors-2020>)
- “Publications: Methods and Standards” (Fed. Reg. 86, no. 53, March 22, 2021, <https://homvee.acf.hhs.gov/publications/methods-standards>)

References

- Brewsaugh, Katrina, and Sarah Prendergast. *Ten Key Design Elements for Rigorous Impact Evaluations in Child Welfare: A Desk Reference for Evaluators*. OPRE Report #2022-171. Washington, DC: Urban Institute; US Department of Health and Human Services (HHS), Administration for Children and Families (ACF), Office of Planning, Research, and Evaluation (OPRE).
- CDC (Centers for Disease Control and Prevention). 2018. *Your Guide to Clear Writing*. Washington, DC: HHS, CDC.
- Sama-Miller, Emily, Julieta Lugo-Gil, Jessica Harding, Lauren Akers, and Rebecca Coughlin. 2020. *Home Visiting Evidence of Effectiveness (HomVee) Systematic Review: Handbook of Procedures and Evidence Standards: Version 2*. OPRE report 2020-151. Washington, DC: HHS, ACF, OPRE.
- Wilson, Sandra Jo, Cristofer S. Price, Suzanne E. U. Kerns, Samuel R. Dastrup, and Scott R. Brown. 2019. *Title IV-E Prevention Services Clearinghouse Handbook of Standards and Procedures, Version 1.0*. OPRE report 2019-56. Washington, DC: HHS, ACF, OPRE.
- WWC (What Works Clearinghouse). 2020. *What Works Clearinghouse Standards Handbook, Version 4.1*. Washington, DC: US Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

About the Authors

Sarah Prendergast is a research associate in the Center on Labor, Human Services, and Population at the Urban Institute. As an applied developmental scientist, her research focuses on programs, policies, and contexts that support child and family well-being. In 2017, Prendergast was awarded the Doris Duke Fellowship for the Promotion of Child Well-Being through Chapin Hall at the University of Chicago. Before joining Urban, Prendergast worked in the Colorado Department of Human Services's Office of Early Childhood as a Society for Research in Child Development Postdoctoral State Policy fellow. Prendergast received her PhD in applied developmental science from Colorado State University.

Katrina Brewsaugh is a senior research associate in the Center on Labor, Human Services, and Population where she focuses on issues related to child welfare. Before joining Urban, she was a senior associate at the Annie E. Casey Foundation providing analytic support to the Child Welfare Strategy Group. She has a PhD in social work from the University of South Florida and a master's in social work from the Florida State University.

STATEMENT OF INDEPENDENCE

The Urban Institute strives to meet the highest standards of integrity and quality in its research and analyses and in the evidence-based policy recommendations offered by its researchers and experts. We believe that operating consistent with the values of independence, rigor, and transparency is essential to maintaining those standards. As an organization, the Urban Institute does not take positions on issues, but it does empower and support its experts in sharing their own evidence-based views and policy recommendations that have been shaped by scholarship. Funders do not determine our research findings or the insights and recommendations of our experts. Urban scholars and experts are expected to be objective and follow the evidence wherever it may lead.



500 L'Enfant Plaza SW
Washington, DC 20024

www.urban.org