Every Child Deserves a Permanent Home: The Permanency Innovations Initiative

Mark Testa, Kristen Woodruff, Roseana Bess, Jerry Milner, and Maria Woolverton

Summary

About one-fifth of children involved in investigations for abuse or neglect are placed in foster care. Although some return to their families quickly, others may remain in foster care for years without permanent family relationships. In this article, Mark Testa, Kristen Woodruff, Roseana Bess, Jerry Milner, and Maria Woolverton examine the Permanency Innovations Initiative (PII), a federally funded effort that tested innovative programs designed to prevent children from experiencing long stays in foster care and to build evidence for strategies that can be brought to scale in child welfare.

PII aimed to follow a four-phase model for selecting, implementing, and testing interventions, including *exploration and installation, initial implementation and formative evaluation, full implementation and summative evaluation, and replication and adaptation*. The results of the initiative weren't encouraging. Some sites were never able to move to the full implementation phase. Others had significant trouble with participation rates. Two sites that were able to experimentally evaluate a fully implemented intervention found no significant differences between the treatment and comparison groups in achieving stable and permanent homes for children, and a third site found that the experimental results actually favored the comparison group.

The authors "principal finding" is that "none of the promising innovations tested in this initiative yielded meaningful improvements in … stable permanence when rigorously evaluated." Discussing the implications for child welfare programs in general, they raise a fundamental issue: Should such programs primarily deal with maltreatment only after it has occurred? Or should they also work to prevent maltreatment from happening in the first place through early, universal interventions that strengthen protective factors within families?

Mark Testa is the Spears-Turner Distinguished Professor in the School of Social Work at the University of North Carolina at Chapel Hill. Kristen Woodruff is senior study director at Westat. Roseana Bess is a senior researcher at Mathematica Policy Research. Jerry Milner is associate commissioner of the Children's Bureau in the Administration for Children and Families (ACF), US Department of Health and Human Services. Maria Woolverton is director of the Division of Family Strengthening in the ACF's Office of Planning, Research and Evaluation.

Research over the past half-century has shown that children's health and emotional wellbeing is best assured in the context of permanent family relationships.[1] This issue of *Future of Children* highlights a range of prevention programs that attempt to provide community supports to parents and children—strengthening parental capacity, increasing child safety, and enhancing child development—so that children can remain safely in their own homes. Child welfare policy leans toward maintaining children at home, and most children who come to the attention of child protective services do remain in the custody of their parents. However, approximately one-fifth of the victims involved in investigations or assessments for maltreatment (that is, abuse or neglect) are placed in foster care.[2] Some children return home quickly, but others remain in foster care for years, without permanent family relationships. As federal policy has shifted to prioritize family permanence, the number of children who stay in foster care for longer than three years has fallen—by 50 percent between 2000 and 2010, from 172,000 to 87,000.[3] Still, many children continue to experience long-term foster care.

The child welfare system reacts to crises, rather than preventing crises from happening. That is, it intervenes in families' lives only after those families are in crisis, rather than helping them avoid crises in the first place. We still have much to learn about how to effectively serve children and families in crisis, particularly those facing the most difficult challenges, so that children can return safely to their birth parents or more quickly achieve other permanent family relationships. Although foster care is needed to protect children and youth from unsafe environments, too many children remain in foster care for years without achieving permanence in the form of reunification, adoption, or guardianship.

Few evidence-supported interventions are geared to the needs of children at risk of long-term foster care. We need more innovations and more well-supported evidence of what works to ensure timely permanence and support children's social and emotional wellbeing in family relationships. Moreover, we need the capacity to generate this evidence. In this article, we present an initiative that's designed to test whether innovative interventions can meet the evidence standards necessary to conclude that the interventions produce positive results for children in foster care.

Permanency Innovations Initiative

The Children's Bureau of the US Administration for Children and Families (ACF) launched the Permanency Innovations Initiative (PII) in 2010 to support implementation of innovative intervention strategies and to evaluate their effectiveness in improving outcomes for children at risk of long-term foster care. The Children's Bureau oversaw the initiative jointly with the ACF's Office of Planning, Research and Evaluation (OPRE). PII was a multi-year, $100 million federal program that funded promising innovations at six sites. The idea was that if reliable implementation and rigorous evaluation showed that any of the interventions effectively improved family permanence and other measures of child wellbeing, those interventions could be scaled up nationally to minimize the number of children who experience long-term foster care.

Box 1. The Six PII Grantees

- Arizona Department of Economic Security
- California Department of Social Services
- Illinois Department of Children and Family Services
- Los Angeles LGBT Center
- University of Kansas Center for Research
- Washoe County, NV, Department of Social Services

Each of the six grantees identified the population in their community of children and youth that faced the most serious barriers to family permanence. Even though these children, youth, and families had already come to the attention of child protective services, PII's aims were preventive in the sense that it sought to avert long-term foster care for traumatized children who were at high risk of remaining in care. The grantees implemented innovative programs that were intended to prevent children in the target population from experiencing long stays in foster care—or, in some cases, from entering foster care at all—and to ensure that when the children exited care, they went to a permanent family home.

PII set high standards for building evidence. It helped the sites conduct rigorous evaluations that could demonstrate a sustained intervention effect when compared to a randomized or matched comparison group, in which "permanent" exits endured beyond the finalization of legal permanence (the benchmark in prior studies). Such high standards meant that fewer than one out of five promising innovations could be expected to pass successfully through all phases of evidence building, but PII's goal was not to advocate for a single cure-all solution.[4] Rather, building on the experimentalist approach advocated by social psychologist Donald Campbell, PII aimed to develop and sustain a continuous cycle of evidence building while testing innovative strategies to reduce long-term foster care. That is, it aimed to systematically explore, reliably implement, and rigorously test strategies to reduce the problem of long-term foster care, and to test alternative solutions should evaluation show that the initial intervention was ineffective or possibly even harmful.[5]

Child welfare lagged well behind in its capacity to generate systematic evidence for what works for whom under what conditions.

Status of Evidence Building in Child Welfare

PII was an example of the federal government's approach to evidence-based policy making, which also included initiatives in education, maternal and child health, teenage pregnancy prevention, community service, and workforce development.[6] Compared with these other human service areas, however, child welfare lagged well behind in its capacity to generate systematic evidence for what works for whom under what conditions.

In 2010, only 20 (9 percent) of the 223 programs cataloged on the California Evidence-Based Clearinghouse for Child Welfare website were well supported by research evidence. As of March 2018, among the 433 programs cataloged, the number was higher in absolute terms at 31, but proportionally lower at 7 percent. Only two of the 31 well-supported interventions were specifically designed or commonly used for children and families served by

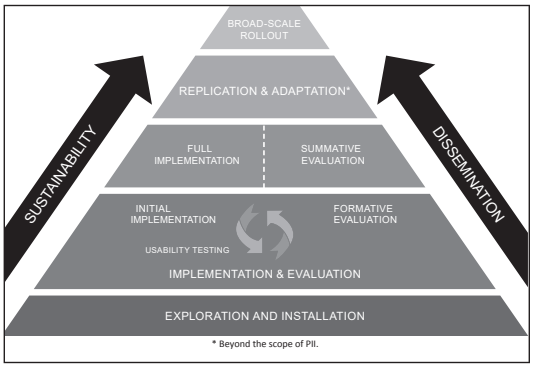
the child welfare system. Still, evidence-supported interventions stemming from the fields of mental health and developmental science held some promise for ameliorating the behavioral and emotional problems of children who come to the attention of the child welfare system.[7] Several PII projects examined how far this promise could extend to children in the usual court-ordered, out-of-home settings.

PII Approach to Evidence Building

In response to the dearth of evidence-supported interventions geared specifically to the needs of children at risk of long-term foster care, the Children's Bureau and OPRE asked the PII Evaluation Team (PII-ET) and the PII Training and Technical Assistance Project to organize a systematic, phased approach to developing, adapting, and implementing interventions with *integrity* (that is, implementing them as planned or as previously tested, in support of their efficacy or effectiveness) and showing empirically that they would work with other similar children and youth beyond those in the studies (what researchers call *external validity*).

In PII's approach, evidence building progresses through four phases, or "tollgates," before a program can move to broad-scale rollout.[8] Figure 1's pyramid illustrates how at each tollgate, many interventions fail to progress to the next phase of evidence building. Thus, when properly evaluated, few interventions prove to be effective or even marginally successful.[9] Given this reality, the earlier in the evidence-building process that a tollgate warning can be sounded, the better. Otherwise, much time and effort may be misspent in implementing and evaluating promising innovations that ultimately fail to produce positive results.

Figure 1. The PII Approach



As figure 1 shows, the four PII tollgates are:

1. *Exploration and installation:* choosing promising innovations to install in real-world settings, based on the best available research evidence of past success.

2. *Initial implementation and formative evaluation:* confirming a program's usability and statistically testing whether its outputs and primary short-term outcomes are trending in the desired direction.

3. *Full implementation and summative evaluation:* supporting implementation as planned (with integrity) and rigorously evaluating whether the intervention creates practical improvements in primary long-term outcomes that can plausibly be attributed to causal effects of the intervention.

4. *Replication and adaptation:* spreading evidence-supported interventions and assessing whether similar positive outcomes can be reproduced with diverse populations at different time frames and in different settings.

Exploration and Installation

The first tollgate involves the *construct validity* (that is, whether a test measures the

concept it's intended to measure) of the research questions and the logic model (a tool that describes the key implementation activities, program outputs, and short-term outcomes each site deems necessary to attain the desired results). Construct validity is strengthened by 1) starting with a clear exposition of the population, intervention, comparison, and outcome (or PICO) constructs of interest; 2) choosing reliable indicators of these higher-order constructs; 3) assessing the fit between the particular indicators and the constructs; and 4) revising and summarizing the PICO construct descriptions in the form of a question.[10]

During the exploration and installation phase, the PII Training and Technical Assistance Project helped the sites set up implementation teams and create a supportive context to solidify child welfare system leadership and stakeholder buy-in and to sustain the site's investment in successful implementation.[11] The long-term outcome measure (the O in PICO), developed by the PII-ET and PII Training and Technical Assistance Project in consultation with the sites, extends the federal measure of permanence. The PII measure stipulates that a child's exit from foster care to reunification, adoption, or guardianship qualifies as *stable* only if it lasts at least six months after exit, without reentry into foster care. The extension helps make sure that sites don't register quick improvements by simply discharging more children than before from state custody without adequately preparing families or offering services to support family permanence.

Even though the Children's Bureau specified that the target population (P) should constitute subgroups of children who experience the most serious barriers to permanence, it left the selection of particular subgroups to the local sites' discretion. To verify that the subgroups proposed in each site's application faced the most serious barriers to permanence, PII-ET extensively mined administrative data. Their analyses ranked the subpopulation characteristics at each site that correlated most strongly with children remaining in foster care for two or more years. In some cases the analysis confirmed the site's original selection; in other cases, it helped the project refocus on risk factors that more strongly predicted long-term foster care.

Several sites had to collect their own data to estimate the target population's size and needs. For example, the Los Angeles LGBT Center funded a survey of foster care youth aged 12 years and older in Los Angeles. A sample of 1,881 youth, split into two groups by age (12–16 and 17–21), was chosen randomly from a population of approximately 7,000 youth in foster care. A total of 786 youth completed telephone surveys, 42 percent of the sample.[12] Based on their responses to a set of questions on sexual attraction and identity, the researchers estimated that approximately 19 percent of youth in foster care in Los Angeles identified as LGBTQ. This was 1.5 to 2 times greater than the percentage of LGBTQ people estimated for the population at large.[13] Extrapolating to the entire population of youth in foster care, researchers estimated that some 1,400 foster care youth in Los Angeles identified as LGBTQ and could potentially benefit from the Los Angeles LGBT Center's services.[14]

Theory of Change, Logic Model, and Research Review

The two technical assistance teams helped each site develop a *theory of change* that 1)

elaborated on the site's basic understanding of the nature of the problem, and 2) outlined a *logic model* specifying the key implementation steps and the underlying causal pathways that were hypothesized to bring about the desired changes. PII-ET then conducted a research review to identify the best available evidence of past success producing the desired outcomes among the interventions the sites were considering.

With the logic model and research reviews in hand, each site selected one or more interventions. Table 1 summarizes the evidence ratings and the interventions (I) selected by each site. Evidence ratings are based on PII-ET's assessment of the rigor and consistency of the evidence of a program's effectiveness: level 1 is the highest rating (well supported, with positive evidence from two or more randomized clinical trials), and level 4 the lowest (promising and acceptable).[15] The other columns identify the remaining PICO components. Also listed are short-term outcomes, which PII-ET evaluated in addition to the primary long-term outcome (O) of time to stable permanence.

Some of the selected interventions were those originally proposed by the sites, but others had to be developed anew. For example, both California sites developed system-focused interventions targeting structural inequities and institutional biases that expose stigmatized groups and racial minorities to a disproportionate risk of removal from their homes and retention in long-term foster care. Because no existing system-focused interventions were geared specifically to the needs of LGBTQ youth in foster care, the Los Angeles LGBT Center created its own program. Similarly, the California Department of Social Services constructed its own Child and Family Practice Model that integrated common elements from a variety of practices with research evidence to support them. The department focused on African American and Native American youth because data mining reinforced its original assessment that these two minority groups were at a particularly high risk of long-term foster care.

Each of the other sites chose one or more existing person-focused or relationship-focused interventions. Kansas selected Parent Management Training–the Oregon Model (PMTO) for reorienting family interactions away from coercive parenting and toward positive parenting practices that help reduce problematic child behavior and reinforce prosocial behaviors.[16] Illinois selected the person-focused Trauma Affect Regulation: Guide for Education and Therapy for Adolescents (TARGET-A) intervention, which helps child maltreatment victims and their caregivers prepare for and manage destructive reactions to neurobiological changes induced by childhood trauma and toxic stress.[17] Arizona chose two relationship-focused interventions that aim to increase the supply of permanency resources: Family Finding, which searches for relatives, neighbors, and other caring adults from a youth's past who might be recruited as legal guardians or adoptive parents; and 3-5-7, which helps children come to terms with unresolved issues of separation, loss, trust, and self-identity and open up to joining a family permanently.[18] Washoe County, NV, integrated two relationship-focused programs into a unified intervention: SAFE, that helps parents accept what must change to protect their children; and Family Connections, which helps parents arrange

Table 1. PICO Components for Grantees: Population, Intervention, Comparison Group, and Outcomes

Target Population	Intervention Selected *(Evidence Level*)*	Comparison Group	Short-Term Outcomes
Arizona Department of Economic Security			
Population 1: Children and youth aged 5–17 years who at one year after removal were deemed at risk of long-term foster care	3-5-7 *(4)*	Randomized comparison group	• Readiness for permanence • Behavioral health • Placement stability
Population 2: Youth aged 13–17.5 years who have been in care two or more years	Family Finding *(4)*		
California Department of Social Services			
African American and Native American children	Child and Family Practice Model (constructed) integrating common elements after reviewing a series of interventions	Matched historical comparison groups	• Family perceptions of changed practice behaviors • Racial disparities in time to family permanence
Illinois Department of Children and Family Services			
Youth aged 11–16 years who are in traditional, relative and specialized foster homes and, after two years of care, are experiencing mental health symptoms and/or have had more than two placement changes	Trauma Affect Regulation: Guide for Education and Therapy for Adolescents (TARGET-A) *(1)*	Randomized comparison group	• Placement stability • Parenting skills in responding to youth's emotional and behavioral dysregulation: CANS and Abbreviated Dysregulation Inventory
Los Angeles LGBT Center			
LA caseworkers and children and youth aged 5–19 years who self-identify as LGBTQ+ and/or gender nonconforming	Recognize, Intervene, Support & Empower (RISE); Outreach & Relationship Building (ORB) *(4)*	One group tested before and after intervention	• Knowledge of LGBTQ+ competencies • Perceived agency support
	RISE–Care Coordination Team (CCT) *(4)*	Randomized comparison group	• Family support • LGBT identity • Integration into LGBT community
University of Kansas Center for Research			
Children and youth aged 3–16 years who meet criteria for serious emotional disturbance	Parent Management Training–Oregon Model (PMTO) *(1)*	Randomized comparison group	• Positive parenting behavior based on parental reports of child's compliance and own parenting behavior, and independent ratings of Family Interaction Task (FIT) observations
Washoe County (Nevada) Department of Social Services			
Population 1: All new cases with a report of child abuse or child neglect who are deemed unsafe	Safety Assessment Family Evaluation (SAFE) *(4)*	Randomized comparison group	• Caregiver readiness for change (Readiness for Parenting Change scale)
Population 2: Children in foster care for at least 12 months with one or more risk factors	Family Connections (FC) *(3)*		• Parenting Stress Index (PSI-Short Form)

**Evidence level* refers to PII-ET's assessment of the level of evidence of program effectiveness using criteria suggested by Barbara Thomlison (2003), where level 1 is the highest rating (well supported, with positive evidence from two or more randomized clinical trials), and level 4 the lowest (promising and acceptable).

for supports and services so their children can either remain safely in their custody or, in cases where removal is necessary, quickly return home.

Initial Implementation and Formative Evaluation

The second PII tollgate involves program usability and the *statistical (*or *conclusion) validity* of differences in program outputs and short-term outcomes. Statistical validity refers to the likelihood that any observed differences in outputs and outcomes between the intervention and comparison groups are greater than what could be expected simply by chance.[19]

Early on, each site tested its intervention's usability with a small sample of the target population. This road testing allowed the sites to quickly revise both the interventions and the implementation activities (such as an assessment form to be completed by program participants) before formative evaluation began. It also gave the sites a chance to reappraise their capacity for the initiative and their interest in participating. After usability testing, the Arizona site underwent a change in leadership and withdrew from PII.

If a program has trouble passing its own logic model, decision makers should think twice before embarking on full implementation.

The remaining five sites then implemented their programs with a larger but still small sample of children and youth as part of the *formative evaluation*. (Formative evaluation tests whether an intervention is associated with expected program outputs and short-term outcomes.) During this phase, sites need to pay close attention to whether what's actually happening follows what was supposed to happen according to the logic model.[20] Small formative samples limit the ability to accurately infer a program's effectiveness. Nonetheless, if a program has trouble passing its own logic model, decision makers should think twice before embarking on full implementation and summative evaluation.[21]

Formative evaluation doesn't require as rigorous a design as summative evaluation does, although such a design can be used. Each of the three sites that progressed to summative evaluation used an "early warning" summative design for their formative evaluation. This type of formative evaluation pilots the same unbiased assignment mechanism to form intervention and comparison groups (for example, random assignment) that will be used at summative evaluation. California's intervention didn't use a similar early warning design because too few counties were enrolled in the demonstration to mount a county-level randomized controlled experiment. Instead, California compared the outcomes for children served by the practice model to a matched historical comparison group—that is, children who, based on administrative data, had been served sometime in the past and were similar to the children receiving the intervention. The Los Angeles LGBT Center planned to randomize child cases to its Care Coordination Team intervention, compared to services as usual. But because they had fewer referrals than anticipated, randomized allocation to intervention and comparison groups wasn't feasible.

Therefore, both California sites relied on comparisons of children and youth who received the interventions to matched historical cohorts who did not, in order to assess the statistical validity of the association between interventions and outcomes and draw tentative inferences about program effectiveness. Neither of the California sites moved on to full implementation and summative evaluation.

The formative results for the LGBT Center's training showed a modest improvement in caseworkers' foundational knowledge and practical use of LGBTQ concepts. However, heavy attrition of participants from the two-month follow-up survey made it hard to measure how well the caseworkers applied the knowledge in practice. Among the 21 percent of respondents who reported not applying the knowledge, nearly 60 percent said they didn't think the information was relevant for their clients. Considering that the exploratory findings had suggested that about 19 percent of surveyed foster youth identify as LGBTQ, it would be important to learn whether the lack of perceived relevance reflected the caseworkers' failure to recognize LGBTQ youth, their discomfort with the issue, or actual differences with respect to the ages and other characteristics of the clients.

The Child and Family Practice Model implemented in five locations in four selected California counties was a system-focused intervention designed to reduce racial disparities in permanency. The formative evaluation tested whether children who were served by caseworkers trained in the new practice model showed evidence of a lower risk of long-term foster care as compared to a similar matched historical group of children in foster care. When children served under the new practice model in one of the five locations were assessed at 12 months after removal from the home, they showed a small but statistically significant improvement (three percentage points) in the time taken to achieve stable permanence. In the other four locations, testing found no differences in the time to stable permanence between the intervention group and the matched historical comparison group. Nor was a statistically significant drop found in disparities among African American and Native American children, compared to other children.

Full Implementation and Summative Evaluation

The person-focused and relationship-focused interventions in Illinois, Kansas, and Washoe County, NV, progressed to full implementation and summative evaluation to test *internal validity*—that is, to determine whether the short- and long-term outcomes were achieved and whether the statistical association observed between intervention and outcomes could plausibly be attributed to the intervention's causal impact.[22] These three sites randomly assigned eligible children and families to intervention and comparison groups. We compared outcomes of all children and families assigned to the intervention (whether or not they had participated fully in their assigned treatment) to outcomes of all children and families assigned to the comparison group. This design, called *intent-to-treat* analysis, provides an internally valid estimate of the intervention's impact in the real world, where some families refuse to participate or don't do so fully. As expected, the extent to which families and youth participated in the treatment at each site varied. Administrative child welfare data made it feasible to

measure time to stable permanence for all children in the study. This approach provides macro-level information to help decide how to invest limited resources in policies and programs.

Program Participation

Illinois

Of the 233 youth assigned to TARGET-A, 48 percent didn't participate in any TARGET sessions. By the end of the two-year summative evaluation period, only 25 percent had completed the full course of ten to 12 sessions. Sixteen percent completed three to nine sessions, and 12 percent completed only one or two. Though the participation rate was lower than organizers had hoped, it was consistent with other TARGET studies of both adults and justice-involved youth.[23]

Kansas

Seventy-eight percent of the Kansas Intensive Permanency Project sample consented to participate in PMTO and data collection. Among those who consented, 73 percent fully complied with and completed the intervention (15 modules in six months or less).

Washoe County, NV

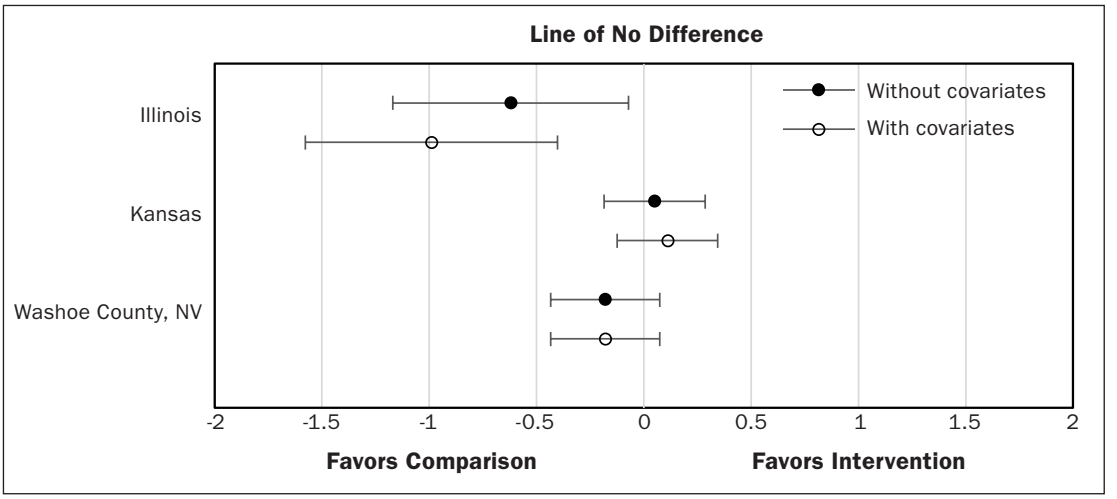
All cases assigned to SAFE-FC received the intervention until the case was closed and services to the families stopped. But even though all cases were exposed to treatment, the proportion of missing data on short-term outcomes ranged from 70 to 80 percent. One-third of the families chose not to provide contact information for primary data collection, and many who had agreed to do so didn't complete one or more of the before-and-after assessments. With so much data missing, the evaluation team couldn't estimate SAFE-FC's effect on short-term outcomes. But because they had administrative data for all cases, the team was able to estimate the intervention's effect on the long-term outcome of timely and stable permanence for all those in the treatment group.

Summative Findings: Timely and Stable Permanence

The summative evaluation estimates how each of the three interventions affected timely and stable permanence, the primary long-term outcome. Figure 2 shows the estimated intervention effect and confidence interval for each site, indicating whether there was a statistically significant difference between intervention and comparison groups in the rates at which children exited foster care to stable permanent homes.[24] In Kansas and Washoe County, the confidence intervals (lines) cross the zero line, indicating no statistically significant difference between intervention and comparison groups in time to stable permanence. In Illinois, the confidence intervals fall below zero, indicating that the comparison group fared better than the intervention group. Confidence intervals entirely above zero would indicate that the intervention had a positive effect, that is, a shorter time to permanence.

Figure 2 presents two estimates for each site: both a crude (simple) analysis and an analysis that's fully adjusted to account for other variables. Including important predictive variables as controls in statistical models can make effect estimates more precise.[25] Crude estimates (the black circles) show the intervention's impact on timely and stable permanence without covariates; adjusted estimates (white circles) account

Figure 2. Timely and Stable Permanence: Intervention Effect and Confidence Intervals



for important covariates—variables selected for their significant association or interaction with the outcome. The adjusted estimates don't differ much from the crude estimates except in the case of Illinois: when important predictive factors are included, that estimate favors the comparison group over the intervention group.

Primary Short-Term Outcomes

Each of these three sites identified a primary short-term outcome of focus from their logic model. In Illinois, it was placement stability; in Kansas, improved parenting behaviors; and in Washoe County, caregivers' readiness to change their own parenting behaviors. The primary short-term outcomes were chosen based on the pivotal role they were hypothesized to play in supporting the long-term outcome of timely and stable permanence.

Illinois

The short-term outcome of placement stability was measured using administrative data that was available for all youth in the intervention group whether or not they received services. Compared to the randomized control group, youth assigned to TARGET-A showed no differences with respect to the number of foster home or institutional placements, or whether running away, detention, or psychiatric hospitalization temporarily interrupted a spell of foster care.

Kansas

Changes in parenting behaviors were measured by the Family Interaction Task, used in previous studies of PMTO to track changes in parenting behaviors.[26] Trained coders, blind to the assigned treatment, viewed videotapes of family members working together on interactive tasks. Coders rated behaviors in six areas: positive involvement, skill encouragement, monitoring, problem solving, inept discipline, and child noncompliance. Only 65 percent of assigned families completed the Family Interaction Task when the study began, and 46 percent did so at follow-up. Given the large amount of missing data, the analysis used imputed data values (meaning estimated or substituted data values) for families who missed the measure.

Contrary to expectations, the families assigned to PMTO fared worse than families in the comparison group on measures of inept discipline and child noncompliance. There were no significant differences for the remaining areas of positive involvement, skill encouragement, monitoring, and problem solving. When all six subscales were added up for an overall measure of parental effectiveness, the results showed no differences between the intervention and comparison groups.

Washoe County, NV

SAFE-FC assessed caregivers' initial readiness for change using a validated instrument called the Readiness for Parenting Change scale.[27] Even though the analysis showed no effect, the results' usefulness is limited by the large amount of missing data that had to be imputed (69 percent of the data was missing at the beginning of the assessment and 77 percent at follow-up).

Promising Directions for Replication and Adaptation

After the average effectiveness of an intervention has been demonstrated through summative evaluation, the PII Approach envisions clearing the fourth tollgate to qualify the intervention as well supported by research evidence—that is, to establish the *external validity* of the intervention's causal impact.[28] External validity refers to whether and how well an intervention's impact can be transferred to other settings or to variations in time frames, populations, and outcomes. The last phase before broad-scale rollout—replication and adaptation—was beyond the scope of PII. But during summative evaluation, researchers explored the statistical associations between secondary short-term outcomes and assignment to the intervention. They also examined outcomes within different subgroups and across settings (such as substance dependence, placement type, and racial group). The purpose was not to cherry-pick confirming results but rather to identify promising directions that could guide future replication and adaptations.

Illinois

Reports from youth indicated that the intervention had the intended effect of increasing in-person monthly visits with fathers and other types of monthly parental contact. Youth self-reports also showed a marginally significant reduction in later exposure to trauma (meaning a single traumatic event). There was no significant impact, however, on later exposure to complex trauma (that is, multiple traumatic events) or on mental health symptoms as documented by caseworkers in the Child and Adolescent Needs and Strengths assessments (a tool that supports decision making, service planning, and outcomes monitoring for children's services). Nor was any association found between assignment to the intervention and a change in the availability of adults as sources of emotional support, or in the youths' capacity to form and maintain relationships. More concerning, assignment to the intervention unexpectedly trended in the wrong direction with respect to behavioral and emotional dysregulation. Youth assigned to the intervention reported a greater increase, on average, in symptoms related to behavioral and emotional/affective dysregulation than did youth assigned to the comparison group. There was no significant intent-to-treat effect on either cognitive dysregulation or the dysregulation

measure in the Child and Adolescent Needs and Strengths assessments.

Kansas

Assignment to treatment showed a significant intent-to-treat effect in the form of improving child functioning scores given by caseworkers, from severe to minimal functional impairment. Parental assessments showed the same positive impact: parents reported that children exhibited fewer problem behaviors and more prosocial skills regarding communication, cooperation, assertion, responsibility, empathy, engagement, and self-control. On the North Carolina Family Assessment Scales, a tool that measures family functioning, a difference in family readiness for reunification wasn't statistically significant, and none of the other domains—parental mental health, substance abuse, or use of community resources and social supports—showed a difference between the intervention and comparison groups.

Discussion and Conclusions

PII's principal finding is discouraging: when rigorously evaluated, none of the promising innovations we tested yielded meaningful improvements in the primary outcome of timely and stable permanence. However, this finding is understandable given past accomplishments in reducing long-term foster care, and the Children's Bureau's focus on achieving permanence for children and families who face the most serious barriers to doing so. PII's limited efficacy in reducing the problems associated with long-term foster care suggests that we need a sustained commitment to developing more evidence-supported interventions in child welfare. The results for Illinois and Kansas, in particular, raise questions about whether person- and relationship-focused interventions from the fields of mental health and developmental science can be transferred to the usual court-ordered, out-of-home settings. It's not entirely clear whether parenting interventions such as PMTO are effective for birth parents who are working toward reunification with their children, particularly given their lack of opportunity to practice the newly learned parenting techniques at home with their children. The same may be said for the effectiveness of TARGET-A in helping children in foster care regulate trauma symptoms, and helping foster caregivers manage child behavior problems that disrupt regular family life. Perhaps to ensure that more children achieve permanence faster, we need to systematically test innovations that cover the spectrum of maltreatment prevention.

The fact that the national foster care population has been increasing since 2012 suggests that we should fundamentally question exactly what the child welfare system is designed to accomplish and what interventions best serve children and families. For example, does child welfare exist only to protect against further harm after a child becomes involved in the system? Or should it also help avoid maltreatment (and thus involvement in the system) in the first place? If its purpose is to help avoid maltreatment, any discussion of evidence building should focus on interventions to strengthen families, such as the Triple P or Family Connects programs discussed elsewhere in this issue, before the need for foster care presents itself—and certainly before children suffer the trauma of neglect and abuse.

We should fundamentally question exactly what the child welfare system is designed to accomplish and what interventions best serve children and families.

Other human service institutions have understood and acted on the belief that developing interventions and safeguards to prevent bad things from happening is far more effective than responding only after the case. The public health system, the food industry, the auto industry, and even our athletic institutions are good examples. In contrast, federal child welfare policy remains largely uncommitted to either funding or seriously tackling true prevention of the initial occurrence of child maltreatment—precisely the social problem for which child welfare systems exist. As long as child welfare systems remain reactive, we will continue to focus on what children need only *after* they're maltreated or removed from the home, and we'll likely spend our energies trying to remedy trauma rather than prevent it, and trying to achieve permanence faster for children who are already in foster care.

We know that lifelong connections to supportive adults are key to improving outcomes for youth in foster care. We also know that however strained or struggling the relationships, the most important sources of such connections are parents and extended family. Thus a critical part of the solution is to develop relationship-focused interventions and supports with parents early and universally to strengthen protective factors. With the vast amount of research now available on trauma, brain science, and wellbeing, child welfare can't just be concerned with securing foster care beds and protecting the physical safety of maltreated children. Yes, physical safety is core to the mission of child welfare, but its presence alone doesn't equate to social and emotional wellbeing, nor does it necessarily guard against repeated cycles of maltreatment across generations.

In 2018, Congress passed and President Trump signed the Family First Prevention Services Act (Family First), a positive step toward preventing some of the damage that can be inflicted by child maltreatment. Family First provides prevention dollars that weren't previously available—an option for states to use the largest pool of federal child welfare funds, Title IV-E, for certain types of prevention services for foster care candidates (that is, children identified in a prevention plan as being at imminent risk of entering foster care). This flexibility will help many children and families avoid foster care placement after becoming involved with the child welfare system. To build on this important initial step of putting families first, however, we must also try to make families better equipped to deal with adversity and to protect their children before they're in crisis and require child protection interventions. Also, we must systematically build evidence about what works and what doesn't work in preventing maltreatment and, ultimately, ensuring that more children grow up in safe and permanent homes.

The PII approach and some of the lessons learned from it are relevant for establishing evidence for the primary prevention of maltreatment and across the child welfare continuum. In the future, as we review potential interventions in the exploration

stage, we must carefully consider whether interventions that work in one relational context would generalize well to other such contexts. For example, and as we saw in Illinois and Kansas, well-supported interventions from behavioral health may not translate well to a family separated by foster care. If implementation as intended isn't feasible, even after considering creative strategies to overcome barriers, then we must move on to another intervention or consider adapting interventions to suit the context. At the initial implementation and formative evaluation stage, we must confront the difficulty of measuring the effectiveness of system-focused interventions that are all-inclusive and not limited to specific person- or relationship-focused interventions. Such broad-based programs often comprise an array of supports to strengthen and assist families and to create environments that are strong in the protective factors that help families avoid the need for child protection. Administrative data make it possible to measure some key outcomes for entire populations so we can assess the impact of a systemwide intervention, even though other data may not be recorded in administrative systems. We shouldn't shy away from keeping the desired outcomes front and center even when they're difficult to measure, such as the short-term outcomes that California and the LGBT Center sought to enhance—namely, treating vulnerable families and youth with respect, compassion, and decency. Primary data collection with targeted populations and carefully planned sampling methods can give us rich information that supplements the more readily available administrative data. Another challenge is that some system-focused interventions, such as those at the California sites, don't lend themselves to randomized controlled trials because too few counties or participants are enrolled. Instead, rigorous quasi-experimental alternatives, such as those described earlier, may be the best alternative for assessing impact.

The difficulty that Illinois, Kansas, and Washoe County faced in reproducing the positive impacts of person- and relationship-focused interventions suggests that interpersonal relationships are key to the success of specific clinical interventions. Evidence-based policy making is largely concerned with improvements on average. Rigorous evaluations may miss the impact made by a single individual—say, a social worker, attorney, judge, or service provider—on the life of a particular child or parent. Further, interventions that work well on average in one relational context may not transfer well to other relational contexts, as we saw in both Illinois and Kansas. Besides building evidence for the generalized efficacy of specific interventions, we need to find a way to add both system-focused interventions and more individualized, relationship-focused interventions to our storehouse of what works for children and families. We must look at the approaches that can really make a difference for families whose needs don't meet the criteria for a specific clinical service that works on average—that is, families who may require system-focused interventions of universal family support and decent treatment of all people.

As we consider building the evidence for primary prevention of child maltreatment, at least three areas should figure into our review of programs and into the question of what constitutes evidence. First, effective primary prevention services should be universal, and offered without regard to which families have demonstrated a risk of maltreating their children. The reason is simple: anyone could become at risk, and

if we wait for the risk to appear, the family may already be in crisis. Also, offering family support services universally reduces the stigma that can prevent a family from seeking and using the very services that might keep it from becoming involved in the child welfare system.

Second, primary prevention services should be designed and offered at the community level, and perhaps be unique to the needs of particular communities, depending on demographics and other characteristics. To reshape the child welfare system so that it focuses on preventing children from being maltreated, we must be able to describe and stand behind successful community-based programs that meet that test—despite the measurement challenges.

Third, primary prevention services and approaches will almost always comprise a combination of services and activities to strengthen and support families' protective capacities. It may be difficult if not impossible to determine what made the difference for a family—whether it was the legal advocacy and services, the housing support, the food pantry, the after-school parent engagement work, or some combination of these. Nevertheless, we should make every effort to understand the value of these approaches in preventing maltreatment, and to consider what evidence can best establish that value to children and families.

The path to well-supported interventions laid down under Family First is admittedly an arduous one. Family First devotes the majority of federal funding to evidence-supported interventions. To qualify as *supported*, the effect must be sustained at least six months beyond the end of treatment, when compared to a control group. To qualify as *well supported*, another rigorous study must replicate the results, and the effect established in one of the studies must be sustained for at least one year beyond the end of treatment. Given the four-to-one odds against showing improvements, we need to learn new and more efficient ways of building evidence, as a routine part of government operations, to achieve the high standards of effectiveness envisioned by Family First.[29]

Showing that an idea doesn't work as intended can be just as valuable as showing that it does. The statistical associations found in post-hoc evaluations of TARGET-A and PMTO point to some promising directions that policy makers and administrators should consider in the next cycle of evidence building. To sustain a supportive, enabling context for evidence building in child welfare, we should heed the advice dispensed by Donald T. Campbell a half-century ago:

> Administrators and parties must advocate the importance of the problem rather than the importance of the answer. They must advocate experimental sequences of reforms, rather than one certain cure-all, advocating Reform A with Alternative B available to try next should an honest evaluation of A prove it worthless or harmful.[30]

It's well established that child maltreatment and the absence of stable family attachments have adverse effects on later physical and mental wellbeing. Despite the challenges, ending the practice of long-term foster care and promoting family permanence to support the future social and emotional wellbeing of children and adolescents are policies that deserve sustained public commitment.

Endnotes

1. Fred Wulczyn et al., *Beyond Common Sense: Child Welfare, Child Well-Being, and the Evidence for Policy Reform* (New Brunswick, NJ: Transaction, 2005).
2. US Department of Health and Human Services, Administration for Children and Families, Children's Bureau, "Child Maltreatment," last updated February 1, 2018, https://www.acf.hhs.gov/cb/research-data-technology/statistics-research/child-maltreatment.
3. US Administration for Children and Families, Children's Bureau, "Adoption & Foster Care Statistics," accessed March 30, 2018, https://www.acf.hhs.gov/cb/research-data-technology/statistics-research/afcars.
4. The one out of five benchmark is based on two reports: Laura and John Arnold Foundation, "How to Solve U.S. Social Problems When Most Rigorous Program Evaluations Find Disappointing Effects (Part One in a Series)," *Straight Talk on Evidence*, accessed March 27, 2018, http://www.straighttalkonevidence.org/2018/03/21/how-to-solve-u-s-social-problems-when-most-rigorous-program-evaluations-find-disappointing-effects-part-one-in-a-series/; and Research and Evaluation Conference on Self-Sufficiency, "Past Conferences: 2016 Session Videos," Jim Manzi, "Rapid Cycle Testing: Lessons from the Private Sector," accessed March 27, 2018, http://www.recsconference.net/2016videos.htm.
5. Donald T. Campbell, "Reforms as Experiments," *American Psychologist* 24 (1969): 409–29, http://dx.doi.org/10.1037/h0027982.
6. Ron Haskins and Greg Margolis, *Show Me the Evidence: Obama's Fight for Rigor and Results in Social Policy* (Washington, DC: Brookings, 2015).
7. Richard P. Barth et al., "Parent-Training Programs in Child Welfare Services: Planning for a More Evidence-Based Approach to Serving Biological Parents," *Research on Social Work Practice* 15 (2005): 353–71, https://dx.doi.org/10.1177/1049731505276321.
8. Diana Epstein and Jacob Alex Klerman, "When is a Program Ready for Rigorous Impact Evaluation? The Role of a Falsifiable Logic Model," *Evaluation Review* 36 (2012): 375–401, https://dx.doi.org/10.1177/0193841X12474275.
9. Peter H. Rossi, "Issues in the Evaluation of Human Services Delivery," *Evaluation Quarterly* 2 (1978): 573–99, https://dx.doi.org/10.1177/0193841X7800200404.
10. William R. Shadish et al., *Experimental and Quasi-Experimental Designs for Generalized Causal Inference* (Boston: Houghton Mifflin, 2002).
11. Dean L. Fixsen et al., *Implementation Research: A Synthesis of the Literature* (Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, National Implementation Research Network, 2005).
12. Bianca D. M. Wilson et al., *Sexual and Gender Minority Youth in Los Angeles Foster Care: Assessing Disproportionality and Disparities in Los Angeles* (Los Angeles: UCLA School of Law, Williams Institute, 2014), https://williamsinstitute.law.ucla.edu/wp-content/uploads/LAFYS_report_final-aug-2014.pdf.
13. Ibid.
14. Ibid.
15. Barbara Thomlison, "Characteristics of Evidence-Based Child Maltreatment Interventions," *Child Welfare* 82 (2003): 541–69.
16. Becci A. Akin et al., "Randomized Trial of PMTO in Foster Care: Six-Month Child Well-Being Outcomes," *Research on Social Work Practice* (2016), https://dx.doi.org/10.1177/1049731516669822.

17. Julian D. Ford and Eileen Russo, "Trauma-Focused, Present-Centered, Emotional Self-Regulation Approach to Integrated Treatment for Posttraumatic Stress and Addiction: Trauma Adaptive Recovery Group Education and Therapy (TARGET)," *American Journal of Psychotherapy* 60 (2006): 335–55.
18. Darla L. Henry, "The 3-5-7 Model: Preparing Children for Permanency," *Children and Youth Services Review* 27 (2005): 197–212, https://dx.doi.org/10.1016/j.childyouth.2004.09.002.
19. Shadish et al., *Experimental and Quasi-Experimental Designs*.
20. Epstein and Klerman, "When Is a Program Ready?"
21. Michael Scriven, *Evaluation Thesaurus*, 4th ed. (Newbury Park, CA: Sage, 1991), 169.
22. Shadish et al., *Experimental and Quasi-Experimental Designs*.
23. Ford and Russo, "Emotional Self-Regulation Approach."
24. Figure is based on findings presented at PII Grantee Meeting, August 29–30, 2016, Washington, DC, and is subject to change based on follow-up data.
25. Michiel R. de Boer et al., "Testing for Baseline Differences in Randomized Controlled Trials: An Unhealthy Research Behavior that is Hard to Eradicate," *International Journal of Behavioral Nutrition and Physical Activity* 12, no. 4 (2015), https://dx.doi.org/10.1186/s12966-015-0162-z.
26. Abigail H. Gewirtz et al., "Effects of Mother's Parenting Practices on Child Internalizing Trajectories Following Partner Violence," *Journal of Family Psychology* 25 (2011): 29–38, https://dx.doi.org/10.1037/a0022195; Gerald R. Patterson et al., "Cascading Effects Following Intervention," *Development and Psychopathology* 22 (2010): 949–70, https://dx.doi.org/10.1017/S0954579410000568.
27. Mark Chaffin et al., "A Combined Motivation and Parent-Child Interaction Therapy Package Reduces Child Welfare Recidivism in a Randomized Dismantling Field Trial," *Journal of Consulting and Clinical Psychology* 79 (2011): 84–95, https://dx.doi.org/10.1037/a0021227.
28. Shadish et al., *Experimental and Quasi-Experimental Designs*.
29. Rossi, "Issues."
30. Campbell, "Reforms as Experiments."