Title IV-E Prevention Services CLEARINGHOUSE

OPRE

Abt

How Does the Prevention Services Clearinghouse Rate the Design and Execution of Studies?

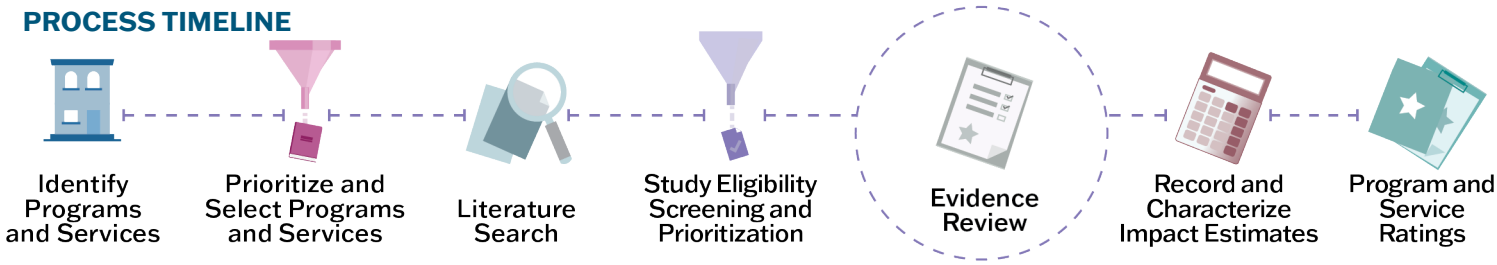
Handbook of Standards and Procedures, Version 2.0

Abt Global: Cara Jackson, Sandra Jo Wilson, and Meaghan Glenn.
OPRE Report 2024-339 | October 2024

After studies are deemed eligible for review, prioritized studies are systematically reviewed using the Prevention Services Clearinghouse design and execution standards. Studies are assigned a rating of high, moderate, or low support of causal evidence, based on the extent to which they meet the standards.

The study design and execution standards assess the extent to which a study was designed and executed in a manner that indicates the program or service, and not any other factors, caused the observed outcomes. Chapter 5 of the Handbook Version 2.0 provides details on the design and execution standards, and the Reporting Guide for Study Authors provides table shells and guidance on how to report information needed to evaluate studies against the design and execution standards.

PROCESS TIMELINE



How often do studies meet the Clearinghouse's study design and execution standards?

Just under a quarter of studies reviewed by the Clearinghouse receive high ratings (22%) and provide rigorous evidence indicating that the program or service caused the outcomes observed. About a quarter of studies receive moderate ratings (26%) and provide some evidence that it was the program or service, and not other factors, that caused the outcomes observed. About half of the studies receive low ratings (51%) and do not provide credible evidence that the program or service caused the outcomes observed. **Figure 1** depicts the distribution of study ratings.

Studies that receive moderate or high ratings **and** that have favorable effects in an eligible outcome domain can contribute to the Clearinghouse's program or service ratings of promising, supported, and well-supported.

Figure 1: Distribution of Study Ratings in the Clearinghouse

Note: data as of June 2024.

What are the most common reasons that studies receive low ratings on the design and execution standards?

Studies may fail to meet design and execution standards for a variety of reasons, and some studies fail for multiple reasons. The most common reasons that studies do not meet design and execution standards are detailed below.

- **The study does not establish baseline equivalence on pre-intervention measures (applicable to QEDs and RCTs with high attrition).** If a study does not use random assignment, or random assignment is compromised due to attrition or other factors, the study must establish that the analytic samples of the intervention and comparison groups were equivalent on baseline measures prior to the implementation of the intervention. If the groups are different at the beginning of a study, it is not clear whether differences observed at the end of a study are due to the program or to pre-existing differences across groups (Handbook Version 2.0 Section 5.7).
- **The impact of the intervention is confounded with another factor that is related to the outcome and only aligns with one group.** In such cases, the study cannot isolate the effect of the intervention from the effect of the confounding factor (Handbook Version 2.0 Section 5.9).

- **The study includes participants who are missing some data,** and the analysis does not use an acceptable approach for addressing missing data. Some approaches to missing data may bias the findings in favor of one group over the other, compromising the ability to assess whether the program or service was responsible for the outcomes observed (Handbook Version 2.0 Section 5.9.4).

Less frequently, studies fail to meet outcome standards or statistical model standards (Handbook Version 2.0 Section 5.9.1).

Can design and execution issues be addressed, and if so, how?

- Some design and execution issues cannot be addressed after the completion of a study, such as when there is a design confound. An example is when an intervention is administered by a single therapist who does not also provide services to the control group (n=1 person-provider confound, Handbook Version 2.0 Section 5.9.3).
- Comprehensive reporting allows the Clearinghouse to assess whether design and execution standards can be met (see the Reporting Guide for Study Authors for advice on how to describe studies completely), or by responding to an author query from the Clearinghouse (e.g., providing internal consistency statistics for study measures upon request).
- **Table 1** presents guidance to address common issues with design and execution standards along with examples of studies that do and do not meet standards.

Table 1. Guidance for Addressing Common Issues with Design and Execution Standards

Design or Execution Standard	Examples of Studies Not Meeting Standards	Approaches to Design and Execute Studies in Alignment with Standards
Baseline Equivalence *Baseline equivalence is the extent to which the analytic intervention and comparison groups appear similar at baseline. Acceptable measures for establishing baseline equivalence include direct pretests, correlated pretests, pretest alternatives, or sociodemographic characteristics. See Section 5.7 of the Handbook for more information.*	The treatment group's average score on the pretest measure for the outcome **is more than 0.25 standard deviations above or below** the comparison group's average score and the pretest is not included in the statistical model to adjust for these differences.	Use matching techniques to ensure that treatment and comparison groups are similar at baseline. If the difference between the groups at baseline is between 0.05 and 0.25 standard deviations, include the pretest in the statistical model to adjust for these differences. See p. 14 of the Reporting Guide for Study Authors for a table shell example.
Confounds: N=1 Person-Provider *Intervention effects cannot be separated from the skills/abilities of the treatment provider when the treatment group has a single provider, and the comparison group receives no treatment or has a different treatment provider.*	A **single therapist** provides treatment to treatment group; comparison group is waitlisted and receives no treatment. The intervention is delivered in 2-person teams. A single team delivers all treatment; comparison group referred to services in the community.	Use two or more treatment provider units (e.g. therapists) in the treatment and comparison group. If only able to conduct a study with a single provider, have the provider also administer business-as-usual treatment to the comparison condition (e.g. placebo, attention control, treatment as usual, or another intervention).
Missing Data *Acceptable approaches to missing data on post-tests, pre-tests, or pre-test alternatives include:* • *Complete case analysis* • *Regression imputation* • *Maximum likelihood* • *Non-response weights** • *Constant replacement**	A study author using a QED matching design **replaces missing outcome measure data with the mean outcome value** for individuals whose data was not missing on the measure.	Use an eligible missing data technique: complete case analysis, regression imputation, maximum likelihood, non-response weights, or constant replacement. If missing data are imputed, include sample counts, means, and standard deviations on imputed and complete case samples for the comparison and intervention groups so the Clearinghouse can assess potential imputation bias. See the Appendix of the Reporting Guide for Study Authors for guidance on reporting missing or imputed data.
Outcome Measurement Standards *To satisfy the reliability standards, the measure must have internal consistency or inter-rater reliability of 0.50 or higher, test-retest reliability of 0.40 or higher, or inter-rater agreement of 0.80 or higher (percent agreement) or 0.60 or higher (kappa).*	Study authors created their own measure of child well-being by adapting questions from an established measure. The authors **did not report any reliability metrics.**	Ensure that the contrast has sufficient internal consistency, test-retest reliability, inter-rater reliability, or inter-rater agreement, either by using measures with known reliability or by checking the reliability of customized measures. Report the reliability metrics of all outcomes in the study.
Statistical Model *Impact models cannot include endogenous measures as covariates.*	Study authors collected data on time spent in therapy sessions during the intervention period and **included this measure in the statistical mode**l of program impact.	Ensure that the statistical model does not include time-variant variables collected or obtained after baseline that could have been influenced by group status, such as implementation fidelity, attendance, or time spent in therapy sessions. Describe all covariates included in the model.

* Please see Handbook Version 2.0, Section 5.9.4 for additional requirements.

How Does the Prevention Services Clearinghouse Rate the Design and Execution of Studies? Handbook of Standards and Procedures, Version 2.0

OPRE Report 2024-339
October 2024
Cara Jackson, Sandra Jo Wilson, and Meaghan Glenn

Submitted to:
Christine Fortunato, Laura Nerenberg, and Jenessa Malin, Project Officers
Office of Planning, Research, and Evaluation
Administration for Children and Families
U.S. Department of Health and Human Services

Contract Numbers: HHS P233201500069I | GS00F252CA

Project Director: Sandra Jo Wilson
Abt Global
6130 Executive Blvd.
Rockville, MD, 20852