Handbook of Standards and Procedures

VERSION 1.0

Title IV-E Prevention Services CLEARINGHOUSE

TITLE IV-E PREVENTION SERVICES CLEARINGHOUSE

HANDBOOK OF STANDARDS AND PROCEDURES, Version 1.0

OPRE Report 2019-56

April 2019

Sandra Jo Wilson, Cristofer S. Price, Suzanne E. U. Kerns, Samuel R. Dastrup, and Scott R. Brown

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

Contract Number: HHS P233201500069I

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

The Prevention Services Clearinghouse is supported by Contract No. HHSP233201500069I from the Administration of Children and Families to Abt Associates. This Handbook was developed by Abt Associates staff with oversight from federal staff from the Administration for Children and Families and informed by consultations with research and practice experts.

Contents

Introduction

This *Prevention Services Clearinghouse Handbook of Standards and Procedures* provides a detailed description of the **standards** used to identify and review programs and services for the Prevention Services Clearinghouse and the **procedures** followed by the Prevention Services Clearinghouse staff. The Prevention Services Clearinghouse periodically provides clarification on topics covered in the Handbook. To learn more, please visit the FAQ page on the Prevention Services Clearinghouse website.

Purpose of the Title IV-E Prevention Services Clearinghouse

The Title IV-E Prevention Services Clearinghouse (hereafter referred to as the Prevention Services Clearinghouse) was established by the Administration for Children and Families (ACF) within the U.S. Department of Health and Human Services (HHS) to systematically review research on programs and services intended to provide enhanced support to children and families and prevent foster care placements. The Prevention Services Clearinghouse, developed in accordance with the Family First Prevention Services Act (FFPSA) of 2018, as codified in Title IV-E of the Social Security Act, rates programs and services as promising, supported, and well-supported practices. These practices include mental health prevention and treatment services, substance abuse prevention and treatment services and in-home parent skill-based programs, as well as kinship navigator programs.

The Prevention Services Clearinghouse was developed to be an objective, rigorous, and transparent source of information on evidence-based programs and services that may be eligible for funding under Title IV-E of the Social Security Act as amended by the FFPSA. The Prevention Services Clearinghouse uses a systematic review process implemented by trained reviewers using consistent, transparent standards and procedures (see Exhibit 1).

On June 22, 2018, HHS published a Federal Register Notice (FRN; 83 FR 29122) requesting public comment on initial criteria and potential programs and services to be considered for systematic review. Commenters included state and local administrators, program and service developers, foundations, non-profit organizations, tribes, researchers and evaluators, and other stakeholders. The Prevention Services Clearinghouse systematic review process was informed by public comments submitted in response to HHS' FRN and the review processes developed and used by other prominent evidence clearinghouses, including the Institute of Education Sciences' What Works Clearinghouse (WWC), the Administration for Children and Families' Home Visiting Evidence of Effectiveness review (HomVEE), and the California Evidence-based Clearinghouse for Child Welfare (CEBC).

On November 30, 2018, the Children's Bureau (CB) released the Program Instruction ACYF-CB-PI-18-09, ACYF-CB-PI-18-10, and ACYF-CB-18-11. This program issuance

provided instructions on the requirements state title IV-E agencies must meet when electing the title IV-E prevention program. Attachment C of the Program Instruction included revised initial criteria for selecting, reviewing, and rating programs and services, as well as the first list of programs and services selected for review by the Prevention Services Clearinghouse.

The Prevention Services Clearinghouse systematic review process, described in detail in the chapters that follow and shown in Exhibit 1, includes the following steps:

1. *Identify programs and services for review.* Candidate programs and services relevant to the mission of the Prevention Services Clearinghouse are identified using an inclusive process that invites recommendations from stakeholders, including states, to ensure broad coverage across program or service areas (Chapter 1).
2. *Select and prioritize programs and services for review.* Candidate programs and services are evaluated against the program or service eligibility criteria and prioritized for review (Chapter 2).
3. *Literature search.* Prevention Services Clearinghouse staff conduct comprehensive literature searches to locate available and relevant research on the prioritized programs and services (Chapter 3).
4. *Study eligibility screening and prioritization.* Studies identified in the literature searches are screened against the study eligibility criteria. Studies determined to be eligible for review are considered against prioritization criteria to determine the order and depth of their review (Chapter 4).
5. *Evidence review.* All eligible studies are reviewed by trained reviewers using the Prevention Services Clearinghouse design and execution standards. Study authors may be queried to request information deemed necessary to assign a rating. One of three ratings is assigned to prioritized studies: **high**, **moderate**, or **low** support of causal evidence (Chapter 5).
6. *Program and service ratings.* Studies that are rated as **high** or **moderate** support of causal evidence are considered in assigning each program or service one of four ratings: ***well-supported***, ***supported***, ***promising***, or ***does not currently meet criteria*** (Chapter 6). These ratings also take into consideration any evidence of risk of harm.

The ratings for all programs and services reviewed for the Prevention Services Clearinghouse, along with other details about the programs and services and about the studies providing evidence, are posted on the Prevention Services Clearinghouse website.

Operational procedures for reviewing programs and services in the Prevention Services Clearinghouse are included in Chapter 7. This includes procedures for re-review of programs and services due to missing information, errors in the original review, emergence of substantial new evidence, or requests by state and local administrators,

program and service developers, tribes, researchers and evaluators, and other stakeholders (Section 7.4).

Exhibit 1. The Prevention Services Clearinghouse Review Process

1. Identify Programs and Services for Review

> The Prevention Services Clearinghouse periodically provides clarification on topics covered in the Handbook. To learn more, please visit the FAQ page on the Prevention Services Clearinghouse website.

Identification of first programs and services for review. The first programs and services selected for systematic review met at least two of the following conditions: (1) recommendation from State or local government administrators in response to the Federal Register Notice 83 FR 29122 (2018 FRN); (2) rated by the California Evidence-Based Clearinghouse; (3) evaluated by Title IV-E Child Welfare Waiver Demonstrations; (4) recipient of a Family Connection Discretionary Grant; and/or (5) recommendation solicited from federal staff in the Administration for Children and Families, Health Resources and Services Administration, the National Institutes of Health, the Centers for Disease Control and Prevention, the Office of the Assistant Secretary for Planning and Evaluation, and the Substance Abuse and Mental Health Services Administration.

Identifying additional programs and services for review. Programs and services for potential review will be identified from:

- Recommendations received in response to the 2018 FRN (including those received from state or local government administrators and tribes), federal partners, and other key stakeholders; and
- A public call for program and service recommendations. At least annually, the Prevention Services Clearinghouse will issue a public call for programs and services and send it to relevant listservs for dissemination. Submissions to the call will be added to the Prevention Services Clearinghouse database of recommended programs and services. The Prevention Services Clearinghouse will retain all submissions that are eligible for review. The public, including state or local government administrators and tribes, will have the opportunity to submit program or service recommendations for potential review through electronic submission or mail directly to the Prevention Services Clearinghouse. Submissions may also include publicly available literature submitted by stakeholders in support of recommended programs and services.

In addition, the Prevention Services Clearinghouse may use an environmental scan or an inventory of the literature or both to identify programs or services.

Particular consideration will be given to programs and services recommended by State or local government administrators and tribes; rated by other clearinghouses (such as CEBC or HomVEE); recommended by federal partners; and/or evaluated as part of any grants supported by the Children's Bureau (such as the Title IV-E Child Welfare Demonstrations or Regional Partnership Grants).

2. Select and Prioritize Programs and Services

2.1 Program or Service Eligibility Criteria

This section describes the criteria for determining whether programs and services under consideration are eligible for inclusion in the Prevention Services Clearinghouse.

The Prevention Services Clearinghouse periodically provides clarification on topics covered in the Handbook. To learn more, please visit the FAQ page on the Prevention Services Clearinghouse website.

2.1.1 Program or Service Areas

Title IV-E of the Social Security Act describes four program or service areas—mental health prevention and treatment programs or services, substance abuse prevention and treatment programs or services, in-home parent skill-based programs or services, and kinship navigator programs. Programs and services may be eligible for Prevention Services Clearinghouse review in more than one of these program or service areas.

Mental Health Prevention and Treatment Programs and Services

Eligible mental health programs and services include those that aim to reduce or eliminate behavioral and emotional disorders or risk for such disorders. Included programs and services may target any mental health issue. It is not required that participants in the program or service have a Diagnostic and Statistical Manual (DSM) or International Statistical Classification of Diseases (ICD) diagnosis. Eligible programs and services can be delivered to children and youth, adults, or families; can employ any therapeutic modality, including individual, family, or group; and, may have any therapeutic orientation, such as cognitive, cognitive-behavioral, psychodynamic, structural, narrative, etc. Programs and services that rely on psychotropic medications or screening procedures without a counseling or behavioral therapeutic component are not eligible (e.g., a treatment that uses methylphenidate or lisdexamfetamine for treatment of Attention Deficit Hyperactivity Disorder without an accompanying therapeutic element).

Substance Abuse Prevention and Treatment Programs and Services

Eligible substance abuse prevention and treatment programs and services include those that have an explicit focus on the prevention, reduction, treatment, remediation, and/or elimination of substance use, misuse, or exposure in general. Included programs and services can target any specific type of substance, multiple substances, or aim to address substance use or misuse in general. Programs and services targeting use or misuse of alcohol, marijuana, illicit drugs, or misuse of prescription or over-the-counter drugs are eligible. Eligible programs and services can be delivered to children and youth, adults, or families. Programs and services aimed solely at reducing, treating, or remediating tobacco use (including smoking, chewing tobacco, and vaping) among adults are not eligible. Eligible programs and services can employ any therapeutic

modality, including individual, family, or group and may have any therapeutic orientation, such as cognitive, cognitive-behavioral, psychodynamic, structural, narrative, etc. Programs and services may include use of pharmacological treatment approaches. Not eligible are programs and services that are directed only at collateral persons or caregivers, or systems interventions that would not generally be recognized as client-oriented substance use treatment. Additionally, programs and services that are pre-clinical programs (e.g., screening or brief programs aimed solely at getting people into treatment) and that do not themselves involve prevention or treatment are not eligible. However, brief programs that do involve prevention or treatment (i.e., make some attempt to address substance use) are eligible. Programs and services that solely rely on pharmacological interventions without a therapeutic component are not eligible (e.g., a treatment that uses methadone for the treatment of opioid addiction without an accompanying therapeutic element).

Exhibit 2.1 provides some examples of eligible and ineligible programs and services in this area.

Exhibit 2.1. Examples of Substance Abuse Prevention and Treatment Programs and Services

Eligible Examples	Not Eligible Examples
A program that is delivered in a group setting for adolescents who were identified as having either marijuana use or prescription pill misuse within the prior 30 days.	A program that does not work directly with the youth but intervenes with the adults in the youth's life to ensure that there is adequate supervision and monitoring to limit access to substances and substance using peers.
A program treating mothers who are misusing opioids using a combination of methadone, cognitive behavioral therapy, and peer support.	A standalone screening program that uses social norming to attempt to motivate people to seek treatment.
A brief, 30 minute motivational intervention that is delivered in emergency rooms after a patient is seen for a drug overdose.	A program that uses the medication acamprosate to reduce withdrawal symptoms for adults with alcohol use disorder without an accompanying therapeutic component.

In-Home Parent Skill-Based Programs and Services

Eligible parent skill-based programs and services include those that are psychological, educational, or behavioral interventions or treatments, broadly defined, that involve direct intervention with a parent or caregiver. Direct intervention contact means that intervention services are provided directly to the parent(s) or caregiver(s); children may be present or involved, but are not required to be present for a program to be eligible. Contact may be face-to-face, over the telephone or video, or online. Programs may be explicitly delivered as in-home interventions or can be interventions for which delivery

in-home is a possible or recommended method to administer the intervention. This may include residential facilities, shelters, or prisons if that is where the parent(s) or caregiver(s) resides.

Exhibit 2.2 provides some examples of eligible and ineligible programs and services in this area.

Exhibit 2.2. Examples of In-Home Parent Skill-Based Programs and Services

Eligible Examples	Not Eligible Examples
A program that is delivered in the family home in individual sessions for 12 weeks. Both the parent and the child attend and the parent is coached to use different skills with the child during the session.	An aggression reduction training for parents of adolescents that is delivered in small groups for 10 weeks and in-home delivery is not possible.*
An on-line parenting program that helps parents set goals and match their parenting goals with evidence-based parenting strategies.	A public service campaign that focuses on positive parenting practices is delivered in a community using television and radio spots, public posters and billboards, and direct mailings.

*This example could be considered within the mental health program or service area.

Kinship Navigator Programs

Eligible kinship navigator programs and services include those focused on assisting kinship caregivers in learning about, finding, and using programs and services to meet the needs of the children and youth they are raising and their own needs, and that promote effective partnerships among public and private agencies to ensure kinship caregiver families are served. Support services may include any combination of financial supports, training or education, support groups, referrals to other social, behavioral, or health services, and assistance with navigating government and other types of assistance, financial or otherwise.

Adaptations to Programs or Services

Many manualized programs have formal adaptations available (i.e., alternative manualized versions of the original program designed to address particular issues or populations). When programs and services that are identified for inclusion in the Prevention Services Clearinghouse have multiple available formal adaptations or multiple treatment manuals, each is reviewed as a separate program or service.

Programs or services that go by different names in different local implementations but that clearly use the same manual are considered to be the same program for purposes of review. Minor modifications to programs or services that are not considered formal adaptations are addressed in Section 4.1.6 below.

In order to maximize the number of different programs reviewed by the Prevention Services Clearinghouse, the Prevention Services Clearinghouse may select one program adaptation for review when multiple formal adaptations are available. In most cases, the Prevention Services Clearinghouse will select the standard, original, or most comprehensive or complete version of a program or service; however, it may also consider other adaptations.

Kinship caregivers may be a grandparent or other relative as well as tribal kin, extended family and friends or other “fictive kin” who are caring for children. Kinship care relationships may be formal or informal.

Programs that involve helping members of the general public access services, irrespective of whether they are caregivers or not, are not eligible.

2.1.2 Book/Manual/Writings Available

To be eligible for the Prevention Services Clearinghouse, programs and services in any of the four program or service areas must be clearly defined and replicable. To meet this criterion, programs and services must have available written protocols, manuals, or other documentation that describes how to implement or administer the practice. Protocols, manuals, or other documentation must be available to the public to download, request, or purchase. Programs and services that require training, certification, or other prerequisites to access manuals or other documentation would meet this criterion.

2.2 Program or Service Prioritization Criteria

For each program or service considered for inclusion in the Prevention Services Clearinghouse, reviewers record whether the program or service explicitly aims to impact each of the target outcomes; whether it is currently in active use; and whether there are implementation and fidelity supports available in addition to a manual or protocol. Reviewers make these determinations by reviewing available documentation and websites, though they may also consult research studies or program developers to gather additional information. The Prevention Services Clearinghouse also prioritizes programs and services in a way that ensures representation of programs and services across the four program or service areas: mental health prevention and treatment programs and services, substance abuse prevention and treatment programs and services, and in-home parent skill-based programs and services, as well as kinship navigator programs.

2.2.1 Target Outcome Domains

Programs and services in the areas of mental health, substance abuse, and in-home parent skills must target outcomes in the domains of (a) Child Safety, (b) Child Permanency, (c), Child Well-Being, and/or (d) Adult Well-Being.

Programs and services in the area of kinship navigator must target outcomes in the domains of (a) Child Safety, (b) Child Permanency, (c) Child Well-Being, (d) Adult Well-Being, (e) Access to Services, (f) Referral to Services, and/or (g) Satisfaction with Programs and Services. Operational definitions for the eligible target outcomes are provided in Section 4.1.5.

2.2.2 In Use/Active

The Prevention Services Clearinghouse prioritizes programs or services that are in active use. This means that they must be currently available or delivered with a book, manual, or other documentation available in English. Programs and services that are no longer actively used, are defunct or discontinued, or are otherwise not currently practiced or delivered would not meet this criterion.

2.2.3 Existence of Implementation and Fidelity Supports

The Prevention Services Clearinghouse prioritizes programs and services for which there are implementation supports, implementation manuals or frameworks, fidelity checklists or other fidelity-monitoring tools, videos, training programs, coaching programs, or any similar resources available for potential program adopters. To meet this criterion, there must be affirmative, documented evidence that such supports are available to the public in English, either at no cost or for purchase.

3. Literature Search

For each program or service identified and prioritized for inclusion, Prevention Services Clearinghouse staff conducts a comprehensive and systematic search for potentially eligible studies of that program or service. All search results are carefully documented in databases to ensure transparency. Duplicate citations are removed before screening them for eligibility.

The Prevention Services Clearinghouse periodically provides clarification on topics covered in the Handbook. To learn more, please visit the FAQ page on the Prevention Services Clearinghouse website.

Other Clearinghouses. The search begins by identifying citations from other evidence clearinghouses or repositories. A number of evidence clearinghouses overlap in content with the Prevention Services Clearinghouse (see Exhibit 3.1). Identifying studies that these other clearinghouses have reviewed is an efficient way of locating studies that may meet Prevention Services Clearinghouse eligibility criteria.

Exhibit 3.1. Clearinghouses Used to Identify Relevant Research

Clearinghouse*	Website
Blueprints for Healthy Youth Development (Blueprints)	www.blueprintsprograms.org
California Evidence-Based Clearinghouse for Child Welfare (CEBC)	www.cebc4cw.org
Home Visiting Evidence of Effectiveness Review (HomVEE)	https://homvee.acf.hhs.gov
Child Trends What Works	https://www.childtrends.org/what-works
CrimeSolutions	www.crimesolutions.gov
Teen Pregnancy Prevention (TPP) Evidence Review	tppevidencereview.aspe.hhs.gov
Washington State Institute for Public Policy (WSIPP)	http://www.wsipp.wa.gov
The Campbell Collaboration	https://campbellcollaboration.org/
The Cochrane Collaboration	https://www.cochrane.org/

*Note: Additional clearinghouses may be used, depending on the program or service selected.

Bibliographic Databases. To ensure that searches are comprehensive, Prevention Services Clearinghouse staff also conduct searches of electronic bibliographic databases to identify additional potentially eligible studies not included on other clearinghouse sites. Trained staff use keywords to execute the searches. Content experts review these search terms for completeness, identify common synonyms, and suggest additional keywords. The following databases are included in all searches, with additional databases added as content experts recommend.

Exhibit 3.2 Bibliographic Databases Used to Identify Relevant Research

Database*	Website
Applied Social Sciences Index and Abstracts (ASSIA)	https://www.proquest.com/products-services/ASSIA-Applied-Social-Sciences-Index-and-Abstracts.html
Cumulative Index to Nursing and Allied Health Literature (CINAHL)	https://health.ebsco.com/products/the-cinahl-database
Education Resources Information Center (ERIC)	https://eric.ed.gov/
MEDLINE Complete (PubMed)	https://www.ncbi.nlm.nih.gov/pubmed/
National Criminal Justice Reference Service (NCJRS)	https://www.ncjrs.gov/
PsycINFO	https://www.apa.org/pubs/databases/psycinfo
Social Sciences Citation Index (SSCI)	http://mjl.clarivate.com/cgi-bin/jrnlst/jloptions.cgi?PC=SS

*Note: Additional databases may be used, depending on the program or service selected.

Grey Literature Scans. Finally, Prevention Services Clearinghouse staff scan the websites of federal, state, foundation, and private agencies who sponsor or conduct relevant research in order to identify any additional potentially eligible studies that may not be indexed in the standard electronic databases.

4. Study Eligibility Screening and Prioritization

4.1 Study Eligibility Criteria

The Prevention Services Clearinghouse defines a "study"[1] as one research investigation of a defined subject sample, and the interventions, measures, and statistical analyses applied to that sample. To be eligible for review for the Prevention Services Clearinghouse, studies must meet all of the eligibility criteria described below.

The Prevention Services Clearinghouse periodically provides clarification on topics covered in the Handbook. To learn more, please visit the FAQ page on the Prevention Services Clearinghouse website.

4.1.1 Date of Publication

Studies must be published or prepared in or after 1990. For studies whose results are reported in multiple documents, the earliest available document must be published or prepared in or after 1990.

4.1.2 Source of Publication

Studies must be publicly available and published in peer-reviewed journals or in reports prepared or commissioned by federal, state, or local government agencies or departments, research institutes, research firms, foundations or other funding entities, or other similar organizations. Dissertations, theses, and conference papers are not eligible.

4.1.3 Language of Publication

Studies must be available in English.

---

[1] Sometimes study results are reported in more than one document, or a single document reports results from multiple studies. Using the Institute of Education Sciences What Works Clearinghouse (WWC) v4.0 convention, two or more impact estimates will be considered as coming from a single study when they share at least three of the following four characteristics:

- The particular sample used to estimate the impact is the same or has a large degree of overlap.
- The process used to assign sample members to intervention and control conditions is the same.
- The data collection and analysis procedures are the same (or nearly the same).
- The research team is the same or has a high degree of overlap.

For additional details and examples see Appendix D: Examples of Study Definition of What Works Clearinghouse Procedures Handbook Version 4.0.

4.1.4 Study Design

Studies must use a randomized or quasi-experimental group design[2] with at least one *intervention* condition and at least one *comparison* condition. Intervention and comparison conditions may be formed through either randomized or non-randomized procedures and the unit of assignment to conditions may be either individuals or groups of individuals (e.g., families, providers, centers). Eligible *intervention* and *comparison* conditions are defined as follows:

- **Intervention Condition.** The intervention group(s) must receive a program or service that is essentially the same for all of the participants in the group (i.e., there may be variation across individuals in what they receive but distinctly different interventions should not be applied to different subsamples that are aggregated into a single study sample).
  - In a study with multiple intervention groups, reviewers determine the eligibility of each intervention based on the Program or Service Eligibility Criteria (Section 2.1). If all intervention groups are eligible, they can be reviewed and compared to the same comparison group.
- **Comparison Condition.** Comparison groups must be "no or minimal intervention" or "treatment as usual" groups. Minimal intervention group members may receive handouts, referrals to available services, or similar nominal interventions. "Treatment as usual" group members may receive services, but those services must be clearly described as the usual or typical services available for that population in the study. Studies that compare one intervention to a second intervention are not eligible for review, even if the second intervention is not eligible under the Program or Service Eligibility Criteria (Section 2.1).
  - In studies with multiple comparison groups, reviewers select one comparison instead of comparing the same intervention group to multiple comparison groups. Selection of comparison group is based on the group that receives the least intensive services in order to maximize the treatment contrast.

4.1.5 Target Outcomes

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 and are defined as follows:

---

[2] Although regression discontinuity designs are group designs, the Prevention Services Clearinghouse plans to apply separate eligibility and review criteria for study designs in which groups are constructed based on a cutoff score. The Prevention Services Clearinghouse may also apply separate standards for single case design studies. These may be forthcoming in future versions of the Handbook.

Eligible Outcomes for Mental Health, Substance Abuse Prevention and Treatment, and In-Home Parent Skill-Based Programs and Services

- **Child Safety.** Child safety refers to a current condition within a home or family and considers whether or not there is an immediate threat of danger to a child. A threat of danger refers to a specific family situation that is out of control, imminent, and likely to have severe physical, psychological, and/or developmental effects on a child. Eligible indicators of child safety for the Prevention Services Clearinghouse pertain to both child maltreatment and risk of maltreatment and include:
  - Evidence of substantiated or unsubstantiated child maltreatment from administrative records.
  - Injuries or ingestions taken from medical records of encounters with health care providers.
  - Measures that assess neglectful, psychologically aggressive, or abusive parenting behavior.
- **Child Permanency.** Child permanency refers to the permanency and stability of a child's living situation (in-home or in foster care) and includes the continuity and preservation of family relationships and connections. Eligible indicators of child permanency for the Prevention Services Clearinghouse include:
  - Length of placements, placement disruptions, stability or permanency of placements, reunification, and use of kinship care.
  - Eligible sources of this information may be reports from child welfare, juvenile justice, or similar administrative databases, including Child and Family Services Reviews. Studies may also obtain placement information from therapist, provider, or parent/caregiver reports.
- **Child Well-being.** Child well-being is a multi-faceted construct that broadly refers to the skills and capacities that enable young people to understand and navigate their world in healthy, positive ways.[3] It is an umbrella term that includes child and youth development in behavioral, social, emotional, physical, and cognitive domains. The Prevention Services Clearinghouse reviews the following domains of child well-being, the specific nature of which may vary with age:
  - *Behavioral and Emotional Functioning.* Characteristics and behaviors relating to the ability to realize one's potential, cope with daily activities, and work and play productively and fruitfully. Both strengths-based and deficit-based indicators are eligible. Examples include measures of externalizing behaviors (e.g., aggressive

[3] ACF Information Memo on Child Well-Being (2012; https://www.acf.hhs.gov/sites/default/files/cb/im1204.pdf)

behavior, disruptiveness, impulsive behavior), internalizing behaviors (e.g., depression, anxiety, mood or thought problems), mental/behavioral health diagnoses, positive behavior, resilience, self-regulation or self-control, and emotional adjustment.

- *Social Functioning.* Skills and capabilities relating to the ability to develop, maintain, and manage interpersonal relationships (e.g., social skills, assertiveness, cooperation, empathy, social adjustment, peer relations, rebelliousness, defiance, and other similar characteristics related to interpersonal interactions and relationships).
- *Cognitive Functions and Abilities.* Abilities related to reasoning, knowledge, problem-solving, mental processing, executive functioning, and the like. Eligible measures include intelligence tests, developmental assessments, measures of visual or spatial processing, and other indicators of cognitive functions and abilities.
- *Educational Achievement and Attainment.* Educational achievement refers to the extent to which students master academic content. Eligible measures include composite or subject-specific (e.g., reading, mathematics) standardized achievement test scores or overall grade point averages. Educational attainment refers to student progress through school or the completion of a degree, certificate, or program. Eligible measures of attainment include grade promotion, high school graduation or dropout rates, certificate or degree completion rates, and other indicators for educational attainment.
- *Physical Development and Health.* Characteristics related to the healthy functioning of the body may include indicators of physical health (e.g., Body Mass Index), physical capabilities (e.g., motor skills), normative indicators of healthy development (e.g., height), and any other measure relating to healthy (or unhealthy) physical development.
- *Substance Use or Misuse.* Measures of substance use or misuse may involve any substances and may be self- or other-reported, clinical tests such as urinalysis, or any other measure that provides an assessment of the participants' substance use behavior. Measures must describe actual use or misuse, such as frequency or quantity of use, type of use, use/no use, time since last use, etc. Substance use diagnoses (e.g., from a clinical interview or DSM criteria) are considered eligible outcomes in this domain. Measures that do not directly index substance use or misuse (e.g., drug-related criminal or delinquency activity such as selling drugs, drug knowledge, behavioral intentions to use or not, etc.) are not eligible in this domain, but may meet the requirements for other outcome domains.
- *Delinquent Behavior.* Delinquent behavior refers to behavior chargeable under applicable laws, whether or not apprehension occurs or charges are brought.

Chargeable offenses also include “status” offenses (e.g., runaway, truancy, curfew violations).

- **Adult Well-being.** Adult well-being refers to the specific skills and capabilities adults need to navigate their world in healthy, positive ways and provide for themselves and their children’s needs. Well-being is an umbrella term that includes outcomes in a range of individual and interpersonal domains. The Prevention Services Clearinghouse reviews the following domains of adult well-being:
  - *Parenting Practices.* Parenting practices include a range of practices and behaviors focused on developing strong, functional relations between parents or caregivers and children and the parents or caregivers’ abilities to successfully manage child socialization and support child development, health, and well-being in an effective and constructive manner. Measures may include items about basic elements of caregiving, such as feeding and physical care; communication and listening; nurturing, loving, or supportive behavior; rules and consequences; setting boundaries; warmth; scaffolding children’s behavior to develop self-discipline; parent-child relationships, and the like. Measures may index either positive parenting practices or negative parenting practices.
  - *Parent/Caregiver Mental or Emotional Health.* Mental or emotional health refers to a parent’s/caregiver’s ability to cope with daily activities, realize his or her potential, and interact productively in the world. Both strengths-based and deficit-based indicators are eligible. Examples include measures of externalizing behaviors (e.g., aggressive behavior), internalizing behaviors (e.g., depression, anxiety, mood or thought problems), mental/behavioral health diagnoses, parent/caregiver stress, relationship stress, positive behavior, resilience, and emotional adjustment.
  - *Parent/Caregiver Substance Use or Misuse.* Measures of substance use or misuse may involve any substances and may be self- or other-reported, clinical tests such as urinalysis, or any other measure that provides an assessment of the participants’ substance use or misuse. Measures must describe actual use or misuse, such as frequency or quantity of use, type of use or misuse, use/no use, time since last use, etc. Substance use diagnoses (e.g., from a clinical interview or DSM criteria) are considered eligible in this domain. Measures that do not directly index substance use or misuse (e.g., drug-related criminal or delinquency activity such as selling drugs, drug knowledge, behavioral intentions, etc.) are not eligible in this domain, but may meet the requirements for other outcome domains.
  - *Parent/Caregiver Criminal Behavior.* Criminal behavior refers to behavior chargeable under applicable laws, whether or not apprehension occurs or charges are brought.
  - *Family Functioning.* Family functioning refers to the capacity or lack of capacity of a family to meet the needs of its members and includes physical care and

maintenance of family members; socialization and education of children; and, economic and financial support of the family.

- *Physical Health.* Refers to the physical health of parents or caregivers and can include a variety of indicators including blood pressure; weight, obesity, or body mass index (BMI); chronic conditions such as asthma or diabetes; and, healthy lifestyle behaviors such as diet and exercise.
- *Economic and Housing Stability.* Economic and housing stability includes indicators of financial or economic stability (e.g., level of income, employment/unemployment, financial assistance) and/or housing stability (e.g., number of moves, quality of housing, homelessness).

Eligible Outcomes for Kinship Navigator Programs

- **Child Safety** (defined as above).
- **Child Permanency** (defined as above).
- **Child Well-Being** (defined as above).
- **Adult Well-Being** (defined as above).
- **Access to Services.** Access to services refers to a parent, caregiver, or family's knowledge of and ability to access, or utilization of services to support the family's financial, legal, social, educational, and/or health needs such as medical care, financial assistance, and social services. Parent/caregiver self-reports, informed collateral reports (e.g., from therapists or case managers), or administrative records are eligible indicators for Prevention Services Clearinghouse reviews.
- **Referral to Services.** Referral to services may include referrals to any needed financial, legal, social, educational, or health services. Measures may be obtained from parent/caregiver self-reports, therapist or provider reports or records, or administrative records. Examples include the presence or absence of referrals or counts/frequencies of referrals.
- **Satisfaction with Programs and Services.** Satisfaction with programs and services refers to parent or caregiver satisfaction with the programs and services to which they are referred or which they receive as part of a kinship navigator program.

4.1.6 Program Adaptations

When multiple formal versions of a program or service are available, the Prevention Services Clearinghouse selects just one version for review at a time and reviews eligible studies only of the version selected. Only studies of the version selected will be eligible for review for that program or service. Other versions may be eligible for review as separate programs or services. Multiple formal versions may be reviewed for the Prevention Services Clearinghouse in the same round of review or in later rounds of review.

To be eligible for review, studies of a program or service must all represent similar implementations of the program under review; that is, programs or services may not be substantially modified or adapted from the manual or version of the program or service selected for review. Adaptations or modifications to processes, such as accelerating program delivery (e.g., from two times/week to three times/week) over a shorter period, are acceptable. But, adaptations or modifications to content (such as adding a new component to an established program or service) or modality (such as changing from in-person to online) are not considered the same for purposes of Prevention Services Clearinghouse reviews (though such programs or services may be eligible themselves for review as separate programs). Reviewers document all adaptations that are reported in studies when screening them for eligibility. Senior content experts on the Prevention Services Clearinghouse staff will be consulted to develop a final decision on whether a particular adaptation is acceptable or not. Exhibit 4.1 provides examples of eligible adaptations as well as adaptations that are considered a different program or service for purposes of review.

Exhibit 4.1. Examples of Program and Service Adaptations within a Study for the Purpose of Study Review

Eligible Adaptations	Adaptations that Result in Different Program or Service
• Modestly changing session frequency or duration • Delivering the intervention in the home compared to office-based delivery • Making small changes to increase the cultural relevancy of the intervention (e.g., changing examples to match the cultural background of subjects; providing the intervention in a different language) without changing program components • Delivering the program by slightly different types of professionals than described in the manual or original research on the program or service (e.g., using social workers instead of counselors to deliver the program)	• Changing from individual to group therapy • Adding any new modules or session content • Subtracting any modules or session content that was part of the original intervention • Radically changing content for different cultural groups, such as to reflect particular issues experienced by those groups • Delivery of the program by substantially different providers than described in the manual (e.g., using para-professionals instead of nurses to deliver the program)

4.2 Study Review Prioritization Criteria

The Prevention Services Clearinghouse will review all eligible studies.

- If a program or service has less than 15 eligible studies, all studies are reviewed using the design and execution standards described in Chapter 5 and assessed for risk of harm, as described in Section 6.2.
- If a program or service has more than 15 eligible studies, all eligible studies will be assessed for risk of harm. Study review prioritization criteria (see below) will be used to determine the order of eligible studies reviewed using the design and execution standards. Once ordered, the first 15 eligible studies will be reviewed using the design and execution standards. If, after review of 15 eligible studies, a program or service has not achieved a rating of well-supported, additional studies will continue to be reviewed in order until the program or service has achieved a rating of well-supported or all eligible studies have been reviewed.

Study review prioritization criteria. As noted above, for programs and services with more than 15 eligible studies, a point system will be used to determine the order of studies reviewed. When a study is determined to be eligible for review using the above-described criteria, reviewers assign points to studies as follows:

- ***Design.*** 3 points for randomized controlled trials (RCTs), 2 points for quasi-experimental designs (QEDs).
- ***Sample Size.*** 1 point for a total sample size of 250 or more participants.
- ***Duration of Sustained Effects Examined.*** 2 points for sustained effects of 12 months or more; 1 point for sustained effects between 6 and 12 months.
- ***Number of Different Outcome Domains Examined.*** 1 point for each different outcome domain examined in the study (maximum of 3 points for Child Safety, Child Permanency, Child Well-Being, or Adult Well-Being).
- ***Pre-Registered Study Designs.*** 3 points for studies that were pre-registered in a trial registry, such as clinicaltrials.gov, or that have published study protocols.

Points are totaled for each study (maximum of 12 points). Studies are then sorted by the summed point total and reviewed in that order.

5. Evidence Review Using the Design and Execution Standards

This chapter describes the design and execution standards that are applied to all studies that receive a full review for the Prevention Services Clearinghouse. The chapter depicts the review process as a sequence of steps to arrive at a design and execution rating, as depicted in two flow charts, one for RCTs and one for QEDs. Definitions of terms are provided in boxes in this chapter as well as in the Glossary in the back of this Handbook.

The Prevention Services Clearinghouse periodically provides clarification on topics covered in the Handbook. To learn more, please visit the FAQ page on the Prevention Services Clearinghouse website.

5.1 Prevention Services Clearinghouse Ratings are applied to Contrasts

Prevention Services Clearinghouse ratings are applied to contrasts. A *contrast* is defined as a comparison of a treated condition to a counterfactual condition on a specific outcome. For example, a study with one intervention group and one comparison group that reports findings on one outcome has a single contrast. A study with one intervention group and one comparison group that reports findings on two outcomes would have two contrasts, one for each of the comparisons between the intervention and comparison group on the two outcomes. Contrasts will be reviewed from randomized controlled trial or quasi-experimental designs.

Most studies report results on more than one outcome and some studies have more than two conditions (e.g., more than one treated condition and/or more than one comparison condition).[4] When studies report results on more than one outcome or compare two or more different intervention groups to a comparison group, the study is reporting results for multiple *contrasts*. Prevention Services Clearinghouse ratings can differ across the contrasts reported in a study; that is, a single study may have multiple design and execution ratings corresponding to each of its reported contrasts.

The design and execution ratings from multiple contrasts and (if available) multiple studies are used to determine the program or service rating. Program or service ratings are described in Chapter 6. The current chapter is focused on the procedures for rating a contrast against the design and execution standards.

[4] Prevention Services Clearinghouse ratings are applied to benchmark full-sample analyses, not full-sample sensitivity analyses or subgroup analyses. Future versions of the Handbook may allow for subgroup results to receive design and execution ratings.

5.2 Design and Execution Rating Categories

For each contrast in an eligible study, Prevention Services Clearinghouse reviewers determine a separate design and execution rating. This assessment results in any of the following ratings, shown in order from strongest to weakest evidence:

- Meets Prevention Services Clearinghouse Standards for **High** Support of Causal Evidence
- Meets Prevention Services Clearinghouse Standards for **Moderate** Support of Causal Evidence
- Meets Prevention Services Clearinghouse Standards for **Low** Support of Causal Evidence

Because the level of evidence can differ among multiple contrasts reported in a study, Prevention Services Clearinghouse reviewers apply design and execution ratings to each contrast separately. Thus, a single study that reports multiple contrasts might be assigned multiple different design and execution ratings. For example, a quasi-experimental design study may report impact estimates for two outcome measures, one of which has a pre-test version of the outcome that satisfies requirements for baseline equivalence, the other of which does not satisfy baseline equivalence requirements. The first contrast may receive a *moderate* rating while the second would receive a *low* rating.

Exhibit 5.1 presents a summary of the designs that are eligible to receive *high* and *moderate* ratings. Details regarding how these ratings are derived are provided in the sections that follow.

Exhibit 5.1. Summary of Designs Eligible to Meet Design and Execution Standards

Meets Prevention Services Clearinghouse Standards for High Support of Causal Evidence	Meets Prevention Services Clearinghouse Standards for Moderate Support of Causal Evidence
Randomized studies that meet: • Standards for integrity of random assignment (Section 5.4) • Standards for low risk of joiner bias (Section 5.5) • Attrition standards (Section 5.6) • Baseline equivalence standards for randomized studies (Sections 5.7 and 5.8) • Statistical model standards (Section 5.9.1) • All measurement standards (Section 5.9.2) • All design confound standards (Section 5.9.3) • Missing data standards (Section 5.9.4)	Randomized studies that fail standards for integrity of random assignment (Section 5.4) or attrition (Section 5.6) and quasi-experimental studies that meet: • Baseline equivalence standards (Sections 5.7 and 5.8) • Statistical model standards (Section 5.9.1) • All measurement standards (Section 5.9.2) • All design confound standards (Section 5.9.3) • Missing data standards (Section 5.9.4)

Meets Prevention Services Clearinghouse Standards for High Support of Causal Evidence	Meets Prevention Services Clearinghouse Standards for Moderate Support of Causal Evidence
Meets Prevention Services Clearinghouse Standards for Low Support of Causal Evidence	
Contrasts that are reviewed and fail to meet *high* or *moderate* standards	

The Review Process Differs for RCTs versus QEDs

When Prevention Services Clearinghouse reviewers rate the evidence produced from a study contrast, they begin by making a determination about whether random assignment was used to create the contrast. Once that determination is made, they follow the sequence of steps in the flow charts depicted in Exhibit 5.2 for RCTs and Exhibit 5.5 for QEDs. The various decisions and standards that apply are described in the accompanying text.

5.3 Method of Assignment

The first step in the review process involves determining whether a contrast was created using a randomized controlled trial (RCT) design or a quasi-experimental design (QED). To be reviewed as a randomized controlled trial, the unit of assignment may be either individuals or groups of individuals (i.e., clusters), but the individuals or clusters must be assigned using a random process and each individual or cluster must have a nonzero probability of being assigned to either condition. The probability of assignment to conditions can differ across individuals or clusters (e.g., it is acceptable for a study to assign 60% of the participants to an intervention group and 40% of the participants to a comparison group).

Randomized Controlled Trial (RCT): A study in which units are assigned to intervention and control conditions via a random process (e.g., a lottery).

Quasi-experimental Design (QED): A study in which units are assigned to intervention and control conditions via a non-random process.

- If assignment to conditions is based on a random process, reviewers first assess the integrity of the randomization and attrition (see Sections 5.4 through 5.6).
- If a contrast does not use random assignment, reviewers follow the steps for QEDs that begin with an assessment of baseline equivalence (see Section 5.7).

5.4 Integrity of Random Assignment

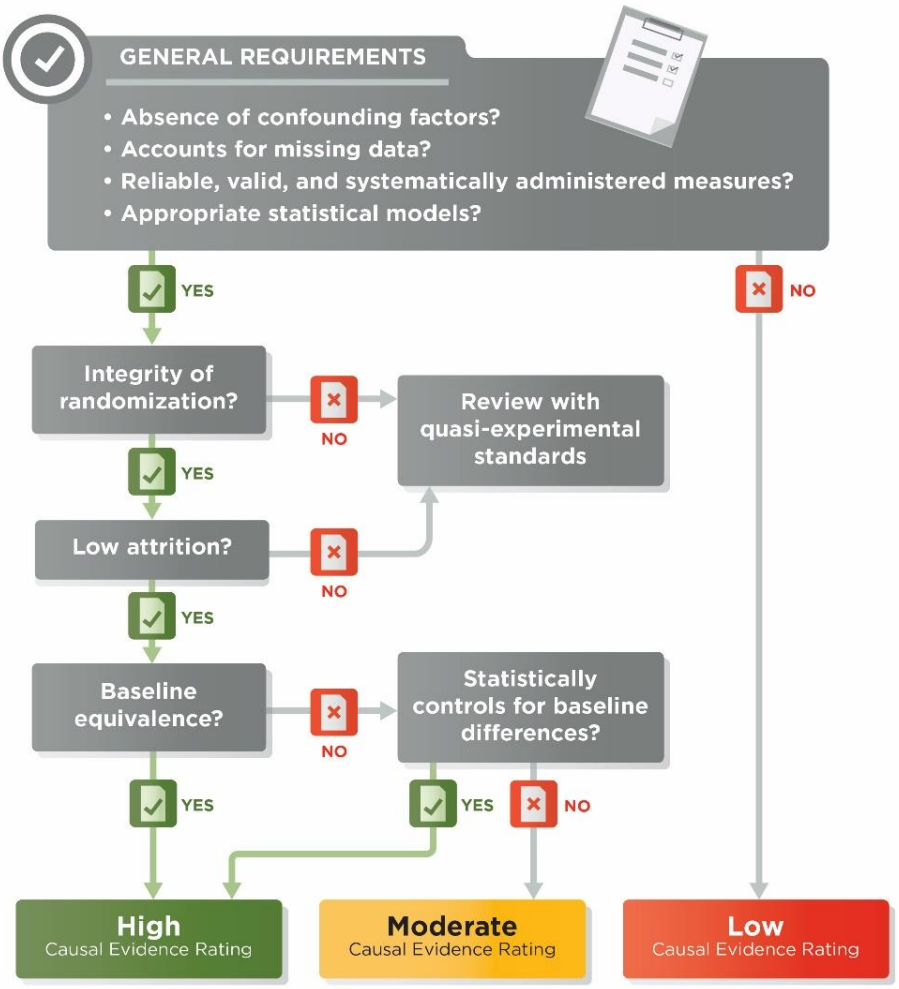
For RCTs, reviewers evaluate the integrity of the random assignment process. The integrity of random assignment is evaluated for both individual and cluster assignment RCTs. Contrasts in which the initial random assignment to intervention or comparison conditions was subsequently compromised fail the criterion for integrity of random assignment. These contrasts may be reviewed by the Prevention Services

Clearinghouse as quasi-experimental designs. The following examples illustrate ways in which random assignment can be compromised.

5.4.1 Examples of Compromised Random Assignment of Individuals

Example 1: In a study where initial assignment to intervention and comparison groups was made by a random process, the researcher identifies individuals who were randomly assigned to the intervention group, but who refused to participate in the intervention. The researcher either reclassifies those individuals as belonging to the comparison group or drops those individuals from the analysis sample. In this example, the randomization has been undermined and the study would not be classified as a study using random assignment of individuals.

Exhibit 5.2. Ratings Flowchart for Contrasts from Randomized Controlled Trials



Example 2: In a multi-site study, individuals are randomly assigned to intervention and comparison groups within 20 sites. One of the sites would not allow randomization, so

assignment to intervention and comparison groups was done by a method other than randomization. Data from all 20 sites are included in the analysis. The site with the non-random assignment has undermined the random assignment for the whole study, and the multi-site study would not be classified as a study using random assignment of individuals.

Example 3: In a study where initial assignment to intervention and comparison groups was made by a random process, the service provider is concerned because many of the individuals who were assigned to the intervention group are refusing treatment. To fill the empty treatment slots, the service provider identifies additional individuals who meet the study eligibility criteria and assigns them to the intervention group to ensure a full sample of participants. The analysis includes individuals originally assigned to the intervention group via randomization and the additional intervention members added later. In this example, the randomization has been undermined and the study would not be classified as a study using random assignment of individuals.

Example 4: In a study where initial assignment to intervention and comparison groups was made by a random process, the service provider is concerned because many of the individuals who were assigned to the intervention group are refusing treatment. To fill the empty treatment slots, the service provider recruits some of the comparison group members to participate in treatment. In the analysis, the researcher includes those treated comparison group members as belonging to the intervention group. In this example, the randomization has been undermined and the study would not be classified as a study using random assignment of individuals.

5.4.2 Examples of Changes to Random Assignment That Are Acceptable

Example 5: In a study where the initial assignment to intervention and comparison groups was made by a random process, the service provider is concerned because many of the individuals who were assigned to the intervention group are refusing treatment. To fill the empty treatment slots, the service provider recruits some individuals who were not assigned to either the intervention or comparison group in the original randomization to fill the empty slots. In the analysis, the researcher maintains the original treatment assignments, and excludes the subsequently recruited individuals from the impact analyses. In this example, the randomization has not been undermined and the study would be classified as a study using random assignment of individuals.

Example 6: Randomization to intervention and comparison groups was conducted within blocks or pairs but all intervention group members or all comparison group members of the pair or block have attrited (no outcome data are available for those members). The integrity of the randomization is not compromised if the entire pair or block is omitted from the impact analysis. For example, groups of three similar clinics were put into randomization blocks; within each block, two clinics were randomized to

the intervention condition and one to the comparison condition. During the study, a clinic closes in one of the blocks, and no outcome data were able to be collected from that clinic. The researcher dropped all three members of the block from the analysis. In this example, randomization has not been compromised.

5.5 Additional Standards for Cluster Randomized Studies

If a contrast was created by randomly assigning clusters to conditions and the randomization has not been compromised, reviewers then evaluate the potential for risk of bias from individuals joining the sample after the randomization occurred. Only cluster randomizations are evaluated for **joiner bias**.

A contrast is created by random assignment of clusters if groups of individuals (e.g., entire communities, clinics, families) are randomly assigned to intervention and control conditions, and all individuals that belong to a cluster are assigned to the intervention status of that cluster.

Cluster randomized contrasts may be subject to risk of bias if individuals can join clusters after the point when they could have known the intervention assignment status of the cluster. The risk exists if individuals can be placed into clusters after the point when the person making the placement knows the intervention assignment status of the clusters.

In such cases, individuals with different characteristics or motivations may be more likely to self-select or be assigned to one condition. When individuals can self-select into or are placed into clusters after the clusters' intervention status is known, any observed difference between the outcomes of intervention and comparison group members could be due not only to the intervention's impact on individuals' outcomes in the cluster, but also to the intervention's impact on the *composition* of the clusters (i.e., the intervention's impact on who joined or was placed into the clusters).

The Prevention Services Clearinghouse design and execution rating standards are focused on assessing the impacts of interventions on the outcomes of individuals. Therefore, if the observed impact of the intervention could be partially due to changes in the composition of clusters (for example, if individuals who are prone to more favorable outcomes are more likely to join or be placed in intervention clusters), then the impact on the composition of the clusters has biased the desired estimate of the intervention's impact on individuals' outcomes.

A cluster randomized contrast has a low risk of joiner bias in two scenarios. The first is if all individuals in a cluster joined or were placed in the cluster prior to the point when they could have plausibly known the intervention assignment status of the cluster. The second is if individuals are placed into clusters before the point when the person making the placement knows the intervention assignment status of clusters.

A cluster randomized contrast could also have low risk of joiner bias if it is very unlikely that knowledge of the intervention status would have influenced the decision to join the cluster. This holds whether individuals could join clusters soon after randomization (early joiners), or even long after randomization (late joiners).

Some contrasts may be created by randomly assigning families to conditions and then evaluating program impacts on multiple parents/caregivers and/or multiple children within those families. The Prevention Services Clearinghouse considers contrasts created this way to be cluster RCTs. Generally, reviewers assume that cluster RCTs in which families are assigned to conditions have low risk of joiner bias. That is, parents/caregivers and/or children who join families during a study are not considered to bias the impact estimates.

Contrasts created by randomly assigning clusters other than families (e.g., clinics, communities) are assumed to have low risk of joiner bias only if they have no joiners (i.e., all individuals were cluster members before knowledge of the intervention assignment status of the clusters). The exceptions to the “only if they have no joiners” requirement may include situations where the availability of the intervention isn’t publicized or isn’t noticeable to likely participants or where transfer from one provider to another is not common or allowed.

The Prevention Services Clearinghouse assumes that there is high risk of joiner bias whenever individuals are placed into clusters after the person making the placement knows the intervention assignment status of clusters. For example, if mental health clinics in a network are randomly assigned to intervention and comparison conditions, and the network administrator places families in clinics after knowing which clinics are in the intervention condition, then the Prevention Services Clearinghouse assumes that there is high risk of joiner bias.

- If reviewers determine that there are no individuals in the sample who joined clusters after assignment or there is low risk for joiner bias, they then assess attrition (see Section 5.6).
- If reviewers determine that there is high risk of joiner bias due to individuals joining clusters after assignment, attrition is not assessed and they move directly to examining baseline equivalence (see Section 5.7).

5.6 Attrition Standards

In RCTs, individuals or clusters that leave the study sample can reduce the credibility of the evidence. When the characteristics of the individuals or clusters who leave are related to the outcomes, this can result in groups that are systematically different from each other and bias the estimate of the impact of an intervention. Therefore, if a

contrast is constructed using individual random assignment or is determined to be cluster randomized with no joiners or low risk of joiner bias, reviewers evaluate attrition.

Because both overall attrition from a sample and differential attrition from intervention and comparison conditions can compromise the integrity of randomization, reviewers evaluate both overall and differential attrition. Attrition is evaluated differently for individual and cluster randomized studies, as described in the subsections below.

The Prevention Services Clearinghouse bases its standards for attrition on those developed by the What Works Clearinghouse (WWC)[5], which applies "optimistic" boundaries for attrition for use with studies where it is less likely that attrition is related to the outcomes, and "cautious" boundaries for use with studies where there is reason to believe that attrition may be more strongly related to the outcomes. The WWC's attrition model is based on assumptions about potential bias as a function of overall and differential attrition. The Prevention Services Clearinghouse uses the cautious boundary for all studies. This reflects the presumption that attrition in studies with the high risk populations of interest to the Prevention Services Clearinghouse may be linked with the outcomes targeted in Clearinghouse reviews. For example, if families at greater risk of entry into the child welfare system are more likely to drop out of a study, this can bias the results; this bias can be even more problematic if there is differential dropout between intervention and comparison groups. Exhibit 5.3 illustrates the combinations of overall and differential attrition that result in tolerable and unacceptable bias using the cautious boundary. Exhibit 5.4 shows the numeric values for the Prevention Services Clearinghouse attrition boundaries.

[5] The selection of the cautious boundary is consistent with other clearinghouses that focus on similar populations (e.g., Home Visiting Evidence of Effectiveness; Strengthening Families; Employment Strategies) See https://ies.ed.gov/ncee/wwc/Docs/ReferenceResources/wwc_attrition_v2.1.pdf and https://ies.ed.gov/ncee/wwc/Docs/ReferenceResources/wwc_attrition_v3.0.pdf for additional information about the derivation of the attrition boundaries.

Exhibit 5.3. Potential Bias Associated with Overall and Differential Attrition

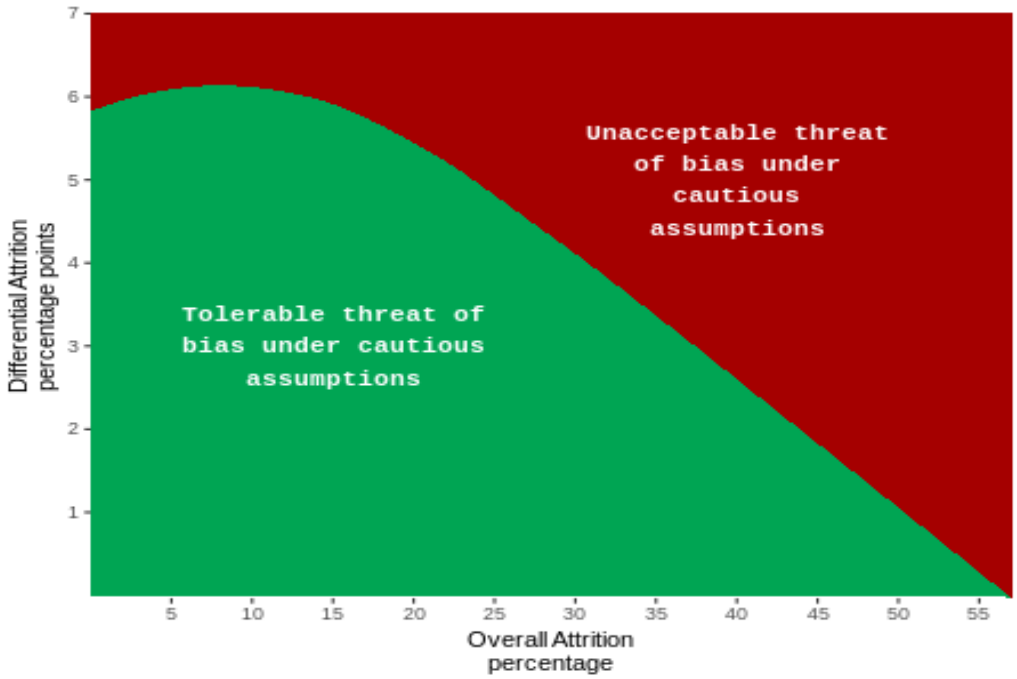


Exhibit 5.4. Prevention Services Clearinghouse Attrition Boundaries

Overall Attrition	Differential Attrition	Overall Attrition	Differential Attrition	Overall Attrition	Differential Attrition
0	5.7	20	5.4	40	2.6
1	5.8	21	5.3	41	2.5
2	5.9	22	5.2	42	2.3
3	5.9	23	5.1	43	2.1
4	6.0	24	4.9	44	2.0
5	6.1	25	4.8	45	1.8
6	6.2	26	4.7	46	1.6
7	6.3	27	4.5	47	1.5
8	6.3	28	4.4	48	1.3
9	6.3	29	4.3	49	1.2
10	6.3	30	4.1	50	1.0
11	6.2	31	4.0	51	0.9
12	6.2	32	3.8	52	0.7
13	6.1	33	3.6	53	0.6
14	6.0	34	3.5	54	0.4
15	5.9	35	3.3	55	0.3
16	5.9	36	3.2	56	0.2
17	5.8	37	3.1	57	0.0
18	5.7	38	2.9		
19	5.5	39	2.8		

Source: What Works Clearinghouse (n.d.)
Note: Overall attrition rates are given as percentages. Differential attrition rates are given as percentage points.

5.6.1 Attrition in Studies with Random Assignment of Individuals

In contrasts with individual random assignment, *overall attrition* is defined as the number of individuals without post-test outcome data as a percentage of the total number of members in the sample at the time that they learned the condition to which they were randomly assigned, specifically:

$$Overall\ Attrition = \frac{Num.of\ individuals\ without\ posttest\ outcome\ data}{Num.of\ individuals\ randomized}$$

Differential attrition is defined as the absolute value of the percentage point difference between the attrition rates for the intervention group and the comparison group, specifically:

$$Differential\ Attrition = \left| \left(\frac{Num.of\ intervention\ group\ members\ without\ posttest\ outcome\ data}{Num.of\ intervention\ group\ members\ randomized} \right) - \left(\frac{Num.of\ comparison\ group\ members\ without\ posttest\ outcome\ data}{Num.of\ comparison\ group\ members\ randomized} \right) \right|$$

The timing of randomization is central to the calculation of attrition for the Prevention Services Clearinghouse. For the purposes of defining the sample for the attrition calculation, randomization of individuals to conditions is considered to have occurred *once individuals learn their assignment condition.* This moment is defined as the earliest point in time at which any of the following occur:

- Individuals are explicitly informed about the condition to which they were assigned, or
- Individuals begin to experience the condition to which they were assigned, or
- Individuals could have plausibly deduced or have been affected by assignment to their condition, or
- Individuals have not yet experienced any of the conditions above, but their counterparts[6] have experienced it.

When eligibility and consent (if needed) is determined prior to the point in time when individuals learn their assignment condition, the Prevention Services Clearinghouse defines attrition of an individual as an individual who learned their assignment condition, but for whom an outcome measurement was not obtained. In this scenario, ineligible and unconsented individuals are not counted in the attrition calculation. This definition reflects an understanding that if an individual did not know, or could not have plausibly known their intervention status before withdrawing from a study, then the intervention assignment could not have affected a decision to participate in the study or not. If

[6] If there is randomization to conditions within strata or blocks, “counterparts” would include the other individuals in the same stratum or block.

consent is obtained after the point that individuals know their assignment condition, and no outcome measures are obtained on unconsented individuals, then the unconsented individuals are counted as attrition.

Occasionally, studies apply exclusionary conditions after the point when individuals learn their assignment condition. For the Prevention Services Clearinghouse, if the study used the same exclusionary conditions in both the intervention and the comparison groups, then eligibility criteria can be applied after that time point and ineligible individuals can be excluded from the attrition calculations and from the analysis.

Example 1: A mental health prevention program is targeted to children at risk for behavior problems. A researcher receives nominations from parents and teachers for 100 children, who are then randomly assigned to conditions. All children in both conditions are then given a diagnostic screening. Those scoring above a criterion on the screener are defined as ineligible and are excluded from the study sample. Because the same exclusion was applied in exactly the same way in both conditions, the excluded children do not need to be counted for the purpose of the attrition calculation.

Example 2: In the same study as in Example 1, a therapist in the intervention condition recognizes that one of her participants does not meet the diagnostic criteria. On her recommendation, the child transfers out of the program. For the purposes of the attrition calculation, this child must be included in the sample and cannot be classified as ineligible, because no similar screen was applied in the comparison group and a similar child who had been randomized to the comparison group would not have been identified and removed. If the researchers continue to identify the child as belonging to the intervention group for the purposes of their impact analysis, and they obtain an outcome measurement for the child, no attrition has occurred. If no outcome measurement is obtained on this child, then attrition has occurred.

Example 3: For a clinic-based study, researchers create a randomized ordering of intervention and comparison assignments and save the list to a secure website. When individual “A” walks into a clinic, an employee of the clinic does an eligibility screen. Individual “A” is determined to be eligible, and the clinic employee successfully recruits the individual to participate in the study. Individual “A” is then asked to complete a baseline survey, which she does. The clinic employee then goes to the secure website and finds the next unused randomization status record and finds that the assignment is to the intervention group. The employee tells the participant her randomization status, at which point she learns that she was randomized to receive services. Although she refuses services, an outcome measure is obtained for her, and the researcher maintains her assignment status as “intervention group” in the analysis and uses her outcome measure in the analysis. (In this example, the researcher utilizes an intent-to-treat

analysis because individuals are analyzed as members of the condition to which they were originally assigned). Individual "A" has not attrited from the study.

Example 4: In the same study as Example 3, individual "B" walks into the clinic and is determined to be eligible, and the clinic employee successfully recruits the individual to participate in the study. Individual "B" is then asked to complete a baseline survey, but does not complete it and says she wants to withdraw from the study. Individual "B" has not attrited from the study because neither she nor the clinic employee knew her randomization status at the time of her withdrawal.

Example 5: In the same study as Examples 3 and 4, individual "C" is determined to be eligible, and the clinic employee successfully recruits the individual to participate in the study. Individual "C" completes her baseline survey and learns her assignment condition. No outcome measurement is obtained for individual "C." Individual "C" has attrited from the study.

5.6.2 Attrition in Studies with Random Assignment of Clusters

In contrasts with randomization of clusters, if the contrasts exhibit low risk of joiner bias or no individuals join the sample, reviewers assess overall and differential attrition at both the cluster and individual levels. In cluster randomized contrasts, individual-level attrition is calculated only in non-attrited clusters. An attrited cluster is one in which no outcome measures were obtained for any members of the cluster. For cluster studies, individual-level overall and differential attrition are calculated as:

$$Overall\ Attrition = \frac{Num.of\ individuals\ without\ posttest\ outcome\ data}{Num.of\ individuals\ in\ nonattrited\ clusters\ at\ randomization}$$

$$Differential\ Attrition = \left| \left(\frac{\begin{matrix} Num.of\ intervention\ group\ members \\ without\ posttest\ outcome\ data \end{matrix}}{\begin{matrix} Num.of\ intervention\ group \\ members\ in\ non-attrited \\ clusters\ at\ randomization \end{matrix}} \right) - \left(\frac{\begin{matrix} Num.of\ comparison\ group\ members \\ without\ posttest\ outcome\ data \end{matrix}}{\begin{matrix} Num.of\ comparison\ group \\ members\ in\ non-attrited \\ clusters\ at\ randomization \end{matrix}} \right) \right|$$

- For cluster randomized contrasts that are deemed to have high risk of joiner bias, attrition is not relevant to the review and the contrasts are required to demonstrate baseline equivalence (see Section 5.7).
- For each contrast in a study for which attrition must be assessed, reviewers determine both overall and differential attrition at the individual level and, if applicable, at the cluster level. If attrition is determined to be below the boundaries shown in Exhibit 5.3, the contrast is said to have low attrition. If attrition is above the boundary, the contrast is said to have high attrition. Baseline equivalence is evaluated for both low and high attrition RCTs, as well as for all QEDs, using the standards described next (see Section 5.7).

5.7 Baseline Equivalence Standards

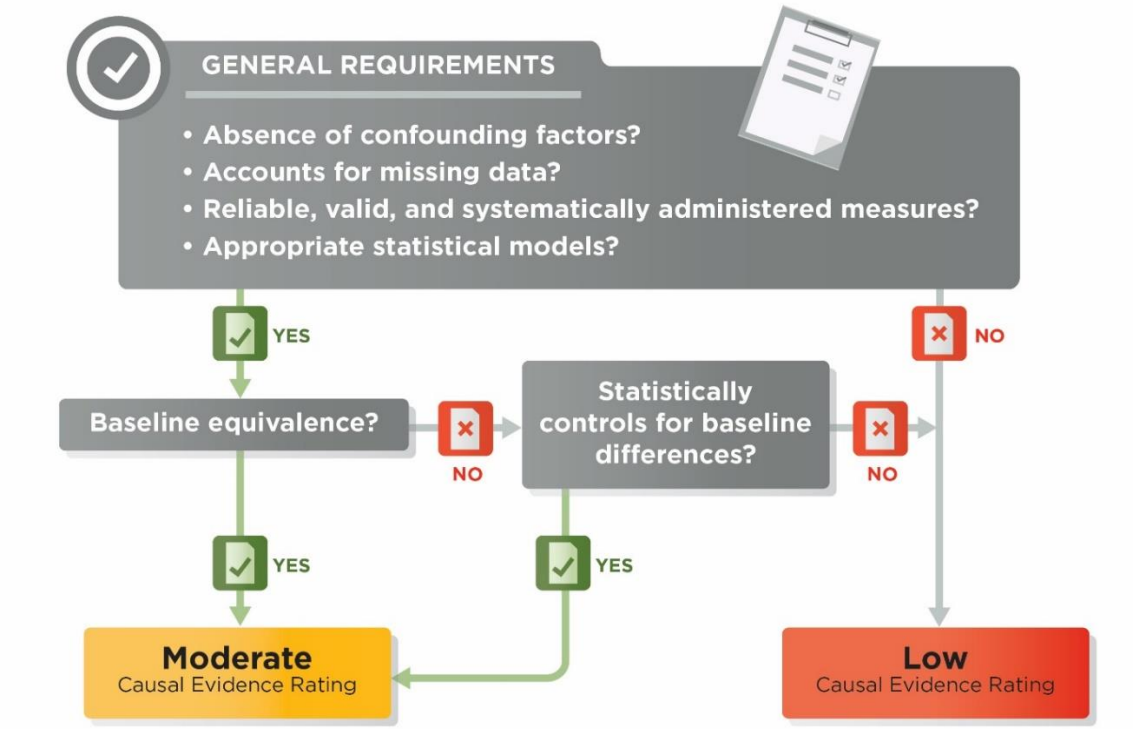
All contrasts from studies that receive full reviews by the Prevention Services Clearinghouse are assessed for baseline equivalence. In some cases, when estimating impacts, contrasts must control for the variables that are out of balance at baseline (see Section 5.7.3). Although the baseline equivalence assessment is applied to all contrasts, the assessment can affect the ratings for those created from RCTs and QEDs differently. The ratings flowchart for RCTs shown in Exhibit 5.2 illustrates how the baseline equivalence standard is applied to RCTs. The ratings flowchart for QEDs shown in Exhibit 5.5 illustrates how the baseline equivalence standard is applied to QEDs.

The Prevention Services Clearinghouse thresholds for baseline equivalence are based on those used by the WWC. Specifically, baseline equivalence is assessed by examining baseline differences expressed in effect size (ES) units. Baseline effect sizes less than 0.05 are considered equivalent and no further covariate adjustments are required.[7] Baseline effect sizes between 0.05 and 0.25 indicate that statistical adjustments in the impact models may be required (see Section 5.8); these baseline effect sizes are said to be in the *adjustment range*. Baseline effect sizes greater than 0.25 are addressed differently for low attrition RCTs versus all other designs. When statistical adjustments are required, the Prevention Services Clearinghouse standards for acceptable adjustment models described in Section 5.8 below are applied.

An exact match between the analytic sample size used to assess baseline equivalence and the analytic sample size used to estimate an impact is preferred for demonstrating baseline equivalence. Whenever there is less than an exact match in sample size between the analytic sample used to assess baseline equivalence and the sample used to estimate an impact, the Prevention Services Clearinghouse applies the WWC v4.0 standards for estimating the largest baseline difference (see Section 5.9.4). If the largest baseline difference is less than 0.25 standard deviation units, the contrast can receive a moderate rating.

[7] Where possible, reviewers record impact estimates with covariate-adjusted estimates or perform difference-in-difference adjustments regardless of whether they are required by baseline equivalence standards. When the baseline effect size is deemed equivalent, reviewers may use unadjusted impact estimates if adjusted estimates are not available.

Exhibit 5.5. Ratings Flowchart for Quasi-experimental Design Studies



5.7.1 Conducting the Baseline Equivalence Assessment

When assessing baseline equivalence, reviewers first determine whether there is a direct pre-test on the outcome variable. In general terms, a pre-test is a pre-intervention measure of the outcome. More specifically, a measure satisfies requirements for being a pre-test if it uses the same or nearly the same measurement instrument as is used for the outcome (i.e., is a direct pre-test), and is measured before the beginning of the intervention, or within a short period after the beginning of the intervention in which little or no effect of the intervention on the pre-test would expected. If there is a direct pre-test available, then that is the variable on which baseline equivalence must be demonstrated.

For some outcomes, a direct pre-test either is impossible (e.g., if the outcome is mortality), or not feasible (e.g., an executive function outcome for 3-year-olds may not be feasible to administer as a pre-test with younger children). In such cases, reviewers have two options for conducting the baseline equivalence assessment. These options are only permitted for contrasts for which it was impossible or infeasible to collect direct pre-test measures on the outcomes.

1. **Pre-test alternative.** A pre-test alternative is defined as a measure in the same or similar domain as the outcome. These are generally correlated with the outcome, and/or may be common precursors to the outcome. When multiple acceptable pre-test alternatives are available, reviewers select the variable that is most conceptually

related to the outcome prior to computing the baseline effect size. The selection of the most appropriate pre-test alternative is documented in the review and confirmed with Prevention Services Clearinghouse leadership.

2. **Race/ethnicity *and* socioeconomic status (SES).** If a suitable pre-test alternative is not available, baseline equivalence must be established on both race/ethnicity and SES.
   a. **Race/ethnicity.** For baseline equivalence on race/ethnicity, reviewers may use the race/ethnicity of either the parents or children in the study. When race/ethnicity is available for both parents and children, reviewers select the race/ethnicity of the individuals who are the primary target of the intervention. In some studies, the race/ethnicity groupings commonly used in the U.S. may not apply (e.g., studies conducted outside the U.S.). In such cases, reviewers perform the baseline equivalence assessment on variables that are appropriate to the particular cultural or national context in the study.
   b. **Socioeconomic Status (SES).** For baseline equivalence on SES, the Prevention Services Clearinghouse prefers income, earnings, federal poverty level in the U.S., or national poverty level in international contexts. If a preferred measure of SES is not available, the Prevention Services Clearinghouse accepts measures of means-tested public assistance (such as AFDC/TANF or food stamps/SNAP receipt), maternal education, employment of a member of the household, child or family Free and Reduced Price Meal Program status, or other similar measures.

In addition, reviewers examine balance on race/ethnicity, SES, and child age, when available, for all contrasts, even those with available pretests or pretest alternatives. If any such characteristics exhibit large imbalances between intervention and comparison groups, Prevention Services Clearinghouse leadership may determine that baseline equivalence is not established. Evidence of large differences (ES > 0.25) in demographic or socioeconomic characteristics can be evidence that the individuals in the intervention and comparison conditions were drawn from very different settings and are not sufficiently comparable for the review. Such cases may be considered to have substantially different characteristics confounds (see Section 5.9.3).

Reviewers examine the following demographic characteristics, when available:

- ***Socioeconomic status.*** Socioeconomic status may be measured with any of the following: income, earnings, federal (or national) poverty levels, means-tested public assistance (such as AFDC/TANF or food stamps/SNAP receipt), maternal education, employment of a member of the household and child, or family Free and Reduced Price Meal Program status.

- ***Race/ethnicity.*** Reviewers may assess child or parent/caregiver race/ethnicity, depending on what data are available in a study.
- ***Age.*** For studies of programs for children and youth, reviewers will assess baseline equivalence on child/youth age.

5.7.2 Other Baseline Equivalence Requirements

Variables that exhibit no variability in a study sample cannot be used to establish baseline equivalence. For example, if a study sample consists entirely of youth with a previous arrest, a binary indicator of that variable cannot be used to establish baseline equivalence because there is no variability on that variable.

5.7.3 How the Baseline Equivalence Assessment Affects Evidence Ratings

Randomized Studies with Low Attrition

For RCTs with low attrition, reviewers examine baseline equivalence on direct pre-tests, or pre-test alternatives or race/ethnicity and SES. If the baseline effect sizes are <0.05 standard deviation units, the contrast can receive a high rating. If the baseline effect sizes are > 0.05 standard deviation units, the contrast can receive a high rating only if the baseline variables are controlled in the impact analyses (see Section 5.8). If baseline effect sizes cannot be computed, but impact analyses clearly include the baseline variables that are required, the contrast can receive a high rating. If the baseline effect sizes are >.05 or appropriate baseline variables are not available *and* statistical controls are not used, the contrast can receive a moderate rating, provided other design and execution standards are met.

Randomized Studies with High Attrition and Quasi-Experimental Design Studies

For RCTs with high attrition and for all QEDs, reviewers examine baseline equivalence on direct pre-tests, or pre-test alternatives or race/ethnicity and SES. If the baseline effect sizes are <0.05 standard deviation units, the contrast can receive a moderate rating. If the baseline effect sizes are between 0.05 and 0.25 standard deviation units, the contrast can receive a moderate rating only if the baseline variables are controlled in the impact analyses (see Section 5.8). If statistical controls are not used, the contrast receives a low rating. If direct pre-tests are not possible or feasible and no pre-test alternatives or race/ethnicity and SES are available, baseline equivalence is not established for the outcome and that contrast receives a low rating.

5.8 Acceptable Methods for Controlling for Pre-tests

When the baseline equivalence assessment determines that an impact model must control for a baseline variable in order to meet evidence standards, any of the following approaches for statistical control are acceptable:

- Regression models with the baseline variables as covariates. This includes all commonly understood forms of regression including ordinary least squares, multi-level or generalized linear models, logistic regression, probit, and analysis of covariance.
- Gain score models where the dependent variable in the regression is a difference score equal to the outcome minus the pre-test.
- Repeated measures analysis of variance models.
- Difference-in-difference models (these must use pre-tests, not other baseline variables).
- Models with fixed effects for individuals (these must use pre-tests, not other baseline variables).

5.9 Other Design and Execution Requirements

All RCTs and QEDs that meet the requirements described above for attrition and baseline equivalence, and use acceptable methods for pre-test controls that are appropriate for the respective design and circumstances must also meet additional requirements to receive a rating of high or moderate. These requirements address issues related to the statistical models used to estimate program impacts, features of the measures and measurement procedures used in the studies, confounding factors, and missing data.

5.9.1 Statistical Model Standards

The Prevention Services Clearinghouse design and execution ratings apply standards for the statistical models that are used to estimate impacts. The statistical model standards include the following:

- When there is unequal allocation to intervention and comparison conditions within randomization blocks the impact model must account for the unequal allocation using any of the three approaches listed below. If impact models do not appropriately account for the unequal allocation, reviewers follow the steps for quasi-experimental designs.
  - Use dummy variables in the impact model to represent the randomization blocks
  - Reweight the observations such that weighted data have equal allocations to intervention and control within each randomization block
  - Conduct separated impact analyses within each block and average the impacts across the blocks.
- Impact models cannot include endogenous measures as covariates. If the impact model for a contrast includes endogenous covariates and alternate model

specifications (without such covariates) or unadjusted means and standard deviations on the outcome variable are not available, the contrast receives a low rating.

  - An endogenous covariate is one that is measured or obtained after baseline and that could have been influenced by the intervention. Inclusion of endogenous covariates results in biased impact estimates.

- The Prevention Services Clearinghouse may, in some cases, determine that a statistical model is invalid for estimating program impacts such as when data are highly skewed or if there are obvious collinearities that make estimates of program impacts suspect or uninterpretable.

5.9.2 Measurement Standards

Prevention Services Clearinghouse standards for outcomes, pre-tests, and pre-test alternatives apply to all eligible outcomes and are aligned with those in use by the WWC. Specifically, there are three outcome standards: **face validity**, **reliability**, and **consistency of measurement between intervention and comparison groups**.

Measures Assumed to be Reliable

- Administrative records obtained from schools, child welfare or other social service agencies, hospitals or clinics.
- Demographic characteristics, such as age, race/ethnicity, education level, SES, employment status, etc.
- Medical or physical tests, such as urinalysis, blood pressure, etc.

Face Validity

To satisfy the criterion for face validity, there must be a sufficient description of the outcome, pre-test, or pre-test alternative measure for the reviewer to determine that the measure is clearly defined, has a direct interpretation, and measures the construct it was designed to measure.

Reliability

Reliability standards apply to all outcome measures and any measure that is used to assess baseline equivalence. They are not applied to other measures that may be used in impact analyses as control covariates. To satisfy the reliability standards, the outcome or pre-test measure either must be a measure which is assumed to be reliable (see the box on the right) or must meet one or more of the following standards for reliability:

- Internal consistency (such as Cronbach's alpha) of 0.50 or higher.
- Test-retest reliability of 0.40 or higher.

- Inter-rater reliability (percentage agreement, correlation, or kappa) of 0.50 or higher.[8]

When required, reliability statistics on the sample of participants in the study under review are preferred, but statistics from test manuals or studies of the psychometric properties of the measures are permitted.

Consistency of Measurement between Intervention and Comparison Groups

The Prevention Services Clearinghouse standard for consistency of measurement requires that:

- Measures are constructed the same way for both intervention and comparison groups.
- The data collectors and data collection modes for data collected from intervention and comparison groups either are the same or are different in ways that would not be expected to have an effect on the measures.
- The time between pre-test (baseline) and post-test (outcome) does not systematically differ between intervention and comparison groups.

Prevention Services Clearinghouse reviewers assume that measures are collected consistently unless there is evidence to the contrary.

Example 1: In a study of a teen pregnancy prevention program, intervention group participants are asked about sexual behavior outcomes in a face-to-face interview with a case worker. Comparison group participants are asked in an online survey. In this example, the study would fail to meet Prevention Services Clearinghouse standards for consistency of measurement.

Example 2: In a mental health program, an anxiety assessment is distributed to youth by an interventionist and collected from the interventionists after the allotted time expires. In the comparison condition, community center staff distribute and collect the assessment using the same procedures. In both conditions, the same anxiety assessment is used and the test forms are sent to the researcher, who scores the results. Although different types of staff distributed and collected the assessment, this would not be expected to affect the test results. In this example, the outcome would not fail to meet Prevention Services Clearinghouse standards for consistency of measurement.

- outcome measures must meet all of the measurement standards for a contrast to receive a moderate or high rating.

[8] These thresholds align with the WWC v4.0 Standards.

- pre-tests or pre-test alternatives that do not meet the measurement standards cannot be used to establish baseline equivalence.

5.9.3 Design Confound Standards

The strength of causal inferences can be affected by the presence of confounding factors. A confounding factor is present if there is any factor, other than the intervention, that is both plausibly related to the outcome measures and also completely or largely aligned with either the intervention group or the comparison group. In such cases, the confounding factor may have a separate effect on the outcome that cannot be eliminated by the study design or isolated from the treatment effect. In such cases, it is impossible to separate how much of the observed effect was related to the intervention and how much to the confounding factor. Thus, the contrast cannot meet evidence standards and will receive a low rating. In QEDs, confounding is almost always a potential issue because study participants are not randomly assigned to intervention and comparison groups and some unobserved factors may be contributing to the outcome. Statistical controls cannot save contrasts from the effect of a confound if one is present.

The Prevention Services Clearinghouse defines two types of confounds: the substantially different characteristics confound, and the n=1 person-provider or administrative unit confound.

Substantially Different Characteristics Confound

Even when intervention and comparison groups are shown to meet standards for equivalence at baseline, or when baseline differences between intervention and comparison groups are adjusted for in analytic models, the effect of an intervention on outcomes can be sometimes be confounded with a characteristic of the treated or comparison units, or with a characteristic of the service providers, especially if that characteristic differs systematically between intervention and comparison groups. The characteristic that differs between the two groups may be related to the expected amount of change between pre-test and post-test measurements, thus confounding the intervention effect.

Prevention Services Clearinghouse defines a “substantially different characteristics confound” to be present if a characteristic of one condition, or a characteristic of the service provider for one condition, is systematically different from that of the other condition. For example, a substantially different characteristics confound may exist if there are large demographic differences between the groups, even if the groups are equivalent on the pre-test. In the case of a systematic difference between a service provider characteristic, the characteristic is not a confound if the characteristic is defined to be a component or requirement of the intervention.

One standard that is applied in Prevention Services Clearinghouse reviews is "*refusal of offer of treatment*." When the intervention group comprises individuals or units that were offered and accepted treatment and most or all of the comparison group[9] comprises individuals or units that were known to have been offered and refused treatment, Prevention Services Clearinghouse defines the design to have a *substantially different characteristics confound*. The Prevention Services Clearinghouse assumes that refusal or willingness to participate in treatment is likely to be related to motivation or need for services, which are likely to be related to outcomes.

Many QED studies will have intervention groups that consist entirely of individuals or units that accepted the offer of treatment. In these circumstances, the strongest designs would limit the comparison group members to those that would have been likely to accept the treatment if offered. The Prevention Services Clearinghouse, however, does not currently differentiate evidence ratings for studies that do and do not limit the comparison group in this manner. Some comparison groups will include individuals for whom it is unknown whether they would have participated in treatment had it been offered. The Prevention Services Clearinghouse does not consider this scenario to have a *substantially different characteristics confound*.

Example 1: A mental health intervention is targeted to families at risk of entry into the child welfare system, and is offered to families who have had at least one unsubstantiated claim of abuse or neglect in the past year. The comparison group consists of at risk families who have been nominated by school social workers in the same community but who have not had any claims of abuse or neglect. A substantially different characteristics confound is present because families in the intervention group have a characteristic that is substantially different from the comparison group that is plausibly related to outcomes.

n=1 Person-Provider Confound or Administrative Unit Confound

When all individuals in the intervention group or all individuals in the comparison group receive intervention or comparison services from a single provider (e.g., a single therapist or a single doctor) the treatment effect is confounded with the skills of the provider. For example, when intervention services are provided by a single therapist and the pre-post gains of her patients on a mental health assessment are compared with the gains of patients of another therapist, it is impossible to disentangle the effect of the intervention from the skills of the therapists. The Prevention Services Clearinghouse calls this type of confound an *n=1 person provider confound* because only one individual person is providing services, and it is impossible to disentangle the provider effects from the treatment effect.

---

9 This is operationally defined as at least 75 percent of the comparison group.

Similar to the n=1 person-provider confound, when all individuals in the intervention group or all individuals in the comparison group receive intervention or comparison services in a single administrative unit (e.g., clinic, community, hospital) the treatment effect may be confounded with the capacity of that administrative unit to produce better outcomes. The Prevention Services Clearinghouse calls this type of confound an *n=1 administrative unit provider confound*.

5.9.4 Missing Data Standards

The Prevention Services Clearinghouse uses the WWC v4.0 standards for missing data with the exception that that the standards are applied only to post-tests on eligible outcome measures, pre-tests, and pre-test alternatives. For other model covariates, any method that is used to address missing data is acceptable.

If a contrast has missing data on post-tests, pre-tests, or pre-test alternatives, reviewers first assess whether the approach to addressing missing data is one of the acceptable approaches described below.

- If a contrast has missing data and does not use one of the acceptable approaches listed below, it receives a rating of low.
- If a contrast has missing data and an acceptable method for addressing the missing data is used, reviewers then proceed based on whether the contrast was created via randomization or not.

The following approaches are acceptable for addressing missing data:

- ***Complete Case Analysis:*** Also known as *listwise deletion.* Refers to the exclusion of observations with missing data from the analysis. For RCTs, cases excluded due to missing data are counted as attrition. For QEDs, if baseline equivalence is established on the exact analytic sample as the impact analyses, there are no further missing data requirements. If the sample for baseline equivalence is not identical to the sample used in the impact analyses, additional requirements to assess potential bias due to missing data apply, as described below.
- ***Regression Imputation.*** Regression-based single or multiple imputation conducted separately for intervention or comparison groups (or that includes an indicator variable for intervention status) in which all covariates in the imputation are included in the impact models and that includes the outcome in the imputation.
- ***Maximum Likelihood:*** Model parameters are estimated using an iterative routine. Standard statistical packages must be used.
- ***Non-Response Weights:*** Weighting based on estimated probabilities of having missing outcome data. Acceptable only for missing post-tests and if the weights are

estimated separately for intervention and comparison groups or if an indicator for intervention group status is included.

- ***Constant Replacement:*** Replacing missing values with a constant value and including an indicator variable in impact estimation models to identify the missing cases. Acceptable only for RCTs with missing pre-tests and pre-test alternatives.

Procedures for Low Attrition RCTs with Missing Data

If a contrast was created from randomization of individuals or clusters, reviewers assess attrition; imputed outcomes are counted as attrition. That is, reviewers count cases with missing outcomes as if they had attrited.

If attrition is low, the contrast can receive an evidence rating of high, provided that all other design and execution standards are met, and an acceptable method of addressing missing data is used. When attrition is low and an acceptable method of addressing missing data is used, impact estimates from models with imputed missing data are acceptable for computing effect sizes and statistical significance.

Procedures for High Attrition RCTs and Quasi-Experiments with Missing Data

If a contrast from a RCT exhibits high attrition (with any imputed cases counted as attrition) or was not created by randomization (i.e., is a QED), reviewers must assess whether the contrast limits the potential bias that may result from using imputed outcome data. If no outcome data are imputed, potential bias from imputed outcome data is not present.

If outcome data are imputed, reviewers calculate an estimate of the potential bias from using imputed outcome data and assess whether that estimate is less than 0.05 standard deviation units of the outcome measure. To estimate the potential bias, reviewers use a pattern-mixture modelling approach, as outlined in Andridge and Little (2011; see also the What Works Clearinghouse Standards, v4.0) and operationalized in a spreadsheet-based tool (Price, 2018).

- If the potential bias is greater than the 0.05 threshold, the contrast receives a low rating.
- If the potential bias is less than the 0.05 standard deviation unit criterion and the contrast is a high attrition RCT that analyzes the full randomized sample using imputed data, then the contrast can receive a moderate evidence rating, provided other design and execution standards are met and an acceptable method of addressing missing data is used.
- If the potential bias is less than the 0.05 standard deviation threshold but the full randomized sample is not used or the contrast is a QED and no pre-test or pre-test alternatives are imputed, reviewers evaluate baseline equivalence for the analytic

sample and proceed as usual for contrasts required to establish baseline equivalence.

If pre-test or pre-test alternative data are imputed or the full analytic sample is not available for baseline equivalence, additional computations to determine the largest baseline difference are applied (Andridge & Little, 2011; What Works Clearinghouse Standards, v4.0). These computations are operationalized in a spreadsheet tool (Price, 2018).

- If the contrast fails to satisfy the largest baseline difference criterion, it receives a low rating.
- If criterion contrast satisfies the largest baseline difference criterion, it may receive an evidence rating of moderate, provided the other design and execution standards are met and an acceptable method of addressing missing data is used.

5.10 Procedures for Recording, Correcting, and Summarizing Impact Estimates

All contrasts in a study are rated against the design and execution standards regardless of the magnitude or statistical significance of the impact estimate. For any contrast that receives a high or moderate design and execution rating, reviewers record or compute an effect size in the form of Hedges' *g*, its sampling variance (or standard error), and statistical significance, correcting as necessary for clustering. If data are not available for such computations, reviewers may send queries to authors requesting such information, in accordance with the author query policies described in Section 7.3.2. If requested data are not obtained in response to author queries, reviewers complete the review with the information available. The Prevention Services Clearinghouse must be able to determine if impact estimates are statistically significant for them to be used to inform program ratings.

In addition, the Prevention Services Clearinghouse applies the following procedures to all effect size computations:

- Because errors and omissions in reporting p-values and statistical significance are common, the Prevention Services Clearinghouse computes the statistical significance for all contrasts, and does not rely on reporting by study authors (Bakker & Wicherts, 2011; Krawczyk, 2015).
- Only aggregate findings are recorded and rated under this Version 1.0 of the Prevention Services Clearinghouse. Future versions of the Handbook may address subgroup findings.
- Impact estimates that are *favorable* (statistically significant and in the desired direction), *unfavorable* (statistically significant and not in the desired direction), or *sustained favorable* (statistically significant and in the desired direction at least 6 or 12 months beyond the end of treatment, see Section 6.2.3) are used to determine

program or service ratings, but all impacts from all contrasts rated as high or moderate are recorded.

5.10.1 Procedures for Computing Effect Sizes

The Prevention Services Clearinghouse uses the standardized mean difference effect size metric for outcomes measured on a continuous scale (e.g., group differences in average scores on an assessment of mental health). All effect sizes are recorded or computed such that larger effect sizes represent positive outcomes for the intervention condition. The basic formulation of the standardized mean difference effect size (*d*) is

$$d = \frac{\bar{X}_{G2} - \bar{X}_{G1}}{s_p}$$

where the numerator is the difference in group means for the intervention and comparison groups, and the denominator is the pooled standard deviation of the intervention and comparison groups. All standardized mean difference effect sizes are adjusted with the small-sample correction factor to provide unbiased estimates of the effect size (Hedges, 1981). This small-sample corrected effect size (Hedges' *g*) can be represented as:

$$g = \left[1 - \left(\frac{3}{4N - 9}\right)\right] * d$$

and the sampling variance of the effect size is represented as

$$var_g = \sqrt{\frac{n_{G1} + n_{G2}}{n_{G1} n_{G2}} + \frac{g^2}{2(n_{G1} + n_{G2})}}$$

where N is the total sample size for the intervention and comparison groups, *g* is the effect size, nG1 is the sample size for the intervention group, and nG2 is the sample size for the comparison group.

For binary outcomes, the Prevention Services Clearinghouse computes effect sizes as odds ratios and then converts them to standardized mean difference effect sizes using the Cox transformation as described in Sánchez-Meca, Marín-Martínez, and Chacón-Moscoso (2003).

Standard formulae for computing effect sizes from common statistical tests are employed, as necessary (Lipsey & Wilson, 2001).

5.10.2 Adjusting for Pre-Tests

Reviewers use statistics that are adjusted for pre-tests and other covariates to compute effect sizes whenever possible. In cases where statistical adjustments are not required

(i.e., low attrition RCTs or studies with baseline effect sizes <0.05) and study authors report only unadjusted pre-test and post-test findings, reviewers compute the effect size of the difference between the intervention and comparison groups at baseline and subtract that value from the post-test effect size.

5.10.3 Procedures for Correcting for Mismatched Analysis

In the event that studies reviewed by the Prevention Services Clearinghouse report findings for clustered data that have not been appropriately corrected for clustering, reviewers apply a clustering correction to the findings. The Prevention Services Clearinghouse applies this correction to findings from mismatched analyses; that is, analysis for which the unit of assignment and unit of analysis are mismatched and not appropriately analyzed using, for example, multi-level models. The intraclass correlations required for this adjustment is taken from the studies under review when possible. When intraclass correlations are not available, a default value of .10 is used, consistent with the conventions used by the What Works Clearinghouse for non-academic measures.

5.10.4 Reporting and Characterizing the Effect Sizes on the Prevention Services Clearinghouse Website

The individual findings from each contrast with a high or moderate rating are reported on the Prevention Services Clearinghouse website. These findings include the effect size, its statistical significance, and a translation of the effect size into percentile units called the *implied percentile effect.* In addition, meta-analysis is used to summarize the findings for each outcome domain. The Prevention Services Clearinghouse uses a fixed effect weighted meta-analysis model using inverse-variance weights (Hedges & Vevea, 1998) to estimate the average effect size for each domain.

Prevention Services Clearinghouse reviewers also convert each effect size into percentile units for reporting on the Prevention Services Clearinghouse website to provide a user-friendly alternative to the effect sizes. This *implied percentile effect* is the average intervention group percentile rank for the outcome minus the comparison group average percentile, which is 50. For example, an implied percentile effect of 4 means that the program or service increased the intervention group performance by 4 percentile points over the comparison group.

6. Program or Service Ratings

The Prevention Services Clearinghouse periodically provides clarification on topics covered in the Handbook. To learn more, please visit the FAQ page on the Prevention Services Clearinghouse website.

This chapter describes the process of translating design and execution ratings from one or more studies of a program or service into ratings for that program or service. As described at the beginning of Chapter 5, Prevention Services Clearinghouse reviewers rate study contrasts, rather than entire studies. To determine program and service ratings, the Prevention Services Clearinghouse combines design and execution ratings from multiple contrasts and (if available) contrasts from multiple studies. To determine the rating for a program or service, all contrasts for each eligible program or service that meet moderate or high evidence standards are examined.

6.1 Four Ratings

Using the qualifying contrasts, reviewers assign one of four ratings to each program or service to characterize the extent of evidence for a particular program or service:

- *Well-supported*. A program or service is rated as a well-supported practice if it has at least two contrasts with non-overlapping samples in studies carried out in *usual care or practice settings* (see Section 6.2.2) that achieve a rating of moderate or high on design and execution and demonstrate favorable effects in a target outcome domain. At least one of the contrasts must demonstrate a sustained favorable effect of at least 12 months *beyond the end of treatment* (see Section 6.2.3) on at least one target outcome.
- *Supported*. A program or service is rated as a supported practice if it has at least one contrast in a study carried out in a usual care or practice setting that achieves a rating of moderate or high on design and execution and demonstrates a sustained favorable effect of at least 6 months beyond the end of treatment on at least one target outcome.
- *Promising*. A program or service is designated as a promising practice if it has at least one contrast in a study that achieves a rating of moderate or high on study design and execution and demonstrates a favorable effect on a target outcome.
- *Does not currently meet criteria*. A program or service that has been reviewed and does not achieve a rating of *well-supported*, *supported*, or *promising* is deemed 'does not currently meet criteria.' This includes (a) programs and services for which all eligible contrasts with moderate or high design and execution ratings have no statistically significant favorable effects and (b) programs and services that do not have any eligible contrasts with moderate or high design and execution ratings.

6.2 Contributing Factors in the Ratings

6.2.1 Risk of Harm

A program or service cannot be not rated as *well-supported*, *supported*, or *promising* if there is an empirical basis, as evidenced by the presence of an unfavorable effect(s) on target or non-target outcomes that suggest that the overall weight of evidence does not support the benefits of the program or service. To be considered, unfavorable effects must be reflected in contrasts that receive a moderate or high rating according to the design and execution standards. To determine whether there is risk of harm, all statistically significant unfavorable impacts on any outcome (whether an eligible target outcome or not) from any studies with contrasts receiving high or moderate evidence ratings are identified. If there is sufficient evidence of risk of harm based on statistically significant unfavorable findings, the program may be deemed 'does not currently meet criteria' by the Prevention Services Clearinghouse. Additionally, programs or services may not be designated as *well-supported, supported,* or *promising* if case data suggests a risk of harm that was probably caused by the treatment and was severe or frequent.

6.2.2 Usual Care or Practice Settings

To receive a rating of *supported* or *well-supported*, the favorable evidence for a program or service must have been obtained from research conducted in a usual care or practice setting. A usual care or practice setting is defined as an existing service agency or provider that delivers mental health services, substance use prevention or treatment services, in-home parent skill-based programs, and/or kinship navigator programs as part of its typical operations.

A usual care setting may use routine personnel who already work for the agency or it may employ outside staff (e.g., researchers, graduate students) if the services themselves are those that would typically be delivered by agency personnel in the absence of a research study. Ad hoc clinics set up expressly for the purposes of research do not constitute usual care or practice settings, even if staffed by personnel who might typically work in a usual care setting.

6.2.3 Beyond the End of Treatment

To receive a rating of *supported* or *well-supported*, programs and services must have sustained favorable effects beyond the end of treatment. The end of treatment is defined as the stated end of treatment by the study or program documentation. If a clear end of treatment is not defined, if treatment extends indefinitely or varies across participants, or if services are staggered, the Prevention Services Clearinghouse selects a time point that corresponds to when the majority of a clearly defined set of services were stated to have been delivered. If that information is not available, but studies provide information about the average or range of service delivery, reviewers will use

the longest program duration (or estimate it from the data provided) as the end of treatment and determine the length of follow-up from that point.

If a study gives the time between pre-test and post-test, but not the time between the end of treatment and measurement of the post-test, reviewers subtract the stated intended duration of treatment from the pre/post interval to estimate the number of months beyond the end of treatment that measurement occurred.

7. Procedures for Reviewing Programs and Services for the Title IV-E Prevention Services Clearinghouse

This chapter summarizes some of the operational procedures the Prevention Services Clearinghouse uses to identify, screen, review, and rate programs and services.

The Prevention Services Clearinghouse periodically provides clarification on topics covered in the Handbook. To learn more, please visit the FAQ page on the Prevention Services Clearinghouse website.

7.1 Prevention Services Clearinghouse Team

The Prevention Services Clearinghouse team includes federal staff, contractors, subcontractors, and consultants, as well as stakeholders and experts who are brought in to advise on various aspects of the clearinghouse operations. All individuals who work on the Prevention Services Clearinghouse are expected to adhere to conflict of interest policies and sign conflict of interest declarations prior to doing any work.

7.2 Procedures for Identifying Eligible Studies of Selected Programs and Services from Search Results

For each program or service identified and prioritized for review by the Prevention Services Clearinghouse, staff conduct a comprehensive and systematic search for potentially eligible studies of that program or service as well as considering publicly available literature submitted by stakeholders in support of recommended programs and services.

7.2.1 Title and Abstract Screening

Once a search is conducted, the titles and abstracts from the citations identified are queued for relevance screening. Two trained screeners independently screen all titles and abstracts. Title and abstract decisions are binary (keep or drop) and documents are marked as relevant if reviewers are able to answer **Yes** or **Not Sure** to both of the following questions:

1. Does the title or abstract describe an evaluation of the program or service under review?
2. Does the study appear to use an experimental or quasi-experimental design?

Senior clearinghouse staff and content experts are on hand for questions throughout the screening process.

Citations which both reviewers agree are irrelevant are documented and dropped from further consideration. All other citations are assigned for retrieval and slated for full-text screening using the full study eligibility criteria described above.

7.2.2 Procedures for Full-text Eligibility Screening

Full-text copies of all citations that are not excluded during the title and abstract screening or that are identified from evidence clearinghouses or other sources are retrieved by Prevention Services Clearinghouse staff. Documents are combined, as necessary, into sets of documents that describe the same study; eligibility screening is performed on each study.[10]

Full-text eligibility screening then proceeds as follows:

- Each study is assigned for eligibility screening to a trained eligibility screener. That screener uses the full Study Eligibility Criteria described in Section 4.1 to determine whether the study is eligible for review.
- If there are more than 15 eligible studies, the screener assigns prioritization points as described in Section 4.2 to each eligible study. All eligible studies are reviewed for risk of harm as described in Section 6.2.1. Unfavorable contrasts in all eligible studies will be reviewed according to the design and execution standards described in Chapter 5.
- Studies prioritized for review are assigned to a trained reviewer. The reviewer's first task is to re-confirm the study's eligibility against the Study Eligibility Criteria. This ensures double-screening, but provides some efficiency in the process.
- If a study is determined by the initial eligibility screener to be ineligible for review, it is assigned to a second screener for confirmation.
  - If the two screeners agree on the disposition and the reason for the decision, the study is dropped from further consideration for the review (but retained in a database with documentation of the reason the study is ineligible).
  - Any disagreements on overall disposition or the reason for the ineligibility are resolved through consensus and in consultation with senior clearinghouse staff.

Content experts and senior clearinghouse staff are available for questions. For example, screeners may have questions about whether a particular measure used in a study represents an eligible target outcome or whether the program described in a study is an eligible adaptation of the program or is an adaptation that must be treated as a separate program. Senior clearinghouse staff may answer methodological questions, such as whether the study design meets the eligibility requirements.

[10] The Prevention Services Clearinghouse defines a study as one research investigation of a defined subject sample, and the interventions, measures, and statistical analyses applied to that sample. It is common for researchers to publish more than one article or manuscript that describes the same study. The Prevention Services Clearinghouse reviews the full set of documents available for each study.

7.3 Procedures for Reviewing Eligible Studies against the Standards

7.3.1 Review and Reconciliation Process

Once a study is deemed eligible, all of its documents are entered into the Prevention Services Clearinghouse Review Database and the study is assigned by the review manager to a trained reviewer.

- The reviewer uses the study ***design and execution standards*** described in Chapter 5 to assign the study one of three ratings: Meets Standards for High, Moderate, or Low Support of Causal Evidence. The review is completed in the database.

The review and reconciliation process differs for studies that receive a low causal evidence rating versus those that receive a moderate or high causal evidence rating.

- If the first reviewer assigns a low causal evidence rating, the review manager assigns the study to a senior reviewer (called a reconciler) for evaluation.
  - If the reconciler confirms the rating, he or she finalizes the review, consulting with the reviewer as necessary.
  - If the reconciler disagrees with the rating, the study is assigned to a second reviewer for evaluation.
    - Once the second review is complete, the reconciler then examines both reviews and finalizes the review, consulting with the two reviewers as necessary.
- If the first reviewer assigns the study a causal evidence rating of high or moderate with the information provided in the study documents, the review manager assigns the study to a second reviewer for evaluation.
  - Once the second review is complete, the study is assigned to a reconciler who examines both reviews and finalizes the review, consulting with the two reviewers as necessary.
- If the first reviewer needs additional information to determine the rating for a study, he or she drafts an Author Query requesting the needed information and submits it to a reconciler for review. The Author Query is then sent to the author.
  - If the author does not respond, the first reviewer completes the review with the available information, following the procedures for reconciliation and/or second review commensurate with the rating of the study.
  - If the author does respond, the first reviewer completes the review with the additional information and follows the procedures for reconciliation and/or second review commensurate with the rating of the study.

Content experts and senior clearinghouse staff are on hand to answer questions and help interpret complicated cases.

7.3.2 Author Query Policies and Procedures

It is the policy of the Prevention Services Clearinghouse to query study authors for information deemed necessary to assign a rating of high, moderate, or low. Author queries may request information about sample sizes, baseline statistics, group formation (e.g., whether randomization was used), and characteristics of the outcome measures required to determine whether outcome requirements are met, and may ask clarifying questions about analytic models (e.g., whether covariates are included in the impact models). Author queries may also request descriptive statistics (e.g., means and standard deviations) needed to compute effect sizes and statistical significance of impacts, as this information may be needed to assign program or service ratings.

7.4 Procedures for Re-review of Programs and Studies

7.4.1 Procedures for Re-review of Programs and Services

Programs and services reviewed for the Prevention Services Clearinghouse may be considered for re-review due to missing information or errors in the original review, or due to the emergence of substantial new evidence that has the potential to change program or service ratings. Prevention Services Clearinghouse staff keep track of the dates that programs and services are reviewed and periodically assess the extent of new evidence available. Periodically, content experts may be consulted to determine if new research is available and if the available research has the potential to change the rating of the program or service. Stakeholders may request a re-review of the program or service rating based on the mis-application or criteria, missing information, or other errors.

7.4.2 Procedures for Re-review of Studies

Individual studies reviewed for the Prevention Services Clearinghouse may also be considered for re-review due to missing information or errors in the original review. If errors or missing information are identified, the Prevention Services Clearinghouse follows standard procedures for re-review. This includes assigning different, blinded reviewers to conduct any re-reviews. If the re-review determines the original review to be in error, the error is corrected on the website. All correspondence regarding re-reviews is logged and maintained by Prevention Services Clearinghouse staff.

Glossary

Exhibit x.1. Definitions

Term	Definition
Randomized Controlled Trial (RCT)	A study in which units are assigned to intervention and control conditions via a random process (e.g., a lottery).
Quasi-Experimental Design (QED)	A study in which units are assigned to intervention and control conditions via a non-random process.
Study	One research investigation of a defined subject sample, and the interventions, measures, and statistical analyses applied to that sample. Sometimes study results are reported in more than one document, or a single document reports results from separate studies.
Contrast	A comparison of a treated condition to a counterfactual untreated condition on an outcome. All Prevention Services Clearinghouse design and execution ratings are applied to contrasts.
Intervention group	The set of units that were assigned randomly or non-randomly to an offer of the intervention condition. Intervention group members may or may not have received treatment, but they were given access to the intervention condition.
Control group	Control group refers to the set of units that were randomly assigned to be embargoed from the offer of the intervention condition; this term is used only in the context of RCTs.
Comparison group	The broader term comparison group refers to the set of units assigned randomly or non-randomly to the comparison condition; this term may be used in the context of RCTs or QEDs.
Joiner bias	If a cluster randomized study permits individuals to join clusters after randomization, the estimate of the effect of the intervention on individual outcomes may be biased if individuals who join the intervention clusters are systematically different from those who join the comparison clusters.
Attrition	This term is used only in the context of RCTs. Attrition refers to the absence of an outcome measurement for a unit that was randomly assigned to an intervention or control condition. A sample member has attrited from the sample if there is no outcome measurement for that sample member.
Outcome	An outcome is the measurement of an eligible target outcome as described in the Study Eligibility Criteria. Outcomes can be measured at pre-test, post-test, or over longer follow-up periods.
Post-test	A post-test is measured at the end of a follow-up period, sometime after units have been exposed to or offered the intervention or comparison

Term	Definition
	conditions. It is a measure on which the impact of the intervention is estimated.
Pre-test	A pre-test is a baseline measure of the outcome variable. Pre-tests, like other baseline measures, are measured before, or just after assignment to intervention and comparison conditions. If they are measured after assignment to conditions, they should be measured before effects of the intervention or comparison conditions would be expected to influence their value.
Other baseline measures	Other baseline measures are measured at baseline, and may be used as covariates in impact models, but they are not the specific measures on which baseline equivalence must be demonstrated in order to satisfy evidence standards. Other baseline measures are measured before, or just after assignment to intervention and comparison conditions. If they are measured after assignment to conditions, they must be measured before effects of the intervention or control conditions would be expected to influence their value, or be time-invariant measures (e.g., gender).
Effect size	An effect size is a standardized, quantitative index representing the magnitude and direction of an empirical relationship. In this context, the effect size is a value that reflects the magnitude of the treatment effect. It may also be employed in Prevention Services Clearinghouse reviews to index the differences between intervention and comparison groups at baseline. The *standardized mean difference effect size* is used for Prevention Services Clearinghouse reviews, in the form of Hedges' *g*.
Intent-to-treat analysis (ITT)	An intent-to-treat or ITT analysis is one in which study authors analyze the participants in a randomized study based on their *original* assignment to conditions, regardless of whether they received the intervention or switched conditions.

Sources

Andridge, R. R., & Little, R. J. A. (2011). Proxy pattern-mixture analysis for survey nonresponse. *Journal of Official Statistics, 27*(2), 153–180.

Bakker, M., & Wicherts, J. M. (2011). The (mis)reporting of statistical results in psychology journals. *Behavior Research Methods, 43*(3), 666–678. https://doi.org/10.3758/s13428-011-0089-5

Decisions Related to the Development of a Clearinghouse of Evidence-Based Practices in Accordance with the Family First Prevention Services Act of 2018, 83 Fed. Reg. 29122.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6,* 107-128.

Hedges, L. V., & Vevea, J. L. (1998). Fixed effect and random effects models in meta-analysis. *Psychological Methods, 3*, 486-504.

Krawczyk, M. (2015). The search for significance: A few peculiarities in the distribution of p values in experimental psychology literature. *PLOS ONE, 10*(6), e0127872. https://doi.org/10.1371/journal.pone.0127872

Lipsey, M. W., & Wilson, D. B. (2001). *Practical Meta-analysis.* Applied Social Research Methods Series, Vol. 49. Thousand Oaks, CA: Sage.

Price, C. (2018). *Spreadsheet tool for computing potential bias due to missing data*. [software] Rockville, MD: Abt Associates.

Sánchez-Meca, J., Marín-Martínez, F., & Chacón-Moscoso, S. (2003). Effect-size indices for dichotomized outcomes in meta-analysis. *Psychological Methods, 8*(4), 448–467. https://doi.org/10.1037/1082-989X.8.4.448

State Requirements for Electing Title IV-E Prevention and Family Services and Programs (PI-18-09). Washington, DC: Children's Bureau, Administration for Children and Families, U. S. Dept. of Health and Human Services.

U. S. Department of Education (2018). *What Works Clearinghouse™ Standards Handbook* (Version 4.0). Washington, DC: Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

What Works Clearinghouse. (n.d.). *Assessing attrition bias.* Retrieved from https://ies.ed.gov/ncee/wwc/Docs/ReferenceResources/wwc_attrition_v2.1.pdf