

## Expert Consultations Discussion Guide

*A tailored introduction will be provided for each of the sessions. Below is a list of anticipated questions to be asked of more than 9 individuals.*

*We expect this session to last about 120 minutes. We estimate that individuals will spend about 45 minutes responding to questions, dependent on the amount of information you choose to share. Your participation in this feedback session is completely voluntary. The data collected in this session will not be shared outside of the federal and project staff directly involved with the project.*

### General Questions

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

1. What clarifications and refinements would you suggest for the current program and service area definitions?
2. 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?

### Topic 2. Eligible Comparison Conditions

1. What clarifications and refinements would you recommend regarding eligible comparison conditions to align with research practices in [topics and program areas of interest]?
2. Are there types of comparison conditions common in the research literature on [topics and program areas of interest] that the Clearinghouse should consider including?
3. 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?
4. 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?

5. 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. Eligible Outcomes & Measurement Standards

1. What clarifications and refinements would you suggest to the definitions for [insert relevant outcome domains/subdomains]?
2. 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

1. 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

1. Baseline Equivalence
  - a. What clarifications and refinements would you suggest with regard to the current baseline equivalence standards?
  - b. 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?
  - c. 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?
  - d. 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?
2. Subgroup Analyses
  - a. If the Clearinghouse were to review subgroup analyses, how should such analyses contribute to ratings?

- b. 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:

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

*Thank you so much for participating in this session and sharing your helpful input. Please send your feedback forms back to {XXX}, who will also be contacting you to arrange for the honorarium payment.*