



Expert Consultation Form

Please complete this form prior to the meeting. If you take additional notes during the meeting, please add them to the form prior to submission. We will ask you to share the form with us at the end of the meeting.

For each topic, the form provides excerpts from the Handbook and, in some cases, from the Family First Prevention Services Act (FFPSA). Following each set of excerpts, you will find our questions and boxes that you can use for your notes.

Please note that completing this form is voluntary. We estimate that it will take one hour to complete this form. The data collected through this information request will not be shared outside of the federal and project staff directly involved with the project. An agency may not conduct or sponsor, and a person is not required to respond to, a collection of information unless it has a currently valid OMB control number. The OMB control number for this collection is 0970-0356

[This is a generic template that includes the questions we will ask of more than 9 individuals]

General Questions

Excerpt from the FFPSA

[insert relevant FFPSA section]

Excerpts from the Handbook

[insert Handbook of Standards and Procedures section]

What adjustments or refinements would you suggest to [insert Handbook of Standards and Procedures section of interest]?

What are the pros and cons of [insert potential update to the Handbook of Standards and Procedures]?

Topic 1: Eligible Programs and Services

Excerpt from the FFPSA

Family First Prevention Services Act

“The purpose of this subtitle is to enable States to use Federal funds available under parts B and E of title IV of the Social Security Act to provide enhanced support to children and families and prevent foster care placements through the provision of mental health and substance abuse prevention and treatment services, in-home parent skill-based programs, and kinship navigator services.” (Family First Prevention Services Act, § 50702, 2018)

Excerpts from the Handbook

2.1 Program or Service Eligibility Criteria

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

2.1.1 Program or Service Areas

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

Mental Health Prevention and Treatment Programs and Services

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

Substance Abuse Prevention and Treatment Programs and Services

Eligible substance abuse prevention and treatment programs and services include those that have an explicit focus on the prevention, reduction, treatment, remediation, and/or elimination of substance use, misuse, or exposure in general. Included programs and services can target any specific type of substance, multiple substances, or aim to address substance use or misuse in general. Programs and services targeting use or misuse of alcohol, marijuana, illicit drugs, or misuse of prescription or over-the-counter drugs are eligible. Eligible programs and services can be delivered to children and youth, adults, or families. Programs and services aimed solely at reducing, treating, or remediating tobacco use (including smoking, chewing tobacco, and vaping) among adults are not eligible. Eligible programs and services can employ any therapeutic modality, including individual, family, or group and may have any therapeutic orientation, such as cognitive, cognitive-behavioral, psychodynamic, structural, narrative, etc. Programs and services may include use of pharmacological treatment approaches. Not eligible are programs and services that are directed only at collateral persons or caregivers, or systems interventions that would not generally be recognized as client-oriented substance use treatment. Additionally, programs and services that are pre-clinical programs (e.g., screening or brief programs aimed solely at getting people into treatment) and that do not themselves involve prevention or treatment are not eligible. However, brief programs that do involve prevention or treatment (i.e., make some attempt to address substance use) are eligible. Programs and services that solely rely on pharmacological interventions without a therapeutic component are not eligible (e.g., a treatment that uses methadone for the treatment of opioid addiction without an accompanying therapeutic element).

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

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

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

In-Home Parent Skill-Based Programs and Services

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

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

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

Eligible Examples	Not Eligible Examples
A program that is delivered in the family home in individual sessions for 12 weeks. Both the parent and the child attend and the parent is coached to use different skills with	An aggression reduction training for parents of adolescents that is delivered in small groups for 10 weeks and in-home delivery is not possible.*

Eligible Examples	Not Eligible Examples
the child during the session.	
An on-line parenting program that helps parents set goals and match their parenting goals with evidence-based parenting strategies.	A public service campaign that focuses on positive parenting practices is delivered in a community using television and radio spots, public posters and billboards, and direct mailings.

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

Kinship Navigator Programs

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

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

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

<p>What clarifications and refinements would you suggest for the current program and service area definitions?</p>
Empty space for user input
<p>How could the Clearinghouse broaden the definitions of the program and service areas as currently defined to include programs and services that are currently ineligible (e.g., housing) while still aligning with FFPSA? If the definitions were broadened, what are examples of programs and services that may fall under these broadened categories?</p>

Topic 2: Eligible Comparison Conditions

Excerpts from the FFPSA:

The FFPSA provided the initial guidance on the types of studies that should contribute evidence to the Clearinghouse. It makes several mentions of control or comparison conditions:

- Specifically, the FFPSA indicates that: “A practice shall be considered to be a ‘promising [supported, well-supported] practice’ if the practice is superior to an *appropriate comparison practice* using convention standards of statistical significance . . .” [italics added]. (Family First Prevention Services Act § 50711, 2018)
- The FFPSA further states that, for promising practices, the superiority of the practice must be “established by the results or outcomes of at least one study that— ... utilized some form of *control (such as an untreated group, a placebo group, or a wait list study)*” [italics added]. (Family First Prevention Services Act § 50711, 2018)
- The definitions for supported and well-supported practices also state that the practices must have “a sustained effect (*when compared to a control group*)...” [italics added]. (Family First Prevention Services Act § 50711, 2018)

Excerpts from the Handbook

4.1.4 Eligible Study Designs

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

- **Intervention Condition.** The intervention group(s) must receive a program or service that is essentially the same for all of the participants in the group (i.e., there may be variation across individuals in what they receive but distinctly different interventions should not be applied to different subsamples that are aggregated into a single study sample).
 - In a study with multiple intervention groups, reviewers determine the eligibility of each intervention based on the Program or Service Eligibility Criteria ([Section 2.1](#)). If all intervention groups are eligible, they can be reviewed and compared to the same comparison group.
- **Comparison Condition.** Comparison groups must be “no or minimal intervention” or “treatment as usual” groups. Minimal intervention group members may receive handouts, referrals to available services, or similar nominal interventions. “Treatment as usual” group members may receive services, but those

services must be clearly described as the usual or typical services available for that population in the study. Studies that compare one intervention to a second intervention are not eligible for review, even if the second intervention is not eligible under the Program or Service Eligibility Criteria ([Section 2.1](#)).

In studies with multiple comparison groups, reviewers select one comparison instead of comparing the same intervention group to multiple comparison groups. Selection of comparison group is based on the group that receives the least intensive services in order to maximize the treatment contrast.

What clarifications and refinements would you recommend regarding eligible comparison conditions to align with research practices in [topics and program or service areas of interest]?

Are there types of comparison conditions common in research literature on [topics and program or service areas of interest] that the Clearinghouse should consider including?

How might the interpretation of a significant favorable finding differ in a study where the comparison condition is: (a) no or minimal treatment; (b) treatment as usual; (c) active comparison with evidence of effectiveness; and (d) active comparison without evidence of effectiveness?

How might the interpretation of a significant unfavorable finding differ in a study where the comparison condition is: (a) no or minimal treatment; (b) treatment as usual; (c) active comparison with evidence of effectiveness; and (d) active comparison without evidence of effectiveness. Should the type of comparison condition be considered in the assessment of risk of harm?

Excerpts from the Handbook

6.1 Four Ratings

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

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

What do you suggest with regard to reviewing multi-arm studies that compare an intervention of interest to two or more comparison arms? How should multiple comparisons within a study contribute to program and service ratings?

Topic 3: Measurement of Child Welfare Outcomes

Excerpts from the FAQ page of the website

The Prevention Services Clearinghouse reviews the following domains of Child Safety:

Child Welfare Administrative Reports. Substantiated or unsubstantiated child maltreatment from administrative records. Eligible indicators include, but are not limited to, substantiated and unsubstantiated reports of abuse or neglect, investigations of abuse and neglect from administrative records, and recurrence of abuse and neglect from administrative records.

Self-Reports of Maltreatment. Eligible indicators include victim and perpetrator reports of abuse or neglect and questionnaire or interview instruments that directly assess abusive behavior or neglect.

Maltreatment Risk Assessment. Eligible indicators include child maltreatment risk assessments.

Medical Indicators of Maltreatment Risk. Eligible indicators include administrative, questionnaire, or interview instruments assessing childhood injuries, ingestions, emergency room visits, hospitalizations and any other indicators of childhood injuries.

The Prevention Services Clearinghouse reviews the following domains of Child Permanency:

Out-of-Home Placement. Any situation where a child is removed from the family home. Eligible indicators include, but are not limited to, any out-of-home placement, placement to foster care, reports of the caregiver relinquishing her or his role, and time to placement in out-of-home care.

Least Restrictive Placement. Included in this subdomain are measures that assess the restrictiveness or disruptiveness of out-of-home placement. This subdomain focuses on improving the environments/settings into which children are placed, including favoring kinship placements over non-kin or institutional placements or placements that maintain connections to the child's community versus those that do not. Outcomes must be operationalized with more than two placement settings, 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 indicators include, but are not limited to, hierarchies of least restrictive preference (e.g., kin placement, family foster care, therapeutic care, group home, residential, hospitalization, and incarceration).

Placement Stability. Placement stability refers to the stability of out-of-home placement (e.g., that children are in placements that are disrupted infrequently). This subdomain focuses on the number of placement disruptions (planned and unplanned) or number of out-of-home placements. Eligible indicators include, but are not limited to, number of placement changes or disruptions of placements, and re-entries or failed exits/reunifications or adoptions.

Planned Permanent Exits. Planned permanent exits from out-of-home care refer to placements or time to placement to a more permanent status, including reunification, guardianship, and adoption. Eligible indicators include, but are not limited to, measures of the amount of time to reunification, guardianship, or adoption and reunification rates.

[NOTE: The text below is an example of information that we might be asking the consultants to comment on.]

Proposed Revisions to the Handbook Definitions of [insert relevant outcome domains/subdomains]

- *[Provide proposed updated text for relevant sections]*

What clarifications and refinements would you suggest to the definitions for [insert relevant outcome domains/subdomains]?

What clarifications and refinements would you suggest to the definitions for [insert relevant outcome domains/subdomains]?

Excerpts from the Handbook

5.9.2 Measurement Standards

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

Face Validity

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

Reliability

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

- Internal consistency (such as Cronbach's alpha) of 0.50 or higher.
- Test-retest reliability of 0.40 or higher.
- Inter-rater reliability (percentage agreement, correlation, or kappa) of 0.50 or higher.

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

Consistency of Measurement between Intervention and Comparison Groups

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

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

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

- outcome measures must meet all of the measurement standards for a contrast to receive a moderate or high rating.
- pre-tests or pre-test alternatives that do not meet the measurement standards cannot be used to establish baseline equivalence.

What clarifications and refinements would you recommend regarding the standards for reliability and validity of measures, especially [particular topics of interest]?

Topic 4: Follow-up Timing

Excerpts from the FFPSA Legislation:

The Family First Prevention Services Act includes the following provisions for determining program and service ratings of supported and well-supported (Family First Prevention Services Act § 50711, 2018):

“(iv) SUPPORTED PRACTICE. A practice shall be considered to be a ‘supported practice’ if—

“(I) the practice is superior to an appropriate comparison practice using conventional standards of statistical significance (in terms of demonstrated meaningful improvements in validated measures of important child and parent outcomes, such as mental health, substance abuse, and child safety and well-being), as established by the results or outcomes of at least one study that—

“(aa) was rated by an independent systematic review for the quality of the study design and execution and determined to be well-designed and well-executed;

“(bb) was a rigorous random-controlled trial (or, if not available, a study using a rigorous quasi-experimental research design); and

“(cc) was carried out in a usual care or practice setting; and

“(II) the study described in subclause (I) established that the practice has a sustained effect (when compared to a control group) for at least 6 months beyond the end of the treatment.

“(v) WELL-SUPPORTED PRACTICE. A practice shall be considered to be a ‘well-supported practice’ if—

“(I) the practice is superior to an appropriate comparison practice using conventional standards of statistical significance (in terms of demonstrated meaningful improvements in validated measures of important child and parent outcomes, such as mental health, substance abuse, and child safety and well-being), as established by the results or outcomes of at least two studies that—

“(aa) were rated by an independent systematic review for the quality of the study design and execution and determined to be well-designed and well-executed;

“(bb) were rigorous random controlled trials (or, if not available, studies using a rigorous quasi-experimental research design); and

“(cc) were carried out in a usual care or practice setting; and

“(II) at least one of the studies described in subclause (I) established that the practice has a sustained effect (when compared to a control group) for at least 1 year beyond the end of treatment.

Excerpts from the Handbook

6.2.3 Beyond the End of Treatment

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

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

Program and service ratings take into consideration the length of the follow-up period after the end of an intervention (as specified in FFPSA). This is difficult to determine when interventions have no clear end point or are designed to continue indefinitely. What suggestions do you have to assess longer term impacts for such interventions that are aligned with FFPSA?

Topic 5: Design & Execution Standards

Excerpts from the Handbook

5.7 Baseline Equivalence Standards

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

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

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

5.7.1 Conducting the Baseline Equivalence Assessment

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

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

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

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

Reviewers examine the following demographic characteristics, when available:

- **Socioeconomic status.** Socioeconomic status may be measured with any of the following: income, earnings, federal (or national) poverty levels, means-tested public assistance (such as AFDC/TANF or food stamps/SNAP receipt), maternal education, employment of a member of the household and child, or family Free and Reduced Price Meal Program status.
- **Race/ethnicity.** Reviewers may assess child or parent/caregiver race/ethnicity, depending on what data are available in a study.
- **Age.** For studies of programs for children and youth, reviewers will assess baseline equivalence on child/youth age.

5.7.2 Other Baseline Equivalence Requirements

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

5.7.3 How the Baseline Equivalence Assessment Affects Evidence Ratings

Randomized Studies with Low Attrition

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

Randomized Studies with High Attrition and Quasi-Experimental Design Studies

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

What clarifications and refinements would experts suggest with regard to the current baseline equivalence standards?

Are there clarifications and refinements to the standards for pre-tests and pre-test alternatives that would be more aligned with research practices in the [insert program or service area] while still maintaining rigor?

What are the tradeoffs of using race/ethnicity and socioeconomic status to establish baseline equivalence when direct pretests or pretest alternatives are not available? Are there alternatives that would be more acceptable in a child welfare context?

The most common reason that studies do not meet design and execution standards is baseline equivalence—either baseline descriptive statistics are not reported or the baseline measures that are reported are out of balance. In addition, the majority of author queries request baseline descriptive statistics needed to establish baseline equivalence. Are there any refinements you would suggest making to the baseline equivalence standard that would continue to provide a moderate level of confidence that a study can produce a defensible causal impact estimate?

If the Clearinghouse were to review subgroup analyses, how should such analyses contribute to ratings?

What additional parameters would you suggest for determining whether or which subgroup analyses should be reviewed by the Clearinghouse (e.g., preregistered subgroup analyses; specification of confirmatory vs. exploratory analyses)?

Topic 6: General Feedback

What research design considerations, beyond those discussed so far today, might need to be part of our standards revision process?

