

# From Evidence to Impact: Strengthening Evaluation in Child Welfare Services

## Context and Landscape

Michael Pergamit, Mark Courtney, Bridgette Lery

URBAN INSTITUTE, UCHICAGO, URBAN INSTITUTE

January 2026

Strong, well-executed evaluations are essential to a child welfare system that truly improves outcomes for children, youth, and families. However, limited evidence and too few rigorous studies prevent the field from identifying what works. Persistent barriers continue to limit the high-quality, systematic evaluations needed to understand program implementation, impact, and cost.

In this series of three briefs, we describe the need for rigorous evaluation of policies, programs, and practices affecting the experiences and outcomes of those served by the child welfare system in the United States, which includes child protective services, family preservation and reunification services, out-of-home care, services for youth transitioning to adulthood from care, and adoption and guardianship services supported by Title IV-E and Title IV-B of the Social Security Act. In this brief, we describe the current evidence on the effectiveness of child welfare programs and services, highlight aspects of the current state of the field that make building evidence difficult, and argue for increased investment in developing capacity for rigorous evaluation. In our [Hurdles for Agencies](#) brief, we discuss the challenges welfare agencies face limiting the establishment of systematic evidence. In our [Design Challenges](#) brief, we discuss the challenges of various evaluation designs. Through these briefs, we hope to begin a conversation in the field about how to overcome the difficulties to establishing evidence-based knowledge of the effectiveness of child welfare services. Our aim is to equip families, child welfare agency staff, program developers, legal representatives, and advocates with insights that can guide systemwide improvements to these services.

## Defining Data, Evidence, and Evaluation in Child Welfare Systems

Rigorous evaluation of child welfare programs is grounded in the same scientific principles that guide medical research in determining the safety and efficacy of health care interventions. Children and families in the child welfare system deserve the same care and evidence-based inquiry to answer a fundamental question, “How do we meet urgent needs now while developing reliable, cost-effective long-term solutions that work for all populations?”

To answer that question, we must clarify what we mean by data, evidence, rigor, and evaluation. *Data*, when collected, organized, and analyzed according to scientific principles, can produce *evidence*. The discipline required to transform data into evidence is what we call *rigor* (Lery, Haight, and Alpert 2016). Evidence, in turn, is information developed using explicit, systematic, and transparent methods, whether qualitative or quantitative, that withstand public scrutiny (Alpert et al., n.d.). The accumulation of such evidence is *knowledge*, which is essential to understanding effectiveness.

Not all meaningful data are quantitative. The lived experience of children, youth, families, and communities involved with the child welfare system offer invaluable insight into what aspects of the system are working, and where services fall short. Their input can provide insight into practical barriers to program implementation, such as when and where programs are offered, how outreach language affects enrollment, and what factors encourage

participation. Involving service beneficiaries in research and evaluation activities can improve those practices as well.

Finally, rigorous evaluation methods, such as randomized controlled trials (RCTs) and well-designed quasi-experimental studies, help determine whether a program causes the outcomes it intends to achieve. Repeated evaluation across settings and populations allow us to understand for whom and under what circumstances interventions are most effective. Together, these elements—comprehensive data, careful evaluation, and scientific rigor—form the foundation of a child welfare system capable of continuous learning and improvement.

## **Hurdles to Establishing Evidence**

### **Many Programs Are Not Ready for Evaluation**

When program staff decide to move data into evidence, it is no small feat. Challenges can occur at multiple steps in the process, and for some, the challenge arises at the first step. To be ready for evaluation, a program must have a well-defined logic model. The outcomes of the program should be clearly identified, as well as the inputs and activities the program needs to produce those outcomes. A theory of change can also be identified, explaining how the inputs and activities are to be implemented to produce outcomes (Epstein and Klerman 2012). Without these critical details, programs may not be able to determine if the service itself, or the degree to which it is implemented (i.e., dosage), is contributing to observed outcomes, or if alternative services might be more effective. When the goal is to confidently determine a program caused an effect, establishing this foundation is essential.

For others, the challenge comes when implementing an evaluation. To design and monitor an evaluation, the size of the eligible population and the extent to which people referred to a program participate and complete it must be understood. Program managers would benefit from this information as well, for program planning and quality monitoring, but they often do not have full access to the data necessary to answer those questions. Program and evaluation staff often must coordinate to identify and access the data needed for the evaluation, which can be a challenge when program managers do not have ready access to the information.

### **Many Evaluations Do Not Demonstrate Whether the Program Caused Observed Outcomes**

Even when programs complete evaluations, the evaluations themselves often lack the rigor needed to infer causal effects. Several clearinghouses rate individual program evaluations to determine the causal relationship between a program and a participant outcome. For child welfare services, the Prevention Services Clearinghouse is one of the most important—it determines which programs and services can be reimbursed by the federal government. However, of the 210 programs the clearinghouse has reviewed, 53 percent did not meet their standards for the rigor of evaluation evidence to even be rated. And only 23 percent achieved at least a rating of “supported,” meaning showing causal evidence.<sup>1</sup>

The California Evidence-Based Clearinghouse for Child Welfare (CEBC) has been in operation much longer than the Prevention Services Clearinghouse and covers a much wider range of program types. Unlike the Prevention Services Clearinghouse, it relies heavily on peer-reviewed journal articles for evidence of program effectiveness. Despite the differences in these clearinghouses, their ratings show similar levels of evidence.

Of the 586 programs the CEBC has reviewed,<sup>2</sup> 46 percent did not meet their standards to be rated, despite most evaluations having gone through peer review before the clearinghouse reviewed them. Only 20.3 percent achieved at least a rating of supported and less than one-quarter of these (23.5 percent) were rated high in relevance to populations receiving child welfare services,<sup>3</sup> making just 4.8 percent of all programs evaluated rated as having both causal evidence and relevance in child welfare settings.

Although clearinghouses are a valuable resource for review and assessment of evaluations, requirements vary in what evaluation designs are acceptable (e.g., RCTs) and how evaluations are carried out. Failing to meet the specific standards of a given clearinghouse may not mean the program is not effective, and relying on clearinghouses as the sole arbiter of a program’s causal evidence may overlook important insights an evaluation may have generated.

## **The Drive for Rigor under Family First Strengthens Methods but Can Narrow What Counts as Evidence**

The federal Family First Prevention Services Act (Family First) allows states to claim Title IV-E funds for using interventions rated as promising, supported, or well-supported by the Prevention Services Clearinghouse, with at least half of those funds being allocated for supported or well-supported programs. The legislation requires states to conduct rigorous evaluations to achieve one of these ratings, and to conduct evaluations even if an intervention already achieved a well-supported rating, which ensures programs are equally effective across the diverse set of state and local contexts in which they operate.<sup>4</sup>

Although the act sparked a wave of evaluations with higher rigorous standards, it had some unintended consequences. First, states turned toward evaluating family preservation programs that qualified under Family First and away from evaluating other programs, such as those meant to increase family reunification, train foster parents, or help children in foster care with mental health problems. Second, evaluating programs with alternative funding sources such as Medicaid, which has no evaluation requirement, decreased after the act passed. Third, the newly established clearinghouse had to focus on the most common types of evaluations—particularly RCTs and matching models—and those conducted using other methods (e.g., when it was not feasible to conduct an RCT or no good matched comparison group existed) were likely deprioritized because they did not meet the new clearinghouse guidelines.

We have also observed that the act can unintentionally create a perverse incentive to divert resources and attention away from developing programs specifically for child welfare populations. For example, some therapeutic programs offer specialized versions designed for families or children involved in the child welfare system. When the Prevention Services clearinghouse assigns a “supported” rating to the general version of such a program, developers often devote less attention to the child welfare–specific version. This dynamic can lead states to prioritize the highest-rated programs rather than those best suited to their target populations or specific contexts. Contributing to the issue, the clearinghouse does not require programs to be evaluated specifically on children who interact with the child welfare system. As a result, a program with a high rating may qualify for Family First funding without evidence of effectiveness for that population. While many evaluations have been conducted with similar populations—especially mental health interventions—others may have been tested on groups that meaningfully differ from those served by the child welfare services system.

## **Impact Evaluations of Small Programs May Not Be Feasible, Especially in Rural States and for Relatively Small Populations with Distinct Needs**

Many programs serve populations that are too small to generate the statistical power needed to identify policy- and practice-relevant program impacts. Some of these are “homegrown” programs that were developed in states with small child welfare services populations and had never been implemented in jurisdictions with larger populations. The inability to rigorously evaluate the impact of small programs poses a particular challenge for programs that serve populations with unique needs. Tribal communities have historically been disparately affected by child welfare services, but very few tribes have large enough populations involved with child welfare services to allow for rigorous impact evaluations.<sup>5</sup> Designing evaluations of programs for other populations with distinct needs (e.g., young parents in foster care) face the same challenge. Evaluators are trained to consider sample sizes that will provide a high level of statistical confidence in the findings. However, although small sample sizes may not produce definitive

findings (i.e., those that achieve statistical significance), by using rigorous methods, an evaluation can still be informative. In addition to using rigorous evaluation methods for these small populations, the field may need to complement these methods with in-depth exploration of the lived experience of these populations to make the case for the effectiveness of policies and programs that seek to improve their outcomes.

## Investing in Evaluation Capacity

Navigating the evaluation process, from determining readiness to completing a study, is complex and requires dedicated time, expertise, and resources. This complexity is compounded by limited access to federal, state, and philanthropic funding. Because the federal government is the largest source of funding for child welfare services, it plays an outsized role in supporting evaluations of programs and services.

However, programs are implemented by states and counties, and Family First now requires states to engage in evaluation activities. For this reason, greater investment in evaluation capacity at the state and county level is warranted. While some state agencies dedicate resources to evaluations through in-house research teams or external contracts, many do not. As a result, states often rely on programs rated by clearinghouses, which may not align with local contexts or meet the distinct needs of the populations they serve.

Beyond financial support, the federal government should provide leadership to advance rigorous evidence-building in child welfare policy and practice, particularly through initiatives supported under Titles IV-E and IV-B of the Social Security Act.

## How Federal Evaluation Funding and Support Can Be Enhanced

Federal, state, and local governments should work together to fund child welfare service evaluations. Because evaluations typically occur within the states and counties they are implemented in, states and counties could increase funding for evaluation so programs could hire evaluators or train current staff on proper methods. The investment in in-house knowledge must come from states and counties because only they understand their specific needs. However, given the need to comply with federal requirements, the federal government will still play a role in supporting states and counties in evaluation. Some ways that the federal government can help states and counties include the following:

- Congress could create an evaluation within Title IV-E funds, similar to the approach used for Chafee funding, but at a level that allows states to directly access and administer these resources rather than restricting them to only be used by a federal agency.
- The federal government could provide technical assistance to states for the development of evaluations through the following:
  - » Investing in development of evaluation capacity in state and county child welfare agencies
    - The Office of Policy, Research, and Evaluation (OPRE) recently sponsored an Evidence Building Academy for child welfare staff and their partners to learn the principles of evaluation. Casey Family Programs has also offered a version of this training. Continuing this service, along with direct technical assistance, would help expand understanding of the value of evaluation and the requirements for rigor.
  - » Incentivizing (through grants or other mechanisms) interagency data linking
  - » Providing guidance and examples of evaluation request-for-proposal requirements, data-sharing agreements, consent language, and other useful tools to support evaluation
- Finally, the federal government could create a structure within the Administration on Children and Families (ACF) for identifying small programs, including those in small-population states, rural areas, and those targeting small, marginalized populations (e.g., tribes), that warrant evaluation. This would facilitate work

across jurisdictions to support evaluation of these programs, a function that few states can take up on their own.

## Conclusion

Building a stronger evidence base for child welfare programs in the US requires coordinated investment, shared responsibility, and sustained leadership. Federal, state, and local governments should work together to expand funding and capacity for evaluation, ensuring programs serving children and families are guided by rigorous, population- and context-specific evidence. By advance planning, coordination, and investing in evaluation capacity, the child welfare community can move toward a cohesive framework for ongoing learning and improvement. Through this shared commitment, agencies at every level can more rapidly identify what works, for whom, and under what circumstances, enhancing rather than relying on clearinghouses, and thus helping every child and family receive services that are effective and grounded in evidence. In the absence of rigorous evidence, the system risks relying on assumptions rather than demonstrated results that affect the most marginalized groups.

In companion briefs, we focus on challenges faced by child welfare agencies limiting their ability to build systematic evidence, and challenges to employing specific evaluation designs. We hope that these documents help facilitate a more vigorous discussion of how to more rapidly build evidence of what works in child welfare services.

## Notes

- <sup>1</sup> As of October 27, 2025.
- <sup>2</sup> As of October 27, 2025.
- <sup>3</sup> A high rating means “the program was designed, or is commonly used, to meet the needs of children, youth, young adults, and/or families receiving child welfare services.” Nearly all remaining programs received a medium rating, meaning “the program was designed, or is commonly used, to serve children, youth, young adults, and/or families who are similar to child welfare populations (i.e., in history, demographics, or presenting problems) and likely include current and former child welfare services recipients.” See “Child Welfare System Relevance Levels,” California Evidence-Based Clearinghouse for Child Welfare, last updated November 11, 2022, <https://www.cebc4cw.org/how-are-programs-on-the-cebc-reviewed/child-welfare-relevance-levels/>.
- <sup>4</sup> This requirement can be waived if the Secretary of HHS deems the program to have clearly shown effectiveness. The Children’s Bureau often does waive the requirement.
- <sup>5</sup> FFPSA does allow the evidence standard to be waived for most tribes; it is only for tribes that operate their own IV-E programs, which are largely only the largest tribes. Most tribal services are contracted to be run by the state, and these tribes are not offered the same exemption even though they are serving tribal communities.

## References

- Alpert, Lily A., Haight, Jennifer M., Orlebeke, Brittany., and Wulczyn, Fred. n.d. *EDGE: Evidence-Driven Growth and Excellence*. Chicago: The Center for State Child Welfare Data, Chapin Hall at the University of Chicago.
- Courtney, Mark E., Pergamit, Michael, Woolverton, Maria, and McDaniel, Marla. 2013. “Challenges to Learning from Experiments: Lessons from Evaluating Independent Living Services.” In *From Evidence to Outcomes in Child Welfare: An International Reader*, edited by Aron Shlonsky and Rami Benbenishty. New York: Oxford University Press.
- Epstein, Diana, and Klerman, Jacob A. 2012. “When Is a Program Ready for Rigorous Impact Evaluation? The Role of a Falsifiable Logic Model.” *Evaluation Review* 36 (5), 375–401. <https://doi.org/10.1177/0193841x12474275>.
- Lery, Bridgette, Haight, Jennifer, and Alpert, Lily. 2016. “Four Principles of Big Data Practice for Effective Child Welfare Decision Making.” *Journal of Public Child Welfare* 10 (4), 466–74. <https://doi.org/10.1080/15548732.2016.1209149>.

## **About the Authors**

Michael Pergamit and Bridgette Lery are senior fellows in the Family and Financial Well-Being Division at the Urban Institute. Mark Courtney is Samuel Deutsch Professor Emeritus at the Crown Family School of Social Work, Policy, and Practice, University of Chicago.

## **Acknowledgments**

This brief was funded by the Annie E Casey Foundation. We are grateful to them and to all our funders, who make it possible for Urban to advance its mission. The views expressed are those of the authors and should not be attributed to the Urban Institute, its trustees, or its funders. Funders do not determine research findings or the insights and recommendations of Urban experts. Further information on the Urban Institute's funding principles is available at [urban.org/fundingprinciples](https://urban.org/fundingprinciples). Copyright © January 2026. Urban Institute. Permission is granted for reproduction of this file, with attribution to the Urban Institute.