

P R E F A C E

Instrument 2: Evaluation Plan Template

PREFACE

Community Collaborations Evaluation Plan Template and Quality Indicators

Authors

Michelle Blocklin, Allison Hyra, Eliza Kean, and Allan Porowski.



Abt Associates | 6130 Executive Boulevard | Rockville, MD 20852

CONTENTS

CONTENTS

Executive Summary.....	iii
1. Introduction.....	1
1.1. Grant Purpose and Scope.....	1
1.2. Defined Target Population.....	1
1.3. Theory of Change.....	1
1.4. Logic Model.....	1
1.5. IRB Approval Plans.....	2
1.6. Evaluation Roles and Responsibilities.....	2
1.7. Feasibility of Evaluation Plan.....	2
2. Process Evaluation.....	3
2.1. Research Questions.....	3
2.2. Fidelity.....	3
2.3. Reach.....	3
2.4. Implementation Drivers, Barriers, and Solutions.....	4
2.5. Timeline.....	5
3. Outcome Evaluation.....	7
3.1. Overall Design.....	7
3.2. Research Questions.....	7
3.3. Treatment Condition.....	9
3.4. Comparison Condition.....	9
3.5. Sample Identification and Selection.....	11
3.6. Data Collection.....	11
3.7. Analysis.....	16
3.8. Timeline.....	19
Appendix A. Evaluation Plan Section Submission and Review Schedule.....	20
Appendix B: Logic Model Template.....	21
Appendix C. Fidelity Matrix.....	22
Appendix D. Contrast Table.....	23

EXECUTIVE SUMMARY

Executive Summary

This template is provided to CWCC grantees to assist in the development of their evaluation plans.¹ It includes all the required components of an evaluation plan as delineated in the FOA and provides a logical flow for describing them. This template also aligns with Children's Bureau's Evaluation Plan Development Tip Sheet (ACYF-CB-IM-19-04). The evaluation plan template includes three major sections: (1) Introduction, (2) Process Evaluation, and (3) Outcome Evaluation.

This template includes the Community Collaborations Evaluation Quality Indicators, which should guide the development of evaluation plans and the execution of those evaluations. These quality indicators provide a foundation for strong grantee evaluations. With rigorous designs, evidence developed under this project will help grantees, ACF, and the broader field make evidence-based decisions and improvements to community-level collaborations. By addressing these quality indicators in their evaluation plans when feasible, grantees also increase the chances their evaluations will be accepted for publication by a peer-reviewed journal or as a conference presentation.

Grantees are encouraged to address as many of the quality indicators as feasible. The TA team will use these indicators as a basis for feedback on grantees' evaluation plans. These quality indicators are also a technical assistance tool, as they provide concrete recommendations for both strengthening evaluation designs and solutions to address evaluation challenges.

The quality indicators outlined in this document were drawn from several sources, including [ACF's Prevention Services Clearinghouse](#) (PSC), the U.S. Department of Education's [What Works Clearinghouse](#) (WWC), the U.S. Department of Labor's [Clearinghouse for Labor Evaluation and Research](#) (CLEAR), and Abt Associates' proprietary EVIRATER™ standards. PSC, WWC and CLEAR standards focus primarily on comparison group designs such as randomized controlled trials and quasi-experimental designs, whereas Abt's EVIRATER standards address the full spectrum of evaluation designs, including pre-post and interrupted time series. Additional community-level quality indicators were drawn from Abt's experience on other projects including the *TPP Scale-Up* project for the Office of Adolescent Health.

This document lays out the structure of the evaluation plan, and then provides the quality indicators relevant for each section. They provide guidance for both participant-level and community-level evaluations. **In the sections below, quality indicators that are optional or relevant only to certain evaluation approaches are marked with an asterisk (*).**

The TA team will support the grantees' development of their evaluation plans and execution of their evaluations to ensure those plans and evaluations align with all feasible quality indicators. Grantees and evaluators should also refer to the resources on [Huddle](#) to support the development and execution of their evaluation plans. Grantees are expected to submit draft versions of sections of their evaluation plans according to the review schedule they developed in partnership with their TA liaisons. Appendix A contains a template for a schedule for the submission and review of evaluation plan sections, for grantees and TA liaisons to agree upon and complete together. This schedule helps ensure that the TA team is able to provide ongoing feedback during the plan development, with the expectation that all or almost all components of draft evaluation plans will have been reviewed at least once by the TA liaisons prior to the complete plan submission by July 31, 2020.

Once complete evaluation plans are submitted to the TA team on July 31, 2020, the TA team will review the plans in coordination with ACF. To be approved, a grantee's evaluation plan should address the recommended quality indicators to the extent possible (and provide a written explanation when an

¹ Use of this template is highly recommended, but not required. Grantees that choose to develop their evaluation plans using different headings still must ensure they provide all the information requested in this document. Use of this template will reduce the time it takes to receive feedback and approval from the TA team and ACF.

Glossary of Terms

To facilitate communication, the TA team used the following terms in specific ways:

- **Initiative/collaboration:** The totality of all partnership efforts, including work previous to, and outside of, the CWCC

Evaluation Plan Components and Quality Indicators

The following sections present a template for your evaluation plan along with corresponding quality indicators that the TA team will use to assess plans. Instructions for completing each section of the evaluation plan are shown in *italics*, each quality indicator is **bolded**, and all include a brief description. Some quality indicators may differ based on whether the evaluation is to be conducted at the individual-level or community-level. In these cases, the level of analysis (individual vs. community) is included in the title of the quality indicator.

The Paperwork Reduction Act Statement: The referenced collection of information is voluntary and will be used to systematically document *Child Welfare Community Collaborations to Strengthen and Preserve Families* (CWCC) grantees evaluation plans Information provided in this collection will be kept private. The time required to complete this collection of information is estimated to average 8 hours per response, including the time to review instructions and complete and review the collection of information. An agency may not conduct or sponsor, and a person is not required to respond to, a collection of information unless it displays a currently valid OMB control number. The valid OMB control number for this collection is 0970-0531, which expires 7/31/2022. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Abt Associates, 6130 Executive Blvd., Rockville, MD 20852, Attn: Allison Hyra.

INTRODUCTION

1. Introduction

Provide a brief summary of your overall initiative/collaboration, initiative/collaboration history, and evaluation plans (process and outcome). This description should encompass all of the work/scope of your collaboration, not just the “added value” of the CWCC grant, if applicable.

Quality Indicator: Initiative/Collaboration background, purpose, and scope description

Grantees should describe their initiative/collaboration history and any previous evaluation efforts. History could include when and why partners joined the group, the collaboration’s goal, mission, and purpose, and other major funding streams. The description should include the totality of the collaboration’s activities and efforts, including work previous to, and outside the scope of, the CWCC grant.

1.1. Grant Purpose and Scope

Briefly describe the grant’s overall approach and goal (i.e., what activities are you implementing as a result of CWCC funding and for what purposes).

Quality Indicator: Grant purpose and scope description

Grantees should briefly describe their CWCC grant activities, including the continuum of family-directed services, planned collaborations, and systems change. Grantees should also describe the challenges (e.g., lack of alignment between agencies, focus on maltreatment rather than prevention) that the grant is designed to address or ameliorate. The purpose and scope should be clearly laid out and each of the components clearly described. Grantees should clearly define how the proposed strategies, practices, policies, or activities will be operationalized. This should include a description of adaptations based on earlier pilots, usability testing, existing evidence from other fields, and input from experts in the field. Grantees should describe the selected activities (e.g., interventions, systems changes, collaborations) and the targeted outcomes.

1.2. Defined Target Population

Describe and define the grant’s target population, including characteristics of targeted families and the targeted communities or geographic catchment area.

Quality Indicator: Defined target population(s)

Grantees should clearly define and describe their target populations. They should provide a detailed description of both the geographic catchment area (including population size) and individuals targeted for the strategies, practices, or activities as well as a sound rationale for their selection (including the characteristics of the youth and families and the targeted number of individuals to whom the strategies, practices, or activities will be provided).

1.3. Theory of Change

Include a narrative describing the grant’s theory of change. After finalizing the theory of change for the implementation plan, please paste here.

Quality Indicator: Clear and thorough theory of change

Grantee process evaluation plans should include a theory of change. The theory of change should provide a broad framework for and narrative accompaniment to the logic model. It should clearly identify the theory that guides the selection of proposed activities (both participant services and systems/collaboration-level efforts) for the desired outcomes, describing the root causes of problems, the pathways to change, and the expected long-term outcomes as a result of these activities. Grantees should describe a clear, data-supported theory of change and relevant assumptions.

1.4. Logic Model

Include a logic model for the grant. A Logic Model Template is included in Appendix B.

INTRODUCTION

Quality Indicator: Clear and thorough logic model

Grantees should submit a grant-level logic model that meets the following criteria:

- Logic model includes key assumptions or contextual information;
- Logic model identifies the key components (or activities) of the grant;
- Logic model documents the inputs necessary to execute grant activities (including relevant activities put in place prior to the grant);
- Logic model identifies the mediators or intermediate outcomes through which the grant activities are expected to have its intended outcomes;
- Logic model identifies the outcome domains that the grant is designed to improve (e.g., reductions in child abuse and neglect and entry into the foster care system); and
- Logic model includes the pathways from key components to outcomes in all necessary steps.

1.5. IRB Approval Plans

Describe your plans for obtaining Institutional Review Board (IRB) approval and identify the IRB. If applicable, describe plans for Tribal review and approvals.

Quality Indicator: Plans for IRB approval

All CWCC evaluations must undergo IRB review. Grantees should describe their plans for receiving IRB approval. They should identify the IRB to be used for the review and note previous experiences with this or other IRBs.

1.6. Evaluation Roles and Responsibilities

Complete the table below identifying key evaluation team members and their roles in the evaluation. Note that both grantee and evaluator staff will likely need to be involved in developing the evaluation plan. The more communication the evaluator has with program staff the better. This communication will ensure a clear understanding of the project and that its goals are reflected in the evaluation plan.

Name	Organization	Role in Evaluation

1.7. Feasibility of Evaluation Plan

Grantees/evaluators should carefully consider and indicate in this section the feasibility of the evaluation plan proposed in the sections below within the constraints of their evaluation budget. If there are particular concerns or potential challenges in carrying out the planned evaluation activities within the evaluation budget, those should be noted here.

PROCESS EVALUATION

2. Process Evaluation

2.1. Research Questions

List the process evaluation research questions. At a minimum, research questions should address fidelity, reach, and implementation drivers, solutions, and barriers.

Quality Indicator: Explicated research questions

Grantees should identify the research questions to be answered through the process evaluation. These research questions do not need to follow the structure for the outcome evaluation outlined in Section 3.2.

2.2. Fidelity

Describe the plans to measure and report on implementation fidelity. We encourage you to complete the Fidelity Matrix in Appendix C to indicate how fidelity will be measured, calculated, and rolled up to the grant/sample level. TA liaisons are available to explain the matrix and help you to complete it.

Quality Indicator: Fidelity documentation

Grantees should describe plans to measure implementation fidelity (i.e., the extent to which activities were implemented according to plan, as designed, or as described in the literature). Four (4) criteria are associated with measurement of implementation fidelity:

1. Fidelity of implementation is measured separately for each key grant activity.
2. The entire sample (or acceptable alternative representation such as a random subsample) receiving the activity is included in implementation reporting.
3. A fidelity threshold is specified for each key component at the level of an individual unit (e.g., child, family, community) and at the project level.
4. A determination of fidelity could be made for each component (activity) at the project level.

Fidelity measurement may be adjusted throughout grant implementation based on Continuous Quality Improvement (CQI).

2.3. Reach

Include your plan for defining and measuring the reach of the grant. Participant-level reach can be defined as the number of people (parents, children, and/or families) that the grant activities will touch. Community-level reach can be defined as the number of communities served by the grant. The geographic unit of the community (e.g., county, ZIP code, census tract) should be documented. (We encourage you to consider the smallest geographic unit that is feasible and appropriate in describing your communities.)

As part of measuring reach, we also encourage you to measure, at the participant-level, some element of services received (types of services received, which organizations provided each service) and dosage (number of hours of services, whether the service was completed, percentage of service completed). If you are unable to collect any participant-level data, describe why. Common challenges include (1) access (e.g., inability to collect and combine data across multiple front-line organizations); (2) quality (e.g., concern that the percentage of target population that will consent to data collection will be too low to generalize to the actual participant population); or (3) capacity (e.g., not enough evaluation resources to support participant-level data collection).

Include your plans for collecting data, sampling, and conducting analyses to answer reach research questions.

PROCESS EVALUATION

Quality Indicator: Calculate initiative reach

Grantees should define and measure the reach or level of uptake of the grant activities (both for participants/families served and for systems, organizations, or agencies affected). For grants with direct services, grantees should track the numbers of individuals served by service type. Grantees should also define and note the number of communities (e.g., ZIP codes, counties, and census tracts) served. Reach should be measured both yearly and as an overall grant period calculation.

*Grantees may conduct geospatial analysis to map the areas served by their grant. Grantees could also use data from the American Community Survey (ACS) to map the community-level reach of their grant on indicators of community need (e.g., rates of child abuse and neglect).

Quality Indicator: Plan for reach data collection

Grantees should describe plans for collecting data on reach. The plan should include data sources, measures, who will obtain informed consent (if applicable), who will collect the data, how the data will be collected, and the frequency of data collection. Attach to the plan any developed data collection instruments, such as surveys, interview protocols, or focus group discussion guides.

Quality Indicator: Reach sample description

Grantees should identify the sample on which they will measure reach. This sample could include participants, staff/professionals (at collaboration organizations and other stakeholder organizations), organizations, and/or communities. If you will use more than one form of data collection, describe the sample for each form of data collection separately. Describe the universe of cases, the evaluation sample (if not the full universe), planned sample sizes, and sampling plan and eligibility criteria for data collection. If grantees are drawing a sample from the universe of cases, they should describe plans to assess sample representativeness of universe. Grantees should also note whether the sampling plan includes vulnerable populations, such as pregnant women, children, cognitively impaired persons, students, minorities, and economically and/or educationally disadvantaged subjects. These special classes of subjects will generally not be exempt from IRB review, and human research with children may be subject to additional state and local laws.

Quality Indicator: Plan for reach data analysis

Grantees should describe how each data element collected will be analyzed, including any checks for data quality, and whether any individually-identifiable responses will be presented in reports. Specify the frequency of analysis and analytic methods and software to be used. Reach can be summarized using descriptive statistics.

Reach research question	Data Source(s)/ Measures	Party responsible for data collection	Frequency of data collection	Sample	Expected sample size

2.4. Implementation Drivers, Barriers, and Solutions

Include the plan for documenting implementation drivers (i.e., facilitators; implementation drivers or facilitators are processes or conditions that aid in the implementation process) to implementation, barriers to implementation, and any solutions to overcoming those barriers. Plan to identify

PROCESS EVALUATION

implementation drivers and barriers at multiple levels, such as federal, cultural, state, local, agency/organization, and staff member/individual.

Quality Indicator: Document implementation drivers, barriers, and solutions

Grantees should describe plans for documenting implementation drivers/facilitators, barriers to implementation, and solutions to those challenges (if available).

Quality Indicator: Plan for implementation drivers, barriers, and solutions data collection

Grantees should describe plans for collecting data on implementation drivers, barriers, and solutions. The plan should include data sources, measures, who will obtain informed consent (if applicable), who will collect the data, how the data will be collected, and the frequency of data collection. Attach to the plan any developed data collection instruments, such as surveys, interview protocols, or focus group discussion guides.

Quality Indicator: Implementation drivers, barriers, and solutions sample description

Grantees should identify the sample on which they will measure implementation drivers, barriers, and solutions. This sample could include participants, staff/professionals at collaboration organizations and other stakeholder organizations, organizations, and/or communities. If you will use more than one form of data collection, describe the sample for each form of data collection separately. Describe the universe of cases, the evaluation sample (if not the full universe), planned sample sizes, and sampling plan and eligibility criteria for data collection. If grantees are drawing a sample from the universe of cases, they should describe plans to assess sample representativeness of universe. Grantees should also note whether the sampling plan includes vulnerable populations, such as pregnant women, children, cognitively impaired persons, students, minorities, and economically and/or educationally disadvantaged subjects. These special classes of subjects will generally not be exempt from IRB review, and human research with children may be subject to additional state and local laws.

Quality Indicator: Plan for implementation drivers, barriers, and solutions data analysis

Grantees should describe how each data element collected will be analyzed, including any checks for data quality, and whether any individually-identifiable responses will be presented in reports. Specify the frequency of analysis and analytic methods and software to be used.

Implementation drivers, barriers, solutions research question	Data Source(s)/ Measures	Party responsible for data collection	Frequency of data collection	Sample	Expected sample size

2.5. Timeline

Include a timeline for all activities of the process evaluation.

Quality Indicator: Implementation evaluation timeline

Grantees should include a timeline for all process evaluation activities, such as data collection and analysis periods.

Process Evaluation Activity	Start Date	End Date
-----------------------------	------------	----------

PROCESS EVALUATION

OUTCOME EVALUATION

3. Outcome Evaluation

3.1. Overall Design

Describe your planned overall design(s); for example, pre-post design, quasi-experimental design (QED), randomized controlled trial (RCT), regression discontinuity design (RDD), interrupted time series (ITS), participant-level, community-level, systems-level².

***³Quality Indicator: Study registration**

Study registration is an increasingly common activity undertaken during evaluation planning. Grantees using RCT or QED outcome evaluation designs may elect to register their evaluation. Registration involves providing your hypotheses and planned analyses to an outside party before starting data collection. It helps to ensure your primary findings are confirmatory (i.e., you're testing what you expected to find) and not exploratory (i.e., you're sifting through data until an interesting relationship or finding appears).⁴

3.2. Research Questions

List your research questions with four components (Target population, Treatment, Comparison condition, Outcome domain).

Note: The Children's Bureau would like all grantees to ask at least one research question addressing the extent to which the initiative was successful in connecting families previously unknown to the child welfare system with a continuum of services. In addition, research questions should reflect the standardized outcome measures agreed upon across grantees, the cross-site evaluation team, and ACF.

Examples of participant-level research questions:

1. Did the collaboration serve more non-system involved families compared to before the grant?
2. Do families who receive services provided by the grant demonstrate lower levels of parental depression after receiving services compared to before receiving services?
3. Do families who are exposed to components of the initiative (specify) have greater knowledge of available services than they did prior to exposure to the initiative?
4. Are families who are exposed to components of the initiative (specify) more likely to enroll in services compared to families who are not exposed to components of the initiative (specify)?

Examples of community-level research questions:

1. Do communities in which the CWCC initiative occurred have lower rates of entry into the child welfare system compared to similar communities where the CWCC initiative did not occur?
2. Do communities in which the CWCC initiative occurred have lower rates of placements into foster care compared to similar communities where the CWCC initiative did not occur?
3. Do communities in which the CWCC initiative occurred have higher rates of enrollment in child abuse and neglect (CAN) prevention services than they did prior to the CWCC initiative?

² Evaluations with systems-level data (e.g., organization-level) should follow guidance for individual-level data, with organizations analyzed as individual actors. If you are conducting a systems-level analysis, please consult your evaluation TA liaison for further guidance.

³ Note that * refers to quality indicators that are optional or relevant only to certain evaluation designs.

⁴ The Center for Open Science is one example of a study registry (<https://cos.io/prereg/>). Evaluation TA liaisons can help grantees select the appropriate registry for their study.

OUTCOME EVALUATION

Examples of systems-level research questions:

1. Do members of the collaborative have a stronger shared vision after the CWCC initiative than they did prior to the initiative?
2. Are collaborative partners more connected after the CWCC initiative than they were prior to the initiative?
3. Are prevention services more aligned in communities in which the CWCC initiative occurred than in communities where the CWCC initiative did not occur?

Quality Indicator: Outcome evaluation research questions

Each outcome evaluation research question should include the following:⁵

1. **Target population.** The population for which the effect of the treatment will be estimated (e.g., the age of a child during the period of exposure to the intervention).
2. **Treatment.** The treatment is the activity or set of activities that the evaluation will test and the treatment group will receive.
3. **Comparison condition.** The condition experienced by the comparison group. At a broad level, this element distinguishes between “business-as-usual” and the specific treatment that the evaluator has selected. For a pre-post design or interrupted time series, the comparison condition would be “pre-treatment.”
4. **Outcome domain.** The general, or high-level outcome that may be affected by the treatment; it can be thought of as a latent construct that can be measured with one or more outcome measures.

All research questions need to align with the project’s logic model. At least one research question should address an intermediate outcome as depicted in the logic model (e.g., changes in risk and protective factors amongst participant families, changes in access to/uptake of services by targeted families). All outcomes indicated in your research questions should be included in your logic model. However, you are not expected to include all outcomes in your logic model in your research questions.

Research questions should also be designated as either confirmatory (i.e., those upon which outcome evaluation conclusions will be drawn) or exploratory (i.e., those that might provide additional suggestive evidence).

Research Question	Target population	Treatment	Comparison condition	Outcome domain
1				
2				
3				
4				
5				
6				

3.3. **Treatment Condition**

Describe the “treatment” that will be tested. That is, what components of the grant will the “treatment group” portion of the evaluation sample (e.g., families, collaboration partner organizations, communities) be exposed to? For sampled families, the “treatment” will be the set of the activities they will be receiving, such as case management services. For sampled organizations, the treatment might be

⁵ https://www.acf.hhs.gov/sites/default/files/opre/pii_approach_to_evaluation_brief_508.pdf

OUTCOME EVALUATION

participating in program eligibility alignment. What is the continuum of services that evaluation participants will be provided? What is the process by which participants will be offered and receive the continuum of services? What collaboration efforts will partner organizations participate in? What components of the initiative will communities be exposed to? For community-level evaluations, you will need to note the proportion of the targeted population that will be exposed to the treatment (your planned “saturation”).

Quality Indicator (Community-level only): Clear definition and description of treatment community (ies)

Grantees should clearly define their treatment communities (e.g., catchment areas). Treatment communities should be conceptually defined as geographic areas in which (1) all components of the grant are available to community members; and (2) a substantial proportion of the targeted population are directly or indirectly affected by these components. Communities could be defined by a variety of geographic units (e.g., counties, cities, ZIP codes, census tracts, school districts). It is best to define communities at the smallest geographic unit possible, although it is important that the selected geographic unit aligns with the available data (e.g., county-level child welfare data). If treatment communities cannot be aligned with available data, the grantee should document this disconnect, including the level of disconnect, and explain why such an alignment was not possible. It is also important for the grantee to describe the treatment community context (e.g., policies, initiatives, legislation related to risk and protective factors for child maltreatment).

Quality Indicator (Community-level only): Saturation calculated

Grantees should calculate saturation for each treatment community, defined as the proportion of the targeted population that has been exposed to one or more components of the grant. Community-wide efforts such as mass media (e.g., radio, newspaper, internet, or TV ads) where saturation is assumed to be 100% should be noted, but excluded from other saturation calculations. Grantees should also calculate saturation for direct services if their grant provides direct services (and this calculation is feasible).

3.4. Comparison Condition

Describe the “compared to what” for each research question. Comparison conditions could include the treatment group pre-treatment (e.g., treatment group at pretest in a pre-post design) or a separate comparison group (e.g., families or organizations similar to the treatment group families or organizations but will not be exposed to the treatment). At the community level, comparisons could include the treatment community(ies) before the grant (e.g., at pretest in a pre-post design) or (an) other similar community(ies) that do not participate in the grant activities. Comparison conditions may be at the individual or community level and include:

- a. Non-treated individuals or communities (randomly chosen)*
- b. Non-treated individuals or communities (not randomly chosen)*
- c. Pre-intervention data from treatment group of individuals or communities*
- d. Benchmark data (published evaluation data that shows gains in similar outcomes from other studies with similar interventions)*
- e. Progress against goal (evidence-based targets for change in an outcome, likely based on evaluation literature)*

Describe how each comparison condition was selected and what (if any/if known) relevant programming the comparison condition may be exposed to or have the opportunity to participate in (e.g., other collaboration efforts, statewide initiatives, or comparison-community CAN prevention efforts).

OUTCOME EVALUATION

****Quality Indicator: Comparing pre-post changes to norms***

The credibility of evidence from a pre-post design can be improved if the pre-post gain of the treated group can be compared to appropriate population norms that correspond to the same time interval between the pretest and posttest measurement (e.g., benchmarking). Grantees could compare evaluation pretest and posttest values to a reference group that approximates a policy-relevant reflection of the evaluation sample. For example, grantees could compare grant activity participants' gains in parenting knowledge to gains demonstrated by similar (non-treatment) populations as documented in published evaluation literature. Note that it is important to ensure that the comparison you make includes a sample that is as similar as possible to your evaluation sample (e.g., samples should reflect a similar risk/protective profile).

****Quality Indicator (Individual-level QEDs only): Clear definition and description of comparison individuals***

To increase the rigor of an individual-level analysis, grantees could include a group of comparison individuals who do not receive the intervention. Individuals in the comparison group should be similar to those who received the intervention, but should not receive the intervention themselves. Statistical methods can be used to select similar comparison individuals or to correct for differences between the groups. Baseline equivalence should be calculated to determine whether the treatment and control groups are comparable at baseline (see baseline equivalence quality indicators in section 3.7).

****Quality Indicator (Community-level QEDs only): Clear definition and description of comparison community (ies)***

To increase the rigor of a community-level analysis, grantees could include data across their state from non-treatment communities as an indication of what might have happened in the absence of the grant. (This approach is recommended, because if grantees are already requesting administrative data for treatment communities, it is likely feasible to request data for communities across the state.) Ideally grantees would use comparison communities similar to the treatment communities but that did not implement a similar community collaboration initiative. This can be accomplished by selecting comparison communities similar to the treatment communities, or by using statistical methods to correct for differences between the two groups. Good comparison communities are *local* (close to the same locale as the treatment communities, ideally in the same state to control for state-level policies and resources available) and *focal* (have similar characteristics as the treatment communities; see baseline equivalence quality indicators in Section 3.7). Grantees must describe comparison community contexts (e.g., policies, initiatives, legislation related to risk, and protective factors for child maltreatment).

****Quality Indicator (QEDs and RCTs only): Avoid confounds***

Even when treatment and comparison groups are relatively similar, there may be other characteristics that fundamentally bias the research. To avoid confounds, aside from treatment status, the comparison group should not share a given characteristic, which is different from the treatment group. For example, if all treatment individuals were teen mothers and all comparison individuals were mothers over age 30, or if all treatment communities were in urban settings and all comparison communities were in rural settings, it will be impossible to disentangle the effect of the intervention from the effect of mother's age or the setting.

****Quality Indicator (RCTs only): Random assignment process***

Random assignment to treatment and comparison groups increases an evaluation's rigor. In order for random assignment to produce two balanced groups, the random assignment process must assign individuals or clusters (e.g., families, communities) entirely by chance and maintain those assignments throughout the study period (e.g., a caseworker cannot decide a family assigned to the control group really should be provided grant-funded activities, and an evaluator cannot take participants assigned to the treatment group who do not receive any treatment out of the analysis or analyze them as part of the control group). The probability of assignment does not need to be 50%; however, each individual or cluster should have the same chance of being assigned to the treatment group versus the comparison

OUTCOME EVALUATION

group as do other individuals or clusters. Randomization may be compromised by researchers or providers in a number of ways. In order to maintain the integrity of the random assignment process, the individuals (or clusters) originally assigned to each condition must remain in that condition for the analysis, regardless of their adherence to the study condition (e.g., even if an individual assigned to the treatment group never receives treatment, they should still be analyzed as part of the treatment group, or if an individual assigned to the control group receives some of the treatment, they should still be analyzed as part of the control group). This is known as an intent-to-treat approach. All randomly assigned individuals or clusters should remain in the sample throughout the study, and individuals or clusters not randomly assigned should not be included in the analytic sample. Action should also be taken to avoid crossovers (e.g., individuals assigned to the comparison group who receive treatment), as this may dilute the treatment effect.

3.5. **Sample Identification and Selection**

Describe the outcome evaluation sample(s) (i.e., the participants/families, organizations, and/or communities that are contributing data to the evaluation) that will be used to address participant-level, community-level, or systems-level research questions. Describe how the sample(s) will be identified, evaluation eligibility criteria, planned sample sizes, and the sampling plan for data collection. Participant-level samples will likely be all or a subset of non-system-involved high-risk families who engage in activities of the initiative. If all families who receive services will be included in the evaluation, describe eligibility for receiving services provided by the grant. For community-level research questions, define the treatment community(ies) and indicate potential comparison communities or plans for identifying and selecting comparison communities. For systems-level research questions, identify the organizations or the types of staff from whom data will be collected.

Quality Indicator: Sample description

Grantees should describe the universe of cases, the evaluation sample (if not the full universe), planned sample sizes, and sampling plan and eligibility criteria for data collection for the outcome evaluation. For individual-level evaluations, grantees should collect data from at least 200 individuals for confirmatory analyses. Grantees should also note whether the sampling plan includes vulnerable populations, such as pregnant women, children, cognitively impaired persons, students, minorities, and economically and/or educationally disadvantaged subjects. These special classes of subjects will generally not be exempt from IRB review, and human research with children may be subject to additional state and local laws.

3.6. **Data Collection**

3.6.1 Outcome Measures and Domains

List and describe outcome domains and constructs, corresponding measures, and their reliability and face validity (or plans to establish reliability and validity). Include citations for existing measures.

To answer the Children's Bureau research question about whether the collaboration was able to serve at-risk families previously unknown to the child welfare system, you should measure some aspect of the following elements: characteristics of families reached prior to the start of the treatment (e.g., demographics such as family composition [number of people in the household and relationship to focal child/ren], race/ethnicity, ages, geographic location, primary language/language spoken at home); risk/protective factors (e.g., measures related to child maltreatment such as parenting attitudes, knowledge, and beliefs; parental stress/resilience; family support/need; parental mental health and/or depression; and parental substance abuse); and whether the family (parent and/or child) had any contact with a child protection agency.

If you are unable to collect any participant-level data, describe why. Common challenges include (1) access (e.g., inability to collect and combine data across multiple front-line organizations); (2) quality (e.g., concern that the percentage of target population that will consent to data collection will be too low

OUTCOME EVALUATION

to generalize to the actual participant population); or (3) capacity (e.g., not enough evaluation resources to support participant-level data collection).

Per grant requirements, grantees will need to include the standardized outcome measures agreed upon across grantees, the cross-site evaluation team, and ACF.

Grantees will also likely collect system-level measures, such as measures of collaboration, cooperation or alignment, and community-level measures, such as community awareness of prevention efforts, rates of reported or substantiated child maltreatment, or rates of entry into foster care. The indicators below apply to measures at the participant, system, and community-level.

Quality Indicator: Outcome measures should be included in logic model

All outcomes measured in the outcome evaluation should be included in the logic model, and at least one intermediate outcome included in the logic model should be measured. However, not all outcomes included in the logic model need to be measured in the outcome evaluation.

Quality Indicator: Outcome measures should be reliable

Grantees should use reliable outcome measures. To be considered *reliable*, each outcome measure should meet one or more of the following criteria: internal consistency (such as Cronbach's alpha) of 0.50 or higher, test-retest reliability of 0.40 or higher, and inter-rater reliability (percentage agreement, correlation, or kappa) of 0.50 or higher.⁶ If a measure does not have documented reliability, the grantee should describe plans for assessing reliability. Standard administrative measures (e.g., substantiated allegations of child maltreatment, entry into foster care) are assumed to be face valid and reliable.

Quality Indicator: Every outcome measure should have face validity

In order for evaluations to draw valid conclusions, grantees should use outcome measures that provide a valid and fair assessment of the initiative's results. A measure with *face validity* is clearly defined, has a direct interpretation, and measures the construct it was designed to measure. When applicable, measures should also demonstrate cultural relevance and valid language translation. Community-level evaluations should also provide a rationale for measuring certain outcomes at the community level (vs. the individual level).

Quality Indicator: Outcome measures should be sensitive to change

Grantees should include outcome measures that are sensitive to change given the timing of measurement and sample size. Outcome measures should not be included if they are not expected to change within the study period. For example, if the evaluation only allows for a period of two weeks in between pre- and post-test data collection, an outcome such as parent-child relationship quality would not be sensitive to change in this time period and should not be included. Similarly, while static measures such as demographic measures (e.g., race and ethnicity) or measures of past experiences (e.g., the Adverse Childhood Experience (ACE) Questionnaire⁷) may provide important information about participants, they are not appropriate for measuring change and should likely not be included as outcome measures. In addition, measures with smaller anticipated change will require a larger sample to detect this change.

***Quality Indicator (RCTs and QEDs only): Outcomes measures should not be over aligned**

Outcome measures should not be too closely aligned or tailored to the intervention being tested. This typically occurs when an outcome measure is created by researchers or intervention developers specifically for a single study. Evidence of over-alignment might include an outcome measure that assessed respondents using some of the same materials that are part of the intervention, which could give the intervention group an unfair advantage over the comparison group. For example, an over aligned

⁶ What Works Clearinghouse Protocol, pp. 22-23.

⁷ Note that the ACEs might be appropriate to use in other aspects of your evaluation such as when you are describing the individuals reached through your initiative in the process study.

OUTCOME EVALUATION

measure of parenting practices would be to ask participants if they used the 1, 2, 3 magic technique to address negative child behaviors (which only program participants would know), rather than using a standardized positive parenting skills scale (which would be applicable to all families whether they participated in the program or not). Standardized measures or measures that have been used in other studies are unlikely to be over-aligned.

3.6.2 Data Collection Plan

Describe your data sources, measures, who will collect the data, how it will be collected, and the timing for pretest and posttest. Provide a brief rationale for the timing of data collection (e.g., Is timing based on time from enrollment, completion, pre-test? If based on completion, how will you know whether/when a participant has completed their engagement with grant activities? Will it allow enough time for change in the outcome?). For individual-level data collection, describe your plan for tracking participants for follow-up data collection. Please attach to the plan any developed data collection instruments, such as surveys, interview protocols, or focus group discussion guides.

Quality Indicator: Pretest and posttest measurements should be identical (and if not identical, highly correlated)

Most evaluations will include a pretest that is the same as the posttest, so it can be assumed in these instances there is near-perfect correlation between the two. If a grantee uses a pretest measure that is not the same as the posttest (e.g., if the posttest measure is not available at pretest or if a measure has changed over time), the pretest measure must be reasonably correlated with the posttest to serve as a proxy. The correlation between pretest (or the collection of baseline covariates used in the analytic model) and posttest measures must be at least .30 (or equivalently, an r-square of 0.09 for posttest regressed on the one or more pretest measures). If documentation is not available from the measure developers, correlation between pretest and posttest should be established during the evaluation. In cases where quantitative data are not available (e.g., pretest parenting measures amongst first-time expectant parents), the pretest should at a minimum have face validity.

Quality Indicator: Consistent outcome measurement

Grantees should use consistent measurement methodologies. Quality indicators for consistency of measurement of the outcome are:

- The same measures must be used at all pre and post time points and across respondents (if the same measures are not used, they will still be acceptable if grantees normalize outcomes via z-scoring using population means and standard deviations).
- The data collectors, data collection modes, and timing of data collection for each measure either are the same across all participants or are different in ways that would not be expected to have an effect on the measures. (Note: Data collectors and modes should be the same at pretest and posttest periods, and timing of data collection should be consistent across respondents within pretesting or post testing.)
- *For RCTs and QEDs only:
 - Measures must be constructed in the same way (i.e., rely on the same questions and be calculated in the same way) for both treatment and comparison groups. For example, you should ensure that reports of child abuse are determined to be substantiated in the same way across treatment and comparison groups.
 - The data collectors and data collection modes for treatment and comparison groups are either the same or are different in ways that would not be expected to have an effect on the measures. For example, you should not collect data via in-person interviews in the treatment group and via online surveys in the control group. However, we would not expect the use of two different online survey platforms to have an effect on the data collected.

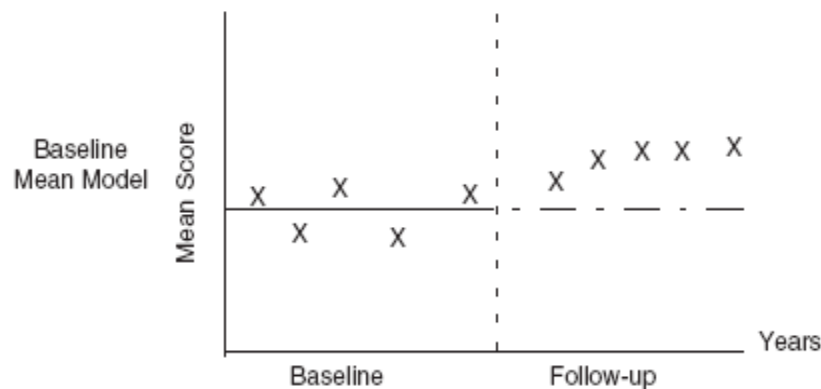
OUTCOME EVALUATION

- The timing of data collection must be consistent across study conditions (i.e., baseline data must be collected at approximately the same time for both the treatment and comparison groups), so that the amount of time between pre-test (baseline) and post-test (outcome) measures does not systematically differ between treatment and comparison groups.

***Quality Indicator (When using administrative data, particularly for community-level studies): Multiple pretest measures**

A pre-post community-level design can be improved if grantees include multiple years of retrospective data prior to the initiative and control for these baseline projections in the analytic model. The design feature described here has been described as a short-interrupted time series (SITS) design.⁸ In Exhibit 2, where there are three or more pre-treatment measurements at appropriate intervals, this quality indicator could be satisfied by demonstrating graphically or statistically that a baseline mean projection is appropriate, and using a model to estimate the impact of the initiative in which actual posttest measurement is compared to the value predicted by the baseline-mean projection.

Exhibit 2: Example of a Baseline-Mean Projection Model from Bloom (2003)



Outcome research question	Data sources (and measures)	Sample	Party responsible for data collection	Data collection method	Frequency/timing
1					
2					
3					

3.6.3 Data Sharing/Data Use Agreements

Describe your plans for obtaining data sharing and data use agreements. Please attach draft, final, and/or executed agreements if available.

⁸ See: Cook, T. D., Campbell, D. T., & Shadish, W. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin. Also see: Bloom, H. S. (2003). Using “short” interrupted time-series analysis to measure the impacts of whole-school reforms: With applications to a study of accelerated schools. *Evaluation Review*, 27(1), 3-49.

OUTCOME EVALUATION

Quality Indicator: Data sharing/data use agreements

Grantees should describe their plans for obtaining data sharing and data use agreements with all relevant organizations to obtain the necessary data to complete the outcome evaluation. Plans should document anticipated problems with and solutions to sharing data and securing final agreements. Data sharing agreements should indicate that aggregated data might be shared with the national evaluator.

3.6.4 Consent/Assent Procedures

Describe your plans and procedures for obtaining consent/assent when needed. You should secure consent/assent from all individuals providing data specifically for the evaluation, including participant families and collaboration partner organization staff.

Quality Indicator: Consent/assent procedures

Grantees should describe plans and procedures for obtaining necessary consent and/or assent for data collection. Procedures and consent/assent forms should ensure all evaluation subjects (e.g., participants, partner organization staff, community stakeholders) know what they are agreeing to, allow them to opt out of the evaluation and still receive services (if appropriate), identify any potential risks of participation, and be translated into other languages as necessary. All consent forms need to document how data may be shared with partner agencies and with Abt Associates as the cross-site process evaluator.

3.6.5 Data Security Procedures

Include plans for establishing and following measures to ensure the security of the data collected, both primary and secondary data (e.g., administrative). Describe any plans to archive the data.

Quality Indicator: Data security procedures

Grantees should indicate how their data will be stored to ensure data security and note procedures should there be a security breach. Grantees should also indicate how they plan to transmit data. The TA team recommends grantees use a secure file transfer system (e.g., secure FTP, MoveItDMZ, or Huddle).

3.6.6 Data Quality

Include your plan for minimizing missing data to achieve a 75%-80% response rate for post testing. Also describe how you will deal with missing data and conduct data quality checks and data cleaning prior to conducting analysis.

Quality Indicator: Data quality

Grantees should propose target response rates (e.g., 75%-80%) for post testing and plans to achieve these target response rates. Grantees should also propose a plan for dealing with missing data (e.g., complete case analysis, dummy variable approach) and plans for minimizing missing data (e.g., ensuring survey completion, minimizing losing sample to follow-up). In addition, grantees should propose data cleaning/quality checks (e.g., outliers, inconsistencies in the data, implausible values for certain variables) and how any issues will be addressed. For example, a grantee might propose a parent-child age check, where parent ages are recoded to missing if they are within 10 years of the child's age.

3.7. Analysis

3.7.1 Analysis plan

Include your plans for conducting analysis, noting models you will run and software you will use.

Quality Indicator: Analysis plan

The analysis plan should include a plan for statistical and qualitative data analysis. Grantees should note a pre-specified cutoff for statistical significance. The TA team suggests $p < .05$ for statistical significance and $p < .10$ as trend-level significance. The analysis plan should also address the confidentiality of respondents, including minimum cell size requirements for data presentation (i.e., to ensure readers cannot deduce the identity of an individual's response). We recommend a minimum cell size of 10 in reporting.

OUTCOME EVALUATION

3.7.2 Contrasts

Describe the test/contrast that will answer each of your research questions, and note whether the test is confirmatory (i.e., those upon which you will draw outcome evaluation conclusions) and or exploratory (i.e., those that might provide additional suggestive evidence).

Quality Indicator: Contrasts

Each contrast should clarify the four components listed for each research questions (Target population, Treatment, Comparison condition, Outcome domain) plus the evaluation design (e.g., RCT, RDD, QED, ITS, pre-post), unit of assignment (units receiving the treatment, such as children, families, systems, communities), outcome measures (instrument, scale, extant data source), sample eligibility criteria (selection criteria or restrictions placed on the analytic sample for each test/contrast), and pretest measure (instrument, scale, measure construction, inclusion in analytic model). To address this quality indicator, we suggest completing the contrast table in Appendix D.

3.7.3 Subgroups (optional)

Describe any subgroups for which you will conduct additional analyses (e.g., teen parents, parents of children under 5, racial/ethnic subgroups, etc.). You will likely not have a large enough organization-level or systems-level sample to conduct subgroup analysis.

3.7.4 Covariates/Decision Rules

Indicate which (pre-treatment) covariates you will include in the model. For example, in testing whether participant family knowledge of available services improved, you may want to control for the number of years the family lived in the target community prior to the initiative. Also describe any decision rules for dropping a covariate from the model (e.g., p-value is greater than .10).

3.7.5 Baseline Equivalence (QEDs only)

Describe your plans for conducting tests of the equivalence of treatment and comparison groups at baseline (pre-treatment). Also describe your plans to increase the likelihood of establishing baseline equivalence between treatment and comparison groups (i.e., treatment and comparison groups should not differ in pretest measures of the outcome variables).

***Quality Indicator: Baseline equivalence between treatment and comparison groups**

Grantees using comparison groups can strengthen their evaluation by documenting differences between the treatment and comparison groups in the analytic sample prior to the implementation of the grant activities (at baseline). Small or nonexistent differences between the treatment and comparison groups prior to grant implementation (baseline equivalence) means the evaluation can better attribute treatment-comparison differences to the grant.

Baseline equivalence should be established on:

- At least one demographic factor (e.g., race/ethnicity, percentage of families headed by single parents); and
- At least one socioeconomic factor (e.g., socioeconomic status indicator, percentage living in poverty); and
 - *For participant-level:* pre-treatment values of the outcome variable when available (e.g., risk or protective factors at baseline); or
 - *For community-level:* at least one community-level indicator of CAN (e.g., rate of foster care entry, rate of abuse/neglect reports, and rate of parental substance use).

Treatment and comparison groups in the analytic sample should not differ by more than 0.25 standard deviations on any of the three categories of baseline equivalence indicators noted above; and at least one

OUTCOME EVALUATION

baseline equivalence indicator variable from each category should be included as a covariate in the final analysis of data.

3.7.6 Attrition (RCTs only)

Describe your plans for calculating attrition. Also describe your plans for minimizing attrition between random assignment and follow-up data collection.

***Quality Indicator: Attrition in individual-level RCTs⁹**

Individual-level RCTs should calculate attrition. Attrition is defined as the number of individuals who are not present for the posttest outcome measurement as a percentage of the total number of individuals in the sample at the time of random assignment. This quality indicator includes an assessment of both overall attrition (total sample loss between randomization and the post-test), and differential attrition (percentage difference in attrition between the treatment and control group). Table X provides the thresholds for both overall and differential attrition rates, which are based on the What Works Clearinghouse¹⁰ and OPRE's Prevention Services Clearinghouse¹¹ standards. If attrition is beyond the threshold, then the study is considered a quasi-experimental design and should establish baseline equivalence (see section 3.7.5 above).

Table X. Highest Differential Attrition Rate for a Sample to Maintain Low Attrition, by Overall Attrition Rate, Under “Optimistic” and “Cautious” Assumptions (What Works Clearinghouse)

Overall Attrition	Differential Attrition		Overall Attrition	Differential Attrition		Overall Attrition	Differential Attrition	
	Cautious Boundary	Optimistic Boundary		Cautious Boundary	Optimistic Boundary		Cautious Boundary	Optimistic Boundary
0	5.7	10.0	22	5.2	9.7	44	2.0	5.1
1	5.8	10.1	23	5.1	9.5	45	1.8	4.9
2	5.9	10.2	24	4.9	9.4	46	1.6	4.6
3	5.9	10.3	25	4.8	9.2	47	1.5	4.4
4	6.0	10.4	26	4.7	9.0	48	1.3	4.2
5	6.1	10.5	27	4.5	8.8	49	1.2	3.9
6	6.2	10.7	28	4.4	8.6	50	1.0	3.7
7	6.3	10.8	29	4.3	8.4	51	0.9	3.5
8	6.3	10.9	30	4.1	8.2	52	0.7	3.2
9	6.3	10.9	31	4.0	8.0	53	0.6	3.0
10	6.3	10.9	32	3.8	7.8	54	0.4	2.8
11	6.2	10.9	33	3.6	7.6	55	0.3	2.6
12	6.2	10.9	34	3.5	7.4	56	0.2	2.3
13	6.1	10.8	35	3.3	7.2	57	0.0	2.1
14	6.0	10.8	36	3.2	7.0	58	-	1.9
15	5.9	10.7	37	3.1	6.7	59	-	1.6
16	5.9	10.6	38	2.9	6.5	60	-	1.4

⁹ If you are conducting a cluster-level RCT, where clusters are assigned and individuals within the cluster are analyzed, please contact your evaluation TA liaison for further guidance.

¹⁰ These attrition thresholds were designed to tolerate a maximum bias of .05 standard deviations. See the WWC Procedures and Standards Handbook, version 2.1 (p. 34) for a discussion of attrition bias:
https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_procedures_v2_1_standards_handbook.pdf#page=38

¹¹ <https://preventionservices.abtsites.com/>

OUTCOME EVALUATION

17	5.8	10.5	39	2.8	6.3	61	-	1.1
18	5.7	10.3	40	2.6	6.0	62	-	0.9
19	5.5	10.2	41	2.5	5.8	63	-	0.7
20	5.4	10.0	42	2.3	5.6	64	-	0.5
21	5.3	9.9	43	2.1	5.3	65	-	0.3

Source: WWC Technical Paper on [Assessing Attrition Bias](#).

Note: Overall attrition rates are given as percentages. Differential attrition rates are given as percentage points. Not every combination of differential and overall attrition is possible for any given study. The evaluation should specific plans to use the cautious or optimistic boundary depending on the anticipated potential for attrition bias.

3.7.7 Social Network Analysis (if planned)

Indicate which entities (e.g., collaborative members, provider organizations) will be in the network, and what their connections will be (e.g., collaboration, referrals). Describe your plans for data collection and analysis. [Note: The TA team can recommend online software that you can use to survey respondents and conduct the network analysis.]

***Quality Indicator: Social network analysis**

As collaboration is a key goal of the Community Collaborations grants, grantees might conduct a social network analysis to assess success in achieving collaboration and to provide additional context for the individual-level and community-level outcome evaluations. A social network analysis allows grantees to capture the level of collaboration achieved between grantees, service providers, and/or other entities in targeted communities. Grantees can survey relevant organizations to ask which other organizations they make referrals to and which other organizations they receive referrals from, or which organizations they collaborate with to accomplish the goals of the grant.

To increase the rigor of the network analysis, grantees can do one or more of the following:

- Analyze networks prior to the grant (either through recall or by asking at the outset of implementation) and then at a later point.
- Ask about and portray the intensity of the connections. Intensity can be reflected in the number of referrals made/received or the amount of formal communications that takes place between partners.
- Link centrality measures (i.e., how central the organization is in the network) of each of the organizations in the network to their numbers served to see whether there is a relationship.
- Test whether centrality measures of an organization where an individual received services moderate pre-post change in individual outcomes (e.g., Do individuals treated by organizations that are more central in the network have greater improvements in outcomes?).

3.8. Timeline

Include a timeline of all your outcome evaluation activities.

Quality Indicator: Outcome evaluation timeline

Grantees should include a timeline for all outcome evaluation activities, such as IRB submission, waves of data collection, analysis, interim and final report writing/submission.

Outcome Evaluation Activity	Start Date	End Date

OUTCOME EVALUATION

APPENDIX A. EVALUATION PLAN SECTION SUBMISSION AND REVIEW SCHEDULE

Appendix A. Evaluation Plan Section Submission and Review Schedule

The TA team is providing this evaluation plan development timeline template to support grantee and evaluator planning for a thorough plan to be submitted on July 31, 2020. Because some of the evaluation plan elements build on one another (e.g., you need to clearly define your project activities before you can complete your logic model); we have ordered sections beginning with those we think are most important to complete early on. Note that we recommend the outcome and process evaluation designs to be developed alongside each other.

Your TA liaisons are prepared to discuss and review portions of your evaluation plan as you draft them and provide you with feedback. We believe that this back and forth/ongoing feedback process is the best way to keep you on track for an on-time submission in July, ensure the plan will be approved by ACF, and ensure the plan will provide a strong foundation for your evaluation.

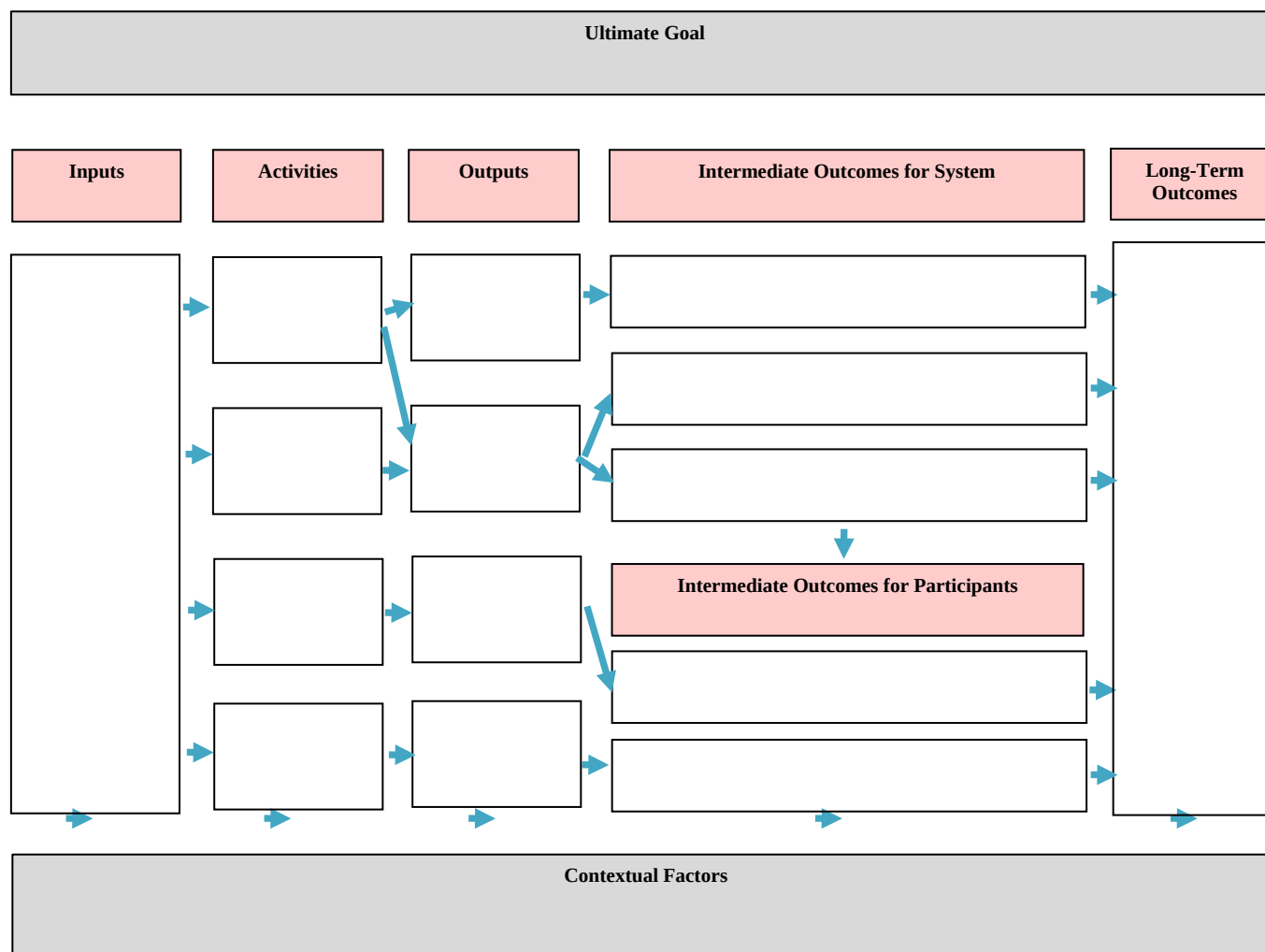
This list of evaluation sections aligns with the evaluation plan template. If you address each of these sections, you will have a completed plan. You should work with your TA liaison to determine a schedule for submitting each of the sections in the table below.

Evaluation Plan Section(s)	Draft Completion Date	Submitted to TA Team? ✓
Introduction and Grant Purpose and Scope		
Revised logic model and theory of change		
Defined target population		
Finalize research questions (process and outcome)		
Treatment and comparison conditions (Outcome Evaluation)		
Fidelity Matrix		
Reach and Implementation Drivers, Barriers, and Solutions		
Outcome Study Sample		
Outcome study measures and domains		
Outcome study data collection plan		
Outcome study analysis and contrast table		
IRB approval plans		
Data sharing/Data use agreements, Consent/assent plans and procedures, data security procedures, data quality		
Process and outcome evaluations timelines		
Complete Evaluation Plan	July 31	

APPENDIX B: LOGIC MODEL TEMPLATE

Appendix B: Logic Model Template

Grant: (name) Logic Model (use text boxes: add/change boxes and arrows as needed)



APPENDIX C. FIDELITY MATRIX

Appendix C. Fidelity Matrix

Indicators	Definition	Unit of implementation	Data source(s)	Data collection (who, when)	Score for levels of implementation at unit level	Threshold for adequate implementation at unit level	Roll-up to next higher level if needed (score and threshold): Indicate level	Roll-up to next higher level if needed (score and threshold): Indicate level	Roll-up to grant level (score and threshold for adequate implementation at sample level)	Expected sample for fidelity measure	Expected years of fidelity measurement
Key Component 1											
Indicator 1											
Indicator 2											
Indicator 2											
Indicator 3											
Indicator 4											
Indicator 5											
All indicators											
Key Component 2											
Indicator 1											
Indicator 2											
Indicator 2											
Indicator 3											
Indicator 4											
Indicator 5											
All indicators											

APPENDIX D. CONTRAST TABLE

Appendix D. Contrast Table

Below we provide a contrast table, including examples for two research questions.

Research Question: Confirmatory/ Exploratory	Design	Target Population*	Sample Eligibility Criteria	Treatment Group	Comparison Group	Outcome			Baseline (if applicable)	
				Treatment Description*	Condition/ Description*	Domain*	Unit of assignment/ observation: Measure [Scale]	Timing of measurement	Unit of assignment/ observation: Measure [Scale]	Timing of measurement
RQ 1	C-ITS	Target zip codes	All zip codes in state	All project/ collaborative activities	Comparable zip codes in state (not served by project/ collaborative)	Child abuse	Zip code: # confirmed cases of child abuse	Spring 2020 Spring 2021 Spring 2022	Zip code: # confirmed cases of child abuse	Spring 2015 Spring 2016 Spring 2017 Spring 2018 Spring 2019
RQ 2	Pre-post	Family Navigation Participants	All families who participate in navigation	Navigation	Navigation participants prior to intervention	Protective Factors	Individual participants: Protective Factors Survey	6 months after first navigation session (Spring 2020 – Spring 2022)	Individual participants: Protective Factors Survey	First navigation session (Fall 2019 – Fall 2021)

* Indicates one of the four components of your outcome evaluation research questions

Example Research Question 1: Did the zip codes targeted by the Initiative/Collaborative have lower rates of confirmed cases of child abuse than comparable zip codes not targeted by the Collaborative (and without a similar intervention)?

Example Research Question 2: To what extent did protective factors improve among recipients of navigation services compared the baseline period?