Handbook of Standards and Procedures

VERSION 2.0

Title IV-E Prevention Services CLEARINGHOUSE

TITLE IV-E PREVENTION SERVICES CLEARINGHOUSE

HANDBOOK OF STANDARDS AND PROCEDURES, VERSION 2.0

OPRE Report 2024-127

June 2024

Sandra Jo Wilson, Scott R. Brown, Suzanne E. U. Kerns, Samuel R. Dastrup, Eric Hedberg, Rebecca Schachtner, Cara Jackson, Jennifer Norvell, Walter Campbell, and Alec Wall

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

Contract Numbers: HHS P233201500069I | GS00F252CA

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

Contents

Introduction

This *Prevention Services Clearinghouse Handbook of Standards and Procedures, Version 2.0* (*Handbook Version 2.0*) 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.

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 in 2019 by the Administration for Children and Families (ACF) within the U.S. Department of Health and Human Services (HHS) to systematically review existing research on programs and services[1] 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 and substance use prevention and treatment programs and services, in-home parent skill-based programs and services, as well as kinship navigator programs.

Handbook Version 2.0

The Prevention Services Clearinghouse Handbook of Standards and Procedures, Version 2.0 replaces the Prevention Services Clearinghouse Handbook, Version 1.0 that has been in use since 2019. Creating Version 2.0 was a multi-year process that included feedback from multiple rounds of public comment and expert input. Experts who advised the Prevention Services Clearinghouse included individuals with lived experience; individuals from communities that programs and services are intended to serve, inclusive of underserved and tribal communities; methodological experts; state, tribal, and local child welfare staff or administrators; policymakers; program or service developers; program or service providers or trainers; study authors; and subject matter experts in the fields of mental health, substance use and misuse, parenting and parent skill-based programs and services, kinship navigator programs, child welfare, implementation science, and cultural responsiveness and equity. The objectives of the revision process were to be responsive to comments shared by the public and suggestions received from expert consultations, enhance the transparency of the Prevention Services Clearinghouse's standards and procedures, be responsive to the needs of underserved communities and families, and update the standards to align with current best practices for systematic evidence reviews while continuing to maintain alignment with the requirements of the authorizing legislation for the Prevention Services Clearinghouse, the Family First Prevention Services Act of 2018, as codified in Title IV-E of the Social Security Act.

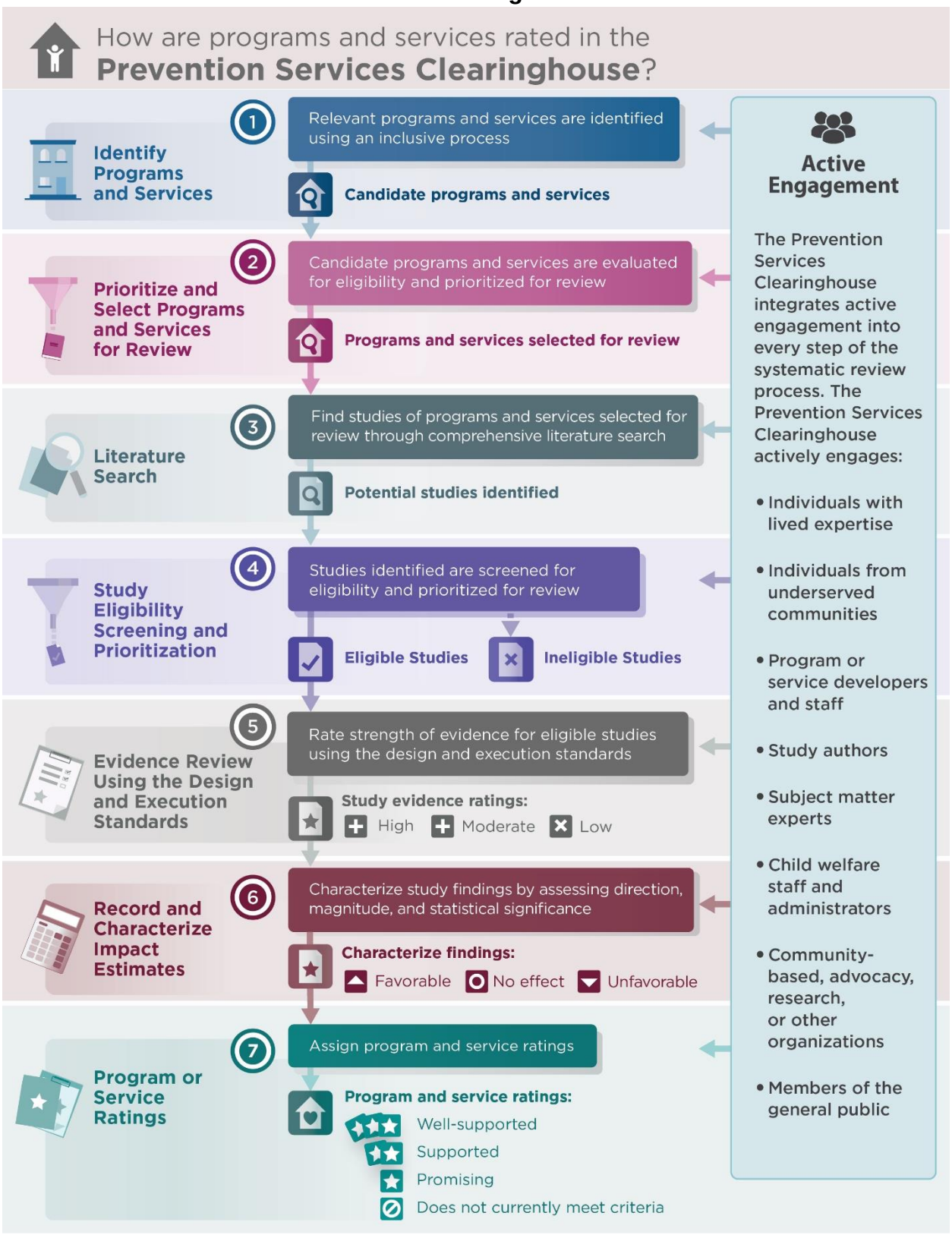
The Prevention Services Clearinghouse periodically provides clarification on topics covered in the Handbook of Standards and Procedures. To learn more, please visit the FAQ page on the Prevention Services Clearinghouse website.

[1] The Prevention Services Clearinghouse does not make an operational distinction between the terms "program" and "service".

Introduction

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 1.1 below).

Exhibit 1.1. The Prevention Services Clearinghouse Review Process



The Prevention Services Clearinghouse Review Process

The Prevention Services Clearinghouse systematic review process, described in detail in the chapters that follow and shown in Exhibit 1.1, includes the following steps:

1. *Identify programs and services.* Candidate programs and services relevant to the mission of the Prevention Services Clearinghouse are identified using an inclusive process that invites recommendations from any member of the public, including state and local government administrators and tribes, to ensure broad coverage across program or service areas and populations served (Chapter 1).

2. *Prioritize and select programs and services.* Candidate programs and services are prioritized from those identified using criteria that include evidence of eligibility and being in active use, recommendations received, child welfare relevance, population(s) served, previous evaluations and studies, implementation supports, and coverage across program or service areas. The Prevention Services Clearinghouse leverages other clearinghouses to gather information relevant to the program and service prioritization criteria. After prioritization, programs or services are selected and added to the working list of programs and services planned for review (Chapter 2).

3. *Literature search.* Prevention Services Clearinghouse staff conduct comprehensive literature searches to locate available and relevant research on programs and services selected for review. In addition to an electronic bibliographic database search, this process includes scans of other clearinghouses and of program or service websites, review of grey literature sources, and review of submissions of research from the public (e.g., citation lists of published research or notification of individual studies published) (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 study prioritization criteria to determine the order and depth of their review (Chapter 4).

5. *Evidence review using the design and execution standards.* All prioritized studies are reviewed by trained reviewers using the Prevention Services Clearinghouse design and execution standards. Additional studies may be reviewed using the design and execution standards, as indicated, according to study prioritization and risk of harm procedures. One of three ratings is assigned to studies reviewed using the design and execution standards (Chapter 5). Study authors may be queried to request information deemed necessary to assign a rating.

6. *Record and characterize impact estimates.* To inform program or service ratings of well-supported, supported, or promising, the Prevention Services

Clearinghouse characterizes impact estimates from high and moderate-rated contrasts as favorable, sustained favorable, unfavorable, or no effect. Characterization of impact estimates as favorable, sustained favorable, no effect, or unfavorable is based on both their direction and statistical significance (Chapter 6).

7. *Program or 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 7). These ratings take into consideration the characterization of the impacts as favorable, sustained favorable, unfavorable, or no effect and any evidence of risk of harm.

Operational procedures for reviewing programs and services in the Prevention Services Clearinghouse are included in Chapter 8. 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 members of the public (Section 8.5).

The Prevention Services Clearinghouse conducts a wide range of active engagement activities to promote transparency, gather input on Prevention Services Clearinghouse standards and procedures, and collect information about programs and services. As shown in Exhibit 1.1, active engagement is incorporated into every step of the systematic review process. This includes, but is not limited to: hosting periodic engagement sessions to gather questions and other feedback from the public; hosting focus groups and consultation sessions with experts, including those with lived expertise; engaging with study authors and developers; and inviting public feedback to improve processes on an ongoing basis. See additional details and examples in Section 8.7.

Planned Pilot Activities to Inform Future Updates to the Review Process

The Prevention Services Clearinghouse plans to conduct several pilots to inform future updates to the Handbook of Standards and Procedures. These pilots aim to

- Understand the feasibility of identifying and reviewing studies only published in Spanish.
- Develop and implement standards and procedures for reviewing studies that use single case designs and determine how such designs may contribute to promising ratings.

- Understand the considerations, including methodological, timeline, and resource related considerations, for the systematic review of subgroup analyses, including how such analyses may contribute to future program or service ratings.

Review Timeline

The Prevention Services Clearinghouse aims to review and rate as many programs and services as quickly as possible to support state and tribal Title IV-E agencies' efforts to improve outcomes for children and families through implementation of the Family First Prevention Services Act. Review timelines are influenced by a number of factors, including:

- Whether the Prevention Services Clearinghouse has questions for program or service developers about manual availability or adaptations, and if so, the time it takes for developers to respond.
- The number of eligible studies evaluating the impact of a particular program or service on at least one target outcome.
- Questions that arise during the review of studies that require internal or external expert consultation.
- Whether eligible studies have all the information necessary to complete the review or whether author queries are needed to obtain additional information – and, if required, the time it takes for study authors to respond.

As a result of these factors, it is not possible to provide a timeline in advance for the review of a particular program or service.

The Prevention Services Clearinghouse Website

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. The website also provides information about the programs and services that are planned for review and a list of programs and services that have been recommended for review.

For each program or service reviewed by the Prevention Services Clearinghouse, the website provides an overview of the program or service, and detailed information about evidence, findings, and studies reviewed, including reasons that studies did not meet design and execution standards or were not eligible for review. The program and service pages also describe who the program or service is designed to serve and provide a detailed description of the demographic characteristics of the participants and settings in the studies reviewed that met design and execution standards. The *Program*

or Service Delivery and Implementation section identifies the book, manual, or other documentation used for the program or service review. It also may include additional information, such as details about program or service dosage and delivery settings; information about education, certifications, and training for each program or service; and other resources or sources for more information. This additional information is provided as a resource and is drawn from existing sources. Program and service developers are provided the opportunity to review this information, but the information does not necessarily reflect the views of the program and service developers.

The *Review Process* section of the website summarizes the Prevention Services Clearinghouse systematic review process and provides information about the standards and procedures. The *Resources* page provides a variety of resources to help the child welfare community access the information they need, including webinars, special topics reports, fact sheets, a reporting guide for study authors, and technical guidance briefs. The website also provides answers to *Frequently Asked Questions* (*FAQs*) to assist providers, administrators, researchers, and other members of the child welfare community in gathering information about Prevention Services Clearinghouse standards and procedures. The *FAQs* section includes information about how to submit a program or service recommendation and how to provide information about studies of programs or services to the Prevention Services Clearinghouse.

Handbook of Standards and Procedures Development and Revision

The first version of the Handbook of Standards and Procedures was informed by comments from state and local child welfare administrators, program and service developers, foundations, non-profit organizations, tribes, researchers and evaluators, and other members of the public submitted in response to an HHS Federal Register Notice (FRN; 83 FR 29122). Development of Handbook Version 1.0 was also informed by consultations with research and practice experts as well as the review standards and 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).

Development of Version 2.0 of the Handbook of Standards and Procedures was informed by public comments in response to HHS Federal Register Notices 86 FR 37332 and 88 FR 73021. Revisions were also informed by an extensive series of consultations with research and practice experts. Experts who advised the Prevention Services Clearinghouse included individuals with lived expertise; individuals from communities that programs and services are intended to serve, inclusive of underserved and tribal communities; methodological experts; state, tribal, and local child welfare department staff or administrators; policymakers; program or service developers;

program or service providers or trainers; study authors; and subject matter experts in the fields of mental health, substance use and misuse, parenting and parent skill-based programs and services, kinship navigator programs, child welfare, implementation science, and cultural responsiveness and equity. We also consulted current review standards and processes used by other prominent evidence clearinghouses, including the WWC, HomVEE, and the CEBC. A detailed summary of revisions can be found in the Summary of Revisions to the Handbook of Standards and Procedures appendix.

1. Identify Programs and Services

The Prevention Services Clearinghouse identifies programs and services using an inclusive process that relies on the following sources:

- Recommendations submitted to the Prevention Services Clearinghouse from public calls. At least annually, the Prevention Services Clearinghouse issues a public call for recommendations of programs and services. These calls are an opportunity for any member of the public, including state, tribal, or local government administrators, to recommend new programs and services for systematic review as the field continues to evolve. The Prevention Services Clearinghouse announces public calls on its website and email list.
- Recommendations submitted to the Prevention Services Clearinghouse via email at any time from any member of the public and other key advisors, including federal partners, and state, tribal, and local administrators.

The public call materials and *FAQs* page on the Prevention Services Clearinghouse website provide detailed information on how to recommend a program or service for review as well as suggested information to include with a program or service recommendation – including how recommendations are responsive to program or service prioritization criteria (see Chapter 2) and any relevant study citations (see Chapter 3).

Prevention Services Clearinghouse staff log all program and service recommendations and respond to submitters notifying them that their recommendations have been received. The Prevention Services Clearinghouse retains all submissions for consideration for future review cycles. This includes recommendations from a 2018 Federal Register Notice (FRN) (83 FR 29122) in addition to all public call and ad hoc submissions.[2]

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

All programs and services identified through these sources are cataloged and posted on the Prevention Services Clearinghouse website. The next section describes the processes used to determine the eligibility of programs and services and how programs and services are prioritized and selected for review.

[2] The first group of programs and services reviewed by the Prevention Services Clearinghouse was identified prior to the release of Handbook v1.0. Those initial programs and services were identified through a public call for recommendations as part of Federal Register Notice 83 FR 29122 (2018 FRN).

2. Prioritize and Select Programs and Services

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

Programs and services identified via the procedures described in the previous chapter are then prioritized for review using the Program or Service Prioritization Criteria. Prioritization includes an initial assessment of program or service eligibility using the Program or Service Eligibility Criteria specified below. Informed by the prioritization process, programs and services are then selected for review, and a final eligibility determination is made. Prevention Services Clearinghouse prioritization and selection processes are informed by active engagement. Experts who advise the Prevention Services Clearinghouse on these processes include, but are not limited to, individuals with lived expertise; individuals from communities that programs and services are intended to serve, inclusive of underserved and tribal communities; methodological experts; state, tribal, and local child welfare department staff or administrators; policymakers; program or service developers; program or service providers or trainers; study authors; and subject matter experts. This chapter also details procedures used when there are multiple manuals or versions of a program or service for the purposes of prioritization and selection of programs and services for review.

2.1 Program or Service Eligibility Criteria

To be eligible for review by the Prevention Services Clearinghouse, programs and services must meet two criteria: (1) they must meet the definition for at least one of the four program or service areas and (2) they must have available books, manuals, or other documentation that describe how to implement or administer the program or service. This section describes these two eligibility criteria.

2.1.1 Program or Service Areas

As specified in the Family First Prevention Services Act, there are four eligible program or service areas – mental health prevention and treatment programs and services, substance use prevention and treatment programs and services, in-home parent skill-based programs and services, and kinship navigator programs. A program or service must meet the definition for one or more of the program or service areas described below.

A wide range of programs and services may be eligible for review if they meet the definition for an eligible program or service area. This includes, but is not limited to:

- economic and other concrete support programs and services,
- domestic violence (DV) programs and services,

- parent partner programs and services,
- peer partner programs and services,
- parent education peer support groups,
- case management programs and services, and
- other programs and services that focus on contextual and systemic risk factors.

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 modality, including individual, family, or group; use any delivery format, including in-person, virtual, or hybrid delivery; may have any psychotherapeutic orientation, such as cognitive, cognitive-behavioral, psychodynamic, structural, narrative, etc., inclusive of those that are culturally adapted or culturally specific; may be offered by any agency or provider (e.g., community-based organizations, public health agencies, private treatment facilities, tribal health agencies, etc.); and may target co-occurring issues (e.g., mental health and substance use). Programs and services that rely solely on psychotropic medications or screening procedures are not eligible (e.g., a treatment that uses methylphenidate or lisdexamfetamine for treatment of attention deficit hyperactivity disorder).

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

Exhibit 2.1. Examples of Mental Health Prevention and Treatment Programs and Services

Eligible Examples	Not Eligible Examples
A multi-component family preservation program that includes cognitive and behavioral psychotherapeutic interventions and provides concrete goods and services related to intervention goals.	A program that aims to promote family reunification by providing a court advocate for parents without any psychotherapeutic component.

Eligible Examples	Not Eligible Examples
A program that targets individuals with severe mental illness that includes a psychiatrist and psychiatric nurses for prescribing and administering medication, mental health professionals to provide individual therapy, and employment supports.	A treatment that solely provides anti-depressant medication.
A program that targets trauma symptoms and substance use among individuals in recovery from substance misuse who also have experienced domestic violence.[a]	
A program that aims to prevent the development of delinquent behavior among at-risk youth by assessing family strengths and needs, providing counseling to enhance motivation, and customizing referrals to appropriate services.	

[a]This example could also be considered within the substance use program and service area.

Substance Use Prevention and Treatment Programs and Services[3]

Eligible substance use prevention and treatment programs and services include those that have an explicit focus on the prevention, reduction, treatment, remediation, recovery from and/or elimination of substance use or misuse. 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 substances including 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. Eligible programs and services can employ any therapeutic modality, including individual, family, or group; use any delivery format, including in person, virtual, or hybrid delivery; may have any therapeutic orientation, such as cognitive, cognitive-behavioral, psychodynamic, structural, narrative, etc., inclusive of those that are culturally adapted or culturally specific; may be offered by any agency or provider (e.g., community-based organizations, public health agencies, private treatment facilities, tribal health agencies, etc.); and may target co-occurring issues (e.g., mental health and substance use). Programs and services may include use of pharmacological treatment approaches, but those that rely solely on pharmacological interventions are not eligible (e.g., a treatment that uses only methadone for the treatment of opioid use disorder).[4] Interventions that do not include client-oriented

[3] The FFPSA legislation refers to this program area as "Substance Abuse Prevention and Treatment."

[4] Programs and services that rely solely on pharmacological interventions are outside of the scope of what is reviewed. This is not a statement about the potential effectiveness of pharmacological interventions.

substance use prevention or treatment components, such as mass communications/media campaigns or interventions that solely target broader policy systems (e.g., a program that solely organizes roundtable meetings of leaders of local agencies and does not provide any direct services to individuals) are also not eligible. Programs and services aimed solely at referring or getting people into treatment are not eligible (i.e., where services end at the point of referral or treatment entry). Also not eligible are programs and services that solely conduct screening for substance use or misuse.

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

Exhibit 2.2. Examples of Substance Use Prevention and Treatment Programs and Services

Eligible Examples	Not Eligible Examples
A program administered through a public health agency 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 school-based support group program for students with substance-using parents that does not directly address prevention or treatment of student substance use.[a]
A program treating parents who are misusing opioids using a combination of methadone, cognitive behavioral therapy, and peer support.	A program that uses a prescription medication to reduce withdrawal symptoms for adults with alcohol use disorder without an accompanying psychotherapeutic component.
A program that trains parents on how to communicate with their children about substance use prevention.	A local program that provides funding for renovating vacant housing units and public improvements to address perceived community-level risk factors for substance use.
A wraparound intervention that aims to connect persons identified with substance use disorders to appropriate treatment and recovery services and provides ongoing case management for achieving treatment goals.	A program in which a hospital staff member walks the patient through a pamphlet of local substance use treatment options.
	A standalone screening program that is used to provide referrals only.

[a]This example could be considered within the mental health program and service area.

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 and target parenting skills or other skills that can be applied to where the child resides, including in the home. Skill-based means

that programs or services must include components targeting parenting skills or other skills that contribute to parental protective capacity and children's safety and well-being. Direct intervention contact means that intervention services are provided directly to the parent(s) or caregiver(s). Programs and services may be delivered in the home or in other settings, and contact may be face-to-face, over the telephone or video, or online.

Exhibit 2.3 provides some examples of eligible and ineligible programs and services in this area.

Exhibit 2.3. 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.	A program that solely teaches adolescents skills for engaging in problem-solving with their parents to improve family functioning.
An online parenting program that helps parents set goals and match their parenting goals with evidence-based parenting strategies.	An online parent education program on child developmental milestones that does not include any content on parenting skills.
A group-based parent training program for child behavior problems that is delivered over 10 weeks in a community setting.	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.
A concrete or economic support program that includes financial counseling and education skills for parents.	A program that provides financial assistance and concrete support services but does not include any parental skill-based component.
A parent partner program that pairs parents with children who have been removed from the home with parents who have achieved reunification to provide mentoring on skills needed to support reunification, social support, and assistance working with social workers to ensure the family is obtaining needed resources.	A program that provides brief weekly text check-ins from a peer parent to provide emotional support without any skill-based content.
A group-based peer parent program that aims to enhance skills to improve family functioning, discuss parenting strategies, prevent and intervene in substance use disorders, promote mental health, and provide mutual support.[a]	A program that aims to increase parental social support by intervening with key persons in the parent's social network but without direct intervention contact with the parent.

Eligible Examples	Not Eligible Examples
A trauma-informed, psycho-educational family acceptance program for LGBTQIA2S+[b] youth and their families that aims to enhance parenting skills, promote youth mental health, and prevent family separations.[c]	

[a]This example could also be considered within the mental health and substance use program and service areas.
[b]Lesbian, Gay, Bisexual, Transgender, Queer or Questioning, Intersex, Asexual, Two-Spirit, and more.
[c]This example could also be considered within the mental health program and service area.

Kinship Navigator Programs

Eligible kinship navigator programs 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.

Support services may include any combination of financial supports, training or education, support groups, referrals to other social, behavioral, legal, or health services, and assistance with navigating government and other types of assistance, financial or otherwise.

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 kinship caregivers or not, are not eligible.

2.1.2 Available Books, Manuals, or Other Documentation

To be eligible for review for the Prevention Services Clearinghouse, programs and services must be clearly defined and replicable. Consistent with the Family First Prevention Services Act, programs and services must have available written or recorded books, manuals, or other documentation that specifies the components of the practice and describes how to implement or administer the practice to be eligible for review (hereafter referred to as a program or service manual). There must be affirmative, documented evidence that the materials that satisfy this requirement exist and are available to the public to download, request, or purchase; materials may be presented in a web-based format. Programs and services that require training, certification, or other prerequisites to access books, manuals or other documentation would meet this criterion provided that the Prevention Services Clearinghouse can verify the existence of the books, manuals, or other documentation provided and how the materials can be accessed by the public. Other documentation can include protocols, practice guidance, recorded videos, or online learning systems as long as the materials

describe how to implement or administer the practice. Books, manuals, or other documentation may be available in English or other languages. There are no requirements associated with the age or longevity of a program or service manual.

As part of ensuring that programs and services are clearly defined and replicable, sufficient information must be available about the program or service content, dosage, modality, providers, and/or other key components of the program or service. The Prevention Services Clearinghouse conducts a comprehensive information gathering process, including, but not limited to the following sources: (1) review of the relevant manual content (if accessible), (2) review of external materials (e.g., history of program or service changes over time in a journal article, descriptions of the variants on a program or service website, description of the program or service in studies citing the manual(s), other clearinghouses that have reviewed the program or service), (3) written queries to manual authors or program or service developers for information about the manual(s), or (4) documented information from any external expert consultations conducted. Experts who advise the Prevention Services Clearinghouse include, but are not limited to, individuals with lived expertise; individuals from communities that programs and services are intended to serve, inclusive of underserved and tribal communities; methodological experts; state, tribal, and local child welfare department staff or administrators; policymakers; program or service developers; program or service providers or trainers; study authors; and subject matter experts.

If, after following the information gathering procedures specified above, the Prevention Services Clearinghouse does not have sufficient information to define the program or service based on content, dosage, modality, providers, and/or other key components of the program or service and apply its adaptation procedures, the Prevention Services Clearinghouse will indicate that the program or service is not eligible for review at this time. Such programs or services may be eligible for a program or service re-review (Section 8.5.1) should sufficient information become available.

2.2 Program or Service Prioritization Criteria

A high volume of programs and services identified are potentially eligible for review. Therefore, the Prevention Services Clearinghouse prioritizes and selects programs and services for review using the prioritization criteria described in this section. 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 use prevention and treatment programs and services, and in-home parent skill-based programs and services, as well as kinship navigator programs.

2.2.1 Available Evidence of Eligibility

The Prevention Services Clearinghouse first prioritizes programs or services that have evidence they would meet the eligibility criteria in Section 2.1 if selected for review. An initial eligibility assessment is made based on publicly available information about program or service content, the availability of an eligible manual, and any materials submitted to the Prevention Services Clearinghouse via email or during public calls or periodic engagement sessions. If it is unclear whether a program or service is potentially eligible, the Prevention Services Clearinghouse may query program and service developers to gather additional information as needed.

2.2.2 In Use/Active

The Prevention Services Clearinghouse then prioritizes programs or services that are in active use. 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. Examples include a manualized program or service pilot tested in a research study that has no evidence of active use since the study, a program or service that was previously offered through a grant but is no longer available due to expiration of funding, or a substantively different prior manual edition (see Section 2.3.2 below) of a program or service that is no longer actively implemented.

2.2.3 Additional Prioritization Criteria

The Prevention Services Clearinghouse also considers the program or service recommendations received, child welfare relevance, population(s) served, information suggesting existing evaluations and studies, and availability of implementation supports in prioritizing programs and services for review.

- **Program or service recommendations received.** The Prevention Services Clearinghouse considers both the total number of recommendations and the recommendation source. Particular consideration is given to programs and services recommended by state, tribal, or local government child welfare administrators and federal partners.
- **Child welfare relevance.** Evidence that the program or service is designed for, or is commonly used to serve, children, youth, young adults, and/or families receiving child welfare services (or populations similar to those receiving child welfare services or at-risk for receiving child welfare services) is considered. Programs and services in any of the four program or service areas can meet this criterion if they target populations experiencing challenges that may put them at-risk for receiving child welfare services.
- **Population(s) served.** The Prevention Services Clearinghouse records the populations the program or service is intended to serve and considers the

particular needs of populations served across programs and services reviewed. Examples of population characteristics that may indicate particular need include status as an underserved community, tribal groups and nations, and needs associated with parent/caregiver or child age. Other considerations related to populations served may include the presence of culturally grounded or adapted programs and services, culturally adapted program or service content, program adaptations for different populations or age groups, studies conducted with diverse or underserved populations, and other information relevant to this criterion submitted to the Prevention Services Clearinghouse (e.g., via program or service recommendations, periodic engagement sessions, or other means).

- **Previous evaluations and studies**. Information suggesting the program or service has been evaluated using an eligible study design is considered. This includes, but is not limited to, information submitted to the Prevention Services Clearinghouse (e.g., via program and service recommendations, periodic engagement sessions, or other means) as well as information from other relevant clearinghouses (such as the CEBC or HomVEE), evaluations funded by ACF or federal partners (such as the Title IV-E Child Welfare Waiver Demonstrations, Regional Partnership Grants, the Prevention Services Evaluation Partnerships), and lists of published studies (e.g., lists submitted with program or service recommendations, bibliographies on program or service websites).
- **Implementation supports.** The Prevention Services Clearinghouse prioritizes programs and services for which there are implementation supports, including (but not limited to) implementation manuals or frameworks, fidelity checklists or other fidelity-monitoring tools (such as regular site supervision meetings or regular submission of video-taped therapist sessions for quality monitoring), videos, training programs, coaching programs, or any similar resources available for potential program or service adopters. To meet this criterion, there must be affirmative, documented evidence that such supports are available to the public, either at no cost or for purchase.

To assess programs and services on these prioritization criteria, the Prevention Services Clearinghouse examines publicly available information (e.g., program or service website), other clearinghouses' websites, materials submitted with program or service recommendations, and information gathered during periodic active engagement sessions or consultations with experts such as those with lived expertise as well as program or service developers.

2.3 Selection of Programs or Services for Review

2.3.1 Working List of Programs or Services Planned for Review

Once programs and services are assessed on the prioritization criteria, the Prevention Services Clearinghouse selects programs and services to add to the working list of *Programs and Services Planned for Review*.[5] The program and service names and program or service area categorizations provided on the working list are based on preliminary information obtained during the program or service eligibility and prioritization process and are subject to change as additional information is obtained during the review process. Final eligibility determinations are made once programs and services are under review. Any programs or services selected for re-review (see Chapter 8) are also added to the working list of programs and services planned for review.

2.3.2 How the Prevention Services Clearinghouse Handles Programs and Services with More Than One Manual

All programs and services reviewed by the Prevention Services Clearinghouse must have publicly available books, manuals, or other documentation that describe how to implement or administer the program or service (referred to subsequently in this section as a "manual"; see Section 2.1.2). Many programs and services have more than one manual edition or version. Some of these are different editions created as a program or service evolves over time or expands beyond its original developer(s) (referred to here as *manual editions*). Other cases with more than one manual represent variants of a program or service, which may be designed to address new issues or different populations or may present alternative approaches to delivering the program or service (referred to here as *manual variants*).[6]

Identifying Manual Editions or Variants

Manual editions and manual variants may be identified during multiple steps of the Prevention Services Clearinghouse review process:

---

[5] The first set of programs and services was selected for systematic review prior to Handbook v1.0. Those programs and services 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.

[6] 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 or service.

- Specific editions or variants may be recommended or otherwise identified at the program and service identification step described in Chapter 1.
- Specific editions or variants may be identified during the program or service eligibility and prioritization steps described in Sections 2.1 and 2.2.
- The Prevention Services Clearinghouse may also identify manual editions or variants during the study eligibility screening process (as described in Chapter 4).[7]
- Manual editions or variants may also be identified when a program or service is re-reviewed, in responses to developer queries, or when information is shared with the Prevention Services Clearinghouse about new manual editions or variants for programs and services that have been reviewed previously (see Chapter 8).

When the Prevention Services Clearinghouse determines that more than one manual edition or manual variant of a program or service exists, the following procedures are used to determine a *focal manual* for the program or service under review and whether other manual editions or manual variants identified have substantial differences from the focal manual.

Selecting a Focal Manual for Programs and Services Under Review

When there is more than one *manual edition* available for a program or service, the Prevention Services Clearinghouse typically selects the most current publicly available manual edition as the focal manual for the program or service, based on the prioritization criterion for programs or services that are in active use (Section 2.2.2).

If more than one *manual variant* of a program or service exists, the Prevention Services Clearinghouse will generally attempt to identify a focal manual that represents the standard or most comprehensive or complete version of the program or service under review. The Prevention Services Clearinghouse will typically treat each manual variant as describing a distinct program or service unless there is indication that the manual variant describes a program or service that is not substantially different from the program or service described in the focal manual identified (as outlined in the processes for identifying and assessing substantial adaptations below).

[7] During study eligibility screening, the Prevention Services Clearinghouse may also identify study-specific adaptations made to a program or service that may not be codified in an alternate manual. The procedures for assessing the eligibility of such studies are described in Section 4.1.9.

Identifying Adaptations in Program or Service Manuals

In situations where multiple manuals are identified, the Prevention Services Clearinghouse may gather information about whether there are any differences in key program or service components between the focal manual identified and any relevant manual editions or variants that are identified.

Sources of information used to identify adaptations or modifications in manual editions or variants to program or service components described in the focal manual may include, but are not limited to: (1) review of the relevant manual content (if accessible), (2) review of external materials (e.g., history of program or service changes over time in a journal article, descriptions of the variants on a program or service website, description of the program or service in studies citing the manual(s), other clearinghouses that have reviewed the program or service), (3) queries to manual authors for information about the manual(s), or (4) external expert consultation. Experts who advise the Prevention Services Clearinghouse on adaptations or modifications include, but are not limited to, individuals with lived expertise; individuals from communities that programs and services are intended to serve, inclusive of underserved and tribal communities; methodological experts; state, tribal, and local child welfare department staff or administrators; policymakers; program or service developers; program or service providers or trainers; study authors; and subject matter experts. Reviewers document any adaptations or modifications observed, focusing on four domains of key program or service components: dosage, modality, content, and providers (see Exhibit 2.5 below for additional detail on these domains).

Assessing Substantial Adaptations

If any adaptations between manual editions or variants are identified, the Prevention Services Clearinghouse assesses whether the adaptations represent a substantial adaptation from the program or service as described in the focal manual.[8] This assessment is conducted using a systematic stepwise process, summarized in Exhibit 2.4 and described below.

Step 1: Is the adaptation explicitly prohibited in the focal program or service manual or the result of adding another program or service to the existing program or service? Any adaptations identified in an alternative manual that are explicitly prohibited in the focal program or service manual are considered to be substantial adaptations. For example, if a program targeting adult depression explicitly states that the use of the program to treat schizophrenia is prohibited, a manual variant targeting individuals with schizophrenia would be considered to be a substantial adaptation.

---

[8] Standards for assessing the eligibility of studies of a selected program or service that exhibit *study-specific* adaptations are similar and are discussed in detail in Section 4.1.9.

Additionally, adaptations that involve adding a separate program or service to the existing program or service (i.e., "bundling") are also considered to be substantial adaptations. For example, a manual variant that adds a separate trauma-informed mental health treatment program to a focal manual that only describes a substance use treatment program would be considered a substantial adaptation.

Step 2: Is the adaptation explicitly allowed by the focal program or service manual? Adaptations identified in an alternative manual that are explicitly allowed in the focal program or service manual are generally considered not to be substantial adaptations. For example, consider a focal manual that includes sample content for a community resources module and specifies that the content should be tailored to the population served and local context. In this case, a variant that adapts the sample content for the community resources module would not be considered a substantial adaptation.

Step 3: Does the adaptation substantially change a key program or service component in the focal program or service manual? When observed adaptations are not explicitly prohibited or allowed in a focal program or service manual, the Prevention Services Clearinghouse assesses whether the adaptations substantially change one or more key components of the program or service as described in the focal manual. Examples of adaptations that are considered to be substantial or not substantial are provided in Exhibit 2.5.

Step 4. After gathering any additional information needed, have experts determined that the adaptation is substantial? If Prevention Services Clearinghouse staff cannot determine if an adaptation in a manual edition or variant is substantial or not after the first three steps, the Prevention Services Clearinghouse may also query study authors, program and service developers, or other outside experts to request information necessary to better understand the adaptations and whether they are substantial or not substantial. Experts who advise the Prevention Services Clearinghouse on adaptations include, but are not limited to, individuals with lived expertise; individuals from communities that programs and services are intended to serve, inclusive of underserved and tribal communities; methodological experts; state, tribal, and local child welfare department staff or administrators; policymakers; program or service developers; program or service providers or trainers; study authors; and subject matter experts. Experts will be consulted to develop a final decision on whether a particular adaptation is substantial or not substantial.

Exhibit 2.4. Process for Assessing Substantial Adaptations

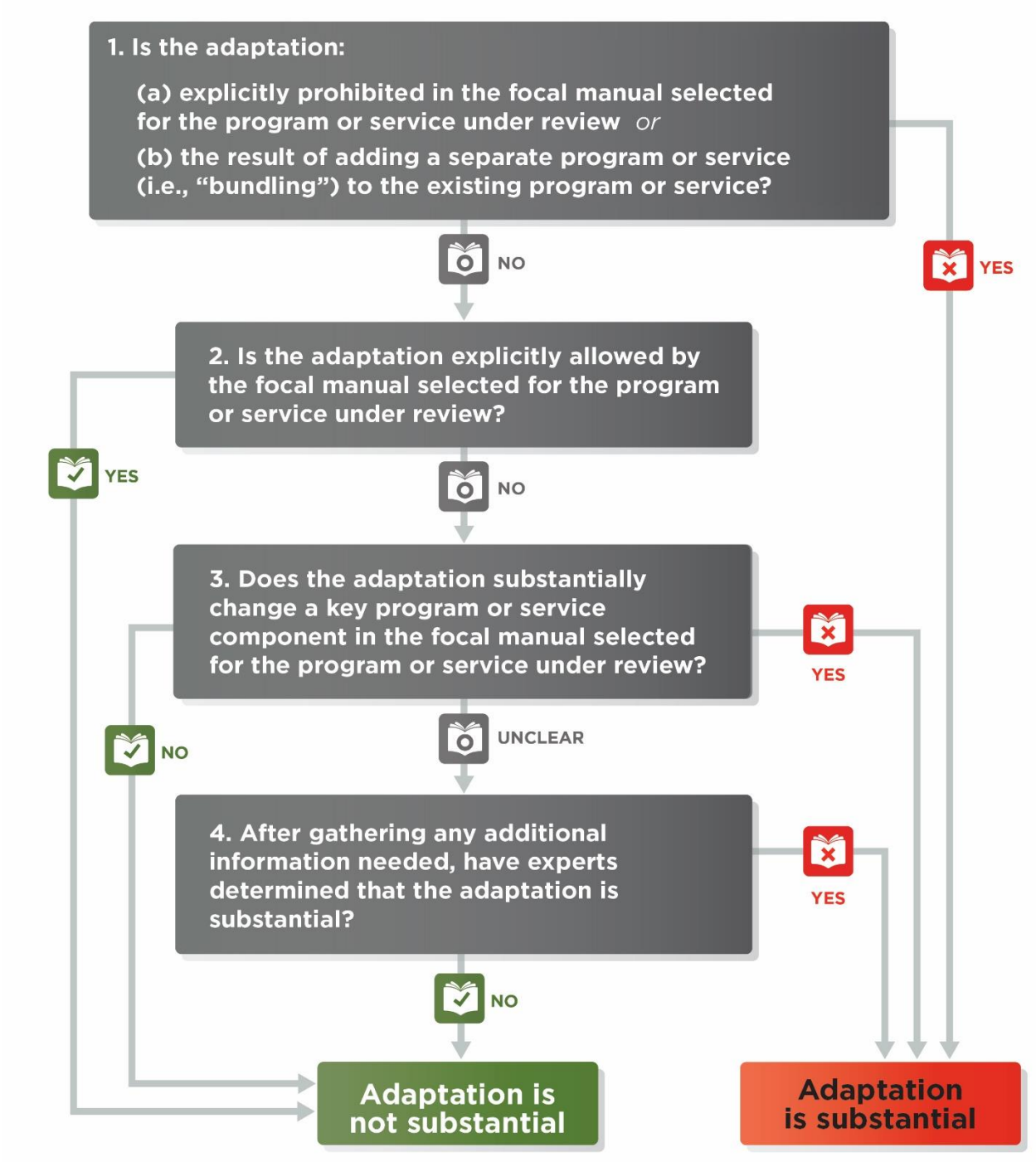


Exhibit 2.5. Examples of Adaptations That Are Substantial and Not Substantial

Program or Service Key Component Domains	Adaptations That Are Not Substantial	Adaptations That Are Substantial
Dosage – The intended quantity, duration, and frequency of services to be delivered. Can include characteristics of individual sessions (e.g., session frequency and length) and of the overall program or service (e.g., treatment duration, total sessions, total hours)	• Modestly changing session frequency (e.g., number of sessions per week or month) or session length (e.g., number of minutes per session), or total number of sessions • Modestly changing the duration of treatment (i.e., the length of time from the start of the program or service to the end) • Modestly changing the total amount of treatment contact time (e.g., number of hours of treatment services delivered over the course of the program or service) • Accelerating or lengthening treatment without changing the total treatment contact time (e.g., switching from 12 weekly sessions to 6 twice-weekly sessions or vice versa but keeping 12 total hours of treatment overall) • Minor differences in session or program or service dosage when these components are defined flexibly for the program or service (e.g., delivering a program in 21 sessions when the number of sessions usually ranges from 15–20 sessions; a treatment duration of 2.5 months when it is typically completed within 2 months)	• Changes in session frequency (e.g., monthly to weekly) or session length (e.g., extending sessions from one hour to four hours) that substantially change the total treatment contact time • Other modifications that substantially change the total treatment contact time (e.g., a brief version of a therapy that reduces the total number of hours spent in therapy from 40 hours to 10 hours)

Program or Service Key Component Domains	Adaptations That Are Not Substantial	Adaptations That Are Substantial
Modality – The delivery setting (e.g., in-home vs. office) and format (e.g., group vs. individual) of the program or service	• Delivering the intervention in the home compared to office-based delivery • Changing the delivery setting or format to meet the needs of a specific population (e.g., using talking circles instead of focus groups in tribal populations, having separate discussion groups for mothers and fathers in populations with gender-specific parenting norms) • An intervention typically delivered one-on-one between a client and a therapist is adapted to allow sessions to be completed either in-person or via videoconference without substantial changes to intervention content	• Changing from individual to group therapy or vice versa • Changing from synchronous (i.e., live) to asynchronous (i.e., self-paced) program or service delivery
Content - The subject matter, themes, activities, examples, skills, methods, and/or goals of a program or service intervention	• Modifying examples or illustrations • Providing the intervention in a different language • Updating activities or exercises to increase the relevancy of the program or service in a particular cultural or contextual setting • Adding an introductory session or a concluding/ graduation session to reinforce existing content	• Making substantial changes to content (e.g., adding a substance use prevention component to a parenting intervention that previously only had content on child anxiety; making substantive content modifications to address developmental differences)
Providers – Characteristics of the providers who are intended to implement the program or service (e.g., education, experience, training)	• Delivering the program or service by slightly different types of providers than described in the manual or original research on the program or service (e.g., using M.S.W. social workers instead of master's-level licensed counselors) • Requiring providers to have relevant language skills and cultural knowledge of and experience working with the population being served	• Delivering the program or service with substantially different providers than described in the manual (e.g., using untrained paraprofessionals instead of trained nurses to deliver a program) • Changing from provider-led sessions to self-led sessions

Note. Adaptations may affect multiple program or service components simultaneously and in this case the Prevention Services Clearinghouse will review changes to each component independently. For example, adding two sessions may not constitute a

substantial adaptation based on dosage alone; however, it may still constitute a substantial adaptation based on content if new subject matter is added.

Determining Whether Manual Editions or Variants Represent a Separate Program or Service

If the Prevention Services Clearinghouse determines that manual editions or variants identified do not have any substantial adaptations from the focal manual identified, studies of the program or service as described in these manual editions or variants will be included in the same review of the program or service as specified in the focal manual. The manual(s) listed under the *Book/Manual/Available documentation* section on the Prevention Services Clearinghouse website for the program or service reviewed reflects the focal manual and any relevant manual editions or variants identified of the program or service determined not to have substantial adaptations.

If the Prevention Services Clearinghouse determines that a *manual edition* has any substantial adaptations from the focal manual identified, the alternative manual edition is considered to describe a separate program or service – and thus subject to program or service eligibility and prioritization criteria (Sections 2.1 and 2.2) for determining its eligibility and prioritization for review. Similarly, if the Prevention Services Clearinghouse determines that a *manual variant* has any substantial adaptations, the manual variant would be treated as describing a separate program or service – and thus subject to program or service eligibility and prioritization criteria (Sections 2.1 and 2.2) for determining its eligibility and prioritization for review. The Prevention Services Clearinghouse may review more than one variant of a program or service at a time.

3. Literature Search

For each program or service selected for review, Prevention Services Clearinghouse staff conduct a comprehensive and systematic search for potentially eligible studies of that program or service. All search results are carefully documented in databases maintained by the Prevention Services Clearinghouse. Duplicate citations are removed before screening them for eligibility.

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

Other Clearinghouses. The search begins by identifying studies 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[a]	Website
Blueprints for Healthy Youth Development (Blueprints)	www.blueprintsprograms.org
California Evidence-Based Clearinghouse for Child Welfare (CEBC)	www.cebc4cw.org
The Campbell Collaboration	https://campbellcollaboration.org/
The Cochrane Collaboration	https://www.cochrane.org/
CrimeSolutions and the Office of Juvenile Justice and Delinquency Prevention Model Programs Guide	www.crimesolutions.gov
Healthy Native Youth	https://www.healthynativeyouth.org/
Home Visiting Evidence of Effectiveness Review (HomVEE)	https://homvee.acf.hhs.gov
National Registry of Evidence-based Programs and Practices (NREPP)	https://www.pewtrusts.org/en/research-and-analysis/data-visualizations/2015/results-first-clearinghouse-database
National Traumatic Stress Network	https://www.nctsn.org/treatments-and-practices/trauma-treatments/interventions
Social Programs that Work (SPTW)	https://evidencebasedprograms.org/programs/
Teen Pregnancy Prevention (TPP) Evidence Review	https://tppevidencereview.youth.gov/
Washington State Institute for Public Policy (WSIPP)	http://www.wsipp.wa.gov

[a]Additional clearinghouses may be used, depending on the program or service selected.

Program or Service Website Scan. The websites of programs and services identified for inclusion often cite or list relevant research as evidence of program or service

effectiveness. For this reason, Prevention Services Clearinghouse staff review relevant program or service website(s), if applicable, to identify potentially eligible studies.

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[a]	Website
Cumulative Index to Nursing and Allied Health Literature (CINAHL)	https://www.ebsco.com/products/research-databases/cinahl-database
Education Resources Information Center (ERIC)	https://eric.ed.gov/
MEDLINE Complete (PubMed)	https://pubmed.ncbi.nlm.nih.gov/
PsycINFO	https://www.apa.org/pubs/databases/psycinfo

[a]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 to identify any additional potentially eligible studies that may not be indexed in the standard electronic databases. The following grey literature sources are included in all searches, with additional sources added as content experts recommend.

Exhibit 3.3. Grey Literature Sources Used to Identify Relevant Research

Source[a]	Website
Abt Global	https://www.abtglobal.com/
Annie E. Casey Foundation	https://www.aecf.org/search/
Chapin Hall Center for Children at the University of Chicago	https://www.chapinhall.org/our-work/
Child Trends	https://www.childtrends.org/publications
Child Welfare Information Gateway Library	https://www.childwelfare.gov/library/
James Bell Associates	https://www.jbassoc.com/resource/#p=1&postType=resource
Mathematica	https://www.mathematica.org/evidence
MDRC	https://www.mdrc.org/publications
RAND Corporation	https://www.rand.org/search/advanced-search.html
Urban Institute	https://www.urban.org/research
Westat	https://www.westat.com/about-us/publications

[a]Additional sources may be used, depending on the program or service selected.

Ad Hoc Submissions of Research. Program or service recommendations sometimes include bibliographies or lists of published research for the program or service. The Prevention Services Clearinghouse also may receive ad hoc submissions of publicly available research via email and may solicit submissions from the public on the website, through requests to the email list, and via periodic engagement sessions. To further ensure that the Prevention Services Clearinghouse is identifying potentially harder-to-reach evidence, such as evaluations conducted by community-based organizations, the Prevention Services Clearinghouse will conduct outreach through the email list after each update to the working list of programs and services planned for review. That outreach will ask the public to submit studies of programs and services that have been added to the working list. All recommendation materials and ad hoc materials submitted are recorded. Submissions made prior to the program or service being added to the working list of programs or services under review (Section 2.3) or in a timely response to the email outreach after the program or service is added to the working list are reviewed to identify potentially eligible studies as part of the literature search. Submissions made after a program or service is added to the working list of programs and services and the email outreach has been conducted are recorded and may be incorporated into the review, depending on the time of receipt in the review process.

4. Study Eligibility Screening and Prioritization

4.1 Study Eligibility Criteria

The Prevention Services Clearinghouse defines a "study"[9] 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 of Standards and Procedures. 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. This requirement pertains only to the date of study publication, not to the date a program or service was developed. There is no requirement related to the age or longevity of a program or service.

4.1.2 Source of Publication

Studies must be publicly available (i.e., available to the public to download, request, or purchase). This includes: (a) studies published in journals; and (b) published studies reported in documents prepared or commissioned by federal, state, tribal, or local government agencies or departments, private agencies or organizations, universities, research institutes, research firms, foundations or other funding entities, or other similar organizations.[10] Dissertations, theses, and conference papers are not eligible.

---

9 Sometimes study results are reported in more than one document, or a single document reports results from multiple studies. The Prevention Services Clearinghouse focuses on whether there is overlap in the research investigation and defined subject sample when assessing whether impact estimates are classified as part of the same study.

10 For studies published in journals, "published" is defined as studies accepted by the journal for publication that are available to the public (online or in print). For studies under the "b" criterion, "published" means a final publicly available version of the prepared/commissioned document is available. For example, a final version of an interim evaluation report submitted to a federal agency that is publicly available would be considered "published". A draft version of the same interim report that had not yet been submitted to the commissioning federal agency would not count as being "published", even if it was publicly available. If multiple versions of the same document are published (e.g., an initial and revised version), the Prevention Services Clearinghouse focuses on the most recently published document that is publicly available. If it is unclear whether a document identified has been published or is publicly available, the Prevention Services Clearinghouse will query study authors to confirm the publication status of the document and how it can be accessed. Documents that cannot be confirmed to be publicly available do not meet the source of publication criterion.

4.1.3 Language of Publication

Studies must be available in English.[11] This can include studies originally published in English or English-language translations of studies originally published in another language.

- The Prevention Services Clearinghouse is planning to conduct a pilot to understand the feasibility of identifying and reviewing studies only published in Spanish.

4.1.4 Location of Study

Studies may be conducted in any country. There are no eligibility requirements related to the number of locations, sites, or jurisdictions represented in a study.

4.1.5 Study Design Criteria

In alignment with the Family First Prevention Services Act, eligible studies must use a design with at least one intervention condition involving the program or service under review and one or more appropriate comparison conditions that utilize some form of control (defined in Section 4.1.7 below). The Prevention Services Clearinghouse currently reviews randomized controlled trials and quasi-experimental designs (as defined below).

- The Prevention Services Clearinghouse is planning to develop and pilot standards for reviewing studies that use single case designs (SCDs).

Design Criteria for Randomized Controlled Trials (RCTs) and Quasi-Experimental Designs (QEDs)

Randomized Controlled Trials (RCTs): Randomized controlled trials are designs in which units (individuals or clusters of individuals) are assigned to conditions using a random process. In other words, each unit must have a nonzero probability of being assigned to each intervention or comparison condition. Assignment probabilities do not have to be equal (e.g., studies may assign 60% to one condition and 40% to another). Eligible randomized designs include simple randomization of individuals or clusters to intervention and comparison conditions. Eligible randomized designs also include those in which individuals or clusters are blocked or stratified and then randomized. For blocked or stratified randomized designs in which different blocks or strata have different assignment probabilities, the different assignment probabilities must be addressed in the impact analysis (see Section 5.9.1 on statistical model standards).

[11] Although studies must be available in English, program or service protocols, manuals, or other materials and implementation supports may be available in other languages.

Quasi-experimental designs (QEDs): The Prevention Services Clearinghouse defines quasi-experimental designs as those in which an intervention condition is compared to a comparison condition, but the units (individuals or clusters of individuals) are not randomly assigned to conditions. The Prevention Services Clearinghouse has no specific requirements for how intervention and comparison conditions are formed for a quasi-experimental design. A wide range of designs can be eligible including, but not limited to, matched and non-matched designs, natural experiments, before-after-control-impact designs, controlled before-and-after designs, and difference-in-difference designs. Conditions may be formed through natural processes or formed by researchers. One example of natural variation would be if a state rolls out a new intervention with natural variation in the start dates across counties, such that there is a potential comparison condition of individuals who did not yet have the intervention available in their county. Researchers may also use statistical techniques to match individuals (or groups of individuals) who participated in an intervention to individuals who did not participate based on their measurable characteristics (e.g., propensity score matching). Intervention and comparison conditions may be formed before or after outcome data are collected.

For both RCTs and QEDs, the unit of assignment to intervention and comparison conditions may be either individuals or clusters of individuals (e.g., families, providers, centers, or geographic areas). The quantitative portion of mixed methods designs may be eligible provided they meet other eligibility criteria, including having at least one intervention condition and one or more appropriate and exclusive comparison conditions. There are no sample size restrictions for study eligibility, meaning the Prevention Services Clearinghouse does not consider sample size in determining the eligibility of studies for review.

Ineligible study designs: Study designs that do not meet the criteria for a randomized controlled trial or quasi-experimental design stated above are not eligible. Single group pretest-posttest designs, in which a single group receives an intervention and the outcomes are assessed before and after the intervention, are not eligible because such designs do not include an appropriate comparison condition. In addition, the Prevention Services Clearinghouse does not currently review studies that use regression discontinuity designs.[12]

[12] Although regression discontinuity designs are types of quasi-experimental designs (Shadish, Cook, & Campbell, 2002), separate eligibility and review criteria for such designs would need to be applied. The Prevention Services Clearinghouse does not currently have standards for rating regression discontinuity, though these may be forthcoming in future versions of the Handbook. Regression discontinuity designs are, therefore, not currently eligible for review.

4.1.6 Intervention Condition

Studies must have intervention condition(s) that are offered the program or service under review that is essentially the same for all participants in the condition. Variation across individuals in what they actually receive is acceptable. Variation as permitted in the program or service manual is also acceptable (e.g., programs or services that provide different manualized services based on results of a screening assessment for need). However, intervention conditions constructed by aggregating subsamples receiving distinctly different intervention conditions into a single study sample are not eligible (e.g., combining a sample offered a brief version of the program with a sample offered an intensive version of the program).

4.1.7 Comparison Conditions

In alignment with the Family First Prevention Services Act, eligible RCT or QED studies must use an appropriate comparison condition. Comparison conditions for RCTs and QEDs must be exclusive (i.e., participants in the comparison condition may not overlap with those in the intervention condition). The Prevention Services Clearinghouse reviews the following types of appropriate comparison conditions:

- *No intervention, untreated group, or wait list.* Participants are offered no services or participants are assigned to receive the intervention under study at a later date.[13]
- *Minimal intervention.* Participants may receive informational materials or psychoeducation,[14] referrals to available services, or similar nominal services. Such additional services must be brief and/or predominantly information-based (e.g., pamphlets about child development, psychoeducation about a specific disorder).
- *Placebo or attention control.* Includes psychological or pharmacological placebos, attention placebos, and nonspecific therapy in which participants receive the same or similar amount of attention or contact as the participants in the intervention condition.[15] Typically, these conditions are designed to account for nonactive effects of treatment, such as participants' expectations, contact

[13] Measurements occurring after a wait list group is offered treatment are not eligible.

[14] Psychoeducation is defined as the provision of information about a health or mental health condition without the provision of therapy.

[15] The term "active control group" is sometimes used in social, educational, and behavioral research to describe interventions that fall under the placebo or attention control category, whereas in medical research, it more commonly describes alternative interventions believed or known to be effective. In both cases, these types of comparison conditions would be eligible for review by the Prevention Services Clearinghouse. Active control groups that receive alternative interventions believed or known to be effective would be categorized as head-to-head comparisons for assessing risk of harm.

time with an interventionist, or the relationship between interventionist and participants (Freedland et al., 2019).

- *Treatment as usual.* Participants in treatment as usual comparison conditions may already be receiving services in their communities or they may be offered services as part of the research study. The Prevention Services Clearinghouse considers a comparison condition to be "treatment as usual" under either of the following two conditions:

  - *Condition 1: Usual or typical services*. This condition refers to comparison conditions in which individuals are already receiving services in their communities or are offered services that they would have received in the absence of the study (i.e., they do not receive anything they would not have been able to receive anyway). In such cases, the study must clearly describe these services as the usual or typical services available for the population included in the study. The amount of contact and type or content of services can differ across participants. It is acceptable for studies with comparison conditions in this category to provide or offer a minimal intervention (as defined above) along with usual services. Studies may also constrain or standardize available usual services that are provided in the context of the study.
  - *Condition 2: Services consistent with usual or typical services*. This condition refers to comparison conditions that are offered services as part of the study that are not offered in the community but are clearly described as consistent with the usual or typical services that would be received by individuals or families similar to those in the study. In such cases, the study must provide specific information to justify that the services offered to the comparison condition participants are consistent with what individuals or families like those in the study could be expected to receive in the absence of a study.[16]

  Treatment as usual services may include therapeutic or pharmacological interventions that meet the criteria for either of these two conditions. If it is unclear from study documentation whether the services described in the study constitute usual or typical services offered in the community (Condition 1) or services consistent with usual or typical services (Condition 2), the Prevention

[16] The National Institutes of Health Office of Behavioral and Social Sciences Research Expert Panel on Comparator Selection in Behavioral and Social Science Clinical Trials (Freedland et al., 2019) refers to such comparison conditions as optimized or standardized care conditions.

Services Clearinghouse may query study authors to gather information needed to make this determination.

- *Head-to-head comparisons.* Head-to-head comparisons can also be referred to as alternative interventions, active interventions, active control interventions, or comparator interventions. Participants are assigned to another intervention that is not a variant of the program or service under review (as defined below) and does not meet the criteria for treatment as usual. This may include studies in which a new or emerging intervention is compared to an established one, studies that compare interventions that ascribe to different or contrasting theoretical orientations, or studies that compare two or more commonly used manualized interventions to each other. Excluded are comparisons to pharmacological interventions or psychotropic medications that do not meet the definition of treatment as usual above.

Comparison Conditions That are Not Eligible for Review:

- *Intervention Variants.* Studies that compare intervention variants to each other may address the components of programs or services that work best or the circumstances in which programs or services may be most effective. However, the primary purpose of the Prevention Services Clearinghouse is to assess the effectiveness of programs and services themselves, not their components or the circumstances in which they work best. Therefore, comparison conditions that are a variant of the intervention under review are not eligible for Prevention Services Clearinghouse review. Examples of such comparisons include:
  - Dismantling studies, which compare a full version of an intervention to a version lacking one or more components of the same intervention;
  - Bundled intervention studies, which compare a full version of an intervention to a version with a second intervention added;
  - Studies that compare different delivery modes (e.g., group vs. individual), provider types (e.g., ethnically matched therapists vs. non-matched therapists), or dosage or fidelity levels for the same intervention;
  - Sequencing studies in which the same intervention is delivered to participants in both conditions but in a different order.
- *Population-level data or benchmarks.* Comparison conditions constructed from population norms or statistics derived from other studies, surveys, censuses, or similar sources are not eligible. In such studies, baseline equivalence on measurable variables related to the outcome does not adequately protect against the potential for large unobserved differences between the intervention and comparison conditions at the beginning of a study.

- *Comprised only of intervention refusers or dropouts.* Comparison conditions composed entirely of individuals who were offered the intervention condition but refused the offer or dropped out of the intervention after being offered the intervention are not eligible. Studies with these types of comparison conditions have inherent design confounds and do not offer credible evidence of the effectiveness of the intervention.

4.1.8 Outcomes

> **Outcome domain:** A broad group of related outcomes.
>
> **Outcome:** A behavior, skill, condition, or other characteristic that is measured to assess the impact of a program or service.
>
> **Outcome measure:** A survey, instrument, device, or other tool used to assess or quantify an outcome.

Consistent with the Family First Prevention Services Act, studies must report program or service impact estimates for at least one eligible *outcome.*[17] The Prevention Services Clearinghouse classifies outcomes into *outcome domains*.[18] There are four outcome domains used by the Prevention Services Clearinghouse for all programs and services: Child Safety, Child Permanency, Child Well-being, and Adult Well-being. Outcomes in three additional outcome domains (Access to Services, Referral to Services, Satisfaction with Programs and Services) are eligible for kinship navigator programs, as described below. Outcomes that are composites of one or more eligible outcomes within these outcome domains, as defined below, are also eligible (e.g., a composite measure of child behavior and social functioning); outcomes that are composites of both eligible and ineligible outcomes are not eligible.

Eligible outcomes and outcome measures may be defined differently across studies to reflect the different ages, backgrounds, cultures, locations, and contexts of the study participants. For example, outcome measures of behaviors and situations that may be experienced differently across cultures or manifested differently for different ages can be eligible as long as the outcome meets one of the definitions for the eligible outcomes below (or more than one for eligible composite outcomes, as described above). As another example, perceptions of outcomes as individual or communal may vary across communities. Communal, community-level, or population-level outcomes can also be eligible as long as they meet one of the definitions for the eligible outcomes below. Additionally, reductions in disparities among Black, American Indian/Alaskan Native, Hispanic or Latino, and other impacted people of color; people who are LGBTQIA2S+,

---

[17] To receive a high or moderate rating on the design and execution standards, *eligible outcomes* must be assessed with an *outcome measure* that meets the measurement standards described in Section 5.9.2.

[18] The Prevention Services Clearinghouse outcome domains were selected to align with the important child and parent outcomes described in the authorizing legislation for the Prevention Services Clearinghouse, the Family First Prevention Services Act of 2018.

persons with disabilities, and other populations who historically have experienced disparity can be eligible for the outcomes listed below. For example, reductions in racial disparities in substantiated child maltreatment administrative records can be eligible.

The eligible outcome measures listed below are examples and the Prevention Services Clearinghouse may engage with those with lived expertise and from relevant populations and communities to ensure the inclusivity of eligible outcome measures used in varying cultural contexts.

Eligible Outcome Domains for Mental Health Prevention and Treatment, Substance Use Prevention and Treatment, and In-Home Parent Skill-Based Programs and Services and Kinship Navigator Programs

- **Child Safety Outcome Domain.** 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. Outcomes in this domain may be assessed in either the positive (e.g., fewer or no records of substantiated maltreatment) or negative (e.g., greater child maltreatment risk) direction. The outcomes that are eligible in the child safety outcome domain are:
  - *Child Welfare Outcomes from Administrative Sources.* Refers to substantiated or unsubstantiated child maltreatment from administrative records. Eligible outcome measures include, but are not limited to, substantiated and unsubstantiated reports of abuse or neglect, investigations of abuse and neglect from administrative records, recurrence of abuse and neglect from administrative records.
  - *Child Welfare Outcomes from Non-Administrative Sources.* Refers to reports of child maltreatment from sources other than administrative records. Eligible outcome measures include, but are not limited to, victim (including youth reports) and perpetrator reports of abuse or neglect, questionnaire or interview instruments that directly assess abusive behavior or neglect.
  - *Child Maltreatment Risk Outcomes from Medical Sources.* Refers to medical reports of child health that may indicate risk of child maltreatment. Eligible outcome measures include, but are not limited to, administrative, questionnaire, or interview instruments assessing childhood injuries, ingestions, emergency room visits, and hospitalizations.
  - *Child Maltreatment Risk Outcomes from Non-Medical Sources.* Refers to the extent to which factors are present that may increase the likelihood of a

child experiencing maltreatment as collected through a systematic assessment that is not a medical report. Eligible outcome measures include, but are not limited to, child maltreatment risk assessments.

- **Child Permanency Outcome Domain.** 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. Outcomes in the child permanency domain may be assessed in either a positive (e.g., family preservation) or negative (e.g., child is removed from the home) direction. The Prevention Services Clearinghouse reviews the following outcomes in the child permanency outcome domain:
  - *Out-of-Home Placement Outcomes.* Refers to any situation where a child is removed from the family home.[19] Eligible outcome measures include, but are not limited to, any out-of-home placement, placement to foster care, reports of the caregivers relinquishing their roles, time to placement in out-of-home care, avoiding out-of-home placement, and remaining in the home.
  - *Least Restrictive Placement Outcomes.* Refers to out-of-home placements that could be considered the least restrictive or disruptive. Least restrictive placement outcomes focus on the environments/settings into which children are placed, favoring kinship, fictive kin, or therapeutic kinship placements over non-kin or institutional placements, or placements that maintain connections to the child's community vs. those that do not. Outcome measures must be operationalized with more than two placement settings, either as binary measures for which the reference category is another out-of-home placement setting or as movement from more restrictive/disruptive to less restrictive/disruptive settings. Eligible outcome measures include, but are not limited to, hierarchies of least restrictive preference (e.g., kin placement, family foster care, therapeutic care, group home, residential, hospitalization).
  - *Placement Stability Outcomes.* Refers to the stability of out-of-home placement (e.g., that children are in placements that are disrupted infrequently). Placement stability outcomes focus on the number of placement disruptions (planned and unplanned) or number of out-of-home placements. Eligible outcome measures include, but are not limited to, number of placement changes or disruptions of placements, and re-entries or failed exits/reunifications or adoptions.

[19] Outcome measures that solely assess a child's incarceration status are classified as delinquent behavior outcomes; outcome measures that include incarceration among multiple out-of-home placement settings may be classified as out-of-home-placement or least restrictive placement outcomes.

  - *Planned Permanent Exit Outcomes.* Refers to placements or time to placement to a more permanent status, including reunification, guardianship, tribal customary adoption, and adoption. Eligible outcome measures include, but are not limited to, time to reunification, guardianship, or adoption and reunification rates.

- **Child Well-being Outcome Domain.** Child well-being is a multi-faceted domain that broadly refers to the skills and capacities that enable young people to understand and navigate their world in healthy, positive ways.[20] It is an umbrella term that includes, but is not limited to, child and youth[21] behavioral, social, emotional, physical, and cognitive development. The Prevention Services Clearinghouse reviews the following outcomes in the child well-being outcome domain, the specific nature of which may vary with age:[22]

  - *Behavioral and Emotional Functioning Outcomes.* Refers to 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 outcome measures are eligible. Eligible outcome measures include, but are not limited to, externalizing behaviors (e.g., aggressive behavior, disruptiveness, impulsive behavior), internalizing behaviors (e.g., depression, anxiety, post-traumatic stress, mood or thought problems), mental/behavioral health diagnoses, positive behavior, resilience, self-regulation or self-control, attachment characterization and quality, and emotional adjustment.

  - *Social Functioning Outcomes.* Refers to skills and capabilities relating to the ability to develop, maintain, and manage interpersonal relationships, including how children understand and characterize important relationships in their lives. Eligible outcome measures include, but are not limited to, measures of social skills, assertiveness, cooperation, empathy, social adjustment, peer relations, relationships to the community, rebelliousness,

---

20 This definition draws from the well-being framework specified in *Promoting social and emotional well-being for children and youth receiving child welfare services* (Administration for Children and Families, 2012).

21 Eligible outcomes for youth may also include those listed in the adult well-being domain.

22 Biomarkers, in which a physiological measure is used as an indicator of a physical**,** psychological or emotional state, are not currently eligible outcomes under the child or adult well-being domains because the Prevention Services Clearinghouse does not have standards appropriate for assessing the reliability, face validity, and consistency of measurement for such outcomes. Examples of biomarkers include, but are not limited to, salivary cortisol as a measure of stress responsivity, respiratory sinus arrhythmia as an indicator of self-control or self-regulation, skin conductance as an indicator of behavioral inhibition, brain wavelength spectra as indicators of cognitive function, or ratios of specific t-cells as an indicator of susceptibility to disease.

defiance, and other similar characteristics related to interpersonal interactions and relationships.

- *Cognitive Functions and Abilities Outcomes.* Refers to abilities related to reasoning, knowledge, problem-solving, mental processing, executive functioning, and the like. Eligible outcome measures include, but are not limited to, intelligence tests, developmental assessments, measures of visual or spatial processing, and other indicators of cognitive functions and abilities.
- *Educational Achievement, Attainment, and Attendance Outcomes.* Educational achievement refers to the extent to which students master academic content. Eligible outcome measures of achievement include, but are not limited to, 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 outcome measures of attainment include, but are not limited to, grade promotion, high school graduation or dropout rates, certificate or degree completion rates, and other indicators for educational attainment. Eligible outcome measures of school attendance include, but are not limited to, absenteeism, attendance records, and tardiness.
- *Physical Development and Health Outcomes.* Refers to characteristics related to the healthy functioning of the body. Eligible outcome measures include, but are not limited to, measures of general physical health (e.g., parent-reported general health), diagnosed health conditions (e.g., asthma), physical capabilities (e.g., motor skills), and normative indicators of healthy development (e.g., early developmental milestones).
- *Substance Use or Misuse Outcomes.* Refers to the use or misuse of any substance. Eligible outcome measures may be self- or other-reported use/misuse, clinical tests such as urinalysis, or any other measure that provides an assessment of the participants' substance use behavior. Outcome measures must index 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), diagnostic assessments, and measures of substance use-related impairment (e.g., severity indices) are considered eligible outcome measures of substance use or misuse. 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, attitudes towards substance

use, etc.) are not eligible as measures of substance use or misuse but may meet the requirements for other outcomes.

  - *Delinquent Behavior Outcomes.* 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). Eligible outcome measures include, but are not limited to, self- or other-reported delinquent behavior and arrests, convictions, or incarcerations.

- **Adult Well-being Domain.** 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 a range of individual and interpersonal outcomes. Outcomes in the adult well-being domain may be assessed on parents, grandparents, kin or non-kin caregivers, and adults without children. The Prevention Services Clearinghouse reviews the following outcomes in the adult well-being domain:

  - *Parenting Practices Outcomes.* Includes 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. Eligible outcome measures may include items about basic elements of caregiving, such as feeding, physical care, and preventive healthcare; communication and listening; nurturing, loving, or supportive behavior; rules and consequences; setting boundaries; warmth; scaffolding children's behavior to develop self-discipline; parental practices associated with child attachment and promoting positive parent-child relationships, and the like. Measures of parenting knowledge or attitudes are not eligible outcome measures of parenting practices.

  - *Mental or Emotional Health Outcomes.* Refers to an adult's/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 outcome measures are eligible. Examples include measures of externalizing behaviors (e.g., aggressive behavior), internalizing behaviors (e.g., depression, anxiety, post-traumatic stress, mood or thought problems), mental/behavioral health diagnoses, parent/caregiver stress, relationship stress, positive behavior, resilience, and emotional adjustment.

  - *Substance Use or Misuse Outcomes.* Refers to the use or misuse of any substance. Eligible outcome measures 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. Outcome 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 diagnostic criteria), diagnostic assessments, and measures of substance use-related impairment (e.g., severity indices) are considered eligible measures of substance use or misuse. 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, attitudes towards substance use, etc.) are not eligible measures of substance use/misuse but may meet the requirements for other outcomes.

- *Criminal Behavior Outcomes.* Refers to behavior chargeable under applicable laws, whether or not apprehension occurs or charges are brought. Eligible outcome measures include, but are not limited to, self- or other-reported criminal behavior and arrests, convictions, or incarcerations.
- *Family Functioning Outcomes.* Family functioning refers to the social, structural, and functional properties of the family as an organized system and environment. Eligible outcome measures include assessments of the quality of relational patterns of interaction within the family, such as adaptability, acceptance, parental attunement, conflict, cohesion, communication, and the quality of the attachment relationship as well as affect and behavior regulation and the capacity to resolve problems to maintain effective functioning. Eligible outcome measures also include assessments of the quality, properties, or characteristics of relationships between individual family members (e.g., adult-child, child-child, or adult-adult), as well as assessments of the global family environment, family ecology, and family organization, including roles, leadership, and alliances.
- *Physical Health Outcomes.* Refers to the physical health of parents, caregivers, or other adults. Eligible outcome measures include, but are not limited to, self-reported general health, diagnosed health conditions (e.g., asthma, diabetes), and healthy lifestyle behaviors (e.g., exercise).
- *Economic and Housing Stability Outcomes.* Refers to financial, economic, or housing stability. Eligible outcome measures include, but are not limited to, level of income, employment/unemployment, financial assistance, food security/insecurity, and housing stability (e.g., number of moves, quality of housing, homelessness).

Additional Eligible Outcomes for Kinship Navigator Programs

All of the above-defined outcomes in the child safety, child permanency, and child and adult well-being domains are eligible for programs and services in the kinship navigator program area. Outcomes in three additional outcome domains are also eligible in the kinship navigator program area: access to services, referral to services, and satisfaction with programs and services.

- **Access to Services Outcome Domain.** Access to services outcomes refer to a parent's, caregiver's, 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. Eligible outcome measures include, but are not limited to, parent/caregiver self-reports, informed collateral reports (e.g., from therapists or case managers), or indicators of service access from administrative records.
- **Referral to Services Outcome Domain.** Referral to services outcomes include referrals to any needed financial, legal, social, educational, or health services. Eligible outcome 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 Outcome Domain.** Satisfaction with programs and services outcomes refer to parent or caregiver satisfaction with the programs and services to which they are referred or which they receive. Eligible outcome measures include, but are not limited to, satisfaction surveys or questionnaires.

4.1.9 Study Adaptations to the Program or Service Under Review

To be eligible for review, studies of a program or service must not indicate substantial differences from the program or service selected for review, as specified in the books, manuals, or other documentation (referred to here as the "manual") that describe how to implement or administer the program or service. The procedures for selecting the manual (or manuals) of the programs or services under review are described in Chapter 2.

Determining if Study Adaptations are Present

To determine if a particular study indicates substantial differences from the program or service under review as described in the focal manual, eligibility screeners must first determine if study adaptations are present. To do this, screeners record any adaptations directly described in the study, any alternative manual editions or variants cited, and whether the study pre-dates the publication of the focal manual for the program or service under review. In each of these circumstances, screeners then attempt to document what (if any) specific adaptations have been made relative to the

focal manual. If insufficient information is available to determine whether the intervention was adapted, screeners may (1) query authors for information about the manual used or intervention delivered, (2) review external materials (e.g., manual cited, history of program changes over time in a journal article), or (3) seek external expert consultation. Experts who advise the Prevention Services Clearinghouse on study adaptations include, but are not limited to, individuals with lived expertise; individuals from communities that programs and services are intended to serve, inclusive of underserved and tribal communities; methodological experts; state, tribal, and local child welfare department staff or administrators; policymakers; program or service developers; program or service providers or trainers; study authors; and subject matter experts. The information gathering process includes any prior work done at the program or service prioritization and selection step to understand differences among other manual editions or manual variants identified and whether they describe substantial adaptations of the program or service selected for review or not (see Section 2.3). At the conclusion of this process, all adaptations present are documented for each study.[23]

Reviewers record adaptations in the same four key program or service component domains as described in Chapter 2 – dosage, modality, content, and providers – and use the same criteria for assessing whether the *study* adaptations are substantial or not substantial as they do for program or service adaptations (see Exhibit 2.4).

Assessing the Eligibility of Studies with Adaptations

If a study exhibits adaptations relative to the focal manual for the program or service under review, a systematic stepwise process (illustrated in Exhibit 4.1 and described in detail below) is then used to determine whether the study is an eligible study of the program or service or is ineligible (i.e., not a study of the program or service under review). Studies with substantial adaptations are ineligible for review for the program or service under review.

Step 1: Is the study adaptation (a) explicitly prohibited in the focal manual for the program or service under review or (b) the result of adding a separate program or service to the program or service under review? Studies that include adaptations explicitly prohibited in the focal manual for the program or service under review are not eligible to be reviewed as a study of the program or service under review. For example,

[23] Any manual editions or manual variants identified during this process are recorded and noted for program or service identification (Chapter 1) and program or service eligibility and prioritization (Chapter 2) purposes. Additional manual editions or variants identified during study eligibility screening are handled using the procedures specified in Chapter 2 for manual editions and variants. Similarly, any additional information about other manuals already identified from program or service identification, eligibility, and prioritization processes obtained during the study eligibility determination process is incorporated into the procedures for how the Prevention Services Clearinghouse handles programs and services with more than one manual specified in Section 2.3.

if the manual for a program targeting parental depression explicitly states that the program should not be used in cases where a parent has an active substance use disorder, a study adapting the program to target parents with substance use disorders would not be eligible for review as a study of that program or service.

Additionally, studies with adaptations that involve adding a separate program or service to the existing program or service (i.e., "bundling") are also not eligible for review as a study of the program or service under review. For example, a study that adds a manualized substance use treatment program to an existing parenting program would not be eligible to be reviewed as a study of the original parenting program.

Step 2: Is the study adaptation explicitly allowed by the focal manual for the program or service under review? Studies with adaptations explicitly allowed in the focal program or service manual are considered eligible studies of the program or service. For example, consider a program or service manual that indicates that the program typically is conducted in 12 group sessions of 3 to 5 parents, but that practitioners have the flexibility to conduct some or all sessions individually if needed due to parental availability, with alternative protocols for conducting individual sessions provided in the manual. A study that indicated all 12 sessions were delivered individually would be considered to be an eligible study of the program or service because this modification is explicitly permitted in the manual.

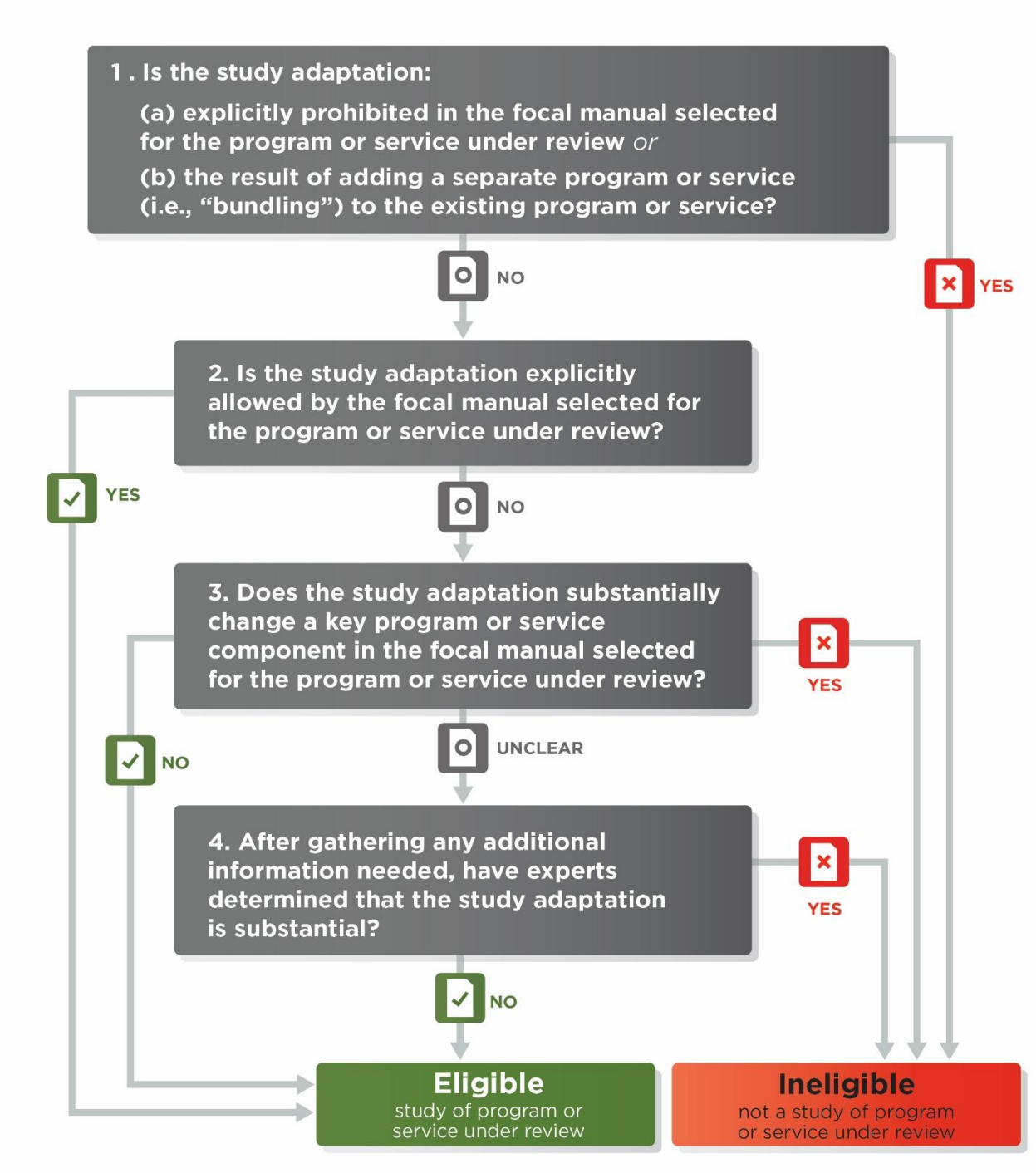
Step 3: Does the study adaptation substantially change a key program or service component in the focal manual for the program or service under review? When adaptations are not explicitly prohibited or allowed in the focal manual for the program or service under review, the Prevention Services Clearinghouse assesses whether the study adaptations substantially change at least one key component of the program or service under review (e.g., dosage, modality, content, or providers). Studies with substantial adaptations to key program or service components are ineligible to be reviewed as studies of the program or service under review, whereas studies with adaptations that are not substantial can be eligible (provided they meet all other study eligibility requirements).[24] Examples of adaptations that are considered to be substantial and that are not considered to be substantial can be found in Section 2.3 (Exhibit 2.5).

Step 4. After gathering any additional information needed, have experts determined that the study adaptation is substantial? If reviewers cannot determine if an adaptation in a study is clearly substantial or not substantial after the first three steps, the Prevention Services Clearinghouse may also query study authors, program and service developers, or other outside experts to request information necessary to better

[24] Adaptations to training processes, implementation fidelity tools, or evaluation tools observed in studies are generally not considered to be substantial adaptations.

understand the study adaptations and whether they are substantial or not substantial. Experts who advise the Prevention Services Clearinghouse on study adaptations include, but are not limited to, individuals with lived expertise; individuals from communities that programs and services are intended to serve, inclusive of underserved and tribal communities; methodological experts; state, tribal, and local child welfare department staff or administrators; policymakers; program or service developers; program or service providers or trainers; study authors; and subject matter experts. Senior content experts on the Prevention Services Clearinghouse staff will be consulted to develop a final decision on whether a particular adaptation is substantial or not substantial.

Exhibit 4.1. Process for Determining Whether Studies with Program or Service Adaptations Are Eligible or Ineligible Studies of the Program or Service Under Review



4.2 Study Review Prioritization Criteria

The Prevention Services Clearinghouse reviews all eligible studies of a program or service prioritized for review.

- If a program or service has 15 or fewer 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 7.2.1.
- If a program or service has more than 15 eligible studies, all eligible studies are assessed for risk of harm. Study review prioritization criteria (see below) are used to determine the order in which studies are reviewed using the design and execution standards. Once ordered, the first 15 eligible studies are 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 are reviewed using the design and execution standards in prioritization order, only if there is potential for the program or service rating to improve.
  - Determination of potential for a program or service rating to improve is based on (1) the rating that would result from studies already reviewed using the design and execution standards and (2) the duration of effects examined in the remaining studies (as assessed for study prioritization). For example, a program or service that has achieved a promising rating after review of the first 15 studies that has no remaining studies reporting impacts measured 6 or more months after the end of the intervention cannot achieve a supported rating. In this case, no additional eligible studies would be reviewed using the design and execution standards (though all studies are still assessed for risk of harm).

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

- ***Design.*** 3 points for randomized controlled designs (RCTs), 2 points for quasi-experimental designs (QEDs).
- ***Statistical Power.*** 1 point for reporting of an analysis of statistical power, indicating the magnitude of effect that the study anticipates being able to detect at a given level of statistical significance with the targeted (or actual) sample size.
- ***Duration of Effects Examined.*** 6 points for any effects measured at 12 months or more after the end of the intervention; 3 points for any effects measured

between 6 and 12 months after the end of the intervention; 0 points for effects only measured less than 6 months after the end of the intervention.

- ***Child Welfare Relevance.*** 1 point for studies whose samples primarily consist of children, youth, young adults, and/or families receiving child welfare services (or populations similar to those receiving child welfare services or at-risk for receiving child welfare services).
- ***Population(s) Served.*** 2 points for studies whose samples are from underserved communities.
- ***Multiple Outcome Domains Examined.*** 1 point if more than one outcome domain is examined in the study.
- ***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 17 points). Studies are then sorted by the summed point total and reviewed in that order.

If there are more than 15 eligible studies and there is a tie in the prioritization point total at the cutoff point total identified for selecting the first 15 studies to review, the following two steps are used to break the tie and select which studies with the same prioritization point total are included among the first 15 studies reviewed:

Step 1. Studies reporting on sustained effects of 12 months or more are identified, if applicable. If there are no studies meeting this criterion, proceed to step 2.

- If including all studies reporting on sustained effects of 12 months or more would result in more than 15 studies to review first, the number of studies from this group needed to reach a total of 15 studies is randomly selected for inclusion in the first 15 studies rated.
- If including all studies reporting on sustained effects of 12 months or more would result in exactly 15 studies to review first, all such studies are included in the first 15 studies reviewed.
- If including all studies reporting on sustained effects of 12 months or more would result in less than 15 studies to review first, all such studies are included, and the remaining studies needed are selected in step 2.

Step 2. No studies with sustained effects of 12 months or more are identified. The number of studies needed to reach a total of 15 studies from those remaining at the tied prioritization score is randomly selected, taking into account any studies already selected in Step 1.

If the program or service rating has the potential to improve after rating 15 studies and there is a tie in the prioritization scores for the additional studies, all studies with the tied prioritization score would be reviewed, provided they have the potential to improve the program or service rating.

5. Evidence Review for RCTs and QEDs Using the Design and Execution Standards

This chapter describes the standards that are used to assign design and execution ratings to RCT and QED studies reviewed by the Prevention Services Clearinghouse.[25] All eligible studies are assessed for risk of harm; study prioritization procedures are used to select which studies are reviewed using the design and execution standards when there are more than 15 eligible studies (see Section 4.2). The chapter presents the review process as a sequence of steps to arrive at a design and execution rating for RCT and QED studies, as depicted in a flow chart (Exhibit 5.2 below). Definitions of terms are provided in boxes in this chapter as well as in the Glossary in the back of the Handbook of Standards and Procedures.

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

Planned Pilot for Single Case Designs (SCDs)
The Prevention Services Clearinghouse plans to develop and pilot design and execution standards for studies with SCDs.

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 an eligible intervention condition to an eligible comparison condition on a specific outcome for a specific posttest measurement. For example, a study with one intervention condition and one comparison condition that reports findings on one outcome measured immediately after treatment has a single contrast. A study with one intervention condition and one comparison condition that reports findings on two outcomes measured immediately after treatment would have two contrasts, one for each of the comparisons between the intervention and comparison conditions on the two outcomes. A study with one intervention condition and one comparison condition that reports findings on one outcome measured at 3, 6, and 12 months after treatment would have three contrasts, one for each outcome measurement period. Contrasts are reviewed from eligible RCTs or QEDs.

Most studies report results on more than one outcome, and some studies have more than two conditions (e.g., more than one intervention condition and/or more than one comparison condition). When studies report results on more than one outcome or compare two or more different intervention conditions to a comparison condition, the

[25] Findings from the planned pilot for SCDs will inform the development of design and execution standards for studies with SCDs and will be reflected in an updated version of the Handbook.

study is reporting results for multiple *contrasts*. In studies with more than one eligible comparison condition, reviewers will review contrasts from all eligible comparison conditions against the eligible intervention condition(s) in the study. Prevention Services Clearinghouse design and execution ratings can differ across the contrasts reported in a study. Consequently, a single study may have multiple design and execution ratings corresponding to each of its reported contrasts.

Prevention Services Clearinghouse ratings are applied to benchmark full-sample analyses, not full-sample sensitivity analyses or subgroup analyses. However, the Prevention Services Clearinghouse will report on the website whether certain subgroup analyses are present in studies where the full-sample analysis receives a high or moderate design and execution rating. The Prevention Services Clearinghouse plans to conduct a pilot to understand the resource requirements and potential implications of reviewing subgroup analyses to contribute to program or service ratings.

The design and execution ratings from all reviewed contrasts in all eligible studies are used to inform the program or service rating, in combination with characterization of impact estimates from reviewed contrasts as favorable, no effect, or unfavorable; assessment of whether favorable effects are sustained beyond the end of treatment; and a risk of harm assessment. Procedures for characterization of impact estimates are described in Chapter 6. Program or service rating criteria and procedures for assessing time beyond the end of treatment and risk of harm are described in Chapter 7. The current chapter is focused on the procedures for rating an RCT or QED contrast against the design and execution standards.

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 one of the following ratings for each contrast, shown in order from strongest to weakest evidence:

- Meets Prevention Services Clearinghouse Standards for **High** Support of Causal Evidence (high rating)
- Meets Prevention Services Clearinghouse Standards for **Moderate** Support of Causal Evidence (moderate rating)
- Meets Prevention Services Clearinghouse Standards for **Low** Support of Causal Evidence (low rating)

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 have

contrasts with 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 pretest 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.[26]

Exhibit 5.1 presents a summary of the designs that are currently eligible to receive high and moderate ratings.[27] 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 controlled trial 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) • 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 controlled trial studies that fail standards for integrity of random assignment (Section 5.4), joiner bias (Section 5.5), or attrition (Section 5.6) and quasi-experimental design 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 Low Support of Causal Evidence	
Contrasts that are reviewed and fail to meet standards for a high or moderate rating	

5.2.1 The Review Process Differs for RCTs versus QEDs

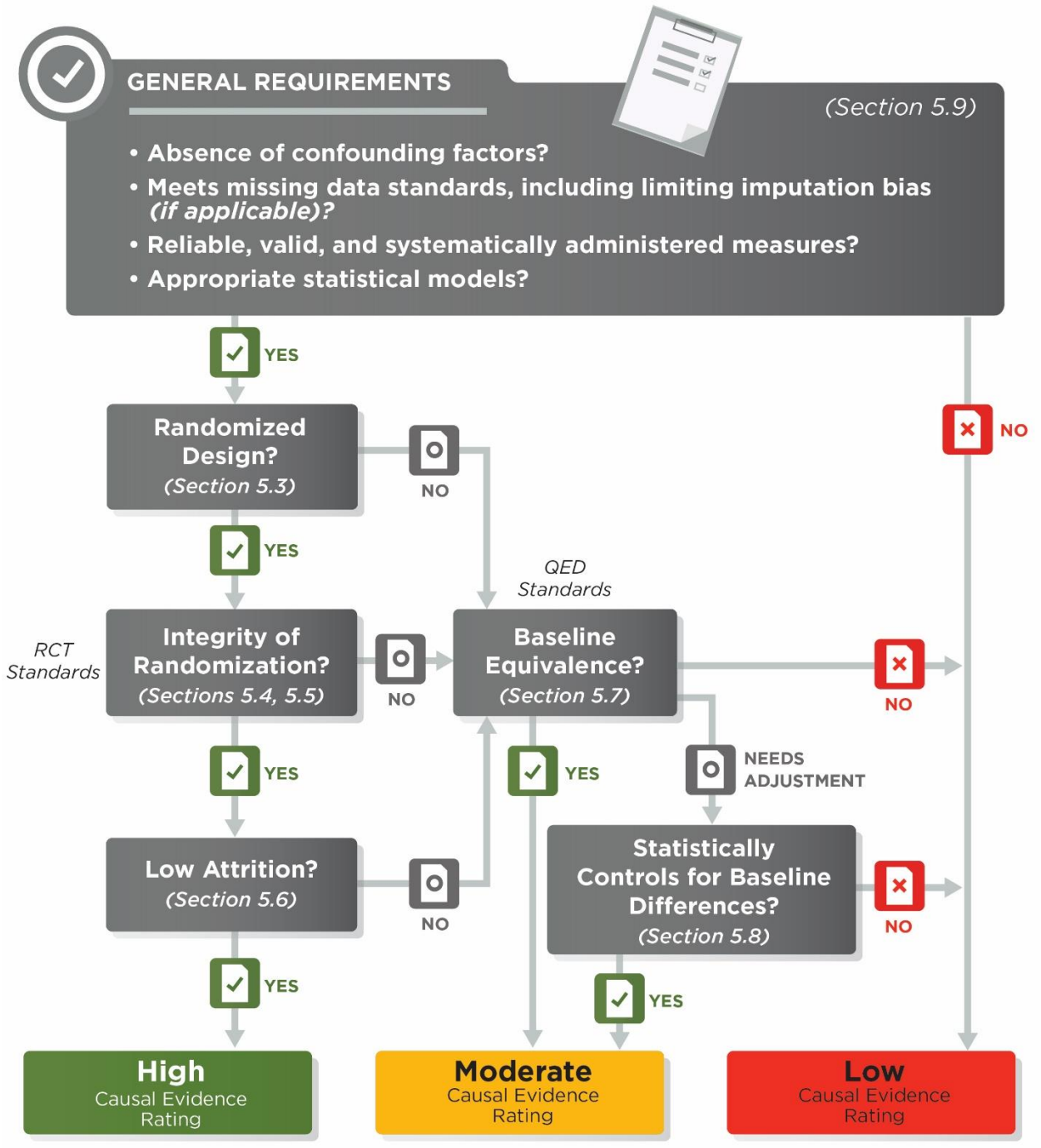
Exhibit 5.2 below provides a summary of the steps of the study review process for RCTs and QEDs. When Prevention Services Clearinghouse reviewers rate the evidence produced from a study contrast, they begin by making a determination about the type of design used to create the contrast. Once that determination is made, they follow the

[26] For brevity, the Prevention Services Clearinghouse website will display a study-level design and execution rating based on the highest-rated contrast in the study, but contrast-level design and execution ratings are what inform program or service ratings (see Chapter 7). In this example, the study would be displayed as having a moderate rating on the basis of having at least one contrast with a moderate rating and none with a high rating.

[27] The Prevention Services Clearinghouse is planning to develop pilot design and execution standards for single case designs. The design and execution criteria in this chapter apply to RCT and QED designs.

sequence of steps in the flow chart depicted in Exhibit 5.2. The various decisions and standards that apply are described in the accompanying text.

Exhibit 5.2. Contrast Rating Flowchart for RCTs and QEDs



Note. RCT refers to a randomized controlled trial; QED refers to a quasi-experimental design.

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) or a quasi-experimental design (QED), as defined in Section 4.1.5.

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

Quasi-experimental Design (QED): A design in which an intervention condition is compared to a comparison condition, but the units are not randomly assigned.

- If assignment to conditions is based on a random process, reviewers assess the integrity of the randomization and, if applicable, 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 using the process for 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 conditions was made by a random process, the researcher identifies individuals who were randomly assigned to the intervention condition but who refused to participate in the intervention. The researcher reclassifies those individuals as belonging to the comparison condition for analysis. In this example, the randomization has been compromised.

Example 2: In a multi-site study, individuals are randomly assigned to intervention and comparison conditions within 20 sites. One of the sites would not allow randomization, so assignment to intervention and comparison conditions was done by a method other than randomization. Data from all 20 sites are included in a single analysis. The site with the non-random assignment has compromised the random assignment for the whole study.

Example 3: In a study where initial assignment to intervention and comparison conditions was made by a random process, many of the individuals who were assigned

to the intervention condition are refusing treatment. To fill the empty treatment slots, the service provider identifies additional individuals who meet the study eligibility criteria and assigns all of them to the intervention condition to ensure a full sample of participants. The analysis includes both individuals originally assigned to the intervention condition via randomization and the additional intervention members added later through a non-random process. In this example, the randomization has been compromised.

***Example 4*:** In a study where initial assignment to intervention and comparison conditions was made by a random process, many of the individuals who were assigned to the intervention condition are refusing treatment. To fill the empty treatment slots, the service provider recruits some of the comparison condition members to participate in the intervention. In the analysis, the researcher includes those comparison condition members who received the intervention condition as belonging to the intervention condition. In this example, the randomization has been compromised.

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 conditions was made by a random process, many of the individuals who were assigned to the intervention condition are refusing treatment. To fill the empty treatment slots, the service provider recruits additional individuals for the intervention condition who were not assigned to either the intervention or comparison condition in the original randomization. In the analysis, the researcher maintains the original intervention and comparison condition assignments and excludes the subsequently recruited individuals from the impact analyses. In this example, the randomization has not been compromised.

Example 6: After randomizing individuals to intervention and comparison conditions, the researchers discovered that some individuals did not meet the symptom threshold for study eligibility measured using baseline data that were collected prior to initiating the intervention. Sample members with symptoms below the eligibility threshold are excluded from analyses. This exclusion is applied to both the intervention and comparison condition members. In the example, the 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 randomized studies 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 comparison 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 also 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 to 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 condition 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). 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.

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.

Some contrasts may be created by randomly assigning families to conditions and then evaluating program or service 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.

For cluster RCTs, the Prevention Services Clearinghouse assesses the risk of joiner bias based on the potential for the intervention to affect joining the cluster, such as

when individuals are placed into clusters after the person making the placement knows the intervention assignment status of the clusters.

- 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, then the reviewers would proceed with the review following the process for quasi-experimental designs and assess baseline equivalence (see Section 5.7).

5.5.1 Examples of High Risk of Joiner Bias

Example 7: In a study, mental health clinics in a network are randomly assigned to an intervention or treatment as usual. As families join the study, the network administrator places them in clinics after learning which clinics are in the intervention condition. If the network administrator takes families' characteristics into account when making placements, then the composition of intervention and comparison conditions may change in ways that favor one group or the other. In this case, the Prevention Services Clearinghouse assumes a high risk of joiner bias exists.

Example 8: In a study, preschools are randomly assigned to offer a parenting intervention or a comparison parenting intervention after families have been enrolled for the academic year. In the second year of the study, the preschools enroll a new group of families, including the parenting programs being offered in their marketing materials. In this case, the families enrolling in preschools after the random assignment in the second study year may choose (or avoid) preschools because of the parenting interventions being offered. In that case, the composition of intervention and comparison conditions may change in ways that favor one group or the other. In this case, if the joiners who enrolled in the second study year are included in the analysis of outcomes, the Prevention Services Clearinghouse assumes a high risk of joiner bias exists.

5.5.2 Examples of Low Risk of Joiner Bias

Example 9: In the same study as Example 7, joiner bias is unlikely to influence study results if all families were placed by the network administrator after clinic assignment was made but prior to knowing which clinics were offering the intervention condition. In this case, reviewers would assume a low risk of joiner bias.

Example 10: In a study, preschools are randomly assigned to offer a parenting intervention or a control parenting intervention after enrollment for the year is finalized. Preschool staff and providers are blinded on which condition they are assigned. The parenting intervention being offered is not publicly advertised. The study reports that

some families in both intervention and comparison clinics at the time of assignment dropped out and were replaced by families on waiting lists in the order in which their applications were received for the preschools prior to the intervention starting. In these cases, reviewers would assume a low risk of joiner bias.

5.6 Attrition Standards

Attrition occurs when individuals or clusters leave the study sample or do not provide data on some or all of the outcomes. In RCTs, attrition can reduce the credibility of the evidence. When the characteristics of the individuals or clusters who attrit 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 WWC[28], 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 WWC's 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 Prevention Services 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 conditions. 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 below shows the numeric values for the Prevention Services Clearinghouse attrition boundaries.

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

[28] The selection of the cautious boundary is consistent with other clearinghouses that focus on similar populations (e.g., HomVEE). 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.

cluster level. If attrition is determined to be below the boundaries shown in Exhibit 5.3, the contrast is assessed as having low attrition. Reviewers would then proceed with evaluating the contrast against the other design and execution standards described in Section 5.9. If attrition is at or above the boundary, the contrast is assessed as having high attrition. High attrition contrasts must then be assessed for baseline equivalence (Section 5.7).

Exhibit 5.3. Potential Bias Associated with Overall and Differential Attrition

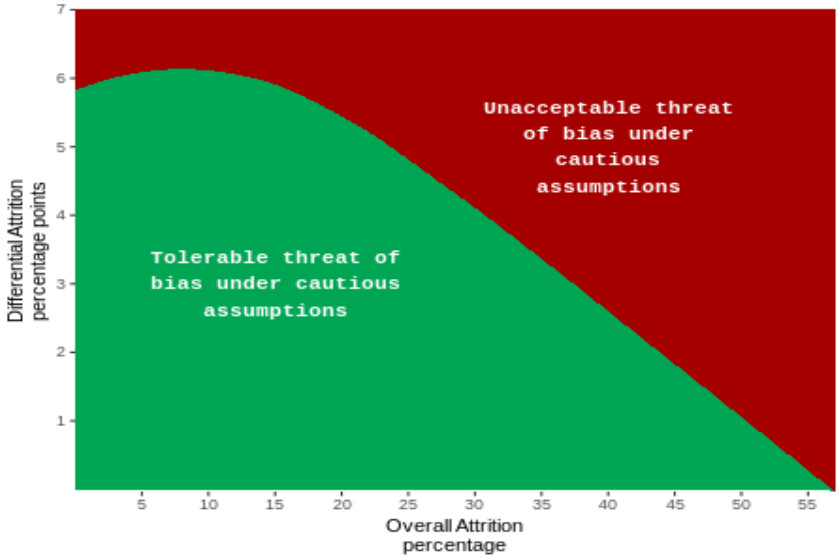


Exhibit 5.4. Prevention Services Clearinghouse Attrition Boundaries

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

Overall Attrition	Differential Attrition	Overall Attrition	Differential Attrition	Overall Attrition	Differential Attrition
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 point differences. Attrition computations are rounded to whole numbers for determining overall attrition and to the nearest hundredth for differential attrition. For example, an overall attrition rate of 15.4% and differential attrition rate of 5.894pp would be rounded to 15% and 5.89pp, respectively. This contrast would be evaluated as low attrition because 5.89pp falls below the boundary of 5.9pp.

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 in the specific intervention condition and comparison condition for the contrast without posttest outcome data as a percentage of the total number of members in the sample for the contrast at the time that they learned the condition to which they were randomly assigned, specifically:

$$Overall\ Attrition = \frac{N\ of\ individuals\ in\ without\ posttest\ outcome\ data}{N\ of\ individuals\ randomized}$$

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

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

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[29] have experienced it.

When eligibility and consent (if needed) are 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 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. If the study applies an exclusionary condition using a measure that may have been affected by assignment condition, this compromises the integrity of randomization. In such cases, attrition is not assessed and affected contrasts are reviewed using the quasi-experimental design standards. If the study (1) used the same exclusionary conditions in both the intervention and the comparison conditions and (2) the exclusionary conditions are based on information collected prior to individuals learning of their assignment condition or measures that would not be affected by assignment condition, then eligibility criteria can be applied after that time point and ineligible individuals can be excluded from the attrition calculations and from the analysis. Otherwise, exclusions that do not meet these criteria and do not compromise randomization are counted as attrition.

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 and was based on information collected prior to assignment to the intervention and comparison conditions, the excluded children do not need to be counted for the purpose of the attrition calculation.

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

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 condition and a similar child who had been randomized to the comparison condition would not have been identified and removed. If the researchers continue to identify the child as belonging to the intervention condition 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 condition. 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 condition" 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 she 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.

Example 6: A study assigns individuals to a 10-session intervention condition or a 10-session attention control comparison condition. The authors exclude any individuals who did not complete at least 2 sessions from study analyses. Though the exclusion condition is applied in the same way to both the intervention and comparison condition,

session completion is measured after individuals learn their assignment condition and could have been affected by assignment status. This exclusion compromises randomization. Therefore, attrition would not be assessed and the contrast is reviewed using the QED standards.

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 within the comparison are calculated as:

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

$$Differential\ Attrition = \left| \left(\frac{N\ of\ intervention\ condition\ members\ without\ posttest\ outcome\ data}{N\ of\ intervention\ condition\ members\ in\ nonattrited\ clusters\ at\ randomization} \right) - \left(\frac{N\ of\ comparison\ condition\ members\ without\ posttest\ outcome\ data}{N\ of\ comparison\ condition\ members\ in\ nonattrited\ clusters\ at\ randomization} \right) \right|$$

5.7 Baseline Equivalence Standards

All contrasts that are QEDs or RCTs that have compromised randomization (including high risk of joiner bias) or high attrition are assessed for baseline equivalence. Low-attrition RCTs that maintain integrity of randomization are not assessed for baseline equivalence.[30]

5.7.1 Selecting a Measure for the Baseline Equivalence Assessment

To assess baseline equivalence, reviewers must first identify a measure to use for conducting the baseline equivalence assessment. Reviewers identify all acceptable baseline measures available for establishing equivalence for the contrast under review and select the preferred baseline measure from among available possibilities.

[30] Random assignment and low attrition result in contrasts that have no expected outcome differences between the intervention and comparison conditions except those caused by the program or service.

General Requirements for Baseline Equivalence Measures

Three general requirements apply for measures to be acceptable for establishing baseline equivalence:

- Measures 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, percentage of youth with any previous arrest cannot be used to establish baseline equivalence because there is no variability on that variable.
- Measures must meet all measurement standard requirements – i.e., reliability, face validity, and consistency in measurement (see Section 5.9.2).
- Measures must occur prior to introduction of the intervention or be as close to the beginning of the intervention as possible. In analyses that include multiple periods of preintervention data, the Prevention Services Clearinghouse will use the preintervention time point closest to the point where individuals were assigned to conditions (or, alternatively, the closest point prior to the introduction of the intervention) to assess baseline equivalence.

Types of Measures for Establishing Baseline Equivalence

The Prevention Services Clearinghouse permits four possible types of measures for establishing baseline equivalence:

1. **Direct pretest**. Defined as the same (or nearly the same) measure used for the outcome.
2. **Correlated pretest**. Defined as a measure in any eligible outcome domain (Section 4.1.8) that has a correlation of 0.60 or higher with the outcome in the analytic sample. (A correlation shown in the comparison condition only is also acceptable). Correlated pretests do not have to be in the same or similar domain as the outcome.
3. **Pretest alternative**. Defined as a measure in the same or similar domain as the outcome. No correlation threshold is specified – pretest alternatives are generally assumed to be correlated with the outcome by virtue of being conceptually related or are common precursors to the outcome.[31] Conditions under which pretest alternatives are acceptable are specified below.

[31] If a study reports a correlation between a pretest measure in the same or similar domain and the outcome that is greater than or equal to 0.6, the measure would also qualify as a correlated pretest. If the study reports a correlation that is less than 0.6, the measure can still be used as a pretest alternative.

4. **Sociodemographic Characteristics.** Under certain conditions specified below, combinations of individual or individual and neighborhood sociodemographic characteristics may be used to establish baseline equivalence. Eligible individual and neighborhood characteristics are summarized below and defined in detail in Exhibit 5.5.

    a. **Individual Sociodemographic Characteristics**. Eligible measures are: (1) Race or ethnicity, (2) Socioeconomic status, (3) Household composition, and (4) Age of sample members.

    b. **Neighborhood Sociodemographic Characteristics.** Neighborhood is defined as a census tract, ZIP Code, or smaller geographic unit (or similarly sized tabulation unit for studies conducted outside of the United States). Eligible measures are: (1) Race or ethnicity, (2) Socioeconomic status, and (3) Household composition.

Exhibit 5.5. Eligible Individual and Neighborhood Sociodemographic Characteristics and Requirements

Characteristic	Eligible Measure(s) and Requirements
Individual Sociodemographic Characteristics	
Race or Ethnicity	May use race or ethnicity of either adults or children in the analytic sample. For studies outside of the U.S., reviewers identify variables appropriate to the cultural or national context of the study, in consultation with senior content experts.
Socioeconomic Status	Ordered preference for income, earnings, federal poverty level (U.S. studies) or applicable national poverty level (studies outside the U.S.), receipt of means-tested public assistance (e.g., Aid to Families with Dependent Children (AFDC)/Temporary Assistance for Needy Families (TANF), food stamps/Supplemental Nutrition Assistance Program (SNAP), Free and Reduced-Price Meal Program status), maternal education, employment of a member of the household, or other similar measures.
Household Composition	Household composition measures include, but are not limited to, total household size, number of children in household, whether children live with a custodial parent, or proportion of one-parent/caregiver headed households.
Age of Sample Members	Age of individuals in the analytic sample (children or adults, as applicable).

Characteristic	Eligible Measure(s) and Requirements
Neighborhood Sociodemographic Characteristics	
Race or Ethnicity	Share of households in different racial or ethnic groups. For studies outside of the U.S., reviewers identify variables appropriate to the cultural or national context of the study, in consultation with senior content experts.
Socioeconomic Status	Ordered preference for average neighborhood income (mean or median), neighborhood poverty rate, neighborhood educational attainment (e.g., share of households with a high school degree or less than a high school degree), or other similar measures. Measures can be based on number of households, families, or individuals in the neighborhood.
Household Composition	Household composition measures include, but are not limited to, share of households with children with a single parent/caregiver, average household size, or average number of children in a household.

Procedures for Selecting Measure for Baseline Equivalence Assessment

Once Prevention Services Clearinghouse reviewers have identified the acceptable baseline measures, they follow an ordered procedure for selecting a measure for the baseline equivalence assessment, as summarized in Exhibit 5.6 below.

Exhibit 5.6. Options for Establishing Baseline Equivalence and Conditions Required for Use, In Order of Preference

Option	Definition and Requirements
If Direct Pretest is Feasible to Measure	
1. Direct Pretest	• Defined as the same (or nearly the same) measure used for the outcome. • If the outcome is binary, reviewers have discretion to select between a direct pretest and correlated pretest (if both are available).
2. Correlated Pretest	• Defined as a measure in any eligible outcome domain (Section 4.1.8) that has a correlation of 0.60 or higher with the outcome in the analytic sample.
Only If Direct Pretest is Not Feasible to Measure or Outcome is Binary	
3. Pretest Alternative	• Defined as a measure in the same or similar domain as the outcome. No correlation threshold.
Only If Direct Pretest is Not Feasible to Measure *and* No Pretest Alternative Available	
2. Sociodemographic Characteristics	• Measures of at least two eligible individual characteristics (preferred) *or* one eligible individual characteristic and all three eligible neighborhood characteristics (see Exhibit 5.5 above). • Requires affirmative evidence that the intervention and comparison conditions overlap substantially on characteristics that determine intervention eligibility or study recruitment criteria.

1. **Direct Pretest.** If a direct pretest is available, it is selected as the pretest measure for continuous and count outcomes. For binary outcomes, reviewers have discretion to select between a direct pretest and a correlated pretest if both are available (see considerations for binary outcomes below).

2. **Correlated Pretest.** If a direct pretest is not available or acceptable, but a correlated pretest is available and acceptable, it is selected as the baseline measure. If multiple correlated pretests are available and acceptable, reviewers typically select the most conceptually related measure. Senior content experts are available to reviewers to assist when selecting among multiple correlated pretests is required.

   If no direct pretests or correlated pretests are available and acceptable, reviewers answer two questions before proceeding: (1) Was it *feasible* for a direct pretest to have been measured at baseline? and (2) Is the outcome binary?

   - **Was it feasible to measure a direct pretest at baseline?** For some outcomes, a direct pretest 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 pretest with younger children). In such cases, use of a pretest alternative is preferred but sociodemographic characteristics are permitted if they meet the requirements in Exhibit 5.5. If a direct pretest was feasible to measure at baseline but was not measured, use of a pretest alternative or sociodemographic characteristics to establish baseline equivalence is not permitted (with an exception for binary and time-to-event outcome measures, described below).
   - **Is the outcome binary?** The Prevention Services Clearinghouse applies additional flexibility for establishing baseline equivalence when outcomes are *binary* (as defined in Chapter 6). Binary measures have different statistical properties than continuous measures that can make continuous correlated pretests or pretest alternatives preferred over direct pretests for establishing equivalence under certain conditions.[32] Specifically, the Prevention Services Clearinghouse permits the use of *continuous* pretest alternative measures when outcomes are binary, even if it was feasible to measure a direct pretest. (For similar reasons, reviewers also have

[32] When events are rare or sample sizes are smaller, effect sizes for binary measures are very sensitive to a difference of one or a few individuals and may thus be more affected by chance variation than continuous measures. This can lead to misleading conclusions about baseline equivalence. Thus, to enhance baseline equivalence conclusions, the Prevention Services Clearinghouse provides greater flexibility to make use of continuous pretests or pretest alternatives for binary outcomes (when available).

discretion to select between a correlated pretest in lieu of a direct pretest for binary outcomes, if one is available.) All decisions are documented, and senior content experts are available to answer questions.

3. **Pretest Alternative.** If multiple acceptable pretest 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 pretest alternative is documented in the review and confirmed with senior content experts on the Prevention Services Clearinghouse.

4. **Sociodemographic Characteristics.** If a direct pretest is not feasible and no pretest alternative is available, multiple sociodemographic characteristics may be used to establish baseline equivalence, subject to an additional requirement regarding group formation.[33]

   – **Requires Demonstration of Similar Group Formation.** To use sociodemographic characteristics for establishing baseline equivalence, the study must include an affirmative description demonstrating that the intervention and comparison conditions substantially overlap on characteristics that determine intervention eligibility or on study recruitment criteria. If an affirmative description is not present or is incomplete, reviewers will query authors for the information needed.

- **4a. Individual Sociodemographic Characteristics Only.** If only individual sociodemographics are used, baseline equivalence must be established by showing equivalence on two of the four eligible individual sociodemographic characteristics (see Exhibit 5.5 above).

- **4b. Individual and Neighborhood Sociodemographic Characteristics.** In some cases, only a single eligible individual sociodemographic characteristic is available, but information is available identifying neighborhoods in which study participants live. For purposes of baseline equivalence, neighborhood characteristics are analyzed as a characteristic of each sample member.

  – In such cases, baseline equivalence must be established using a combination of a single individual sociodemographic characteristic and measures of all three eligible neighborhood characteristics (see Exhibit 5.5

[33] Requirements for demonstrating equivalence on multiple sociodemographic characteristics reflects considerations on the extent to which such characteristics are correlated individually and collectively with outcomes of interest and how correlations may vary by the unit of analysis (i.e., individual versus neighborhood). The requirements also reflect that failure to meet equivalence on any of these measures would be concerning with respect to potential bias in impact estimates.

above). Neighborhood characteristics would typically be obtained from national statistical agencies.

5.7.2 Assessing Baseline Equivalence

Baseline equivalence must be demonstrated on the *analytic sample* for each contrast reviewed (i.e., individuals included in the analysis that provides impact estimates for the contrast). Baseline equivalence is assessed by examining baseline differences expressed in effect size (ES) units (computed using the appropriate formulae for effect sizes specified in Chapter 6). In instances where authors apply weights to study data for analysis (for example, propensity score weights from a matching process), baseline equivalence must be assessed using the same weights used to measure impacts.

The Prevention Services Clearinghouse uses the following thresholds for baseline equivalence to determine whether a contrast meets the baseline equivalence standard, as illustrated in Exhibit 5.2 above:

- Baseline effect sizes less than 0.05 are considered equivalent and no further covariate adjustments are required.[34]
- Baseline effect sizes between 0.05 and 0.25 indicate that statistical adjustments in the impact analysis models are required (see Section 5.8). These baseline effect sizes are said to be in the *adjustment range* and impact estimates must control for the variables that are out of balance at baseline. When statistical adjustments are required, the Prevention Services Clearinghouse standards for acceptable adjustment models described in Section 5.8 below are applied.
- Baseline effect sizes greater than 0.25 are considered non-equivalent, and the contrast receives a low rating.

When sociodemographic characteristics (individual only or individual and neighborhood) are used for establishing baseline equivalence, *all* of the applicable sociodemographic measures used must meet these baseline equivalence thresholds and applicable standards. That is, if any characteristic used for establishing baseline equivalence has a baseline effect size greater than 0.25, baseline equivalence is not established, and the contrast receives a low rating. Similarly, all sociodemographic measures in the adjustment range must have acceptable statistical adjustments in the impact model.

[34] 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.

If direct pretests are not possible or feasible and no eligible correlated pretests, pretest alternatives, or eligible sociodemographic characteristics are available, baseline equivalence is not established and the contrast receives a low rating.

Procedures for Assessing Baseline Equivalence When Information Is Not Reported on the Baseline Measure for the Sample Analyzed

If baseline equivalence information is not reported for the exact analytic sample or is not reported at all, the Prevention Services Clearinghouse will send an author query requesting baseline equivalence information on the exact analytic sample. For example, consider a high attrition RCT where study authors report baseline information for the full randomized sample but only include respondents who provided outcome data at posttest in their analyses. In this case, baseline information is not reported for the analytic sample and an author query would be sent.

If the baseline equivalence information is reported for some sample other than the exact analytic sample and is not provided in response to an author query, the Prevention Services Clearinghouse may attempt to assess whether the contrast would meet baseline equivalence standards based on an estimated largest (or worst-case) baseline difference. The Prevention Services Clearinghouse standards for estimating the largest baseline difference are based on those used by the WWC (see Section 5.9.4). If information required to estimate the largest baseline difference is available, baseline equivalence is assessed against the specified effect size thresholds using the largest baseline difference effect size computed according to procedures used by the WWC (What Works Clearinghouse, 2022). Baseline equivalence is then assessed against the specified effect size thresholds using the worst-case baseline effect size.

5.8 Acceptable Methods for Controlling for Pretests

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

- Regression models with the baseline measure(s) 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 pretest.
- Repeated measures analysis of variance models.

- Difference-in-differences models (these must use direct pretests or correlated pretests, not other baseline measures). Reviewers may use reported unadjusted statistics (e.g., means or differences from pre- and posttests) to calculate a difference-in-differences estimate to satisfy this requirement.
- Models with fixed effects for individuals (these must use pretests, not other baseline measures).

5.9 Other Design and Execution Requirements

All RCTs and QEDs that meet the requirements described above for attrition and baseline equivalence and that use acceptable methods for pretest controls that are appropriate for the respective design and circumstances must also meet several additional requirements to receive a design and execution rating of high or moderate. These requirements address issues related to the statistical models used to estimate program or service 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 the impacts of the programs and services under review. The statistical model standards include the following:

- For RCTs, when there is unequal allocation to intervention and comparison conditions across randomization blocks (i.e., the probability of being assigned to the treatment or comparison condition is different for different 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 group 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.
  - An endogenous covariate is one that is measured or obtained after baseline and that could have been influenced by the intervention. Covariates obtained after baseline that cannot change as a result of the intervention, such as birth dates or race/ethnicity, are not considered endogenous.

Inclusion of endogenous covariates in impact models results in biased impact estimates.

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

Applying the Statistical Model Standards

If the impact model for a contrast includes endogenous covariates, then the Prevention Services Clearinghouse will seek to apply its standards to alternate model specifications reported without such covariates, if available. Likewise, if the statistical model is invalid for estimating program or service impacts, the Prevention Services Clearinghouse will seek to apply its standards to alternate reported results from valid statistical models. If in either of these situations no results from alternate models are available that meet the statistical model standards, then the Prevention Services Clearinghouse will seek to review the contrast based upon unadjusted means and standard deviations of the outcome variable (querying authors if this information is not available). If no results from valid alternate models are reported and unadjusted means and standard deviations are not available, then the contrast receives a low rating.

5.9.2 Measurement Standards

In alignment with the Family First Prevention Services Act, the Prevention Services Clearinghouse applies three outcome standards: **face validity**, **reliability**, and **consistency of measurement between intervention and comparison conditions**. Prevention Services Clearinghouse measurement standards apply to all outcome measures and any measure used to determine baseline equivalence.

Face Validity

To satisfy the criterion for face validity, there must be a sufficient description of the outcome or baseline 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.

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, weight measurement, etc.

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 baseline measure either must be a measure which is assumed to be reliable (see the box above) 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 (correlation) of 0.50 or higher
- Inter-rater agreement (percentage agreement or kappa) of 0.80 or higher for percent agreement and 0.60 or higher for kappa.

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 Conditions

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

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

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 condition participants are asked about sexual behavior outcomes in a face-to-face interview with a case worker. Comparison condition 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.

Example 3: Participants in a study of a substance use intervention at a 28-day withdrawal management clinic are randomly assigned at clinic entry to a treatment as usual condition or a study intervention plus treatment as usual condition. Pretest data are collected immediately prior to random assignment for both conditions. Intervention plus treatment as usual condition participants are assessed on substance use outcomes at the end of intervention, which occurs on average 12 weeks after the pretest assessment. Participants in the treatment as usual condition are assessed at discharge from the withdrawal management clinic, four weeks after the pretest assessment. The time between the intervention and comparison conditions systematically differs (by 8 weeks) and could plausibly affect the outcomes. The outcomes would 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 high or moderate rating.
- Baseline measures 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 one of the study conditions. Confounds can be present in intervention conditions or comparison conditions. 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. When a confound is present, 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 the design and execution 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 conditions 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 conditions are shown to meet standards for equivalence at baseline, or when baseline differences between intervention and comparison conditions are adjusted for in analytic models, the effect of an intervention on outcomes can 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 conditions. The characteristic that differs between the two groups may be related to the expected amount of change between pretest and posttest measurements, thus confounding the intervention effect.

The Prevention Services Clearinghouse determines 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 pretest. Reviewers examine balance on race/ethnicity, SES, and child age, when available, for all contrasts. If any such characteristics exhibit large imbalances between intervention and comparison conditions, they may be considered to have substantially different characteristics confounds. 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.

Example 1: In a quasi-experimental study of a substance use treatment program, the intervention condition consists of participants in the community who were enrolled in the program. The comparison condition is constructed of individuals from the same community who met the eligibility criteria for the treatment program. However, the participants in the substance use treatment program all had incomes below the federal poverty threshold and had Medicaid coverage. The participants in the comparison condition were above the poverty threshold and not covered by Medicaid. The socioeconomic status of individuals in the comparison condition is substantially higher than that for the individuals in the treatment program. A substantially different characteristics confound is present because it would be impossible to tell if observed differences between the groups after treatment were due to socioeconomic status differences or to the substance use treatment program.

One standard that is applied in Prevention Services Clearinghouse reviews is "*refusal of offer of treatment*." When the intervention condition comprises individuals or units that

were offered and accepted treatment and most of the comparison condition[35] comprises individuals or units that were known to have been offered and refused or dropped out of 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 conditions that consist entirely of individuals or units that accepted the offer of treatment. In these circumstances, the strongest designs would limit the comparison condition 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 condition in this manner. Some comparison conditions 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 2: A mental health intervention is designed for 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 condition 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 condition have a characteristic that is not present in the comparison condition that is plausibly related to outcomes.

n=1 Person-Provider Confound or Administrative Unit Confound

When all individuals in the intervention condition or all individuals in the comparison condition 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. 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.

Example 3: When intervention services are provided by a single therapist and the outcome for her patients on a mental health assessment is compared with the outcome for patients of another therapist, it is impossible to disentangle the effect of the intervention from the skills of the therapists and a confound is present. If, however, multiple therapists are present in both conditions, a confound would not be present.

[35] This is operationally defined as at least 75 percent of the comparison condition.

Example 4: If a single therapist provides treatment and the comparison condition is waitlisted, receiving no intervention services from the therapist, the treatment effect is still confounded with the skills of the provider.

Example 5: If a single provider provides the program or service to the intervention condition participants and provides treatment as usual or some other services to comparison condition participants, this confound is *not* present.

Example 6: Another example of a non-confounding factor would be if a single therapist provides an attention control to the comparison condition and the intervention to the intervention condition.

Similar to the n=1 person-provider confound, when all individuals in the intervention condition or all individuals in the comparison condition receive intervention or comparison services in a single administrative unit (e.g., clinic, hospital, organization, or community) the treatment effect may be confounded with the capacity of that administrative unit to produce better outcomes or with characteristics of the administrative unit not directly associated with delivery of the intervention. The Prevention Services Clearinghouse calls this type of confound an *n=1 administrative unit provider confound*. When individuals in the intervention condition receive services from a single administrative unit and individuals in the comparison condition receive no services from the administrative unit (e.g., a waitlist comparison condition that does not receive treatment as usual during the waiting period), this confound is present. However, if a single provider location serves at least some participants in both the intervention condition and the comparison condition, this confound is not present.

Example 7: If an investigator randomly assigned two therapists in a single clinic to provide the intervention and two other therapists in the same single clinic to provide treatment as usual, there is no confounding factor present.

***Example 8*:** No confound would be present for a study where all participants in the intervention condition received services from a single child welfare agency if some or all participants in the control condition also received other services from the same agency (e.g., some agency caseworkers delivered intervention services and others delivered treatment as usual).

Example 9: If a study randomly assigned one clinic to provide the program or service and one clinic to continue providing treatment as usual, a confound would be present because it would not be possible to disentangle the effect of the program or service from the effect of either clinic. This study could avoid the confound by randomly assigning therapists within each clinic to provide either the program or service or continue providing treatment as usual.

5.9.4 Missing Data Standards

The Prevention Services Clearinghouse missing data standards are based on those used by the WWC (What Works Clearinghouse, 2022) and are applied only to posttests on eligible outcome measures, pretests, correlated pretests, and pretest alternatives. For other model covariates, any method that is used to address missing data is acceptable. The missing data standards are intended to ensure that missing or imputed data on outcome measures, pretests, correlated pretests, and pretest alternatives does not introduce bias in impact estimates that could affect the credibility of the findings.

If a contrast has missing data on posttests, pretests, correlated pretests, or pretest alternatives, reviewers first assess whether the approach to addressing missing data is acceptable. Examples of the acceptable approaches are described below.

- 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 are examples of acceptable approaches 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 conditions (or, if together, includes an indicator variable for intervention status) in which all covariates in the impact model and the outcome are included 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 posttests and if the weights are estimated separately for intervention and comparison conditions or if an indicator for intervention condition status is included.

- ***Constant Replacement:*** Replacing missing values with a constant value and including an indicator variable in impact estimation models to identify the cases

with missing values. Acceptable only for RCTs with missing pretests or pretest 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 a design and execution 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 assessing statistical significance and computing effect sizes.

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 What Works Clearinghouse, 2022).

- 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 pretest or pretest alternatives are imputed, reviewers evaluate baseline equivalence for the analytic sample and proceed as usual for contrasts required to establish baseline equivalence.

If pretest, correlated pretest, or pretest 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, 2022).

- 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 a moderate rating, provided the other design and execution standards are met and an acceptable method of addressing missing data is used.

6. Record and Characterize Impact Estimates

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

The Prevention Services Clearinghouse design and execution standards focus on the causal or internal validity of the findings reported in a study. That is, they focus on whether the study produces *credible evidence* about the causal impact of the intervention on the outcomes for the individuals in the intervention condition. The design and execution standards do not address the estimate of the *effectiveness* of the intervention ("impact estimate") or the statistical significance of this estimate. The Prevention Services Clearinghouse rates all eligible contrasts in a study against the design and execution standards to determine if the evidence they provide is credible. Contrasts that receive a design and execution rating of high or moderate are deemed to provide credible evidence. Only contrasts that provide credible evidence are then used to estimate the effectiveness of the intervention and inform program or service ratings of well-supported, supported, or promising.

6.1 Four Characterizations of Impact Estimates

To inform program or service ratings of well-supported, supported, or promising (as described in Chapter 7), the Prevention Services Clearinghouse must have sufficient information to permit characterization of impact estimates from contrasts as favorable, sustained favorable, unfavorable, or no effect. Characterization of impact estimates as favorable, sustained favorable, no effect, or unfavorable is based on both their *direction* and *statistical significance*[36], as follows:

- *Favorable* impact estimates are those in the desired direction (i.e., the intervention condition has better outcomes than the comparison condition) and are statistically significant (i.e., different from zero).
- *Sustained favorable* impact estimates are those in the desired direction and are statistically significant and that are observed at least 6 or 12 months beyond the end of treatment (see Section 7.2.3).

[36] The Prevention Services Clearinghouse defines statistical significance based on the conventional threshold of a *p* value of less than 0.05 based on a two-tailed statistical test. An equivalent 95% confidence interval that does not contain the value indicating no effect (e.g., 0 for a confidence interval for the mean difference of a continuous outcome, 1 for a confidence interval for an odds ratio) is also considered to meet the threshold for statistical significance.

- *Unfavorable* impact estimates are those that are not in the desired direction (i.e., the intervention condition has worse outcomes than the comparison condition) and are statistically significant.
- Impact estimates that are not statistically significant in either the desired or undesired direction are labeled as *no effect,* meaning that the contrast does not provide adequate statistical evidence that the intervention and comparison conditions' average outcomes are different.

The Prevention Services Clearinghouse must be able to determine the direction and statistical significance of an impact for it to be used to inform program or service ratings. If information needed to make these determinations is not available in a study, reviewers may send queries to authors requesting such information, in accordance with the author query policies described in Section 8.4.2. If requested data are not obtained in response to author queries, reviewers complete the review with the information available.

6.2 Procedures for Establishing Direction of Impact Estimates

The direction of an impact estimate indicates whether the intervention condition is, on average, better off or worse off relative to the comparison condition. Reviewers must be able to determine whether a higher value on the measure of the outcome indicates better or worse outcomes, which permits them to determine whether the direction of the impact estimate favors the intervention condition or the comparison condition. Outcomes for which direction cannot be determined or is ambiguous receive a low design and execution rating and cannot be used to inform program or service ratings. Such outcomes do not meet the face validity standard (Section 5.9.2) because they do not have a direct interpretation. Reviewers use descriptions of measures and contextual information about the direction of impacts provided in studies and may obtain copies of measures or measure manuals or query study authors to assist in establishing direction.

6.3 Procedures for Establishing Statistical Significance of Impact Estimates

The Prevention Services Clearinghouse typically relies on study-reported p-values to form the basis of the assessment of statistical significance for a finding.[37] To be used to establish statistical significance, p-values from models reported in studies must include any required adjustments (e.g., for pretests or for clustering) and must be *aligned* with the specific contrast under review – that is, reflecting a statistical test of the difference

[37] Although p-values can be replicated for straightforward models, such as independent t tests or simple models with a single control variable that is perfectly balanced between conditions, p-values are difficult to replicate for models which include several covariates that are not perfectly baseline equivalent, include numerous random effects, or adjust standard errors using robust variance methods (Ryan, 2013).

between the intervention and comparison conditions on a single outcome at a single measurement period (e.g., a p-value from a statistical test of a time by treatment interaction combining two intervention conditions and one comparison condition would not be aligned; neither would the p-value from a time by treatment interaction test for a single intervention condition compared to a single comparison condition that included three or more time points). Models (or statistics reported from such models) also must meet the statistical model standards (Section 5.9.1) and missing data standards (Section 5.9.4), or they cannot be used to establish the statistical significance of a finding.

When study-reported analyses are not adjusted (and adjustment is required), are not aligned with the contrast under review, do not meet the statistical model or missing data standards, or statistical significance is not reported, the Prevention Services Clearinghouse will attempt to perform its own statistical test of a finding using any available study information or author queries, as needed.[38] In such cases, the Prevention Services Clearinghouse uses a calculated t-test value is compared to a critical value associated with a 5%, two-tailed, Type I error rate for a given degrees of freedom in a two-tailed test (Pearson, 1928). This null hypothesis test is used to make a determination of whether the finding is statistically significant.[39] The specific procedures used to assess statistical significance depend on the type of the outcome measure, as described below.

6.3.1 Statistical Significance Calculation Procedures by Outcome Type

Continuous Outcomes

For outcomes measured on a continuous or approximately continuous scale (e.g., differences in average scores between an intervention condition and a comparison condition on an assessment of mental health where total scores can range from 0 to

[38] This approach follows other clearinghouses which seek to utilize the best model available from study authors who have access to the raw data (e.g., What Works Clearinghouse v5.0). However, errors and omissions in reporting p-values and statistical significance are common (Bakker & Wicherts, 2011; Krawczyk, 2015). If potential errors or omissions are identified, the Prevention Services Clearinghouse first attempts to conduct an author query to resolve the apparent error or omitted information. If apparent errors or omissions in reporting cannot be resolved via an author query, the Prevention Services Clearinghouse will give preference its own statistical test based on available study information when the information necessary to do so is available, inclusive of any required adjustments.

[39] A small p-value is not an indicator of the magnitude of the impact; it represents the probability of obtaining at least the impact observed on the outcome if the null hypothesis was true (i.e., that the impact of the intervention is 0). When p-values are greater than the conventional value of .05, this suggests that it is more likely that the null hypothesis is the correct answer (i.e., there is no real difference between the conditions). Statistics that produce a *p*-value smaller than .05 indicate that there is less than a 5 percent chance that the impact is a false positive.

40), the impact estimate Δ is defined in terms of the difference in means $\bar{Y}_{G2}$ for the intervention condition and $\bar{Y}_{G1}$ for the comparison condition

$$\Delta = \bar{Y}_{G2} - \bar{Y}_{G1}.$$

The standard error of this difference is estimated with

$$se_{\Delta} = \sqrt{\frac{s_p^2}{n_{G1}+n_{G2}}},$$

where n_{G1} and n_{G2} are the sample sizes for the comparison and intervention conditions, respectively, and s_p^2 is the squared pooled standard deviation of the two conditions.

The statistical test is then computed as

$$t = \frac{\Delta}{SE_{\Delta}},$$

for individual assignment. The t statistic is evaluated against a critical value from the student's t-distribution with

$$df_{SRS} = n_{G1} + n_{G2} - 2$$

degrees of freedom and an alpha=.05 in a two-tailed test.

In the case of cluster randomized designs, a correction factor based on the intraclass correlation (ICC, ρ_{intra}) is employed where a value of .10 is assumed if the ICC is not available.[40] This correction factor is the inverse of the square-root of the cluster-sampling design effect

$$1 + (\bar{n} - 1)\rho_{intra},$$

noted in survey research (Kish, 1965). The clustered value of t is

$$t = \frac{\Delta}{SE_{\Delta}} \times \frac{1}{\sqrt{1+(\bar{n}-1)\rho_{intra}}},$$

where $\bar{n}$ is the average number of units per cluster. This statistic is evaluated against a student's t-distribution with

$$df_{Cluster} = \frac{([n_{G1}+n_{G2}-2\,]-2(\bar{n}-1)\rho_{intra})^2}{[n_{G1}+n_{G2}-2\,](1-\rho_{intra})^2+\bar{n}(n_{G1}+n_{G2}-2\bar{n})\rho_{intra}^2+2(n_{G1}+n_{G2}-2\bar{n})\rho_{intra}(1-\rho_{intra})}$$

[40] This assumption is consistent with conventions used by the What Works Clearinghouse (2022) for non-academic measures.

degrees of freedom via a Satterthwaite approximation.[41]

Binary Outcomes

For binary outcomes (i.e., outcomes that can take one of exactly two values, such as "employed" or "not employed"), the impact estimate is defined as a logged odds ratio computed using the proportion of events in the intervention condition and the proportion of events in the comparison condition.[42] The statistical significance of the odds ratio is then tested. The odds ratio is defined as

$$OR = \frac{\bar{Y}_{G2}(1-\bar{Y}_{G1})}{\bar{Y}_{G1}(1-\bar{Y}_{G2})},$$

where $\bar{Y}_{G2}$ is the observed or predicted probability of the outcome for intervention condition (G_2) and $\bar{Y}_{G1}$ is the observed or predicted probability of the outcome for the comparison condition (G_1). The impact estimate is then computed as the natural logarithm of the odds ratio:

$$\ln(OR) = \ln\left(\frac{\bar{Y}_{G2}(1-\bar{Y}_{G1})}{\bar{Y}_{G1}(1-\bar{Y}_{G2})}\right).$$

The standard error of the log of the odds ratio (Woolf, 1955) is defined as

$$se_{\ln(OR)} = \sqrt{\frac{1}{n_{G1}\bar{Y}_{G1}(1-\bar{Y}_{G1})} + \frac{1}{n_{G2}\bar{Y}_{G2}(1-\bar{Y}_{G2})}},$$

and the statistic used to test the statistical significance of the impact estimate is

$$t = \frac{\ln(OR)}{SE_{\ln(OR)}},$$

for individual assignment and

$$t = \frac{\ln(OR)}{SE_{\ln(OR)}} \times \frac{1}{\sqrt{1+(\bar{n}-1)\rho_{intra}}},$$

---

41 This is also the method used by the What Works Clearinghouse (2022).

42 Odds ratios are needed to standardize impact estimates for binary outcomes because the same absolute difference in the probability of experiencing an event (e.g., intervention effect of 0.01) is proportionally different depending on the probability of the comparison condition experiencing the event. For example, an intervention that reduces the probability of an event, such as an out-of-home placement, by one percentage point is proportionally different for an event that 4% of the comparison condition experiences (25% relative reduction in probability of event occurring) than for an event 50% of the comparison condition experiences (2% relative reduction in probability of event).

for clustered assignment. The t statistics are evaluated against a critical value from the student's t-distribution with df_{SRS} (defined above) for individual assignment and $df_{Cluster}$ (defined above) for clustered assignment, each with an alpha=.05 in a two-tailed test.

Count Outcomes

For outcomes measured as a number of discrete events (e.g., a count of the number of times an event occurred per participant, if at all) the impact estimate is defined as the natural log of the incidence rate ratio (IRR) of intervention condition (G_2) to comparison condition (G_1) means and tests the statistical significance of that. The incidence rate ratio is

$$IRR = \frac{\bar{Y}_{G2}}{\bar{Y}_{G1}},$$

And the impact estimate is its natural logarithm

$$\ln(IRR) = \ln\left(\frac{\bar{Y}_{G2}}{\bar{Y}_{G1}}\right).$$

The asymptotic standard error of the log incidence rate ratio is defined as (Bakbergenuly et al., 2020; Hedges et al.,1999)[43]

$$se_{\ln(IRR)} = \sqrt{\frac{s_{G1}^2}{n_{G1}\bar{Y}^2{}_{G1}} + \frac{s_{G2}^2}{n_{G2}\bar{Y}^2{}_{G2}}},$$

where s_{G1}^2 is the variance of the count outcome for comparison condition and s_{G2}^2 is the variance of the count outcome for the intervention condition. The statistical test is

$$t = \frac{\ln(RR)}{SE_{\ln(RR)}}$$

for individual assignment and

$$t = \frac{\ln(RR)}{SE_{\ln(RR)}} \times \frac{1}{\sqrt{1+(\bar{n}-1)\rho_{intra}}}$$

for clustered assignment. These are evaluated against critical values with the degrees of freedom for the individual and cluster assignment cases as defined above.

---

[43] Count outcomes (outcomes measured positive integers, 0s, with a long right-side tail) are often analyzed with a Poisson distribution, which includes the assumption that the mean equals the variance, $s^2 = \bar{Y}$. In this case, the variance formula for the log of the rate ratio, $\ln(RR)$, simplifies to the variance of a Poisson regression coefficient as s^2 and $\bar{Y}$ cancel (displayed with strikethroughs),

$$var_{\ln(IRR)} \frac{1}{n_{G1}\bar{Y}_{G1}} + \frac{1}{n_{G2}\bar{Y}_{G2}} = \frac{1\times\cancel{s_{G1}^2}}{n_{G1}\bar{Y}_{G1}\times\cancel{\bar{Y}_{G1}}} + \frac{1\times\cancel{s_{G2}^2}}{n_{G2}\bar{Y}_{G2}\times\cancel{\bar{Y}_{G2}}}$$

which is the log of the incidence rate ratio, $\ln(IRR)$, noted in several texts (such as Ryan, 2013).

Time-To-Event Outcomes

Time-to-event outcomes (e.g., *when* a specific event happened, if at all, during the study period) are often analyzed using Cox hazard models (Collett, 2003) or log-rank tests that summarize the ratio of observed events to expected events across several time periods. The Prevention Services Clearinghouse generally relies on author-reported hazard ratios (and similar estimates) and their standard errors to establish statistical significance. If such information is not available or the analysis is not appropriate or not aligned with the contrast under review, the Prevention Services Clearinghouse may query for additional information. In some cases where sufficient information is available in the study (or provided by study authors), the Prevention Services Clearinghouse may compute its own log-rank test, as detailed in Collett (2003), to establish statistical significance.

6.3.2 Statistical Significance Procedures for Repeated Measures

Some studies take repeated measurements of outcomes during the study and analyze the impact of the intervention on the outcomes across those repeated measurements (e.g., a growth model with pretest, end of intervention, and 12-month post-intervention follow-up measures). When multiple outcome periods are included in the same model, the statistical significance of the time by intervention interaction must be reported for a contrast that aligns with the contrast under review (i.e., is specific to a comparison of an intervention condition to a comparison condition on an outcome *for a specific posttest measurement*). For example, consider a growth model specification that permits testing separate p-values for the end-of-intervention and 12-month follow-up predicted mean differences, respectively, for the intervention and comparison conditions in the example above based on the interaction of the specific posttest measurement with the intervention. A report of model-derived p-values for each respective contrast (end-of-intervention and 12-month follow-up) would be aligned.

If the necessary point-in-time information for model results is not reported or if the model does not meet statistical model standards, the Prevention Services Clearinghouse will use (or request, if not available) unadjusted means (and standard deviations, if applicable) at each measurement time point, applying the relevant statistical test based on the type of outcome using the formulae described above, and attempt to perform any required adjustments for baseline differences or clustering.

6.4 Procedures for Computing Magnitude of Impact Estimates (Effect Sizes)

To assist readers in understanding the *magnitude* of the observed impact estimates, the Prevention Services Clearinghouse attempts to compute and display effect sizes for all contrasts that are rated high or moderate. The Prevention Services Clearinghouse uses the standardized mean difference effect size, which is the difference between the

intervention and comparison conditions at each measurement time point in standard deviation units. Formulae for computing effect sizes are presented in this section.

The Prevention Services Clearinghouse displays all effect sizes so that positive values indicate that the intervention condition had better outcomes than the comparison condition and negative values indicate that the intervention condition had worse outcomes than the comparison condition. For outcomes where higher values indicate worse outcomes, the effect size is multiplied by -1. For outcomes where higher values indicate better outcomes, the effect sizes are displayed in the observed direction from the study, implicitly multiplied by 1.

6.4.1 Effect Size Calculation Procedures by Outcome Type

Continuous Outcomes

For continuous and approximately continuous outcomes, the Prevention Services Clearinghouse first computes the standardized mean difference effect size (*d*) as

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

where the numerator ($\bar{Y}_{G2} - \bar{Y}_{G1}$) is the difference in means for the intervention (G_2) and comparison (G_1) conditions, and the denominator (s_p) is the pooled standard deviation of the intervention and comparison conditions. All standardized mean difference effect sizes are adjusted with the small-sample correction factor, ω, to provide unbiased estimates of the effect size (Hedges, 1981). The correction factor ω is computed as:

$$\omega = 1 - \left(\frac{3}{4(n_{G1} + n_{G2}) - 9}\right),$$

where n_{G1} is the sample size for the comparison condition and n_{G2} is the sample size for the intervention condition. The small-sample corrected effect size (referred to as Hedges' *g*) is then represented as:

$$g = \omega \times d.$$

Standard formulae for computing standardized mean difference effect sizes from common statistics (e.g., t-tests, F-tests, regression coefficients) are employed, as necessary (Borenstein & Hedges, 2019; Lipsey & Wilson, 2001; Shadish, Robinson, & Lu, 1997).

Binary Outcomes

For binary outcomes, the Prevention Services Clearinghouse performs conversions to transform impact estimates based on binary variables from the log-odds metric into an

approximate standardized mean difference. Log odds ratios are converted into the *d*-metric via a Cox transformation (Cox, 1970)[44] as follows:

$$d_{Cox} = .607 \times \ln(OR) = .607 \times \ln\left(\frac{\bar{Y}_{G2}(1-\bar{Y}_{G1})}{\bar{Y}_{G1}(1-\bar{Y}_{G2})}\right).$$

In cases where the binary outcomes impact estimate is reported as a rate-ratio, the rate ratio is first converted into a log odds ratio and are then converted into the *d*-metric via the Cox transformation defined above. The odds ratio is computed from the rate ratio as follows (see Zhang et al., 1998):

$$OR = \frac{\frac{\bar{Y}_{G2}}{\bar{Y}_{G1}} - \bar{Y}_{G2}}{1-\bar{Y}_{G2}},$$

then computes the natural logarithm of the odds ratio and performs the Cox transformation on the result as described above.

Count Outcomes

For count outcomes, the standardized mean difference is computed using the same procedures as continuous outcomes described above. To compute the denominator, the number of events is treated as a continuous outcome for which we compute a pooled standard deviation between the intervention and comparison conditions. The effect size is then the difference between the intervention and comparison condition means divided by the pooled standard deviation adjusted with the small-sample correction factor, ω.

Time-To-Event Outcomes

Time-to-event outcomes are often analyzed using Cox hazard models (Collett, 2003) or log-rank tests which summarize the ratio of observed events to expected events across several time periods. The original data required to compute hazard ratios are not typically reported in studies. Thus, the Prevention Services Clearinghouse generally cannot compute effect sizes directly for such outcomes. Such effect sizes are displayed as “not calculated” on the Prevention Services Clearinghouse website.

In cases where data are sufficient to compute effect sizes from time-to-event outcomes, the Prevention Services Clearinghouse will convert hazard ratios into log odds ratios. These are then converted into the *d*-metric via the Cox transformation as described

[44] The value of .607 is the ratio of the quantile associated with a cumulative standard normal probability associated with .8 to the log-odds of .8 (Cox, 1970). Although different probabilities produce different ratios, this value has been accepted by convention as a normalizing constant (see Sanchez-Meca et al., 2003), and we adopt this value to maintain correspondence with other clearinghouses (Sama-Miller et al., 2021; What Works Clearinghouse, 2022). The inverse, 1.65, also appears as a denominator to the difference in log-odds to produce the equivalent transformation.

above. Hazard ratios (which are rate ratios over time that are averaged; see Collett 2003) are first converted to odds ratios using the rate-ratio, $\frac{\bar{Y}_{G2}}{\bar{Y}_{G1}}$, and the intervention condition rate, $\bar{Y}_{G2}$ (see Zhang et al., 1998),

$$OR = \frac{\frac{\bar{Y}_{G2}}{\bar{Y}_{G1}} - \bar{Y}_{G2}}{1 - \bar{Y}_{G2}},$$

and then converted into a log-odds ratio for the Cox transformation.

6.4.2 Effect Size Calculation Procedures for Repeated Measures

The Prevention Services Clearinghouse will estimate effect sizes based on the values of the predicted or observed means of the outcome for each measurement time period for each group, if such data are reported or available from study authors. For continuous or count outcomes, the effect size will also utilize the contemporaneous posttest measurement standard deviation for each group to compute a pooled standard deviation for each respective measurement time period.

6.4.3 Conventions for Statistics Used in Effect Size Calculations

For effect size calculations, the Prevention Services Clearinghouse will generally seek to use statistics derived from study-reported models used to establish the statistical significance of a finding and prefers statistics that are adjusted for pretests and other covariates 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 pretest and posttest findings, reviewers compute the effect size of the difference between the intervention and comparison conditions at baseline and subtract that value from the posttest effect size.

For many outcomes or models with multiple covariates, the covariate influence on the adjusted regression parameter is difficult to reproduce exactly. However, the adjusted predictions in average outcomes for intervention and comparison conditions are sufficient to produce an accurate estimate of the adjusted difference in means. The Prevention Services Clearinghouse follows the WWC (What Works Clearinghouse, 2022) in our preferences for computing effect sizes derived from study materials when multiple options are available, as follows:

1. Average (or marginal) predicted outcome values from generalized models (e.g., predicted probabilities from a logistic regression)
2. Average (or marginal) predicted outcome values from linear (OLS) models
3. Treatment impact estimates from coefficients from generalized models

6.5 Display of Impact Estimates on the Prevention Services Clearinghouse Website

Each contrast with a high or moderate rating is displayed on the Prevention Services Clearinghouse website. Information about contrasts displayed include the effect size in standard deviation units (as defined above), 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 outcome domain.

Prevention Services Clearinghouse reviewers 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 condition percentile rank for the outcome minus the comparison condition average percentile, which is 50. For example, an implied percentile effect of 4 means that the program or service increased the intervention condition performance by 4 percentile points over the comparison condition.

7. Program or Service Ratings

The Prevention Services Clearinghouse periodically provides clarification on topics covered in the Handbook of Standards and Procedures. 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 and characterization of impact estimates from one or more RCT or QED studies of a program or service into ratings for that program or service.[45] As described at the beginning of Chapter 5, Prevention Services Clearinghouse reviewers rate each study contrast in an eligible study. Chapter 6 describes how impact estimates from contrasts that receive a design and execution rating of high or moderate are recorded and characterized. To determine program or service ratings, the Prevention Services Clearinghouse combines design and execution ratings and characterizations of impact estimates 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 with a high or moderate rating and characterizations of their impact estimates are examined.

7.1 Four Ratings

Using the eligible contrasts that meet design and execution standards, reviewers assign one of four ratings to each program or service to characterize the extent of evidence for a particular program or service in alignment with the Family First Prevention Services Act requirements for each rating category, as summarized in Exhibit 7.1 below:

- ***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 RCT or QED studies carried out in *usual care or practice settings* (see Section 7.2.2) that achieve a design and execution rating of high or moderate and demonstrate *favorable* effects (see Section 6.1) in a target outcome domain. At least one of the contrasts must demonstrate a *sustained favorable* effect (Section 6.1) of at least 12 months *beyond the end of treatment* (see Section 7.2.3) on at least one eligible outcome.

- ***Supported***. A program or service is rated as a supported practice if it has at least one contrast in an RCT or QED study carried out in a usual care or practice setting that achieves a design and execution rating of high or moderate and demonstrates a sustained favorable effect of at least 6 months beyond the end of treatment on at least one eligible outcome.

[45] The Prevention Services Clearinghouse is planning to develop pilot standards for single case designs. How contrasts from single case designs may contribute to program or service ratings will be determined during the pilot activities.

- ***Promising***. A program or service is designated as a promising practice if it has at least one contrast in a study that utilizes some form of control, achieves a design and execution rating of high or moderate, and demonstrates a favorable effect on at least one eligible outcome. The Prevention Services Clearinghouse is planning to develop and pilot standards for studies with single case designs and how such designs can contribute to the promising rating.
- ***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 rating includes programs or services (a) that do not have any eligible contrasts with design and execution ratings of high or moderate with favorable effects, (b) that do not have any eligible contrasts with design and execution ratings of high or moderate, (c) that have no eligible studies, or (d) that do not meet risk of harm standards.

Exhibit 7.1. Summary of Criteria for a Well-supported, Supported, or Promising Program or Service Rating

Program or Service Rating	Number of Studies With Favorable Effects	Length of Sustained Effects	Usual Care or Practice Setting	Has Risk of Harm
Well-supported	At least 2 eligible RCT or QED studies each with 1 or more favorable effects from contrasts with a design and execution rating of high or moderate.	At least 1 sustained favorable effect occurs 12+ months after the end of treatment	Required for contrasts with favorable effects to contribute to program or service rating.	No
Supported	At least 1 eligible RCT or QED study has 1 or more favorable effects from contrasts with a design and execution rating of high or moderate.	At least 1 sustained favorable effect occurs 6+ months after the end of treatment.	Required for contrasts with favorable effects to contribute to program or service rating.	No
Promising	At least 1 eligible study has 1 or more favorable effects from contrasts with a design and execution rating of high or moderate.	n/a	Not required for contrasts with favorable effects to contribute to program or service rating.	No

Note. Favorable effects are those in the desired direction that are statistically significant. Contrasts determined to have an empirical risk of harm cannot contribute to a well-supported, supported, or promising rating. See Section 7.2.3 for definition of "end of treatment" and standards for computing the time since end of treatment. See Section 7.2.2 for definition of usual care or practice setting. "Has risk of harm" refers to the program or service having a risk of harm as determined according to the standards in Section 7.2.1.

Programs or services on the working list that are deemed to be ineligible after a final eligibility determination (Section 2.3) do not receive one of the four ratings listed above. Instead, the program or service is indicated as being ineligible for review.

Program or service ratings assigned under any prior versions of the Handbook of Standards and Procedures remain in effect until such time that a re-review occurs (see Section 8.2 regarding procedures for selection of the version of the Handbook of Standards and Procedures used to conduct program or service reviews and re-reviews and Section 8.5 for re-review procedures).

7.2 Contributing Factors in the Ratings

7.2.1 Risk of Harm

Consistent with the Family First Prevention Services Act, a program or service cannot be 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 suggests that the overall weight of evidence does not support the benefits of the program or service. 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 program or service and was severe or frequent.

To be considered in the risk of harm assessment, unfavorable effects must be reflected in contrasts that receive a high or moderate rating according to the design and execution standards. Factors the Prevention Services Clearinghouse considers when assessing the empirical evidence of a risk of harm from all contrasts that received a high or moderate rating according to the design and execution standards include (but are not limited to):

- Direction and magnitude of the mean effect size for each type of outcome within an outcome domain (e.g., out-of-home placement outcomes within the child permanency outcome domain), with particular attention to those where any individual unfavorable effects are present;
- Statistical significance of the mean effect size for each type of outcome within an outcome domain;
- The number of contrasts characterized as favorable, unfavorable, and no effect for each type of outcome within an outcome domain;
- The number of studies reporting unfavorable effects and their sample sizes;
- Magnitude of individual unfavorable effect sizes and the specific outcome measures assessed.

If, after all available information about risk of harm is considered, the weight of the evidence for a program or services suggests that a risk of harm is likely, the program or service will receive a rating of 'Does not currently meet criteria' by the Prevention Services Clearinghouse.

Additional considerations for contrasts where the comparison condition is a head-to-head comparison or placebo or attention control comparison.

For contrasts with head-to-head or placebo or attention control comparison conditions, the Prevention Services Clearinghouse will assess whether there is any evidence for the risk of harm in the comparison condition. If risk of harm is present in these kinds of comparison conditions, findings are not clearly interpretable as evidence of intervention effectiveness. When risk of harm is not present in the comparison intervention, favorable findings can be interpreted as the intervention condition being at least better off than they would have been if no intervention had been offered at all and can potentially contribute as evidence of effectiveness.

For contrasts with a design and execution rating of high or moderate where either of these two kinds of comparison conditions are present, the Prevention Services Clearinghouse first assesses available information from the studies reviewed on whether there is any evidence of risk of harm within the studies reviewed (e.g., if participants receiving the comparison intervention were faring worse over time relative to the beginning of the study or if there were any known risks of the comparison intervention described in the trial protocol or review of literature). If there is potential evidence of risk of harm in comparison condition in the studies reviewed, the Prevention Services Clearinghouse may consult additional sources of publicly available information or experts to develop an assessment of the potential for risk of harm in the comparison intervention. Experts who advise the Prevention Services Clearinghouse on risk of harm include, but are not limited to, individuals with lived expertise; individuals from communities that programs and services are intended to serve, inclusive of underserved and tribal communities; methodological experts; state, tribal, and local child welfare department staff or administrators; policymakers; program or service developers; program or service providers or trainers; study authors; and subject matter experts. Additionally, any comparison conditions with interventions that have been rated by the Prevention Services Clearinghouse and indicated as having a risk of harm would be assumed to have a risk of harm.

Contrasts with a high or moderate rating where there is an empirical basis or case data indicating that the comparison condition constitutes a risk of harm to those receiving it cannot contribute to a well-supported, supported, or promising program or service rating. When risk of harm is not present in the comparison intervention, favorable findings can contribute as evidence of effectiveness.

7.2.2 Usual Care or Practice Settings

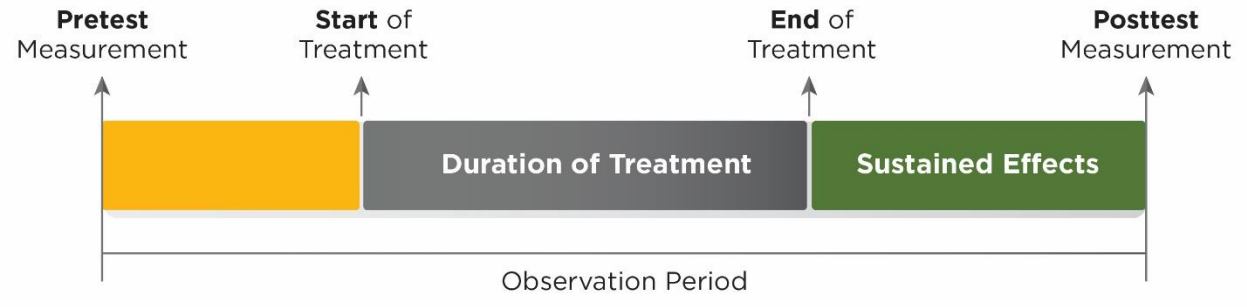
As specified in the Family First Prevention Services Act, to receive a rating of *supported* or *well-supported*, the favorable evidence for a program or service must have been obtained from research conducted in a usual care or practice setting. A usual care or practice setting is defined as an existing service agency or provider that delivers mental health prevention or treatment programs or services, substance use prevention or treatment programs or services, in-home parent skill-based programs or services, and/or kinship navigator programs as part of its typical operations. Also included are community settings, such as schools, with embedded service providers that may provide such services as part of their typical operations (e.g., school counselors).

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 or clinics that provide services solely for participants in research studies or clinical trials (i.e., that do not provide any services to persons not participating in research studies as part of their typical operations) do not constitute usual care or practice settings, even if staffed by personnel who might typically work in a usual care setting.

7.2.3 Beyond the End of Treatment

As specified in the Family First Prevention Services Act, to receive a rating of supported or well-supported, a program or service must have sustained favorable effects beyond the end of treatment. The *length of sustained effects* is defined as the time between the *end of treatment* and a *posttest measurement* of an eligible target outcome (see Exhibit 7.2).

Exhibit 7.2. Determining the Length of Sustained Effects



Reviewers may use several sources of information to determine the length of sustained effects following an ordered process. The order of preference among different sources of information for determining the length of sustained effects in a study is as follows:

1. **Information reported in the study itself.**

    a. The Prevention Services Clearinghouse prefers to determine the length of sustained effects from explicit statements in studies about the time between a defined end of treatment and a posttest measurement.

    b. If the study documentation does not explicitly report the time between a defined end of treatment and a posttest measurement, reviewers may use other study-reported information about the duration of treatment, the end of treatment, or timing of pretest and posttest measurement to estimate the length of sustained effects.[46]

2. **Study authors.**

    a. If the study does not report the length of sustained effects directly and other study-reported information is not sufficient to estimate the length of sustained effects (or if such information is unclear or inconsistent with other information about timing reported in the study or program or service documentation), reviewers will request additional information from study authors to determine the length of sustained effects.

3. **Program or service documentation.**

    a. If study-reported information is not sufficient and is not provided by authors upon request, reviewers may use program or service documentation (including, but not limited, to the program or service manual) to estimate the duration of treatment and then derive the length of sustained effects using information reported in study documentation about the timing of pretest and posttest measurement.

If the length of sustained effects cannot be determined from the above-mentioned sources, reviewers will assume that posttest measurement occurred at the end of treatment.

[46] This assumes that the start of treatment is closely aligned with when a pretest measurement is collected. As noted in Exhibit 7.2, pretest measurement may occur before the start of the intervention. If the posttest measurement timing is reported relative to pretest measurement instead of the start of treatment, and time elapses between pretest measurement and the start of treatment, the Prevention Services Clearinghouse will subtract the time from pretest measurement to the end of treatment from the time from pretest measurement to posttest measurement to estimate the length of sustained effects. Standards for consistency of measurement (Section 5.9.2) apply.

Using Program or Service Duration and Measurement Timing to Estimate Length of Sustained Effects

In some cases, reviewers may need to determine the duration of treatment to estimate the length of sustained effects (e.g., studies that only report a posttest measurement relative to the pretest measurement or the start of treatment). Across the range of programs and services that may be eligible for review, the duration of treatment can be the same across participants or it may vary across participants. Some programs and services, such as those intended to address chronic conditions, may not have a defined duration according to the program or service documentation. Studies of such programs or services may, however, set an end of treatment, depending on factors including (but not limited to) study design, research funding, or local conditions (e.g., a study of a program targeting a chronic condition that does not have a specified duration in the program or service manual defines an end of treatment at two years after the start of services based on expiration of research funding for treatment provision).

In situations where reviewers need to determine the duration of treatment in order to estimate the length of sustained effects, they will use the following conventions.

Study has a defined duration of treatment that is the same for all participants. When the duration of treatment is clearly defined in the study and expected or intended to be the same for all participants (e.g., 12 weekly sessions, a 3-month service delivery period), the Prevention Services Clearinghouse will use the study-reported information.

If such information is not available in the study or upon request from study authors and the duration of treatment is described as being the same for all participants in the program or service documentation, the duration of treatment stated in the program or service documentation can be used. If there are multiple versions of the program or service with different durations of treatment described in the program or service manual, and it is unclear after an author query which version was used in the study, reviewers will use the longest duration of treatment described in the program or service documentation.

- ***Example 1:*** A study of a program designed to be delivered in 16 weekly sessions indicates that it planned to deliver the program as designed. The duration of treatment is defined as 16 weeks.
- ***Example 2:*** A therapy program can be delivered in 12 sessions in either a weekly format lasting 12 weeks or twice-weekly format lasting 6 weeks. A study indicates that it planned to deliver the twice-weekly format lasting 6 weeks. The duration of treatment for this study is defined as 6 weeks.
- ***Example 3:*** A program can be delivered over 10 weeks when the intervention is targeting a lower-risk population or over 16 weeks when the intervention is

targeting a higher-risk population. A study of this program does not provide information on the planned duration of treatment or indicate whether a lower or higher-risk population is being served. Study authors do not respond to a query for information on the duration of treatment. The duration of treatment is defined as 16 weeks for this study based on the longest duration of treatment in available documentation for the program or service (e.g., the manual, program or service website description of dosage).

Study has a defined duration of treatment that varies across participants. When the duration of treatment is clearly defined in the study and varies across study participants (e.g., a 6-session online program that participants complete at their own pace, typically taking 4–8 weeks to complete; a program in which treatment duration lasts from 3–6 months, depending on need), the Prevention Services Clearinghouse prefers the average treatment duration in the study (favoring the median duration, though a mean duration is also acceptable if the median is not available).

In some cases, not all participants will have completed treatment by the time the posttest measurement is conducted. In these cases, the average treatment duration is assessed using all participants included in the analysis of the outcome (with a preference for the median duration, but a mean treatment duration is also acceptable; either must include the current duration of treatment for those still in treatment at the time of the posttest measurement).

If information needed to assess the average treatment duration is not available in the study or upon request from study authors, reviewers will consult program or service documentation to identify whether there is an average or typical duration for the program or service delivered in the study. If so, the average duration from the program or service documentation may be used. If the program or service documentation does not provide an average but describes a planned or typical range (e.g., 3–6 months), the upper end of the applicable range is used. If a program or service manual describes multiple potential durations of treatment and the study and authors do not provide sufficient information to determine which version was used in the study, the Prevention Services Clearinghouse uses the longest duration of treatment described in the program or service documentation.

Study has an undefined duration of treatment. When the duration of treatment is not described in a study, reviewers will first query study authors to determine whether there was a defined end of treatment and whether the duration was the same across all participants. Reviewers will also request the average duration of treatment (preferring a median but the average is acceptable). If authors respond, reviewers use the conventions described above based on whether the duration of treatment in the study is the same for all participants or varies across participants.

If authors do not respond, reviewers will consult program or service documentation to determine whether there is a clearly defined duration of treatment. If the program or service has a clearly defined duration of treatment, reviewers attempt to determine whether it is the same for all participants or varies across participants and apply the conventions described above.

If a program or service does not have a clearly defined duration of treatment (e.g., is intended to be administered indefinitely to help manage symptoms of a chronic condition), reviewers assess whether there is sufficient information in the study and program or service documentation to determine a time point when a majority of a clearly defined set of services were delivered and uses that to define the duration of treatment. This may include, but is not limited to, information about a time point when participants have received a majority of services as defined in the manual, a time point when a majority of participants achieve stability during a program or service, or a time point at which services taper off or transition into a maintenance phase. The Prevention Services Clearinghouse may also confer with experts to assist with determining an appropriate end of treatment for estimating the duration of sustained effects, including, but not limited to: program or service developers; individuals with lived expertise; individuals from communities that programs and services are intended to serve, inclusive of underserved and tribal communities; methodological experts; state, tribal, and local child welfare department staff or administrators; policymakers; program or service providers or trainers; study authors; and subject matter experts. If the duration of treatment cannot be defined using these procedures, reviewers assume the posttest measurement occurred at the end of treatment.

Boosters. Some programs and services or studies may offer "booster" interventions after the end of a clearly defined set of services. In defining the end of treatment when boosters are offered, reviewers evaluate booster content, dosage, and intensity relative to the content, dosage, and intensity of the clearly defined set of services to determine whether they are minimal or non-minimal in nature:

- **Minimal.** Boosters are considered to be "minimal" if they (a) meet the standard of a "minimal intervention" (as defined in Section 4.1.7) or (b) are clearly minimal in content, dosage, and intensity relative to the defined set of services (e.g., quarterly 1-hour therapy check-in sessions for up to one year after a participant exits an intensive six-month in-home intervention with 120 hours of planned services).
- **Non-minimal.** Boosters that do not meet the "minimal" criterion are considered to be "non-minimal". In this case, reviewers apply the procedures specified above, inclusive of boosters, to define the duration of treatment.

- ***Example 1:*** After administering a brief in-person motivational intervention consisting of two 2-hour sessions, a study adds monthly 30-minute booster sessions delivered by telephone over a two-year period after the initial session. The content, dosage, and intensity of the boosters is more than minimal relative to the brief in-person motivational intervention. Reviewers would, therefore, include the booster sessions when determining the duration of treatment.

If it not clear whether boosters are minimal or non-minimal, reviewers may engage with experts to inform this determination. Experts consulted include, but are not limited to: program or service developers; individuals with lived expertise; individuals from communities that programs and services are intended to serve, inclusive of underserved and tribal communities; methodological experts; state, tribal, and local child welfare department staff or administrators; policymakers; program or service providers or trainers; study authors; and subject matter experts. Reviewers will use all available information from the study, program or service documentation, and, if applicable, queries or expert consultations to apply the booster criteria specified above.

8. Prevention Services Clearinghouse Procedures

This chapter summarizes the operational procedures the Prevention Services Clearinghouse uses to identify, screen, review, and rate programs and services. It also summarizes how the Prevention Services Clearinghouse communicates and engages with the field, including program and service developers, study authors, and other experts.

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

8.1 Prevention Services Clearinghouse Team

The Prevention Services Clearinghouse team includes federal staff, contractors, subcontractors, and consultants, as well as experts who are engaged throughout the review process. Experts who advise the Prevention Services Clearinghouse include, but are not limited to, individuals with lived expertise; individuals from communities that programs and services are intended to serve, inclusive of underserved and tribal communities; methodological experts; state, tribal, and local child welfare department staff or administrators; policymakers; program or service developers; program or service providers or trainers; study authors; and subject matter experts. All individuals who work on Prevention Services Clearinghouse reviews are expected to adhere to conflict of interest policies and sign conflict of interest declarations prior to doing any work.

8.2 Selection of Handbook of Standards and Procedures Version to Use for Review of Programs and Services

For each program or service identified and prioritized for review by the Prevention Services Clearinghouse, the version of the Handbook of Standards and Procedures used to conduct the review is clearly identified. Reviews of programs and services not previously reviewed and re-reviews of programs and services previously reviewed (see Section 8.5.1 below) that are added to the working list of programs and services planned for review (see Section 2.3) after publication of the current version of the Handbook of Standards and Procedures will be conducted using the current version of the Handbook of Standards and Procedures. Programs and services that are included on the working list prior to the publication of the current version of the Handbook of Standards and Procedures may reviewed under either a prior version or the current version, with the working list clearly identifying which version of the Handbook of Standards and Procedures is being used to conduct the review.

Study re-reviews to address errors or new information about individual studies already reviewed are conducted under the same version of the Handbook of Standards and Procedures used to conduct the currently published review of the program or service

(see Section 8.5.2 below). For example, if a study re-review is warranted for a program or service that has not been re-reviewed under the current version of the Handbook of Standards and Procedures, the previous version of the Handbook of Standards and Procedures used to conduct the original study review will be used for the study re-review.

8.3 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 the public in support of recommended programs and services (see Chapter 3). Search results are de-duplicated prior to title and abstract screening.

8.3.1 Title and Abstract Screening

The literature search produces a set of potentially relevant documents (i.e., individual articles or reports). Each document identified in the literature search is assigned to two trained staff ("screeners") to make independent assessments of its relevance to the program or service under review. Relevance screening is based on document titles and abstracts only.

Documents are marked as *relevant* (i.e., keep) if screeners 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 title or abstract appear to indicate a randomized controlled trial or quasi-experimental design was used?

Otherwise, documents are marked as *irrelevant* (i.e., drop). Senior Prevention Services Clearinghouse content experts are on hand for questions throughout the screening process.

If both screeners agree the document is irrelevant (e.g., clearly studying a different program or service, study design is clearly not eligible), it is no longer considered in the review and does not appear in the Studies Reviewed section of the program or service review on the website. If screeners disagree about the relevance of a document, senior Prevention Services Clearinghouse staff members may make a final decision on relevance.

8.3.2 Procedures for Full-text Eligibility Screening

Full-text copies of all documents deemed relevant are retrieved by Prevention Services Clearinghouse staff. Documents are combined, as necessary, into sets of documents that describe the same study.[47] Eligibility screening is performed on each study using all study eligibility criteria (as specified in Section 4.1).

Full-text eligibility screening then proceeds as follows:

- Each study is assigned to an initial screener to determine whether the study is eligible for review (see Section 4.1). If the screener needs additional information to determine the eligibility of a study, Prevention Services Clearinghouse staff send an Author Query.
- If a study is determined by the initial screener to be ineligible for review, it is assigned to a second screener for confirmation.
  - If the two screeners agree the study is ineligible and on 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 the eligibility decision or the reason a study was deemed ineligible are resolved through consensus and in consultation with senior Prevention Services Clearinghouse staff. Experts are consulted as needed (see Section 8.1 for the types of experts that may be consulted).
- If a study is eligible, screeners also assign study review prioritization points (see Section 4.2). Eligible studies are prioritized for review using the design and execution standards when there are more than 15 studies; if there are 15 or fewer eligible, all of them are reviewed using the design and execution standards. All eligible studies are reviewed for risk of harm as described in Section 7.2.1. Unfavorable contrasts in all eligible studies are 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. This ensures double screening but provides some efficiency in the process.

[47] 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 (see Section 4.1). 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.

Content experts and senior Prevention Services 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 or service described in a study is substantially adapted or not from the program or service under review. Senior Prevention Services Clearinghouse staff are available to answer methodological questions, such as whether the study design meets the eligibility requirements. Following the procedures stated in the Handbook of Standards and Procedures, content experts may identify that additional information is needed to answer questions, including querying or consulting with study authors, program or service developers, or other experts. Other experts may include, but are not limited to, individuals with lived expertise; individuals from communities that programs and services are intended to serve, inclusive of underserved and tribal communities; methodological experts; state, tribal, and local child welfare department staff or administrators; policymakers; program or service providers or trainers; and subject matter experts.

8.4 Procedures for Reviewing Eligible Studies against the Standards

8.4.1 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 a 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: high, moderate, or low. The review is completed in the database.

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

- If the first reviewer assigns a low 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 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 a reviewer needs additional information to determine the rating for a study, the reviewer drafts an author query requesting the needed information and submits it to a reconciler for review (see Section 8.4.2 below). The author query is then sent to the author.
  - If the author does not respond, the 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 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 Prevention Services Clearinghouse staff are on hand to answer questions and help interpret complicated cases.

8.4.2 Author Query Policies and Procedures

It is the policy of the Prevention Services Clearinghouse to query study authors for information deemed necessary (1) to determine the eligibility of a study for review; (2) to assign a design and execution rating of high, moderate, or low; (3) to assess statistical significance; (4) to determine the length of sustained effects; or (5) to assess risk of harm. Author queries may request information about the intervention or comparison conditions, sample sizes, baseline statistics, group formation (e.g., whether randomization was used), 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 models estimating impacts). Author queries may also request information needed to assign program or service ratings. For example, author queries may request descriptive statistics (e.g., means and standard deviations) needed to compute effect sizes and statistical significance of impact estimates, as this information may be needed to assign program or service ratings. All queries provide clearly defined timelines for authors to respond in order to facilitate timely completion of reviews. If authors do not respond to queries, the Prevention Services Clearinghouse proceeds using available information. Information provided after the defined timeline is retained by the Prevention Services Clearinghouse but is not guaranteed to be included in the review.

8.4.3 Program and Service Developer Query Policies and Procedures

The Prevention Services Clearinghouse engages program and service developers at multiple steps of the review process. If information is needed to apply the program or service eligibility, prioritization, or selection criteria specified in Chapter 2, program or service developers may receive a query requesting information or clarification. Program or service developers may subscribe to the Prevention Services Clearinghouse e-mail listserv to receive notifications of programs and services selected for addition to the working list (Section 2.3). Program and service developers are provided with a copy of the proposed public description of the program or service, inclusive of the manual citations, for comment in advance of the publication of the results of the review of the program or service. As indicated in Sections 2.3 and 4.1.9, program or service developers may be consulted as part of a process of gathering additional information if it is not clear from the program or service manual whether adaptations identified are substantial. Program and service developers are also notified if their program or service has been identified for a program or service re-review or a study re-review. Program or service developers are notified of the results of a program or service review (or re-review) through the publication of the review results on the Prevention Services Clearinghouse website. Program or service developers may submit questions to the Prevention Services Clearinghouse via email (see Section 8.6) for technical clarifications on the results of a review of a program or service.

It is the policy of the Prevention Services Clearinghouse to query program and service developers for information deemed necessary (1) to verify program or service eligibility and manual citations; (2) to understand the program or service as it is intended to be delivered; (3) to understand alternative manual editions, manual variants, or other changes in the program or service over time; or (4) to inform a public description of the program or service on the Prevention Services Clearinghouse website. All queries provide clearly defined timelines for developers to respond in order to facilitate timely completion of reviews. If developers do not respond to queries, the Prevention Services Clearinghouse proceeds using available information. Information provided after the defined timeline is retained by the Prevention Services Clearinghouse but is not guaranteed to be included in the review.

8.5 Procedures for Re-review of Programs and Services and Studies and for Updating Manual Citations

8.5.1 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 – including keeping track of programs or services where no eligible studies were identified during the review and considering ad hoc submissions of research to the Prevention Services Clearinghouse. Study authors and program or service developers can submit new evidence or studies as part of requests to re-review a program or service. Periodically, 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. These experts may include, but are not limited to, individuals with lived expertise; individuals from communities that programs and services are intended to serve, inclusive of underserved and tribal communities; methodological experts; state, tribal, and local child welfare department staff or administrators; policymakers; program or service developers; program or service providers or trainers; study authors; and subject matter experts. The public may request a re-review of the program or service rating based on the misapplication of criteria, missing information, or other errors.

Programs and services reviewed by the Prevention Services Clearinghouse under a prior version of the Handbook of Standards and Procedures also may be considered for re-review under the current version of the Handbook of Standards and Procedures if a re-review has the potential to change the program or service rating.

8.5.2 Procedures for Re-review of Studies

Individual studies reviewed by the Prevention Services Clearinghouse may also be considered for re-review due to missing information or errors in the currently published review. Requests for study re-reviews should be sent via email to the Prevention Services Clearinghouse (Section 8.6 below). The Prevention Services Clearinghouse recommends that individuals requesting a re-review due to a perceived error in the review provide supporting information to assist with review of the request, including:

- The Prevention Services Clearinghouse study number and citation(s)
- The perceived missing information or error(s) in the current review
- The relevant specific section(s) of the Handbook of Standards and Procedures to which the missing information or error(s) applies (e.g., baseline equivalence, confounds, effect size calculation, etc.)

If errors or missing information are identified, the Prevention Services Clearinghouse follows standard procedures to re-review the study. This includes assigning different, blinded reviewers to conduct any re-reviews. If the re-review determines the current 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. The specific timeline to determine if a re-review will be conducted may vary depending on the nature and complexity of the information provided and the volume of requests received.

8.5.3 Procedures for Updating Manual Citations

As specified in Chapter 2, the Prevention Services Clearinghouse selects a focal manual for its review of each program or service. The Prevention Services Clearinghouse follows standard procedures for identifying the focal program or service manual and whether alternative manual editions or manual variants have substantial adaptations from the focal manual (see Section 2.3).

Over time, newer manual editions may emerge for a program or service that has been reviewed. If updated manual editions do not have substantial modifications or adaptations from the manual reviewed, a manual citation on the website may be updated to reflect that a newer manual edition is in active use that is substantially similar to the original focal manual for the program or service. If a manual citation is updated, the Prevention Services Clearinghouse will note the original focal manual citation that was the basis for conducting the review of the program or service. However, manual editions with substantial adaptations are considered to be a separate program or service, subject to the criteria specified in Chapter 2 for program or service eligibility, prioritization, and selection.

In considering whether an update to a manual citation is warranted, the Prevention Services Clearinghouse must have sufficient information available to be able to apply the procedures specified in Section 2.3 for determining whether any substantial adaptations are present in the newer manual edition compared to the original edition reviewed. If there is insufficient information available to apply these procedures, the Prevention Services Clearinghouse will default to retaining the existing manual citation. The public may request an update of manual citation for the program or service on the basis of missing information, errors in citation, or the emergence of new information (e.g., a new manual edition being published).

Manual citations also may be updated as part of a re-review of a program or service. The same procedures specified above and criteria for assessing program or service adaptations in Section 2.3 apply in considering whether the manual citation is updated as part of the re-review of the program or service.

8.6 External Communications

To support transparency, the Prevention Services Clearinghouse communicates with study authors, program and service developers, experts, and individual members of the public via the Prevention Services Clearinghouse email account. All questions,

comments, and suggestions should be submitted via email to the Prevention Services Clearinghouse at PreventionServices@abtglobal.com. The Prevention Services Clearinghouse then forwards questions to the appropriate staff member and responds via email.

The Prevention Services Clearinghouse encourages interested parties, including, but not limited to, study authors and program or service developers to submit technical clarification questions pertaining to the results of a study or program or service review via email. The Prevention Services Clearinghouse responds to all such requests, though the timeline for response may vary depending on the number and complexity of questions received and the volume of requests received. Because of the volume of correspondence the Prevention Services Clearinghouse receives and the need to maintain a written record of key decisions, individual briefings with members of the public are not possible.

8.7 Active Engagement

In addition to communications via the Prevention Services Clearinghouse email account, the Prevention Services Clearinghouse conducts a wide range of active engagement activities to promote transparency, gather input on Prevention Services Clearinghouse standards and procedures, and collect information about programs and services relevant to the Prevention Services Clearinghouse. Active engagement of experts is incorporated throughout the systematic review process. Experts who advise the Prevention Services Clearinghouse include, but are not limited to, individuals with lived expertise; individuals from communities that programs and services are intended to serve, inclusive of underserved and tribal communities; methodological experts; state, tribal, and local child welfare department staff or administrators; policymakers; program or service developers; program or service providers or trainers; study authors; and subject matter experts in the fields of mental health, substance use and misuse, parenting and parent skill-based programs and services, kinship navigator programs, child welfare, implementation science, and cultural responsiveness.

Active engagement activities include, but are not limited to, the following:

- Engaging the public in the development of new versions of the Handbook of Standards and Procedures.
- Hosting periodic engagement sessions to gather questions and other feedback from the public to inform the Prevention Services Clearinghouse's activities.
- Conducting targeted outreach to organizations and groups to expand the reach of the Prevention Services Clearinghouse's dissemination efforts.
- Releasing annual public calls to solicit program and service recommendations.

- Hosting focus groups and consultation sessions with experts, including those with lived expertise, to discuss improvements to the website, dissemination activities, strategies for soliciting program and service recommendations, and additional resources that may be of interest to the Prevention Services Clearinghouse audience. These activities are intended to ensure that the Prevention Services Clearinghouse website and resources are accessible and actionable for diverse audiences.
- Engaging with study authors for information deemed necessary (1) to determine the eligibility of a study for review; (2) to assign a design and execution rating of high, moderate, or low; (3) to assess statistical significance; (4) to determine the length of sustained effects; or (5) to assess risk of harm.
- Engaging with program and service developers for information needed to apply the program or service eligibility, prioritization, or selection criteria specified in Chapter 2. Program and service developers are also provided with a copy of the proposed public description of the program or service, inclusive of the manual citations, for comment in advance of the publication of the results of the review of the program or service.
- Inviting public feedback on how we can continue to improve our processes and considers all suggestions received on an ongoing basis. The Prevention Services Clearinghouse recognizes that active engagement requires the collaboration and input of many individuals and organizations.

Glossary

Exhibit G1. Definitions.

Term	Definition
Adaptation	An adaptation or modification to program or service components that may be described in different manuals or in studies. The Prevention Services Clearinghouse assesses adaptations in four domains: dosage, modality, content, and providers.
Analytic sample	Individuals (or units) included in the analysis that provides impact estimates for the contrast.
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 comparison condition. A sample member has attrited from the sample if there is no outcome measurement for that sample member.
Biomarker	A physiological measure used as an indicator of a physical, psychological or emotional state.
Bundling	Adaptations that involve adding a separate program or service to an existing program or service
Comparison condition	A set of services that differ in some way from the intervention condition offered to study participants (inclusive of conditions that do not offer any planned services or delay offers of services).
Contrast	A comparison of an eligible intervention condition to an eligible comparison condition on a specific outcome for a specific posttest measurement. All Prevention Services Clearinghouse design and execution ratings are applied to contrasts.
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 intervention effect. It may also be employed in Prevention Services Clearinghouse reviews to index the differences between intervention and comparison conditions at baseline. The *standardized mean difference effect size* is used for Prevention Services Clearinghouse reviews, in the form of Hedges' *g*.
Endogenous covariate	An endogenous covariate is one that is measured or obtained after baseline and that could have been influenced by the intervention.
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.

Term	Definition
Intervention condition	The set of services offered to study participants, inclusive of, but not necessarily limited to, those related to the program or service under review.
Impact estimate	An estimate of the difference on an outcome measure between an intervention condition and a comparison condition.
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.
Manual	Refers to written or recorded books, manuals, or other documentation for a program or service that describes how to implement or administer the practice. Other documentation can include protocols, practice guidance, recorded videos, or online learning systems as long as the materials describe how to implement or administer the practice.
Manual edition	Different editions of a manual created as a program or service evolves over time or expand beyond the original developer(s) of a program or service.
Manual variant	Variants of a program or service manual designed to address new issues or different populations or that present alternative approaches to delivering the program or service.
Neighborhood	Defined as a census tract, ZIP Code, or smaller geographic unit (or similarly sized tabulation unit for studies conducted outside of the United States).
Other baseline measures	Other baseline measures are measured at baseline, and may be used as covariates in models estimating impacts, but they are not the specific measures on which baseline equivalence must be demonstrated in order to satisfy design and execution 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 comparison conditions would be expected to influence their value, or be time-invariant measures (e.g., gender).
Outcome	A behavior, skill, condition, or other characteristic that is measured to assess the impact of a program or service.
Outcome domain	A broad group of related outcomes.
Outcome measure	A survey, instrument, device, or other tool used to assess or quantify an outcome. Outcome measures must be of eligible outcomes to be reviewed. Measurement standard requirements are applied to eligible outcome measures in a study.

Term	Definition
Posttest	A posttest is measured at the end of a follow-up period, sometime after units have been exposed to or offered the intervention or comparison conditions. It is a measure on which the impact of the intervention is estimated.
Pretest	A pretest is a baseline measure of the outcome variable. Pretests, 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.
Quasi-Experimental Design (QED)	A study in which units are assigned to intervention and comparison conditions via a non-random process.
Randomized Controlled Trial (RCT)	A study in which units are assigned to intervention and comparison conditions via a random process (e.g., a lottery).
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.
Underserved communities	Underserved communities may include Black, Latino, Indigenous and Native American, Asian American, Native Hawaiian, and Pacific Islander persons and other persons of color; members of religious minorities; women and girls; LGBTQIA2S+ persons; persons with disabilities; persons who live in rural areas; persons who live in United States Territories; persons otherwise adversely affected by persistent poverty or inequality; and individuals who belong to multiple such communities.

Appendix: Summary of Revisions to the Handbook of Standards and Procedures

Development of Version 2.0 of the Handbook of Standards and Procedures was informed by public comments in response to HHS Federal Register Notices 86 FR 37332 and 88 FR 73021. Revisions were also informed by an extensive series of consultations. Experts who advised the Prevention Services Clearinghouse, either through consultations or through the submission of public comments, included individuals with lived expertise; individuals from communities that programs and services are intended to serve, inclusive of underserved and tribal communities; methodological experts; state, tribal, and local child welfare department staff or administrators; policymakers; program or service developers; program or service providers or trainers; study authors; and subject matter experts in the fields of mental health, substance use and misuse, parenting and parent skill-based programs and services, kinship navigator programs, child welfare, implementation science, and cultural responsiveness and equity. We also consulted current review standards and processes used by other prominent evidence clearinghouses, including the WWC, HomVEE, and the CEBC.

The objectives of the revision process were to be responsive to comments shared by the public and suggestions received from expert consultations, enhance the transparency of the Prevention Services Clearinghouse's standards and procedures, be responsive to the needs of underserved communities and families, and update the standards and procedures to align with current best practices for systematic evidence reviews while continuing to maintain alignment with the requirements of the authorizing legislation for the Prevention Services Clearinghouse, the Family First Prevention Services Act of 2018, as codified in Title IV-E of the Social Security Act.

The revision process also identified three areas where pilot activities will be conducted to inform future updates to the Handbook of Standards and Procedures. These pilots aim to:

- Understand the feasibility of identifying and reviewing studies only published in Spanish.
- Develop and implement standards and procedures for reviewing studies that use single case designs and determine how such designs may contribute to promising ratings.
- Understand the considerations, including methodological, timeline, and resource related considerations, for the systematic review of subgroup analyses, including how such analyses may contribute to future program or service ratings.

The remainder of this section describes the specific revisions and clarifications, organized by the chapters in the Handbook of Standards and Procedures, Version 2.0 (Handbook Version 2.0).

Introduction

The Introduction of Handbook Version 2.0 now includes a description of the review timeline and of the content included on Prevention Services Clearinghouse website. The systematic review process description and infographic have been updated to reflect the standards and procedures described in this Handbook Version 2.0 and how active engagement conducted by the Prevention Services Clearinghouse informs the review process.

Chapter 1. Identify Programs and Services

Revisions clarify that programs and services may be identified from ad hoc recommendations submitted to the Prevention Services Clearinghouse via email (PreventionServices@abtglobal.com) in addition to those submitted during public calls and that all recommendations received are retained for consideration.

Chapter 2. Prioritize and Select Programs and Services

Revisions and Clarifications to Program or Service Area Definitions (Section 2.1.1). Clarifications were made to the kinship navigator and mental health prevention and treatment programs and services definitions, with updated examples provided for mental health prevention and treatment programs and services. Revisions clarify the in-home parent-skill based and substance use prevention and treatment program or service area definitions, as noted below.

- *In-home parent skill-based programs and services*. The revised definition indicates that programs and services that involve direct intervention with a parent or caregiver and target parenting skills or other skills that can be applied wherever the child resides, including in the home, are eligible. The revised definition also clarifies that delivery of services can occur in the home or in other settings that provide skills applicable in the home. Further, the revised definition describes necessary content for a program or service to be considered “skill-based.” Revised examples of eligible and ineligible programs and services are also given to clarify the intent of revisions to the definition.
- *Substance use prevention and treatment programs and services*. The revised definition clarifies that programs or services targeting parents or caregivers to prevent substance use or misuse among children are eligible and that programs or services targeting recovery from substance use or misuse are eligible.

Revised examples of eligible and ineligible programs and services are also given to clarify the intent of revisions to the definition.

Clarifications to Available Books, Manuals, or Other Documentation (Section 2.1.2). Revisions to this section clarify that materials to satisfy this requirement may be presented in a web-based format and that manuals can include recorded videos or online learning systems if these materials describe how to implement or administer the practice. Revisions also clarify that sufficient information must be available about the program or service content, dosage, modality, providers, and/or other key components of the program or service for the program or service to be eligible for review and describes how the Prevention Services Clearinghouse conducts a comprehensive search for this information.

Revisions and Clarifications to Program or Service Prioritization Criteria (Section 2.2). Revisions clarify how program or service prioritization is informed by program or service recommendations received, evidence the program or service has been evaluated, implementation supports, and variation in program and service areas (Section 2.2). There are also two new prioritization criteria: (1) child welfare relevance – evidence that the program or service is designed for, or is commonly used to serve, children, youth, young adults, and/or families receiving child welfare services (or populations similar to those receiving child welfare services or at-risk for receiving child welfare services); (2) populations served – considering the particular needs of populations served across programs and services reviewed.

Clarifications on Program or Service Selection (Section 2.3). A new section (Section 2.3.1) clarifies how specific programs and services are selected for review and published on the Prevention Services Clearinghouse website's *working list of programs and services planned for review* ("working list").

Clarifications on How the Prevention Services Clearinghouse Handles Programs and Services with More than One Manual (*Section 2.3.2*). The Handbook Version 2.0 includes an expanded section on how the Prevention Services Clearinghouse selects a focal manual when multiple manuals for a program or service are identified, including other manual editions (i.e., editions of a manual as a program or service evolves over time or expands) and manual variants (i.e., new variants of a program or service or a manual to address new issues, different populations, or alternative approaches to delivering the program or service). This section also clarifies the standard process by which the Prevention Services Clearinghouse assesses whether manual editions or variants have any substantial adaptations relative to the focal manual and presents more inclusive criteria. A revised table of examples classifies program or service components and gives examples of adaptations considered to be substantial and those considered to be not substantial.

Chapter 3. Literature Search

To help ensure identification of studies conducted with tribal populations, the Handbook Version 2.0 now includes the Healthy Native Youth database on the clearinghouse list of consulted databases. The National Traumatic Stress Network database has also been added to the list of consulted databases. The original list of bibliographic databases has been trimmed for efficiency and resource considerations, based on a determination that some databases in Handbook Version 1.0 largely provided duplicative results. This section also clarifies that any publicly available research from program or service websites is also incorporated into the literature search. Procedures for incorporation of ad hoc submissions of research to the Prevention Services Clearinghouse via email or lists of published studies submitted during public calls are also clarified.

Chapter 4. Study Eligibility Screening and Prioritization

Minor clarifications have been made to the study eligibility criteria for source of publication (Section 4.1.2), language of publication (Section 4.1.3), and study location (Section 4.1.4). The Prevention Services Clearinghouse is planning to conduct a pilot to understand the feasibility of identifying and reviewing studies only published in Spanish.

Revisions to Study Design Criteria and Intervention Condition (Sections 4.1.5, 4.1.6). Revisions include expanded and clarified definitions for eligible *randomized controlled trials* and *quasi-experimental designs* (Section 4.1.5). The Prevention Services Clearinghouse is planning to develop and pilot standards for reviewing studies that use single case designs (SCDs). The criterion for eligible intervention conditions – that the intervention condition is offered an eligible program or service that is essentially the same for all participants in the group – remains the same as in Handbook Version 1.0, with minor clarifications, but is presented as a distinct subsection in Handbook Version 2.0 (Section 4.1.6) for clarity.

Revisions to Eligible Comparison Conditions (Section 4.1.7). The eligibility criteria for comparison conditions have been expanded and now allow for five types of eligible comparison conditions:

- *No intervention, untreated group, or wait list.* Participants are offered no services or are assigned to receive services at a later date (clarifying that outcomes measured after a wait list group is offered the intervention are not eligible).
- *Minimal intervention.* Participants may receive informational materials or psychoeducation, referrals to available services, or similar nominal services.
- *Placebo or attention control.* Includes psychological or pharmacological placebos, attention placebos, and nonspecific therapy in which participants

receive the same or similar amount of attention or contact as the participants in the intervention condition.

- *Treatment as usual.* Both "usual or typical services" (i.e., participants do not receive anything they would not have been able to receive anyway) or "services consistent with usual or typical services" (i.e., services offered as part of the study that are not offered in the community but are clearly described as consistent with the usual or typical services that would be received by individuals or families similar to those in the study) are considered eligible under treatment as usual. Pharmacological interventions that meet the definition of treatment as usual are also eligible.

- *Head-to-head comparisons*. Participants are assigned to another intervention that is not a variant of the program or service under review and does not meet the criteria for treatment as usual. Excluded are comparisons to pharmacological interventions or psychotropic medications that do not meet the definition of treatment as usual.

The Handbook Version 2.0 also clarifies the types of comparison conditions that are not eligible for review.

Revisions to Eligible Outcomes (Section 4.1.8). Clarifications to the definitions of outcomes, outcome domains, and outcome measures are provided. Clarification is provided that communal, community-level, or population-level outcomes; reductions in disparities; and attachment outcomes are eligible. Educational achievement and attainment outcomes in the child well-being domain has been expanded to include school attendance and absenteeism as eligible outcomes. Clarifications to eligible outcomes within the child safety and child permanency outcome domains already in effect and described in the *FAQ* section of the Prevention Services Clearinghouse website were incorporated directly into the Handbook Version 2.0. The definition of family functioning outcomes within the adult well-being outcome domain was revised for clarity and alignment with common definitions used in research literature. The Handbook Version 2.0 also indicates that biomarkers are not currently eligible for review.

Revisions to Study Adaptations to the Program or Service Under Review Criteria (Section 4.1.9). Handbook Version 2.0 clarifies the standard process used to identify whether adaptations to the program or service under review are present in the *studies* being screened for eligibility and how it is determined if they are substantial.

Revisions to Study Prioritization Criteria (Section 4.2). Handbook Version 2.0 clarifies that study prioritization criteria apply only when there are 15 or more eligible studies of a program or service and are intended to order the review of eligible studies so that the

highest possible program or service rating can be achieved efficiently. Multiple modifications were made to the process of assigning prioritization points to identify the order in which studies are reviewed. The criteria now give greater weight to studies that include outcomes measured 6 or 12 months beyond the end of treatment to ensure that these studies are reviewed earlier, when present. New prioritization criteria added include: two prioritization points for studies whose samples are from underserved communities, one prioritization point for child welfare relevance of the study sample, and one prioritization point for studies that report an analysis of statistical power. The prioritization criterion for study sample size was removed, and the number of prioritization points for studies with multiple outcome domains examined was reduced from three points to one point. The study prioritization criteria were also revised to improve the efficiency of reviews for programs or services with more than 15 eligible studies.

Chapter 5. Evidence Review for RCTs and QEDs Using the Design and Execution Standards

Revisions and Clarifications to Contrasts Rated, Design and Execution Rating Categories, Method of Assignment, and Integrity of Random Assignment (Sections 5.1 to 5.5). This section now indicates that contrasts from *all eligible comparison conditions* (Section 5.1) are reviewed against the design and execution standards, whereas under Handbook Version 1.0 only contrasts from the least intensive eligible comparison condition for a particular contrast were rated if multiple comparison conditions were eligible for review (Handbook Version 1.0, Section 4.1.4). Handbook Version 2.0 retains the policy from Handbook Version 1.0 of only reviewing full-sample analyses and not reviewing subgroup or sensitivity analyses, although the Prevention Services Clearinghouse plans to conduct a pilot to understand the resource requirements and potential implications of reviewing subgroup analyses using the design and execution standards and assess how such analyses may contribute to future program or service ratings. As in Handbook Version 1.0, this does not exclude review of full sample analyses that examine a specific population. The Prevention Services Clearinghouse website now displays demographic characteristics of participants and settings in studies that receive a high or moderate rating on the design and execution standards. New and revised examples are provided to clarify integrity of randomization standards for individual and cluster-assignment designs.

Revisions and clarifications to attrition, baseline equivalence, and pretest standards (Sections 5.6 to 5.8). Handbook Version 2.0 no longer requires baseline equivalence to be established for a contrast from a low attrition randomized controlled trial to receive a high support of causal evidence rating. The baseline equivalence standards now offer greater flexibility on options for demonstrating baseline equivalence on participant sociodemographic characteristics when a pretest alternative is not available. *Handbook*

Version 2.0 also describes revised procedures for baseline equivalence when establishing baseline equivalence for a binary outcome.

Revisions and clarifications to other design and execution standards (Section 5.9). The statistical model standards (Section 5.9.1) have been revised to clarify procedures used when statistical models do not meet standards and alternative statistical models are not available or do not meet standards. Clarifications to measurement reliability standards (Section 5.9.2), design confound standards (Section 5.9.3), and missing data standards (Section 5.9.4) have also been implemented.

Chapter 6. Record and Characterize Impact Estimates

Handbook Version 2.0 now includes a separate chapter that outlines the procedures used to record and characterize impact estimates to inform program and service ratings, including procedures for testing the statistical significance of impact estimates and computing effect sizes. Clarification is provided on information needed and procedures used to assess statistical significance and compute effect sizes from repeated measures models (e.g., growth curve models). In alignment with other federal clearinghouses (e.g., WWC, HomVEE), point-in-time estimates for each measurement time period are required.

Chapter 7. Program or Service Ratings

Revisions and Clarifications to Program or Service Ratings (Section 7.1) and Risk of Harm (Section 7.2.1). Minor clarifications were made to the criteria for promising, supported, and well-supported program or service ratings (Section 7.1) consistent with the Family First Prevention Services Act. This section also clarifies that the Prevention Services Clearinghouse will retain program or service ratings from reviews conducted under Handbook Version 1.0 until such time as a program or service is re-reviewed under Handbook Version 2.0. Revisions to the procedures for assessing risk of harm (Section 7.2.1) were made that address how contrasts from head-to-head comparisons are handled in the assessment.

Revisions and Clarifications to Usual Care or Practice Settings Definition (Section 7.2.2). The definition of usual care or practice settings (Section 7.2.2) has been clarified to indicate that community settings, such as schools, with embedded service providers who may provide eligible programs or services as part of their typical operations (e.g., school counselors), are also considered usual care or practice settings. It also clarifies that clinics that provide services solely for participants in research studies or clinical trials (i.e., that do not provide any services to persons not participating in research studies as part of their typical operations) do not constitute usual care or practice settings.

Revisions and Clarifications to Beyond the End of Treatment (Section 7.2.3). Revisions to this section clarify the order of preference for information that may be provided in studies about the end of treatment. Procedures for computing the duration of sustained effects when the duration of treatment is fixed, when the duration of treatment is defined but varies across participants, and when the duration of treatment is undefined are described. Treatment of boosters for computing the duration of sustained effects is also now explicitly addressed.

Chapter 8. Prevention Services Clearinghouse Procedures

The procedures for identifying eligible studies (Section 8.3) and reviewing studies against the design and execution standards (Section 8.4) remain essentially the same, with minor clarifications of operational procedures. Policies for selection of the version of the Handbook of Standards and Procedures used to conduct a review of a program or service are provided (Section 8.2). Author query policies (Section 8.4.2) have been clarified; new content clarifying the reasons the Prevention Services Clearinghouse may query program and service developers for information about programs or services has been added (Section 8.4.3). A new section describes how new manual editions are handled in the review process and procedures for updating manual citations (Section 8.5.3). This chapter now describes procedures for which version of the Handbook of Standards and Procedures will be used for reviews, program or service re-reviews, and study re-reviews.

A new section details how active engagement informs the entire review process and work of the Prevention Services Clearinghouse (Section 8.7), including how active engagement of experts is incorporated throughout the systematic review process. This section indicates that experts who advise the Prevention Services Clearinghouse include, but are not limited to, individuals with lived expertise; individuals from communities that programs and services are intended to serve, inclusive of underserved and tribal communities; methodological experts; state, tribal, and local child welfare department staff or administrators; policymakers; program or service developers; program or service providers or trainers; study authors; and subject matter experts in the fields of mental health, substance use and misuse, parenting and parent skill-based programs and services, kinship navigator programs, child welfare, implementation science, and cultural responsiveness.

Active engagement activities, include, but are not limited to, the following:

- Opportunities for the public to inform revisions to the Handbook of Standards and Procedures;
- Periodic engagement sessions to share out information and gather questions and feedback from the public;

- Targeted outreach to expand the reach of dissemination efforts;
- Annual public calls to solicit program and service recommendations;
- Focus groups and consultation sessions with experts about the Prevention Services Clearinghouse website, resources, and other activities;
- Communication with study authors and program or service developers throughout the review process; and
- Solicitation of public feedback to support continued improvement to Prevention Services Clearinghouse processes.

Sources

Administration for Children and Families. (2012). *Promoting social and emotional well-being for children and youth receiving child welfare services.* ACF Information Memorandum IM-12-04. U.S. Department of Health and Human Services, Administration for Children and Families, Children's Bureau. https://www.acf.hhs.gov/sites/default/files/documents/cb/im1204.pdf

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

Bakbergenuly, I., Hoaglin, D.C. & Kulinskaya, E. (2020). Estimation in meta-analyses of response ratios. *BMC Medical Research Methodology, 20*, 263 https://doi.org/10.1186/s12874-020-01137-1

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

Borenstein, M., & Hedges, L. V. (2019). Effect sizes for meta-analysis. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *Handbook of research synthesis and meta-analysis* (3rd ed., pp. 207–243). Russell Sage Foundation.

Collett, D. (2003). *Modeling Survival Data in Medical Research* (2nd ed.). CRC Press.

Cox, D. R. (1970). *The analysis of binary data.* Methuen & Co.

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.

Freedland, K. E., King, A. C., Ambrosius, W. T., Mayo-Wilson, E., Mohr, D. C., Czajkowski, S. M., Thabane, L., Collins, L. M., Rebok, G. W., Treweek, S. P., Cook, T. D., Edinger, J. D., Stoney, C. M., Campo, R. A., Young-Hyman, D., Riley, W. T., & National Institutes of Health Office of Behavioral and Social Sciences Research Expert Panel on Comparator Selection in Behavioral and Social Science Clinical Trials (2019). The selection of comparators for randomized controlled trials of health-related behavioral interventions: Recommendations of an NIH expert panel. *Journal of Clinical Epidemiology*, *110*, 74–81. https://doi.org/10.1016/j.jclinepi.2019.02.011

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6,* 107–128. https://doi.org/10.2307/1164588

Hedges, L. V., Gurevitch, J., & Curtis, P. S. (1999). The meta-analysis of response ratios in experimental ecology. *Ecology, 80*(4), 1150–1156. https://doi.org/10.1890/0012-9658(1999)080[1150:TMAORR]2.0.CO;2

Hedges, L. V., & Vevea, J. L. (1998). Fixed effect and random effects models in meta-analysis. *Psychological Methods, 3*, 486–504. https://psycnet.apa.org/doi/10.1037/1082-989X.3.4.486

Kish, L. (1965). *Survey Sampling*. Wiley.

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. Sage.

Ryan, T. P. (2013). *Sample size determination and power*. John Wiley & Sons.

Neyman, J., & Pearson, E. S. (1928). On the use and interpretation of certain test criteria for purposes of statistical inference: Part I. *Biometrika, 20,* 175–240. https://doi.org/10.2307/2331945

Sama-Miller, E., Lugo-Gil, J., Harding, J., Akers, L., & Coughlin, R. (2021). *Home Visiting Evidence of Effectiveness (HomVEE) systematic review: Handbook of procedures and evidence standards: Version 2.1*. OPRE Report 2021-195. Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

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

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton, Mifflin and Company.

Shadish, W. R., Robinson, L., & Lu, C. (1997). *ES: A computer program and manual for effect size calculation*. University of Memphis.

What Works Clearinghouse. (2022). *What Works Clearinghouse procedures and standards handbook, version 5.0*. U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

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

Woolf, B. (1955). On estimating the relation between blood group and disease. *Annals of Human Genetics, 19*(4), 251–253. https://doi.org/10.1111/j.1469-1809.1955.tb01348.x

Zhang, J., & Yu, K. F. (1998). What's the relative risk?: A method of correcting the odds ratio in cohort studies of common outcomes. *JAMA*, *280*(19), 1690–1691. https://doi.org/10.1001/jama.280.19.1690