

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

## Hurdles for Agencies

Michael Pergamit, Mark Courtney, Bridgette Lery

URBAN INSTITUTE, UCHICAGO, URBAN INSTITUTE

January 2026

Building evaluation capacity in child welfare agencies helps staff understand the value of evaluation and rationale behind specific methodological requirements. This preparation equips agency staff to fulfill their roles effectively throughout the evaluation process. Evaluating child welfare programs and services requires cooperation and coordination with agency administrators, program managers, frontline workers, families, and funders. Such efforts often require agency staff to take on additional responsibilities, adjust workflows, and collaborate with other organizations. Although evaluation can initially seem burdensome, its value lies in evidence-based insights that improve how agencies serve children and families. When agencies commit to understanding the effectiveness of their programs or services, these challenges become a part of shared learning and improvement.

This brief is second in a series of three briefs describing 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. In the first brief, [Context and Landscape](#), we identify the current evidence on the effectiveness of child welfare programs and services. In this brief, we discuss the challenges child welfare agencies face limiting the development of systematic evidence. In our final brief, [Design Challenges](#), we discuss hurdles to using various evaluation designs. Through these briefs, we hope to begin a conversation in the field about how to overcome the challenges 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.

### Evaluation Planning Should Be Part of Agency Planning

Over the past two decades, most child welfare services agencies have significantly improved their ability to generate and analyze data on specific populations and services, and some have implemented continuous quality-improvement processes that rely on those data to inform child welfare services policy and practice. Attention should also focus on implementing new policies and practices that maximize opportunities to rigorously evaluate innovation. Instead of building plans for evaluation into the policy and practice planning process, agencies and legislatures too often begin implementing changes before considering how to evaluate the effectiveness of those implementations.

Randomized controlled trials (RCTs) are considered the “gold standard” to determine casual effects in program evaluation. However, RCTs are not always practical or appropriate (see Pergamit, Courtney, Lery 2025a) and lack of planning poses limitations for rigorous evaluation designs of policies and programs. For example, fundamental changes in practice, such as child welfare practice models, cannot be evaluated using random assignment of individuals or families to an intervention and control group. Changes in practice models generally involve training or retraining all frontline and supervisory staff in a jurisdiction. Once that training has been disseminated widely enough, and with fidelity, randomly assigning families to a control group is impossible because all key staff have adopted the new practices.

More generally, a random assignment of individuals or families is inappropriate if the intervention is likely to significantly change the context of service provision. For example, programs that enhance legal representation of parents in family courts by improving training for attorneys and reducing their caseloads could lead to more zealous and effective representation of parents. This in turn could influence the behavior of others involved in court proceedings (e.g., attorneys and judges), making them generally more sympathetic to parents involved in dependency proceedings. If that were the case, then randomly assigning some parents to having enhanced representation and others to lawyering-as-usual would not be a good test of the effectiveness of enhanced parental representation because the context in which that representation occurs would have been altered in ways that would also improve outcomes for the control group.

Rigorous quasi-experimental evaluation designs often depend on variation in when and where interventions are implemented. However, decisions made by child welfare agencies about policy and program rollouts are not always aligned with rigorous evaluation designs. When designed well, staggered implementation across sites over time can create an opportunity to measure program impact. But, in practice, jurisdictions often choose to implement programs systemwide at once, eliminating the natural comparison points needed for rigorous analysis.

Even when agencies do stagger implementation, the rollout often reflects political or operational considerations rather than evaluation strategy. For example, early sites may be selected because they fall within the district of a program's legislative sponsor or because certain offices are perceived as more capable of implementation. These choices, while practical, make it difficult for evaluators to identify comparable locations or isolate the program's true effects.

Ultimately, evaluation must be built into planning from the outset. Without intentional design that balances implementation needs with research rigor, promising initiatives become missed opportunities for learning how to better serve children and families.

## **Challenges with Agencies**

### **Evaluation Activities Can Burden Agency Staff**

Evaluation startup requires an upfront staffing commitment to operating a program that includes reliable evaluable procedures. This often looks different than typical processes for administration, especially when data collection or tracking is not required for ongoing program implementation. Evaluation procedures may include participant eligibility screening (who could potentially participate in a program), enrollment (who actually participates), and tracking practices (collecting information about participants over time). A standard question to prepare for evaluation is, "How many people are eligible for the program in a given period?" But program staff cannot answer this question without the requisite data, which is not always available from standard program administration procedures. Even if the data exist, they may be hard to access when the agency does not have a dedicated child welfare services analytics team and instead uses their IT department or shares an analytics team with other departments.

Data management challenges in program administration and evaluation become increasingly complex when multiple partners are involved. State laws and regulations often restrict sharing data, particularly with nongovernmental organizations, and even when sharing is permitted, challenges remain. For example, difficulties arise when the child welfare agency refers parents or families to a service provider that also receives referrals from other service systems, such as behavioral health or community-based organizations. These overlapping referral pathways make it essential to carefully manage eligibility and tracking. Ensuring that everyone eligible is referred, and only those eligible are referred, is important for effective program administration, resource management, and

evaluations, and it requires staff to establish and maintain a robust referral process. Establishing and sustaining such a robust referral process requires sufficient staff capacity, clear procedures, and secure data systems.

The need for precise tracking becomes even more important when an evaluation involves random assignment, such as dividing participants between an intervention group and a services-as-usual group. If eligible families are missed during randomization, the study's sample size shrinks, limiting the ability to detect program effects. Conversely, if ineligible participants are randomized, the study's statistical power and validity are undermined. In both cases, weak referral management can compromise both service delivery and the strength of the evidence generated. Reliable estimates of the number of eligible families helps ensure that program slots are filled and allows evaluators to assess the level of confidence associated with impacts observed. However, most programs do not track take-up. Thus, it can be difficult to estimate how many families will participate for a given number of referrals.

While any evaluation is likely to require extra effort for program staff, rigorous approaches designed to attribute cause to outcomes often come with increased complexity. One vulnerability specific to RCTs is contamination. Those assigned to services-as-usual (the control group) could end up participating in the evaluated program, creating what are called *crossovers* in the RCT sample. Measures must be in place to prevent this from happening and often involve complex data integration across different systems, which may require staff and resources agencies do not have. Especially problematic is if a family is assigned to the control group but gets referred to the program via another service provider they are also working with. Establishing a method for checking if referrals from other sources are for families in the control group can be a major undertaking. Even if the child welfare services agency can track this information, it may present too much burden on the service provider.

## Data Collection and Sharing Is Challenging

Data needed for evaluation, particularly for measuring outcomes, can be gathered in different ways. Parents or youth may be interviewed to capture their perspectives and report on outcomes. For some outcomes, this may be the best source. However, collecting information directly from participants adds strain on people already dealing with the issues involved in their child welfare services case. Furthermore, participants may not be able to provide accurate information needed to determine if the program was effective. For some complex or detailed outcomes as well as some sensitive outcomes (e.g., child protection reports, investigations, and foster care placements, and Temporary Assistance for Needy Families and other public benefits receipt), administrative data may be a better option. And for evaluation designs where the evaluators do not have contact with the comparison group, administrative data systems are the only reliable source. In either case, the needed data may reside outside the child welfare services agency.

Program evaluations often require sharing data among the referring agency, the service provider, and an evaluator. In many cases, they also require distinct public institutions to share data with each other to provide to an evaluator. Often, a public agency other than the child welfare agency has access to relevant outcome data, and obtaining those data generally requires child welfare agencies to establish and maintain data-sharing agreements with other public agencies (e.g., secondary and postsecondary education systems, public assistance systems, and employment development departments) or to support evaluators' efforts to negotiate project-specific data-sharing agreements with those public agencies. For example, programs intended to improve the prospects of youth transitioning to adulthood from foster care are often designed to increase youth educational attainment, employment and earnings, and decrease reliance on need-based public benefit programs (e.g., Temporary Assistance for Needy Families and Supplemental Nutrition Assistance Program), outcomes not typically captured by child welfare system data.

Often agencies lack experience sharing data with each other, requiring considerable organizational coordination and effort such as the following:

- establishing contacts with other agencies
- understanding agency data and how to facilitate matching records
- gaining knowledge of federal and state laws and regulations, and agency-specific rules
- working with institutional review boards
- determining logistics and technical details (e.g., which records will be used, by whom, when)

Furthermore, partnering agencies may have little incentive to share their data if they are not also sponsoring the evaluation, and even when an agency is experienced in data sharing, the time required to negotiate new agreements and the public agency's staff capacity can make such sharing difficult.

## **Conclusion**

Despite the challenges, rigorous program evaluation in child welfare services settings is both possible and essential. We have had many successful partnerships with child welfare services agencies resulting in sound evaluations with robust results. In these cases, agency leaders and staff recognized the value of evaluation, understood the resources and processes required, and remained committed over time, even through leadership turnover.

Rigorous evaluation demands sustained investment and dedication, but the payoff is significant: confidence that services are truly helping children and families. Without credible evidence, agencies risk delivering programs that are ineffective or worse, inadvertently harmful, by diverting families from interventions that could meet their needs.

Agencies best positioned to improve outcomes foster a culture that views evaluations as integral to responsible practice. Increased federal, state, and local investments in evaluations infrastructure would not only identify what works but also what needs to change. Redirecting funds from ineffective legacy programs toward proven, evidence-based services strengthens accountability, improves outcomes, and ensures every dollar spent moves the system closer to helping children and families.

## **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.