12-2020

Technical Reviewing for the Family First Prevention Services Act: Strategies and Recommendations

Antonio R. Garcia
University of Kentucky, antonio.garcia@uky.edu

Peter J. Pecora
University of Washington

Audrey H. Schnell
The Analysis Factor

Cynthia Burnson
National Council on Crime and Delinquency Children's Research Center

Elizabeth Harris
National Council on Crime and Delinquency Children's Research Center

See next page for additional authors

Technical Reviewing for the Family First Prevention Services Act: Strategies and Recommendations

Digital Object Identifier (DOI)
https://doi.org/10.1016/j.childyouth.2020.105597

Notes/Citation Information

Published in *Children and Youth Services Review,* v. 119, 105597.

Authors

Antonio R. Garcia, Peter J. Pecora, Audrey H. Schnell, Cynthia Burnson, Elizabeth Harris, and Allison Finseth

Children and Youth Services Review 119 (2020) 105597

Contents lists available at ScienceDirect

Children and Youth Services Review

journal homepage: www.elsevier.com/locate/childyouth

Technical reviewing for the Family First Prevention Services Act: Strategies and recommendations

Antonio R. Garcia[a,*], Peter J. Pecora[b], Audrey H. Schnell[c], Cynthia Burnson[d], Elizabeth Harris[d], Allison Finseth[d]

[a] University of Kentucky, College of Social Work, 655 Patterson Office Tower, Lexington, KY 40506, United States
[b] School of Social Work, University of Washington, 2001 8th Ave, Floor 25, Seattle, WA 98121, United States
[c] The Analysis Factor, 430 W State St, Suite #204, Ithaca, NY 14850, United States
[d] National Council on Crime and Delinquency Children's Research Center, United States

ARTICLE INFO

Keywords:
Child welfare
Family First Prevention Services Act (FFPSA)
Technical review
Evidence-based practice
Foster care
SafeCare
Family-Centered Treatment

ABSTRACT

The Family First Prevention Services Act (FFPSA) has compelled states to expand their priorities to implement evidence-based practices (EBPs) as a means to prevent foster care placement. While the states may opt to include EBPs already approved by the Administration for Children and Families (ACF), some state leaders are opting to commission an independent technical review for the EBP they would prefer to implement as part of their FFPSA plan. While the goal is for ACF to approve their plan and issue a temporary license, little guidance is provided on how to conduct technical reviews. Relying upon the expectations that ACF has outlined for each state, we illustrate the process for conducting reviews of SafeCare in Iowa and Utah and of Family-Centered Treatment in Arkansas. Despite FFPSA and ACF guidance, rendering an evidence rating was difficult given the variability in how some studies measured baseline equivalence, lack of robust testing methods, and conflicting findings across studies. We conclude with recommendations on addressing these challenges and strategies for conducting high-quality technical reviews. The review process offers an opportunity to synthesize a large body of research to inform child welfare practice.

1. Introduction

1.1. Context

The U.S. child welfare system is inundated with reports of child maltreatment every 10 s (Childhelp, n.d.). This amounts to roughly 3.5 million reports of suspected maltreatment, of which nearly 675,000 are substantiated (ACF, 2019). The system and its dedicated workers are often taxed as they grapple with how best to address the needs of children who experienced maltreatment and the overwhelming number who remain at imminent risk in their home environment. Child welfare scholars have long debated whether efforts should focus primarily on being reactive or proactive, with the former focusing on intervention after a maltreatment incident. The recently enacted Family First Prevention Services Act (FFPSA) represents a major attempt to focus more on prevention than on reactive responses. The federal law, passed on February 9, 2018, as part of the Bipartisan Budget Act, reforms the federal child welfare financing streams (Title IV-E and Title IV-B of the Social Security Act) by allowing federal reimbursement for evidence-based mental health services, substance use treatment, and in-home parenting skills training as a way of preventing children from entering foster care (Torres & Mathur, 2018). The law requires that to be eligible for funding, states may only use interventions that have proven to be effective in rigorously designed studies and includes provisions for establishing the Title IV-E Prevention Services Clearinghouse (the Clearinghouse) to determine which programs meet the inclusionary threshold as outlined by the Administration for Children and Families (ACF). These standards are codified in the Title IV-E Prevention Services Clearinghouse Handbook of Standards and Procedures, hereafter referred to as the Handbook (Wilson, Price, Kerns, Dastrup, & Brown, 2019).

For some states, leaders are opting to commission an independent technical review for the evidence-based practice (EBP) they would prefer to implement, hoping that ACF will issue a temporary evidence rating until such time that the Clearinghouse completes their lengthy evidence review. For example, the Arkansas Division of Children and Family Services submitted independent technical reviews for Family-Centered Treatment (FCT) and Youth Villages Intercept, which were approved as Well-Supported and Supported, respectively. Independent reviews are therefore

* Corresponding author.
E-mail addresses: antonio.garcia@uky.edu (A.R. Garcia), PPecora@casey.org (P.J. Pecora), Audrey@theanalysisfactor.coms (A.H. Schnell), CBurnson@nccdglobal.org (C. Burnson), EHarris@nccdglobal.org (A. Finseth).

https://doi.org/10.1016/j.childyouth.2020.105597
Received 15 June 2020; Received in revised form 7 October 2020; Accepted 7 October 2020
Available online 16 October 2020

becoming an important pathway for states to obtain support for programs that are tailored to their community's needs.

However, other than the Handbook and a few other sources, there is little guidance on how to conduct valid, reliable technical reviews. What types of knowledge, insights, and collaborations help facilitate the review process? In this article, we examine the technical evidence review process for (1) SafeCare for Iowa and Utah and (2) FCT for inclusion in the Arkansas FFPSA plan. We highlight ACF expectations, the additional information included in the reviews, and the collaborative decision-making process. While by no means a prescriptive manualized process, our case study of each technical review may shed light on (1) how reviewers select, critique, and rate studies in relation to the federal FFPSA criteria; (2) how reviewers justify a final designation for a program or service intervention; (3) challenges that reviewers experience; and (4) strategies they employed. We conclude with recommendations for addressing challenges as teams navigate the technical review process.

This article is not intended as a critique of the independent review process or of the Clearinghouse's criteria. Rather, as the first study of the process that independent reviewers are using to evaluate programs for FFPSA, we aim to describe and document the process. The article is intended both as a pragmatic tool for states and researchers who are considering an independent review and as documentation of the child welfare field's response to a radical shift in funding allocation and of how evidence-based practices are identified and scrutinized.

1.2. Overview of FFPSA eligibility and stipulations

Unlike other federal legislation, FFPSA allows states to use federal monies for prevention services in addition to adoption, foster care, and kinship supports. However, funding is limited and states must adhere to stringent conditions. Listed below are those who are eligible for FFPSA benefits. States that do not comply lose access to the relevant funding.

- A child identified in a prevention plan as being at imminent risk of entering foster care who could safely remain at home or in a kinship placement if services (i.e., mental health services, substance abuse prevention and treatment services, or in-home parent skill-based programs) are provided to prevent the child from going into foster care. This includes children whose adoption or guardianship arrangement is at risk of disruption or dissolution, which would then result in a foster care placement.
- A child in foster care who is pregnant or parenting a child.
- The parents or kin caregivers of such a child.
- Under FFPSA, Title IV-E funding is limited to 12 months from the date the child is identified in a prevention plan as either a "candidate" or a "pregnant or parenting foster youth." However, services can be reauthorized for the same family if still needed to prevent placement after 12 months (Federal Register, 2018).

In addition, states must select interventions or services that meet the following criteria.

- Trauma-informed.
- Comply with general practice requirements, inclusive of descriptions of components and administration of the practice (e.g., a manual), does not constitute risk of harm, and multiple studies support benefits of practice.
- Meet the requirements for a "Promising, Supported, or Well-Supported practice." Beginning in fiscal year (FY) 2024, at least 50% of FFPSA funds need to be spent on Well-Supported interventions. A follow-up piece of legislation signed into law at the end of 2019, the Family First Transition Act, temporarily loosens these requirements. For FY 2020 and 2021, FFPSA funds may be used for Promising, Supported, or Well-Supported prevention programs. For FY 2022 and 2023, 50% of FFPSA funds must be for Supported or Well-Supported programs (H.R. Resolution 2020).

1.3. Brief Overview of the technical review process

A number of states are commissioning technical reviews to support or justify their selection of EBPs that meet the aforementioned criteria. Factors such as accessibility, costs, consideration of a special target population, major service gaps (e.g., parenting skills, domestic violence, or trauma), communities without services ("service deserts"), and other areas of concern may need to be prioritized. Some states may have experienced success with a particular program and wish to continue funding it. To that end, some states may commission technical reviews for interventions that the Clearinghouse chose not to prioritize at this time or has not completed a review of. As noted in Table 4, some of these states include Arkansas, Colorado, Iowa, Kentucky, Nebraska, and Utah.

Consulting firms, foundations, and research institutes, including Casey Family Programs, the National Council on Crime and Delinquency (NCCD), and the Social Research Institute at the University of Utah are spearheading technical reviews of several EBPs, including but not limited to 1–2–3 Magic, cognitive behavioral therapy, dialectical behavior therapy, FCT, Youth Villages Intercept, High Fidelity Wraparound, Homebuilders, and SafeCare. While a limited number of technical reviews have been approved (i.e., FCT and Youth Villages Intercept in the spring of 2020), not much is known about how the technical review process is being implemented. The Clearinghouse set forth some broad guidelines that technical reviewers must adhere to, as explained in the *HHS Initial Practice Criteria and First List of Services and Programs Selected for Review as part of the Title IV-E Prevention Services Clearinghouse*. The guidelines include the following conditions.

Service or Program Eligibility: Eligibility is limited to mental health, substance abuse prevention and treatment services, and in-home parent skill-based programs, as well as kinship navigator programs. Services or programs must have a book, manual, or other available documentation that specifies the components of the practice protocol and describes how to administer the practice.

Service or Program Prioritization Criteria. Timing and resources may not allow for the Clearinghouse to conduct a detailed review of all services and programs that meet the Service or Program Eligibility Criteria. Thus, services or programs are prioritized based on whether outcomes address child safety, child permanency, child well-being, and adult (parent and kin caregiver) well-being and on whether they are kinship navigator programs that include the aforementioned outcome domains. Reviews are also limited to services or programs currently in use with a book, manual, or other documentation available in English and to programs that have implementation training and staff support and/or fidelity monitoring tools and resources available in English. Although the Clearinghouse publicizes what is being reviewed, it does not provide information about projected completion date, underscoring one reason why states have undertaken technical reviews.

Study Prioritization Criteria. Timing and resources may not allow the Clearinghouse to review all studies within a selected service or program determined to be eligible according to Handbook criteria. The order and depth of study reviews are determined on the basis of study features that may include sample size, duration of sustained effects examined, and type of study design.

Study Rating Criteria. The Clearinghouse rates studies using the following criteria:

- Study design and execution. Building from the standards of existing evidence reviews such as the What Works Clearinghouse (WWC) and Home Visiting Evidence of Effectiveness (HomVEE), the Clearinghouse will assess studies on the basis of study design, overall and differential sample attrition, the equivalence of intervention and comparison groups at baseline (as applicable), and when necessary, procedures accounting for clustering. In addition, the study must account for confounding factors and examine at least one target outcome using a measure that is reliable and achieves

face validity. Inconsistencies in systematic administration, as noted in study text, will also be considered. Studies will be rated as "high," "moderate," or "low." The study-level ratings will provide an indicator of the extent to which a study provides unbiased estimates of model impacts.

- Effects. The following effects, defined using conventional standards of statistical significance, will be examined in the full analysis sample for studies that achieve a "high" or "moderate" rating on Study Design and Execution:
  - Favorable Effects. Studies will be rated based on whether they demonstrate at least one meaningful favorable effect (i.e., positive significant effect) on a 'target outcome.'
  - Unfavorable Effects. Studies will be rated based on the number of unfavorable effects (i.e., negative significant effects) on either 'target' or non-target outcomes.
  - Sustained Favorable Effect. Studies with at least one meaningful favorable effect on a 'target outcome' will be rated on whether or not they demonstrate a favorable effect sustained beyond the end of treatment. Is there a customary criteria (statistically) for a sustained favorable rating? Does the effect have to remain statistically significant at follow-up?
  - Studies will be classified as not demonstrating a sustained favorable effect (i.e., effects are demonstrated for less than 6 months), demonstrating a sustained favorable effect of 6 months or more (but less than 12 months), or demonstrating a sustained favorable effect of 12 months or more. Initially, due to time and resource constraints, the Clearinghouse will use only effects resulting from analyses of the full study sample for rating. This decision may be reconsidered in the future.

Service or Program Rating Criteria. The Clearinghouse will rate a service or program as a 'promising,' 'supported,' or 'well-supported' practice if it meets the below criteria:

- *Promising Practice:* A service or program has at least one study that achieves a rating of 'moderate' or 'high' on Study Design and Execution, and demonstrates a favorable effect on at least one 'target outcome.'
- *Supported Practice:* A service or program has at least one study carried out in a usual care or practice setting that achieves a rating of 'moderate' or 'high' on Study 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.
- *Well-Supported Practice:* A service or program has at least two studies with non-overlapping analytic samples carried out in a usual care or practice setting that achieve a rating of 'moderate' or 'high' on Study Design and Execution. At least one of the studies must demonstrate a sustained favorable effect of at least 12 months beyond the end of treatment on at least one target outcome.
- *Does Not Currently Meet Criteria:* A service or program will be rated as 'does not currently meet criteria' if the service or program has been reviewed and does not currently meet the evidence criteria for 'promising,' 'supported,' or 'well-supported" practices.

1.4. Research Aims: Parameters for conducting a technical review

In a document called *Attachment B: Checklist for Program or Service Designation for HHS Consideration,* ACF has outlined what is expected of each state during this transition period if they would like to review and claim a particular intervention whose research evidence has not yet been rated by the Clearinghouse (See https://www.acf.hhs.gov/cb/resource/pi1906). States are permitted to commission a technical review of the research evidence of an intervention, as long as they document the review process by completing the 11 tables outlined by the Clearinghouse.

The technical review is lengthy and arduous. How states and research teams grapple with adhering to the expectations is largely unknown, leaving many wondering how to interpret, process, and implement the criteria set forth by the Clearinghouse. The best strategies and techniques for conducting valid and reliable technical reviews for FFPSA have not been illuminated. Relying upon two technical reviews, we illustrate the process of conducting these reviews for (1) SafeCare in Iowa and Utah and (2) FCT in Arkansas. In doing so, we address the following questions.

1. What does an empirically robust technical review entail? In other words, what processes are involved with justifying a final evidence rating designation for a program or intervention?
2. What are the major hurdles one might encounter during the review process?
3. How might some of the challenges be addressed?
4. What strategies might be helpful to consider for other new technical reviews?

2. Methods

2.1. Recent federal clarifications

Some key provisions of the technical review process are listed below, adapted from Pecora, Garcia, & Schnell, 2020.

The June 22, 2018 Federal Notice reiterates that the intervention evaluation focus on impact. The intervention must have been published in "government reports and peer-reviewed journal articles that assess effectiveness (i.e., impact) using quantitative methods" (Federal Register, 2018). These guidelines specify that HHS does not intend to include "access to services, or satisfaction with programs and services" as target outcomes. This might have a negative impact on kinship navigator programs, given that the evaluation process focuses on kinship navigator program goals. While an exception may be made, these programs may want to broaden their outcomes measurement to include rates of child maltreatment, runaway episodes, placement disruptions, attainment of legal permanency, child emotional and behavioral health, and other measures of child well-being.

Technical reports provided to government or peer-reviewed journal articles are needed as documentation for an intervention's effectiveness. This allows inclusion of waiver evaluation reports and other government-funded initiatives with evaluation requirements. Note that states have requested that "government reports" include reports submitted to city, county, state, and tribal governments in addition to the federal government and other funders such as foundations.

The notice requests comment on populations that may be considered "similar" to those involved in the child welfare system (Federal Register, 2018, Section 2.2.2). Because children and parents may enter the child welfare system due to emotional or behavioral diagnoses, interventions developed for and tested in mental health and related programs should be considered "similar." This should include drug and alcohol treatment programs, including medically assisted substance abuse treatment programs.

Intervention studies are restricted to certain countries and must have been published during 1990 and later. The study must have been conducted in the United States, United Kingdom, Canada, New Zealand, or Australia and published/prepared in English during or after 1990 (Federal Register, 2018, Section 2.3.3).

2.2. Study selection process

A fairly broad literature review is necessary to (a) identify which studies are eligible to be included in the technical review and (b) to uncover whether the intervention has a study where negative effects on clients have been found. A more focused set of studies is then examined to see if the intervention meets the Clearinghouse standards for a particular evidence level in certain outcome areas. According to the

Handbook, inclusion criteria include the following.

- Peer-reviewed journal articles and/or publicly available literature that may include, but is not limited to federal, state, and local government and foundation reports.
- Study designs that assess effectiveness (i.e., impact) using quantitative methods and utilize an appropriate control. Eligible study designs include randomized controlled trials (RCT), quasi-experimental designs (QED), and other non-experimental designs that use an appropriate control.
- Studies that examine the impact of the service or program on at least one of the target outcome domains: child safety, child permanency, child well-being, and adult (parent and kin caregiver) well-being. Target outcomes for studies of kinship navigator programs will include all outcome domains listed above, as well as access to, referral to, and satisfaction with services and programs.
- Studies available in English.
- Studies published or prepared during or after 1990 (Wilson et al., 2019, pp. 5–8).

From the list above, one might wonder how a study is defined. Consider these instructions from Section 4 of the Handbook where a "study" is defined as:

> …. one research investigation of a defined subject sample, and the interventions, measures, and statistical analyses applied to that sample. For example, sometimes study results are reported in more than one document. Therefore, the Prevention Services Clearinghouse uses the Institute of Education Sciences What Works Clearinghouse (WWC) v4.0 convention. This convention states that two or more impact estimates will be considered as coming from a single study when they share at least three of the following four characteristics:

a. The particular sample used to estimate the impact is the same or has a large degree of overlap.
b. The process used to assign sample members to intervention and control conditions is the same.
c. The data collection and analysis procedures are the same (or nearly the same).
d. The research team is the same or has a high degree of overlap (page 9 and footnote no. 1).

The Handbook lists the bibliographic databases and Clearinghouse repositories that must be searched to identify possible publications to review. Given that some of these sources have been shut down (e.g., Child Trends What Works), for this study of technical reviews, we searched all of the required sites and noted which ones were still operating and what studies or publications were found on each site. We found overlap in what each site includes and some unique listings as well. After all possible publications were located, we reviewed them to identify which publications report data from studies that would be eligible for Clearinghouse review versus other areas of focus such as intervention modifications, small case studies of how consumers responded to the intervention, strategies for implementation, and use of the model with a particular age, gender, or racial/ethnic group.

3. Results

3.1. Section overview

We conducted two of the first independent case studies of technical reviews to be submitted by states and approved by ACF while following Handbook standards as closely as possible. We use the process and results from these studies to demonstrate adherence to the Handbook standards for each of the 11 required tables, all separated into one of the following five sections.

- Summary of Programs and Services Reviewed and Their Designations (Table 1)
- Standards and Procedures for a Systematic Review (Tables 2 and 3)
- Review of Programs and Services (Tables 4 and 5)
- Review of "Well-Designed" and "Well-Executed Studies (Tables 6–10)
- Program or Service Designation (Table 11)

3.2. Section i

Overview. In Table 1, the program or intervention is named and the recommended evidence rating is provided.

SafeCare Review. In this case, we found and cited a website that listed many of the SafeCare publications,[1] constructed an appendix of the 116 SafeCare publications we found, and noted reasons why we excluded most of them, such as the study focused on an adaptation of the main model or focused on implementation issues. The proposed rating, based on the careful review, was Well-Supported. The justification to support this rating was spelled out in subsequent sections and tables.

FCT Review. The same basic process was followed. The program's website was helpful for identifying studies, as was searching the reference list of the most recent studies on FCT. This process was easier in that there are only a few major studies of FCT. The most common reason for exclusion for quantitative studies was the lack of a comparison group. One challenge was determining whether publications were about FCT or other programs with similar or identical names or key words. Carefully reading methods sections and determining key authors and research groups was helpful in discerning which studies to keep.

3.3. Section II

Overview. Table 2 is a key summary table for many of the questions that need to be considered and answered after study review and analysis. Tables 2 and 4 contain concise lists of the specific criteria to be used, based on the Handbook. While specific levels are not provided for judging effect sizes, a table and figure with boundaries are provided in the Handbook to assess subject attrition.

Table 2 contains this question for reviewers: *"Author or Developer Queries:* Were systematic standards and procedures used to query study authors or program or service developers? (Applicable if author or developer queries made)." This refers to the Handbook section on page 49 that describes what topics commonly might be asked of developers or evaluators. Unfortunately, the Handbook does not prescribe *how* those requests should be made.

Documentation of reviewer qualifications is recorded here. We summarized that information in bio-sketch summaries that include information about the reviewer's program evaluation or meta-analysis experience, and our curricula vitae.

SafeCare Review. In Table 2, using a checklist, we verified the standards and procedures used to conduct the review. For example, we verified that the review included conducting a comprehensive literature review, determining study eligibility, assessing study design and execution, examining study effects in usual care or practice settings, considering risk of harm, and providing a final evidence level designation for the intervention. We also provided bio-sketch summaries that include information about the reviewer program evaluation experience and curricula vitae. In Table 3, documentation of the independent status of the reviewers and conflict of interest assurances were included. Note that the state must have each technical reviewer sign a conflict of interest form, so allow time for that form to be located or

[1] https://safecare.publichealth.gsu.edu/safecare/safecare-research/publications/.

developed.

FCT Review. The same process was followed. The reviewers verified that they had not been evaluators of the program, nor been in any kind of financial relationship with the program developers.

3.4. Section III

Overview. In this section, the Clearinghouse standards require confirmation that a manual is available, including a complete reference for that manual and how it can be obtained. All the versions of the manual that were used in each of the studies must be included. For example, only a manual published in 2016 is cited, but one or more of the studies was conducted in 2012, it will be unclear which manual was used to guide the practitioners in that 2012 study. The manual used in the 2012 study should be provided as well. Table 5 is where design and intervention fidelity are described for each study; and also, how the treatment and comparison groups differ.

SafeCare Review. In Table 4, we confirmed the manual and where it is published. Table 5 included critical elements of the technical review: study title/authors, where articles can be obtained, language verification, study design, whether the manual was used with those assigned to receive the intervention condition (i.e., fidelity), comparison condition, target outcome, year published, and if the study is eligible for review. Three QEDs (Beachy-Quick, Lee, McConnell, Orsi, Timpe, & Winokur, 2017; Chaffin et al., 2012; Gershater-Molko et al., 2002; 2003) met inclusion criteria for review.

All of the elements in Table 5 were verified. We included footnotes and reprinted tables and figures from the actual article or report to more thoroughly provide details about the design, intervention group, comparison group, and fidelity. For example, we provided details about (1) how the intervention and comparison groups differed (i.e., comparison groups did not receive SafeCare or they received services as usual); and (2) how study protocol adhered to intervention fidelity by relying upon certified coaches who used fidelity checklists or monitoring assessments.

FCT Review. Program manual and fidelity processes were reviewed and confirmed by accessing the FCT website, referencing descriptions in the research publications, and through follow-up conversations with program developers and researchers. Study design and other eligibility criteria were confirmed. In contrast to the SafeCare approach, our review verified these elements and provided citations and a brief description in Attachment B, but did not provide extensive documentation as part of the submission.

3.5. Section IV

Overview. In Table 6, documentation is provided that verifies that the literature review used the same systematic standards and procedures across all programs if focusing on more than one program. Verification of the comprehensive screening of studies to see which ones met all of the eligibility criteria in accordance with the Handbook, is documented in Table 6 (Sections 3 and 4). Electronic copies of all the study reports and articles must be provided.

Baseline equivalence and subject attrition are documented in Table 7. The two groups being compared must demonstrate equivalence at baseline to establish that the comparison is valid and that any difference in outcomes is due to the intervention, not pre-existing differences. If there is a direct pre-test available, then that is the variable on which baseline equivalence must be demonstrated. If a direct pre-test is not available (which is often the case for child welfare outcomes such as reunification), an alternative pre-test may be used. This must be a variable conceptually or known to be associated with the outcome. If a suitable pre-test alternative is not available, baseline equivalence must be established on both race/ethnicity and socioeconomic status. In addition, reviewers must examine balance on child age, when available, for all contrasts (defined as a comparison of a treated condition to a counterfactual condition on an outcome) (the Handbook, Section 5.1). Those with available pretests or pretest alternatives must also be assessed.

While the guidance is not explicit, it appears that effect size differences at baseline in excess of the 0.25 standard deviation threshold for these three contrasts (when available) may be a basis for a low rating or may require that a study be excluded from review as these may be viewed as evidence of confounds in the design or execution.

While a "Notes" column is provided to the far right in the table, it is too narrow and notes may be included below the table. Major study confounds are also identified in this section. The Handbook defines two types of confounds: the substantially different characteristics confound between the treatment and control groups, and the n = 1 person-provider or administrative unit confound (pp. 35–37). Tables may need to be extracted from the original study publications to document entries for this table.

Table 8 instructions contain a typo: columns ii and vi must be answered with a "Yes" (not column v). Tables may need to be extracted from the original study publications to document entries for this table. Information on the sample demographics and characteristics of the comparison group are provided for each study.

Table 9 should only include target outcomes with favorable effects. Section 5.10 in the Handbook defines favorable effects as statistically significant and in the desired direction. Note that one of the entries in the ACF sample table appears to be wrong in that the example includes a non-significant outcome: *CBCL (Anxious/Depressed Scale)*. A reviewer might wonder if positive effects for the same outcome(s) need to be detected by two different studies. After consulting with the Clearinghouse, we learned that one study could detect positive effects with one target outcome, and another study may find positive effects with a different outcome or set of outcomes. But the reviewers must confirm that they reviewed data from at least two *independent* studies, rather than from publications or papers that relied on the same dataset.

If data on instrument reliability were not provided, instrument reliability coefficients might be available from *other* (hopefully similar) studies.

The specific effect sizes must be listed, along with the length of treatment beyond the end of treatment. Note that in some studies, the consumer follow-up period began at the *start* of treatment, and so that follow-up period length would need to be noted in this table, and in fact, for the entire technical review. Specific effect size statistics are requested. Some of these may require re-analyzing study data. The individual findings from each contrast (i.e., observed difference between a control and treatment group on a specific outcome measure) with a high or moderate rating should be reported. These findings include the effect size **and in some cases** its statistical significance. A Hedges' *g* adjustment must be applied to effect sizes calculated with standardized mean differences scores. This adjustment uses a "small sample size corrections factor" (Handbook, pp. 40–41).

The length of follow-up should be calculated not from the point of randomization or the start of the intervention but rather from the end of treatment. In accordance with the FFPSA and ACYF-CB-PI-18-09, and as indicated in the Handbook, a Well-Supported designation requires at least one of its studies demonstrate a sustained favorable effect of at least 12 months *beyond the end of treatment* on at least one target outcome.

Finally, in Table 10, non-significant results are reported. Note that this is not referring to statistically non-significant results, but rather whether the study found significantly ***unfavorable*** outcomes (statistically significant and not in the desired direction, see p. 40). In these circumstances, unfavorable outcomes would include findings that indicate an intervention places children at risk of harm, or has a detrimental impact on their well-being, compared to services as usual.

To determine whether there is risk of harm, all statistically significant unfavorable impacts on any outcome (whether an eligible target outcome or not) from any studies with contrasts receiving high or

moderate evidence ratings are identified. If there is sufficient evidence of risk of harm based on statistically significant unfavorable findings, the program may be deemed 'does not currently meet criteria' by the Prevention Services Clearinghouse. Additionally, programs or services may not be designated as *well-supported, supported,* or *promising* if case data suggests a risk of harm that was probably caused by the treatment and was severe or frequent. (Handbook, p. 44)

SafeCare Review. In Table 6 the team listed all eligible studies that were well-designed and well-executed. Table 7 is especially important because this is where we documented the following: whether or not the study controlled for confounding factors, measures that did and did not achieve baseline equivalence between comparison and intervention groups, overall and differential attrition for RCTs only (which did not apply herein), and whether each study met attrition standards (not applicable due to QED design).

While Gershater-Molko, Lutzker, and Wesch (2002) did not control for confounding factors due to incomplete reporting of demographic information, we verified absence of confounds for two other studies. Both Chaffin et al. (2012) and Beachy-Quick et al. (2017) relied upon propensity score matching; thus they achieved baseline equivalence. However, standardized assessment data were not reported for treatment and comparison groups by Gershater-Molko et al.

In Table 8, we reported the sample size, sample demographics of the intervention and comparison groups, and confirmed that none of these studies modified or adapted the program manual (i.e., they all adhered to model fidelity). We underscored that while Gershater-Molko et al. matched intervention and comparison groups on three variables (age, geographic location, and involvement with the child welfare agency), they did not match groups relative to two of the three variables (ethnicity and SES) highlighted in the FFPSA Clearinghouse Handbook. An abbreviated example of this table is included as Table 5, with only the Beachy-Quick et al. (2017) information included.

In Table 9, target outcomes, all of which examined the occurrence of child maltreatment across studies, are described. How they operationalized or measured the outcomes are to be discussed. In this situation, we examined the number of months without a report of child abuse or neglect between the two groups, and frequency of post-contact recidivism of reports of maltreatment (Gershater-Molko et al., 2002); recurrence of child maltreatment (Chaffin et al. (2012)); and subsequent referrals, assessments, and out-of-home placements (Beachy-Quick et al., 2017). Gershater-Molko reported an inter-reliability coefficient of 0.98; while other studies did not report them due to use of administrative data. We confirmed that each of the outcome measures were valid (i.e., demonstrates face, content, construct, and/or criterion validity) and systematically measured (i.e., measured in similar conditions and circumstances).

P values, effect sizes for outcome measures, and length of effect beyond the end of treatment (in months) were also documented in Table 9. In this case, each study reported improved outcomes across the board for the intervention group, with one exception. That is, while Beachy-Quick et al. reported that the percentage of subsequent out-of-home placements for the intervention group was lower than the comparison group, non-significant findings were detected with respect to subsequent reports of abuse and neglect and intake assessments. Effect sizes were calculated for Gershater-Molko et al. ($d = 0.89$) and Chaffin et al. (0.74–0.83, reported as hazard risk ratios). However, only percentages were reported for Beachy-Quick et al., and thus effect sizes could not be calculated. Per FFPSA, at least one study needs to demonstrate favorable effects for at least 12 months after the "end" of treatment. All of the studies met or exceeded this threshold, with Gershater-Molko et al. at 18 months, Beachy-Quick et al. at 12 months, and Chaffin et al. at six years.

An abbreviated example of this table is included here as Table 6, with only the Beachy-Quick et al. (2017) information included. This section ends with Table 10, where we noted that target outcomes with unfavorable effects were not detected.

FCT Review. The same process was followed for this technical review. In our initial review of studies, eight publications were identified as related to FCT. The technical review team determined that five studies examined an adapted version of FCT, were published after 1990, were available in English, were either randomized controlled designs (RCTs) or quasi-experimental designs (QEDs), had intervention and comparison conditions, and examined impacts using one of the FFPSA target outcomes. However, two publications were determined to be the same study, as they shared authors, sample, processes to assign sample members to intervention and control conditions, and data collection and analysis procedures. This resulted in four eligible studies.

One study was an RCT and therefore was subject to an attrition analysis. However, attrition was not assessed as not all the information needed (standard deviations in this case) to assess baseline equivalence was available in the report, and the authors declined to provide them. Therefore, this study was excluded from review as it would not have met the threshold of well-executed and well-designed, regardless of attrition findings. The rest of the studies were QEDs.

To assess baseline equivalence, each study was examined for direct pre-tests, pre-test alternatives, or age, race and SES on which to establish baseline equivalence. All 51 contrasts examined used direct pre-tests or pre-test alternatives. When selecting pre-test alternatives, researchers considered variables that are known to be correlated the outcome from past research and were most conceptually related to the outcome. For example, when examining frequency of pending placement status in the year following treatment, frequency of adjudications in the year prior to treatment was used as a pre-test alternative. For another example, when repeat investigation was an outcome, high risk level prior to beginning treatment was used as a pre-test alternative. Risk assessment at the close of an investigation has been shown to be related to repeat investigations.

Once direct pre-tests or pre-test alternatives were identified, a standardized mean effect size was calculated for the difference between the comparison and treatment group on the baseline variable. If the effect size was less than 0.05, the contrast received a moderate rating and no adjustment was needed. If the baseline effect size was between 0.05 and 0.25, the contrast was deemed moderate only if the baseline variables were also controlled for in the impact analysis using one of the methods listed in Section 5.8 of the Handbook. All contrasts in one study had baseline effect sizes in the 0.05–0.25 range, and no control method was used; therefore, the study was rated as low causal evidence. None of the baseline variables were statistically significantly different from one another. Non-statistically significant differences is a common method of demonstrating baseline equivalence but it is not a guarantee that the baseline effect size is in the required range. If the baseline effect size was greater than 0.25, the contrast received a low causal evidence rating. Baseline effect sizes for contrasts with moderate causal evidence ratings were subtracted from the outcome effect size. No contrasts that were ultimately designated with a moderate causal rating had baseline effect sizes over 0.05.

Race and age were examined for baseline equivalence when available, to establish that the two groups were not drawn from substantially different groups and therefore may have a substantially different characteristics confound. All were below the threshold of 0.25.

In most cases, baseline equivalence was established on the exact analytic sample. The Sullivan, Bennear, Honess, Painter, and Wood (2012) evaluation study was an exception. In this study, there was missing data at two-year follow-up, and baseline equivalence had only been established for the sample with full data at baseline and the one-year follow-up. In the Indiana waiver report, days until reunification also had missing data as the contrast only applied to those who had been placed in out-of-home care. [Author] requested access to the spreadsheet-based tool referenced in the Handbook to assess the largest baseline difference (Section 5.9.4, p. 39) but did not receive a response; thus baseline equivalence was assumed. Had the spreadsheet-based tool indicated that baseline equivalence was not achieved, and therefore the

contrast was of low causal evidence, the overall designation of the program of Well-Supported would not have changed.

Each contrast with a moderate design and execution rating was analyzed using chi-square tests for categorical variables and t-tests for continuous variables. Chi-square effect sizes were assessed for statistical significance ($p < .05$) and converted into standardized mean effect sizes following the analytic approach outlined by Lipsey and Wilson (2001), t-tests were assessed for statistical significance ($p < .05$) and standardized mean effect sizes in the form of Hedges' g were calculated. All calculations followed the Handbook guidelines, and the Lipsey and Wilson analytic methods.

In some studies, the outcomes were clearly defined by length of follow-up time. The treatment occurred during a specified time, and the follow-up window was one year and two years following the treatment period. In instances when the end of the treatment was not immediately apparent, the longest length of treatment was calculated, and the follow-up period was designated as the time following that date. For example, one study defined treatment as occurring during the child welfare case. The latest case closure date was determined from case length statistics, and the follow-up period was defined as the data collection window spanning from the latest possible date a case could have closed to the end of the follow-up period.

3.6. Section v

Overview. Table 11 is an important summary table. This is where whether and how risk of harm was assessed for all eligible studies in accordance with Section 6.2.3 of the Handbook is documented There must be an affirmative statement if any risk of harm or actual harm was found in the review. Note that there are specific questions to respond to for each rating level that must be answered.

SafeCare Review. In Table 11, we reported the following: "We found no studies of SafeCare that reported any increased risk of harm to a parent or child, or found actual harmful effects." We also confirmed that at least two studies (Beachy-Quick et al., 2017; Chaffin et al., 2012) were "well-designed" and "well-executed" studies with non-overlapping samples that were implemented in usual care or practice settings. Gershater-Molko et al. (2002), however had several limitations, including (1) lack of baseline equivalence for two key variables (SES and ethnicity); (2) adjustment for staff characteristics and training was not described; and (3) a modest amount of subject attrition information was provided. Regarding subject attrition, we relied on their 2003 study to understand how they derived their sample. Of the 205 families that consented, 92 received no services, 49 received some services, and 41 completed all the SafeCare training components. Those 41 families constitute the sample for their 2002 publication (Gershater-Molko, Lutzker, & Wesch, 2003, p. 379). In Table 11, we confirmed again that each study sustained a favorable effect for at least 12 months beyond the end of treatment.

One example of how to report the "harm to subjects" information in Table 11 is below.

Each of the studies that measured the outcomes of SafeCare were reviewed, not only to see if there were significant positive differences, but to determine if risk of harm or actual harm to any client was found. We found no studies of SafeCare that reported any increased risk of harm to a parent or child, or found actual harmful effects (Utah Department of Human Services, 2020, p. 54).

Designation Rating and Justification for Rating. In judging all the information, we proposed a Well-Supported rating for ACF consideration. Findings show that SafeCare is more effective in increasing the number of months without a report of abuse or neglect for up to three years (compared to families enrolled in Family Preservation). SafeCare was also more successful in suppressing child abuse and neglect during post-contact periods with families who had high rates of child abuse and neglect at baseline—compared to the families in the Family Preservation program (Gershater-Molko et al., 2002, pp. 281–282). Chaffin et al. (2012) statewide trial in Oklahoma demonstrated significant reduction in maltreatment recidivism for six years when SafeCare is implemented to fidelity. Moreover, Beachy-Quick et al. (2017) found that 0% of the children in the SafeCare-Colorado dyads experienced an out-of-home (OOH) placement whereas seven percent of the comparison children experienced an OOH placement during the 12 months following the completion of the program. To further justify our SafeCare evidence rating, we added Table 12, Methods Rating and Justification, which included the following dimensions.

- Baseline equivalence: Low for Gershater-Molko et al. (2002) and high for Chaffin et al. (2012) and Beachy-Quick et al. (2017).
- Casual evidence rating (highest rating per the Clearinghouse for a QED is moderate): Moderate for Chaffin et al. (2012) and for Beachy-Quick et al. (2017).
- Integrity of randomization: N/A; all studies were QED.
- Low attrition: N/A; attrition is only considered with RCTs.

Finally, we confirmed that we found at least one statistically significant effect of SafeCare relevant to FFPSA's guidelines.

FCT Review. We followed the same process. All studies were examined for evidence of risk of harm, and no evidence of harm was found. A total of 23 contrasts from two studies received moderate causal evidence ratings, and outcome effect sizes were calculated. This resulted in eight favorable, 13 no effects, and two unfavorable effects. Six of the favorable effects were sustained for at least 12 months. Favorable contrasts were identified from at least two separate studies and in at least two outcome domains. Following Handbook guidelines, the NCCD technical review team designated FCT as Well-Supported.

Extensive correspondence with study authors and program developers was necessary to gather the information needed to complete the review. Additional information was collected about study design, program implementation, statistics needed to calculate baseline equivalence assessments, and measurement approaches. Every effort was made to follow the Handbook guidelines exactly. When faced with areas of ambiguity, the NCCD researchers conducted consensus meetings and consulted with study authors to resolve them.

4. Discussion and recommendations

4.1. Technical review challenges and potential solutions

In reviewing the process for these two interventions, one can see the complexity of the process—especially for documenting baseline equivalence, effect size estimates, and judging what is a low, moderate, or high quality study—as well as selecting the proper set of contrasts. Depending upon the array of eligible studies available, the review team may need to grapple with making an evidence rating recommendation based upon (1) variability in whether or not interventions were implemented to fidelity; (2) studies that may not be as empirically robust as other studies because of baseline equivalence differences or subject attrition; and (3) conflicting findings across studies. To address these challenges, we engaged in a collaborative decision-making process; acknowledging that some stipulations need to be outlined when these situations inevitably occur. Ultimately, in our concerted efforts to be transparent and justify our evidence ratings, we developed the criterion in Table 12 (Methods Rating and Justification). This helped us to focus on pertinent criteria, thereby standardizing the process for rendering a fair rating for SafeCare and FCT.

Due to the fact that studies must have been conducted in the United States, United Kingdom, Canada, New Zealand, or Australia; and published/prepared in English during or after 1990,[2] states may want to

[2] See Federal Register, Sections 2.3.3 and 2.3.4 at https://www.federalregister.gov/d/2018–13420.

request that the country restriction be expanded for certain interventions. For example, as mentioned earlier, certain countries in Europe have conducted relevant research on medically assisted substance abuse treatment (K. Stack, personal communication, June 22, 2018).

In recent federal guidelines, target outcomes are outlined, albeit they are not constrained or restrictive. States may want to confirm with ACF on how to apply a broader approach to their FFPSA plans. At the very least, FFPSA guidelines stipulate that states must elucidate a coherent theory of change. In other words, states might want to consider how and under what conditions their selected interventions—or more broadly, their FFPSA plan—will achieve intended outcomes. Justifying their selection by citing the empirical evidence supporting the use and implementation of an EBP, and elaborating upon why their FFPSA plan would lead to FFPSA target outcomes within their respective agency setting are critical steps in the technical review process. For example, we know that maternal depression can sometimes lead to child placement because of child emotional or physical neglect and that unchecked teen anger can lead to foster care reentry. Studies that show that the interventions contribute to these desired effects should be considered and rendered relevant for FFPSA target outcomes.

Another challenge was that some of the guidance related to summarizing and comparing the study effect sizes was unclear. Most of the documented and required effect size statistics (e.g., improvement index) are based on continuous outcomes, making it less clear how to calculate required effect sizes for dichotomous or categorical outcomes (What Works Clearinghouse, n.d.). In addition, studies may not provide the necessary information to calculate an effect size due to only reporting *p* values or other statistics (e.g., hazard ratios). This level of complexity may require review teams to have a statistician collaborate on the review.

As we engaged in the technical review process, we also learned that there are practical limitations related to the scope and relevancy of the FFPSA technical review and implementation process. Namely, more efforts should be devoted to expanding the relevance and feasibility of adhering to FFPSA in "real world" practice. For example, the numerous stipulations set forth by the ACF FFPSA Clearinghouse may limit what types of interventions are considered and implemented. The extent to which an approved EBP aligns with the needs, culture, and context of both agencies and the clients they serve remains unknown. While there are some EBPs that have been tested among a racially and ethnically diverse pool of maltreated youth (e.g., parent-child interaction therapy, trauma-focused cognitive behavioral therapy, Level 4 Pathways Positive Parenting Program, and multi-systemic therapy), agencies face numerous barriers to implementing EBPs (Garcia, DeNard, Morones, & Eldeeb, 2019). Some of these barriers include: lack of funding and clarity about roles and responsibilities, lack of attention toward assessing implementation readiness, balancing crisis oriented work with the demands of completing implementation activities, limited supports and training for leaders and supervisors to engage in transformational leaders and instill a climate that embraces inter and intra-organizational collaboration and communication (Aarons, Hurlburt, & Horwitz, 2011; Garcia et al., 2019, 2020; Palinkas & Aarons, 2009). Beyond implementation, EBPs do not holistically address the needs of clients served in the child welfare system. Clients often grapple with presenting issues of concern that EBPs alone do not address, such as poverty, homelessness, and lack of access to culturally relevant resources. To that end, agency leaders and workers must grapple with the complex reality of aligning services and prevention and intervention programs with clients' needs and strengths. With very few programs approved by the ACF FFPSA Clearinghouse, states often have no choice but to select other programs and allocate additional time and resources to engage in the technical review process for approval to include optimal programs in their FFPSA plan. In the meantime, optimal service delivery might be delayed. In light of these challenges, the Clearinghouse might want to consider the time it takes for states to submit a FFPSA plan that is both evidence-informed and culturally applicable to the needs of the clients they serve.

4.2. Recommendations for conducting a high-quality technical review

Based upon our reviews of SafeCare and FCT we have a number of observations and strategies to recommend. We found it very helpful to talk with the intervention developers to help us find publications for the literature review, and to select which studies to most closely consider for review. While the FFPSA Handbook (Wilson et al., 2019) is detailed, one will likely need to consult some of the WWC manuals (What Works Clearinghouse, n.d., 2017). These manuals provide some of the background details that underpin the Prevention Clearinghouse approaches to calculating baseline equivalence and effect size calculations.

There is no clear documentation of what qualification independent reviewers must have. In the FCT review, the lead reviewers both hold PhDs and have direct experience with conducting experimental studies. Two master's-level reviewers assisted with reviewing the literature and interns and junior researchers with bachelor's degrees conducted the original literature searches. A master's-level statistical consultant helped ensure that the procedures used were reviewed properly, and that review was sound and clear. The team had credentials in sociology, human development, social work, public policy, statistics and public health. Having a multidisciplinary team was useful as articles appear in a range of types of publications.

For FCT, every study reviewed required contact with the original study authors for clarification and additional information. Contacting study authors early in the process, even before all the questions are compiled, begins the communication and reduces the likelihood that author queries will slow the review process. Identifying whether study authors have access to original data or additional statistics early is important. Many studies, because of institutional review board requirements, cannot retain raw data indefinitely—and without such information, a low causal evidence rating may be assigned.

Establishing internal processes such as checklists, spreadsheets, and repositories helps organize and document the effort. Conducting the review with two researchers in parallel and then holding consensus meetings to resolve any differences of opinion not only is required by the Handbook but increases reliability and confidence in the final rating.

As illustrated by the review section table samples, we found it helpful to add notes sections below many of the tables so that additional details could be provided, including in some cases, reprinting entire tables from a report or journal article. This is because the federal table columns do not easily enable the inclusion of lengthy material, tables or figures. This strategy may be especially useful for Table 11, which asks for an explanation of the evidence rating being recommended. For at least one intervention (SafeCare), we created a 12th table to provide some of the key information requested by the questions in Table 11 and to increase transparency for justifying our evidence rating.

It is essential that a close working relationship be established with the state FFPSA planning managers as they need to confirm what version of the intervention is being implemented, and when the technical review must be completed by so that it can be included in the State FFPSA prevention plan. The draft technical review should be shared, and then discussed in a work session with the state child welfare agency leaders, to help ensure that the agency management and research team understand the review, and that the review is complete and objective. One might also have the intervention purveyor review the technical review to help note factual errors or blind spots.

4.3. Implications of the FFPSA evidence standards and reviews for the field

The new FFPSA law reinforces an evidence-based approach to selecting which interventions will be reimbursed by the Federal government. While delayed by the FFPSA transition act, the requirement by FY 2022–2023 is that 50% of the funds reimbursed go to Supported and Well-Supported interventions. And in FY 2024, at least of 50% of the funds must be spent on interventions with the highest evidence rating (Well-Supported). The importance of taking into account demographics

of population served, validity and reliability of measures, and robust and transparent statistical methods are all reinforced by the Clearinghouse standards.

In our conversations with state/county child welfare leaders and some child welfare evaluators, we are seeing how parts of the new FFPSA law and Prevention Services Clearinghouse evaluation study criteria are beginning to influence certain program evaluation design aspects, certain methodological components, and research standards (e.g., establishing adequate comparison groups, measuring and addressing baseline equivalence, minimizing and tracking subject attrition carefully, using standardized measures, estimating statistical power to see if one has an adequate sample size, and instituting six- and 12-month follow-up periods that begin at the time of intervention closure).

One of the concerns with this situation is that we need a robust evaluation pipeline where formative evaluation studies are conducted for programs with less research evidence, so that communities of color, other marginalized child welfare populations, and special communities can have promising interventions that they have found effective moved up the evidence ladder to Supported and then Well-Supported. This is one of the major consequences of the FFPSA legislation that federal, state, county and philanthropic organizations should actively address (Bell, 2019; Casey Family Programs, 2018).

In conclusion, this article should inform how other states commission and carry out their technical reviews. Greater consistency in how states conduct the technical review process will lessen process variability so there is greater confidence in the evidence ratings. Looking forward, the technical review process offers a unique opportunity to compile and critique existing knowledge about the efficacy of child-focused interventions or programs, and to rely on those findings to inform what services youth in the child welfare system receive.

CRediT authorship contribution statement

Antonio R. Garcia: Conceptualization, Methodology, Writing - original draft, Writing - review & editing, Supervision, Project administration. **Peter J. Pecora:** Conceptualization, Methodology, Writing - original draft, Writing - review & editing. **Audrey H. Schnell:** Methodology, Writing - original draft, Writing - review & editing, Data curation. **Cynthia Burnson:** Writing - original draft, Writing - review & editing. **Elizabeth Harris:** Writing - original draft, Writing - review & editing. **Allison Finseth:** .

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

We wish to thank Allison Finseth for her professional editing.

Appendix A

See Table 1.

Table 1
Technical Reviews of Interventions That States Plan to Conduct or Have Conducted as of June 11, 2020.

State	Interventions Being Reviewed (Lead Reviewers Known at This Time)
Arkansas	*All EBPs are being reviewed by the National Council on Crime and Delinquency (NCCD):* • FCT (This technical review has been approved by ACF, with FCT being rated as Well-Supported.) • Youth Villages Intercept (This technical review has been approved by ACF, with YV Intercept rated as Supported.) Note that Arkansas included SafeCare in their state FFPSA plan. Shortly after Utah submitted their technical review for SafeCare to the Federal government, the Clearinghouse announced it had completed its review of that intervention and rated SafeCare as Supported. That rating is being appealed by Utah and SafeCare.
Colorado	(Reviewers TBD. Contact Elysia Clemens, *Colorado Evaluation Lab at the University of Denver*) • Colorado Community Response • High-Fidelity Wraparound • SafeCare (Note that the Clearinghouse recently rated SafeCare as Supported.)
Iowa	• SafeCare *(Casey Family Programs: Peter J. Pecora; University of Kentucky: Antonio Garcia, The Analysis Factor: Audrey Schnell)* Note that the Clearinghouse recently rated SafeCare as Supported.
Kentucky	*All EBPs are being reviewed by the Public Consulting Group (PCG):* • 1-2-3 Magic • Cognitive Behavioral Therapy (CBT) • Homebuilders (Note that the Clearinghouse recently rated Homebuilders as Well-Supported.) • Sobriety Treatment and Recovery Team (START)
Nebraska	• FCT[a] (*The Stephen's Group)* (The Arkansas technical review for FCT has been approved by ACF, with FCT being rated as Well-Supported.)
Utah	Utah is contracting with the Social Research Institute at the University of Utah to conduct most of the independent systematic reviews. Note that shortly after Utah submitted its the technical review for SafeCare to the Federal government, the Clearinghouse announced it had completed its review of that intervention and rated SafeCare as *Supported*. That rating was appealed by Utah and SafeCare, and is currently being reviewed by the Clearinghouse. To recap, these are the completed technical reviews that were submitted in February: • SafeCare *(Casey Family Programs: Peter J. Pecora; University of Kentucky: Antonio Garcia, The Analysis Factor: Audrey Schnell)* • Seeking Safety *(Social Research Institute at the University of Utah)* [Was subsequently judged as *"Unable to be Rated"* by the Clearinghouse.] The services Utah plans to review next are as follows (in order of priority). • Acceptance and Commitment Therapy (ACT) • Dialectical Behavioral Therapy (DBT) • Helping Women Recover/Helping Men Recover • High-Fidelity Wraparound (Utah will work from the Colorado technical review results)
Washington	Washington State Department of Children, Youth, and Families is contracting with Angelique Day of the University of Washington, School of Social Work to conduct technical reviews of the following interventions. • Canoe Journey • Family Spirit • Talking Circles

[a] Nebraska currently has an FCT contract, and it was included with the transitional payments documentation in their plan.

Appendix B. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.childyouth.2020.105597.

References

Administration for Children and Families. (2019). Child abuse, neglect data released: 28th edition of the Child Maltreatment Report. https://www.acf.hhs.gov/media/press/2019/child-abuse-neglect-data-released.

Aarons, G. A., Hurlburt, M., & Horwitz, S. M. (2011). Advancing a conceptual model of evidence-based practice implementation in public service sectors. *Administration and Policy in Mental Health and Mental Health Services Research, 38*(1), 4–23.

Beachy-Quick, K., Lee, C., McConnell, L., Orsi, R., Timpe, Z., & Winokur, M. (2017). SafeCare Colorado Program Evaluation Report 2014-2017. Fort Collins, CO: Colorado State University, College of Health and Human Sciences, School of Social Work. Prepared for the Office of Early Childhood, Colorado Department of Human Services. Retrieved from: http://www.chhs.colostate.edu/ssw/wp-content/uploads/sites/7/2019/10/SafeCare-Colorado-Project-Evaluation-Report-2014-2017_final_corrected.pdf.

Bell, W. C. (2019). *Engaging philanthropy in improving child well-being.* Seattle: Casey Family Programs.

Chaffin, M., Hecht, D., Bard, D., Silovsky, J. F., & Beasley, W. H. (2012). A statewide trial of the SafeCare home-based services model with parents in Child Protective Services. *Pediatrics, 129*(3), 509–515.

Casey Family Programs (2018). *Engaging philanthropy in improving child well-being outcomes.* Seattle: Author.

Childhelp. (n.d.) Child abuse statistics and facts. https://www.childhelp.org/child-abuse-statistics/.

Register, Federal (2018). Administration for children and families, HHS. Decisions related to the development of a clearinghouse of evidence-based practices in accordance with the family first prevention services act of 2018. *Federal Register, 83*(121), 29122–29124.

Garcia, A. R., DeNard, C., Morones, S., & Eldeeb, N. (2019). Mitigating barriers to implementing evidence-based interventions: Lessons learned from scholars and agency directors. *Children and Youth Services Review, 100*, 313–331.

Garcia, A.R., Myers, C., Morones, S., Ohene, S., & Kim, M. (2020). "It starts from the top": Caseworkers, leaders, and TripleP providers' perceptions of implementation processes and contexts. Human Service Organizations: Management, Leadership, & Governance. doi.org/10.1080/23303131.2020.1755759.

Gershater-Molko, R. M., Lutzker, J. R., & Wesch, D. (2002). Using recidivism data to evaluate Project SafeCare: Teaching "bonding", safety, and health care skills to parents. *Child Maltreatment, 7*, 277–285.

Gershater-Molko, R. M., Lutzker, J. R., & Wesch, D. (2003). Project SafeCare: Improving health, safety, and parenting skills in families reported for, and at risk for child maltreatment. *Journal of Family Violence, 18*, 377–386.

H.R. esolution. 1865, 116th Cong., Cong. Rec. 2534 Further Consolidated Appropriations Act (2020) (enacted). https://www.congress.gov/bill/116th-congress/house-bill/1865/titles.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis.* Thousand Oaks, Calif.: Sage Publications.

Palinkas, A., & Aarons, G. A. (2009). A view from the top: Executive and management challenges in a statewide implementation of an evidence-based practice to reduce child neglect. *Int. J. Child Health Human Devel. 2*(1), 47–55.

Pecora, P.J., Garcia, A.R., Schnell, A. (2020). Family First Prevention Services Act (FFPSA) technical review overview and introduction to the SafeCare review for Iowa and Utah. Report presented to Iowa and. Utah Department of Human Services.

Sullivan, M. B., Bennear, L. S., Honess, B. S., Painter, W. E., & Wood, T. J. (2012). Family Centered Treatment—An alternative to residential placements for adjudicated youth: Outcomes and cost effectiveness. *OJJDP Journal of Juvenile Justice, 2*(1), 25–40.

Torres, K., & Mathur, R. (2018). *Fact sheet: Family first prevention services act.* Washington, DC: First Focus Campaign for Children.

Utah Department of Human Services (2020). *FFPSA technical review for SafeCare Attachment B: Checklist for program or service designation for HHS consideration.* Salt Lake City: Author.

What Works Clearinghouse. (n.d.). What Works Clearinghouse Procedures Handbook Version 4.0 https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_procedures_handbook_v4.pdf.

What Works Clearinghouse. (2017). Reviewer guidance. https://ies.ed.gov/ncee/wwc/Docs/ReferenceResources/wwc_reviewer_guidance_103017.pdf.

Wilson, S. J., Price, C. S., Kerns, S. E. U., Dastrup, S. D., & Brown, S. R. (2019). Title IV-E prevention services clearinghouse handbook of standards and procedures, version 1.0, OPRE Report # 2019-56. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.