Mathematica
Progress Together

Impact Evaluation of the Nurturing Parenting Program Nurturing Skills for Families

October 11, 2022

Max Gross, Aden Bhagwat, Russell Cole

Impact Evaluation of the Nurturing Parenting Program Nurturing Skills for Families

October 11, 2022

Max Gross, Aden Bhagwat, Russell Cole

Submitted to:

Arizona Department of Child Safety
3003 N. Central Avenue
Phoenix, AZ 85012
Attention: Katherine Guffey, MSW
Executive Consultant to the Director

Submitted by:

Mathematica
P.O. Box 2393
Princeton, NJ 08543-2393
Phone: (609) 799-3535
Fax: (609) 799-0005

Acknowledgements

We thank the many people who contributed to and made this impact report possible.

We are especially grateful to the staff at the Arizona Department of Child Safety. We particularly thank our project officer, Katherine Guffey, for her guidance and support. We also thank Layli Milden and Victoria Mills, who provided valuable context about Arizona's array of prevention services, and the many staff who prepared data for the study, including Andrew Egge, Andrew Weller, and Rick Moody.

We also thank the service providers who provided data and program information for the study. The evaluation would not have been possible without their vital support. In particular, we thank Jane Ramirez and Joanne Karolzak of Casa de los Niños; Necole Bell-Harper, Harriet Earley, and Jessica Varney of Arizona's Children Association; and Linda Scott, Tim Allen, Megan Lipman, Melina Harro, Jacob Kies, and Steven Wong of Jewish Family & Children's Service.

We acknowledge Peter Pecora of Casey Family Programs and the University of Washington, who provided insightful feedback about evaluation design and the Nurturing Parenting Programs.

At Mathematica, Joanne Lee and Allon Kalisher provided careful review of this report. Charles Tilley and Jeani Choe provided excellent programming support. John Kennedy and Bridget Gutierrez provided editorial assistance; Colleen Fitts provided production support; and Ana Realpe, Gabe Lee Villanueva, Paula Conner, and LaTia Downing provided contracts and budget support.

Max Gross
Aden Bhagwat
Russell Cole

This page has been left blank for double-sided copying.

Contents

TECHNICAL APPENDIX

Abstract

Introduction

Parent education programs, which are common in the United States, can improve parent, child, and family outcomes. The Nurturing Parenting Program Nurturing Skills for Families (NPP) aims to strengthen parenting skills and reduce child maltreatment through a flexible sequence of lessons that are tailored to each family's needs. A quasi-experimental design was used to examine the effectiveness of NPP in improving children's outcomes in Arizona.

Methods

The study compared the safety and permanency outcomes of children whose families were referred to NPP to those of children who were referred to a broad range of other family preservation services (N = 7,947). Outcomes included investigations, substantiated investigations, and removals, which were measured in administrative data at three follow-up time points: (1) after the end of the program, (2) 6 months after, and (3) 12 months after. The study examined both the effects of being referred to NPP, regardless of a family's level of program participation, and the effects of completing NPP. Both of these full-sample analyses meet review criteria for the Title IV-E Prevention Services Clearinghouse (PSC), and baseline equivalence was established for each analysis. Impacts were calculated by comparing regression-adjusted differences between the study groups.

Results

The study found no evidence of impacts of being referred to NPP (regardless of a family's level of program participation). However, the study found several favorable and statistically significant impacts of completing NPP on children's safety and permanency. Children whose families completed NPP were less likely than the comparison group to experience an investigation or substantiated investigation immediately after the program ended ($p < 0.05$). These children were also less likely to experience a removal up to 12 months after the end of the program ($p < 0.01$).

Conclusions

When NPP is experienced as it is intended—that is, families receive a full dose of the program and complete it—the program has favorable effects on child welfare outcomes among children and families in Arizona. Under PSC review criteria, the analysis and findings for program completion are consistent with a rating of "supported" for NPP. Additional research is necessary to understand the supports that enable families to complete the program and the specific components of NPP that are particularly effective.

This page has been left blank for double-sided copying.

I. Introduction

Parent education programs to enhance parenting practices and promote child well-being are common in the United States. About 800,000 families receive parent education and training each year through the child welfare system alone (Barth et al. 2005). These programs offer parents information, resources, and support to improve their skills, yet they vary in the specific services they provide. Common activities include lessons on positive parenting behaviors, modeling of healthy interactions with children, role-playing of specific parenting skills, and parent support groups (Centers for Disease Control and Prevention 2009; Child Welfare Information Gateway [CWIG] 2019). Some programs also provide separate training for children or referrals to supplemental services, such as mental health and substance abuse services.

Research has shown that parent education programs can improve parent, child, and family outcomes. Studies have found that these programs can improve parents' mental health and parenting style as well as increase children's social and emotional competence and reduce their problem behaviors (Barth and Creel 2014; National Center on Parent, Family, and Community Engagement 2015; CWIG 2019). Consequently, some programs also have shown promise for reducing rates of child maltreatment and out-of-home placements (Chaiyachati et al. 2018; Quick-Beachy et al. 2018; Burnson et al. 2021).

Recognizing the potential of these programs to improve child welfare, the federal government funds certain evidence-based parent education programs through the Family First Prevention Services Act (FFPSA). FFPSA provides up to 12 months of funding for three types of child welfare prevention services: (1) in-home, skills-based programs for parents, such as parent education programs; (2) mental health services; and (3) substance abuse treatment. Programs must show evidence of effectiveness to be eligible for funding. The Title IV-E Prevention Services Clearinghouse (PSC) rates evidence for prevention programs, based on their study design and findings, by using four categories: (1) does not currently meet criteria, (2) promising, (3) supported, or (4) well-supported (Wilson et al. 2019). States can be reimbursed for programs that receive at least a rating of promising. However, beginning in 2026, at least half of the federal reimbursements must be used for well-supported programs.

To build evidence on its prevention services, the Arizona Department of Child Safety (DCS) partnered with Mathematica to design and conduct an independent impact evaluation of the Nurturing Parenting Program Nurturing Skills for Families (henceforth referred to as NPP). In 2021, DCS redesigned its array of prevention services to be trauma-informed, consistent across providers, and aligned with FFPSA (DCS 2021a, 2021b). As part of the redesign, DCS expanded NPP, which had been used by certain service providers, to be implemented NPP statewide. The program seeks to strengthen parenting skills primarily through individualized lessons offered in the home (Family Development Resources Inc. 2007). NPP is one of many parent education programs developed by Nurturing Parenting® that are collectively called the Nurturing Parenting Programs; other programs include those tailored for specific groups, such as by children's age, for children with special needs, and for military families.

This report presents findings from the impact evaluation of NPP. The report first details the program and prior research on NPP. It then describes the study design, sample, NPP dosage and outcome trends, and program impacts. The report concludes with a discussion of lessons learned for policy and future research. The appendices to this report provide a technical overview of the analytic approach and supplemental information on baseline equivalence and program impacts.

A. Description of NPP

NPP is a parent education program designed for families in the child welfare system. The program aims to strengthen parenting skills and reduce child maltreatment. It includes lessons guided by five parenting beliefs: (1) have appropriate expectations of children, (2) develop empathy, (3) discipline with dignity, (4) establish appropriate family roles, and (5) empower children.

The program is designed to be flexible based on each family's needs. In practice, families in Arizona typically spend about three to four months in the program, depending upon their level of risk and need. Facilitators meet with families in the home once or twice per week over this period. During the first four sessions, facilitators introduce families to the program and assess their parenting strengths using the Adolescent Parenting Inventory (AAPI-2), a tool to assess the risk of child maltreatment. Facilitators also work with families to create an individualized Family Nurturing Plan, based on the assessment results, which outlines the content that facilitators will cover with the family.

NPP facilitators choose from more than 90 lessons to develop a sequence uniquely tailored to each family's needs. Lessons are organized around 19 parenting competency areas—for example, making good choices and alternatives to spanking. Each competency area contains about three to seven lessons. A typical sequence will focus on several competency areas. There are also additional, tailored lessons for certain populations, such as Native American families. This customization means that program content intentionally varies across families to meet their specific needs, yet the overarching goal to reduce maltreatment through improved parenting is the same.

Throughout the program, NPP facilitators meet with families in the home for one-hour sessions to deliver the parenting lessons determined in the Family Nurturing Plan. Sessions also involve parent-child activities for parents to apply the lesson. There are several handbooks, lesson guides, and resources available to deliver the program, such as a Facilitator Lesson Guide for Parents and a Parent Handbook. Simpler resources are also available, such as the Easy Reader Parent Handbook, which includes plain language and illustrations for each lesson, and a lesson guide developed for families with younger children.[1]

NPP is similar to other Nurturing Parenting Programs. All programs are based on the same theoretical foundations, use the same validated assessment instrument, and focus on the same five parenting constructs. Each program has its own manual containing only lessons relevant to the intended population. In particular, NPP contains similar content to two age-specific programs: Parents & Their Infants, Toddlers, and Preschoolers® and Parents & Their School-Age Children 5–11 Years®. NPP provides facilitators with flexibility to select and sequence lessons according to each family's needs, while the age-specific programs are tailored to the child's age and less flexible. The infants, toddlers, and preschoolers model can be delivered as weekly 2.5 hour group-based sessions for 16 weeks, weekly 60- to 90-minute individual sessions in the home for seven weeks, or a combination of group- and home-based sessions. The school-age children model requires a weekly 2.5 hour group-based session for 15 weeks.

[1] Information on the program handbooks, lesson guides, and resources are available at https://www.nurturingparenting.com/ecommerce/category/1:2:1/.

B. Prior research on Nurturing Parenting Programs

Relatively few studies have been conducted on NPP that have used a comparison group design. But some studies of Nurturing Parenting Programs have shown promise. This section describes findings from studies of the programs for infants, toddlers, and preschoolers and for school-age children.

Several pre-post studies have found a positive association between participation in the age-specific Nurturing Parenting Programs and family outcomes. Brock and colleagues (2013) studied the school-age children program and Greeno and colleagues (2021) examined both age-specific versions. Each study found that caregivers exhibited better parenting attitudes and knowledge after the program than before it, including greater empathy and increased knowledge of positive discipline techniques (Brock et al. 2013; Greeno et al. 2021). Some evidence also suggests that program participation is associated with reduced child maltreatment. Greeno and colleagues (2021) also found that among 34 caregivers in a mid-Atlantic state, 29 percent were subject to a child maltreatment investigation and 21 percent were subject to a substantiated investigation in the year before the program; these numbers dropped to just 15 percent and 3 percent in the year after, respectively. Using a larger sample of more than 500 caregivers in Louisiana, Maher and colleagues (2011) found that higher program attendance in the infants, toddlers, and preschoolers program was associated with a reduced likelihood of being reported for child maltreatment. However, because these studies lacked a comparison group, it is possible that the changes in outcomes were not attributable to the program.

Two studies have used experimental or quasi-experimental designs (QEDs) to evaluate Nurturing Parenting Programs, and both found favorable impacts. A randomized controlled trial in Cook County, Illinois, found that families who were offered the infants, toddlers, and preschoolers program after a child was removed from their home spent less time in foster care and had higher rates of family reunification and kinship guardianship (Illinois Department of Children and Family Services 2018). Likewise, an earlier QED study compared families in Florida that participated in (1) the infants, toddlers, and preschoolers program, (2) the school-age children program, and (3) other parent education programs that were not Nurturing Parenting Programs (Weikert et al. 2007). The study found that parents who completed either of the Nurturing Parenting Programs had statistically significantly higher scores on parenting attitudes and practices than the comparison group.

These findings highlight the potential benefits of Nurturing Parenting Programs. This study adds to the prior research by producing the first evidence on NPP that employed a comparison group in Arizona.

This page has been left blank for double-sided copying.

II. Study Design

This study tested the effectiveness of NPP using a QED. This chapter outlines the research questions, details the study's data sources, describes the comparison condition, and discusses how we defined each study condition in the data. It then details the outcomes, timeframe of NPP, and the two types of policy-relevant effects that we estimated in analyses for review by the PSC.

A. Research questions

This study answered three research questions. The first analysis provided information about the effects of being referred to NPP on children's outcomes at three points in time after the program. The second analysis explored the effects of actually completing NPP. Both of these analyses use QEDs that are eligible for review by the PSC and satisfy criteria in Wilson et al. (2019) for study design ratings of "moderate support of causal evidence." The third analysis investigated how these impacts varied. [Note that the PSC does not currently review subgroup analyses (Wilson et al. 2019).]

The research questions were as follows:

1. What are the impacts of referrals to NPP on child safety and child permanency at the following time points?
    a. Immediately after the end of the program
    b. Six months after the end of the program
    c. Twelve months after the end of the program
2. What are the impacts of completing NPP on child safety and child permanency at each of the three time points?
3. How do the impacts of referrals to NPP and completing NPP vary for key subgroups of the program's intended population, by age, gender, race and ethnicity, program provider, and level of services?

B. Data sources

This study relied on data from two sources:

- **Arizona Department of Child Safety (DCS).** We obtained comprehensive administrative data from DCS covering all child welfare investigations from January 1, 2015, to December 31, 2020, spanning more than 280,000 investigations. The data listed all children and caregivers involved in each investigation and included detailed demographic characteristics, such as gender, age, and race and ethnicity. For each investigation, the data also included the type of maltreatment allegation that DCS coded as most severe (for example, neglect or physical abuse) and whether any allegation was substantiated. For children placed in out-of-home care, the data included when the removal occurred and the dates associated with placement changes. For caregivers referred to services, the data included the referral date; a broad category for the type of services, such as in-home service; and the name of the service provider. We used these data to form study groups, construct outcome measures and background characteristics, and identify and adjust for possible differences in background characteristics between groups.
- **NPP service providers.** We obtained rosters of all cases referred to NPP from two providers of in-home family preservation services in Arizona. Arizona's Children Association is a child welfare and behavioral health agency that serves families across the state. Casa de los Niños is a community

organization that offers parenting classes and mental health and family support services in Pima County. For both providers, we obtained detailed information about all cases referred to NPP from July 1, 2018, to December 31, 2020, including the dates of service referral and discharge, and the reason for discharge (for example, because NPP was completed). We linked the NPP rosters to the DCS data using a common case identifier.

C. Comparison condition

The comparison condition consisted of all in-home family preservation services other than NPP that DCS refers families to following an investigation of child maltreatment. The PSC refers to this type of comparison condition as treatment as usual (Wilson et al. 2019). NPP is one of many in-home family preservation services that DCS may refer families to after an investigation. Other services include crisis intervention, counseling, domestic violence education, linkages to community resources, and more (DCS 2021a). For this study, these treatment-as-usual services were offered by 21 different providers. According to email correspondence with DCS and providers, none of these service providers offered NPP during the study sample period, thus mitigating concerns of potential contamination across study conditions.

The comparison condition differed from the NPP condition in several ways. Only staff in the NPP condition were formally trained to deliver Nurturing Parenting Programs. As a result, staff in the comparison condition did not use assessment tools from the Nurturing Parenting Programs or follow other program protocols regarding the use of lessons and activities, although it is possible that they may have used one-off lessons on rare occasions.

The comparison condition was similar to the NPP condition in terms of the intensity of services and the professional backgrounds of service staff. For all in-home family preservation services, including NPP, DCS offers two service levels. Intensive services last up to four months and are offered when children are at significant risk of removal from the home. Moderate services last up to three months and are offered when children face some level of risk of subsequent abuse or neglect. Likewise, services for both conditions are provided by a team consisting of a team lead or therapist and a family support worker.

D. Assignment to NPP and the comparison condition

Both the NPP and comparison conditions included children whose families were referred to in-home family preservation services. Following an investigation of alleged child abuse or neglect, child welfare staff determine whether children are safe to remain in their homes (that is, whether to remove children) and whether families demonstrate a need for services. Staff may refer families to in-home services who show a need and whose children are safe to continue living in their homes, regardless of whether the investigation is substantiated.

We defined the NPP and comparison groups based on whether children were involved in cases that were referred by DCS to providers who delivered NPP or other treatment-as-usual services. If any caregiver listed on a child's case was referred by DCS to a provider who delivered NPP within the DCS in-home services contract, we defined the child as being in the NPP group. For simplicity, we reference this group as being "referred to NPP" throughout the report. We defined the comparison group as children in which at least one caregiver was referred to other in-home services and none were referred to NPP.

Local availability plays a key role in whether families are referred to NPP. To make a referral for in-home services, child welfare staff submit a request to a centralized referral unit. The unit examines availability

in the family's area and refers the family to a service provider that has an opening. Although NPP is offered in every Arizona county, only certain providers offer it. While the referral unit considers the family's needs in identifying a provider, provider availability plays a primary role.

Because of constraints due to local availability, there is an element of chance in whether DCS refers a case to a provider who used NPP, which supports the study design. Differences in outcomes between the two study groups can be causally attributed to NPP as long as the groups were similar before the program. The supply of and demand for services in a family's local area introduces an element of randomness into whether the family's case was referred to NPP or comparison services. For example, two families from the same area with similar needs could have been referred to different providers—one that offers NPP and one that does not—solely because the NPP provider had only one opening. Therefore, whether a family was referred to NPP was in part unrelated to the family's own circumstances, which supported the similarity of the two study groups. We present a formal assessment of the similarity of study groups in Chapter III.

E. Outcomes

This study focused on two policy-relevant outcome domains that were eligible for PSC review and available in administrative data: (1) child safety and (2) child permanency. The PSC defines child safety outcomes as those that examine whether there is a threat of danger to the child and defines child permanency outcomes as those measuring the stability of a child's living situation, including whether the child was removed (Wilson et al. 2019).

We examined the impacts of NPP on main and secondary outcomes. Program impacts on the main outcomes served as the basis for drawing substantive conclusions about the program's effectiveness. Impacts on secondary outcomes added context to the main findings or addressed questions of substantive, practical, or policy significance that extended beyond the main test of program effectiveness. We selected the main outcomes before beginning the analysis to prevent focusing the assessment of program effectiveness on outcomes where impacts emerged as statistically significant.

The primary goal of NPP is to strengthen parenting practices and promote child safety. Therefore, we examined two main outcomes related to child safety: (1) child welfare investigations and (2) substantiated child welfare investigations (Table II.1). These outcomes are common indicators of child safety available in administrative data.

A key objective of FFPSA is to prevent out-of-home placement among children who are candidates for foster care. Thus, as the main measure of child permanency, we examined the extent to which NPP influenced whether children were removed.

Table II.1. Main outcomes

Outcome	Measure
Child safety	
Investigation	Whether the child was subject to any maltreatment investigation
Substantiated investigation	Whether the child was subject to any substantiated maltreatment investigation
Child permanency	
Removal	Whether the child was removed from the home

As secondary outcomes, we separately studied investigations for neglect and physical abuse. These outcomes shed light on the channels through which NPP influences parenting practices.

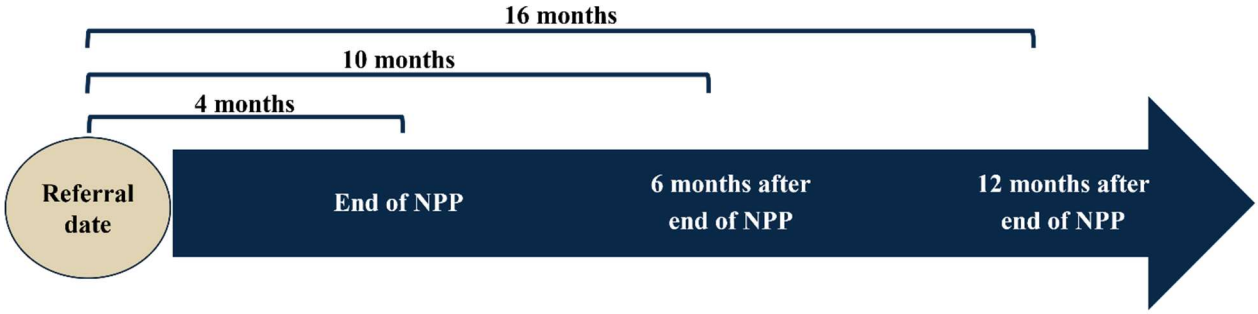
F. Time frame of NPP and outcomes

We estimated program impacts at three follow-up time points: (1) immediately after the end of NPP, (2) 6 months after, and (3) 12 months after. For example, we examined whether children in the NPP group were less likely than those in the comparison group to be subject to an investigation (1) between the start and end of the program, (2) between the start and 6 months after the end of the program, and (3) between the start and 12 months after the end of the program (Figure II.1).

The start of services was well-defined for both study groups by the service referral date. However, the definition of the end of NPP was less clear because the length of participation was designed to be responsive to a family's needs, and thus varied across families. Consistent with the PSC's recommendations, we defined the end of NPP as a time point in which the majority of services were stated to have been delivered (Wilson et al. 2019). In-home services last for 3 to 4 months, depending upon the service level (DCS 2021a). Accordingly, we observed that the median amount of time between referral to NPP and discharge was 84 days and that close to 95 percent of NPP cases lasted for 4 months or less. Therefore, we used 4 months as a standardized length of treatment for both study groups. This meant, for example, that we defined the investigations outcome at the 6-month follow-up time point as whether an investigation occurred between the referral date and 10 months after the referral date (4 months of services + 6 months of follow-up = 10 months).

Children whose families were referred to in-home services in late 2019 or 2020 may be missing 6- and 12-month follow-up data because outcomes were only available through December 31, 2020. This means that the samples used to estimate impacts were different at each follow-up time point. Appendix A provides further details on missing data and Appendix B shows that the impact results were similar when using a consistent sample at each time point.

Figure II.1. Timing of NPP and follow-up time points



G. Two types of policy-relevant effects

To provide a comprehensive overview of the impacts of NPP, we examined two types of policy-relevant effects: the impacts of referrals to NPP (Research Question 1) and the impacts of completing NPP (Research Question 2). The first represented the average impact for all children whose families were referred to NPP, which incorporated the effects for those who received a high level of program dosage as well as those with lower levels of program participation. We also examined the impacts of NPP for children whose families completed the program. As described in greater detail in Chapter IV, about two-

thirds of children whose families were referred to NPP completed it. Additional information on the analytic approach used to estimate these two types of effects is provided in Appendix A.

III. Study Sample

This chapter describes how we constructed the study sample, discusses the characteristics of children in the sample, and details our approach for establishing baseline equivalence (that is, assessing the similarity of the NPP and comparison groups before the program).

A. Sample construction

The study sample consisted of children from birth to 17 years old whose families were referred either to NPP or to other in-home family preservation services between July 2018 and December 2020.[2] The NPP group was composed of children who were involved in cases referred to either Arizona's Children Association or Casa de los Niños for NPP. The comparison group was composed of children who were involved in cases referred to in-home services with providers that did not offer NPP. There were 2 service providers in the NPP group and 21 service providers in the comparison group. Therefore, the study did not suffer from the "n = 1 administrative unit confound" because both study groups consisted of multiple providers (Wilson et al. 2019). Because cases can be referred for several different services over time, for both groups, we retained the earliest referral to either NPP or to other in-home services during the sample period.

We further refined the sample in two ways. First, we focused on children who were living in the home at the time of the service referral (that is, those who were not in out-of-home care), because a key policy-relevant question was whether NPP influenced removals. Second, we focused on referrals in which local availability might have factored into whether the child's family was referred to NPP, by restricting the sample to counties with at least one child in each of the NPP and comparison groups. Overall, the sample consisted of 7,947 children from 4,335 cases: 1,102 children (615 cases) in the NPP group and 6,845 children (3,720 cases) in the comparison group.

B. Characteristics of children in the study sample

Children in the study sample tended to be young and Hispanic or White, with a history of involvement in the child welfare system. The average age of children in the sample was 7 years old, and 49 percent of the children were female (Table III.1). Three in four children in the study sample were either Hispanic (38 percent) or White (37 percent). This tracked closely with the overall makeup of children in Arizona (Kids Count Data Center 2021). Seventeen percent of children in the sample were African American, more than twice the proportion of African American children statewide, which reflected the overrepresentation of these children in child welfare systems (CWIG 2021). Smaller percentages of children in the study sample identified as American Indian (7 percent) and Asian or Pacific Islander (1 percent). In terms of

[2] The out-of-home care and in-home services populations in Arizona were decreasing before, during, and after the sample period. The number of children in out-of-home care steadily declined in Arizona from 11 per 1,000 children in 2015 to 8 per 1,000 children in 2021 (Kids Count Data Center 2022). Preliminary estimates suggest a continued decrease in 2022 (DCS 2022). Based on the study data, the number of families referred to in-home family preservation services also decreased during this period, but not by as much as the out-of-home care population. These trends suggest that this study evaluated NPP during a time when there was a growing share of families who were referred to in-home services as an alternative to child removal.

characteristics of children's families, their cases with DCS included about three total children (a proxy for family size) and caregivers were 35 years old on average.

The data included records of any Arizona child welfare involvement in the three years before the service referral for all children in the sample, which we used to examine children's histories of investigations and removals. By design, all children (100 percent) were subject to at least one child welfare investigation in Arizona before the service referral. Many experienced more than one; on average, children were subject to about 1.7 investigations in the three years before the referral. Neglect was the most common reason for investigations (79 percent). About one-third of children were subject to an investigation for physical abuse (35 percent). About 38 percent of children experienced a substantiated investigation in the three years before the service referral and 21 percent experienced a removal. Among those 21 percent who experienced a removal, the average removal episode lasted about eight months, nearly all children reunified with their families (93 percent), and in-home services were referred about four months after children exited out-of-home care, on average.

Table III.1. Background characteristics of children in the study sample

Characteristic	Percentage or mean
Child demographics	
Age (years)	7.2
Female	49
Race and ethnicity	
Hispanic	38
White	37
African American	17
American Indian	7
Asian/Pacific Islander	1
Family characteristics	
Family size[a]	3.2
Caregiver age (years)[b]	34.9
Child welfare history[c]	
Investigation	100
Average # of investigations	1.7
Investigation for	
Neglect	79
Physical abuse	35
Substantiated investigation	38
Removal	21
Number of children	7,947

Source: Arizona Department of Child Safety (DCS) administrative data.

Note: Race and ethnicity categories are mutually exclusive. Children of more than one race and ethnicity are categorized into one category using the following hierarchy: American Indian, African American, Asian or Pacific Islander, Hispanic, White, Other (DCS 2022). Investigations for neglect and physical abuse are not mutually exclusive.

[a] Number of children on the child's case.

[b] Average age of all perpetrators on the child's case.

[c] DCS involvement in the three years before the family's service referral.

C. Baseline equivalence

The NPP and comparison groups were generally similar before the program, which makes sense given how families were referred to NPP or other in-home services. We assessed baseline equivalence on the characteristics related to child demographics, family characteristics, and child welfare histories in Table III.1. We consolidated the five race and ethnicity categories into three mutually exclusive groups to focus the equivalence assessment on those with larger sample sizes: (1) Hispanic, (2) White, and (3) Other (which includes African American, American Indian, and Asian or Pacific Islander). We did not assess equivalence on whether children had an investigation before the referral, because all children in both groups had an investigation, by design.

Across 10 characteristics, race and ethnicity was the only one that showed statistically significant differences (Table III.2). There were a larger percentage of White children in the NPP group, and a larger percentage of children in the Other category in the comparison group. The Other category also had an absolute effect size difference of 0.25, which was at the threshold considered acceptable by the PSC.

Because of the differences in children's race and ethnicity between the two study groups, we used entropy balancing to form more similar groups. This method produces weights so that the comparison group has similar background characteristics to the NPP group. Entropy balancing has several advantages relative to other matching or weighting procedures used to establish baseline equivalence, such as propensity score methods. Namely, unlike other methods, entropy balancing constructs weights by imposing that specified background characteristics be similar across research groups, thus ensuring equivalence along those characteristics (Hainmueller 2012). In this sense, it is a more direct approach to achieve baseline equivalence than propensity score methods, which often require iterating on the propensity score model to achieve equivalence. Importantly, the weights were constructed solely using the child demographics, family characteristics, and child welfare history variables observed before the referral. We applied the weights produced from entropy balancing to estimate program impacts.

After reweighting the comparison group, the two study groups had nearly identical background characteristics (Table III.2). None of the characteristics showed statistically significant differences, and all effect size differences were smaller than 0.25 standard deviations. For example, without weights, there was a 5 percentage point difference in the share of White children between the two study groups (42 percent in the NPP group and 37 percent in the comparison group). With weights, the percentage of White children in both groups was 42 percent.

We followed the same approach for assessing baseline equivalence and reweighting the comparison groups for the samples with outcome data 6 and 12 months after the program, and found similar equivalence results for these samples. As discussed in greater detail in Appendix A, we accounted for background characteristics that showed effect size differences in the PSC's adjustment range of 0.05 to 0.25 standard deviations by controlling for them in the models used to estimate program impacts. Appendix A also provides technical details on the weighting approach and presents supplemental baseline equivalence results.

Table III.2. Background characteristics of the full sample with and without weights, by study group

Characteristic	Without weights				With weights			
			Difference				Difference	
	NPP group	Comparison group	Mean	Effect size	NPP group	Comparison group	Mean	Effect size
Child demographics								
Age (years)	7.4	7.1	0.3	0.05	7.4	7.4	0.0	0.00
Female	51	49	1	0.04	51	49	1	0.03
Race and ethnicity								
Hispanic	40	38	2	0.05	40	40	0	0.00
White	42	37	5*	0.13	42	42	0	0.00
Other	19	26	-7**	-0.25	19	19	0	0.00
Family characteristics								
Family size[a]	3.0	3.2	-0.2	-0.10	3.0	3.0	0.0	0.00
Caregiver age (years)[b]	35.2	34.9	0.4	0.04	35.2	35.1	0.2	0.02
Child welfare history[c]								
Average # of investigations	1.8	1.7	0.0	0.03	1.8	1.8	0.0	0.00
Investigation for								
Neglect	79	79	0	0.00	79	79	0	0.00
Physical abuse	36	35	1	0.02	36	36	1	0.01
Substantiated investigation	39	38	1	0.02	39	39	0	0.00
Removal	19	21	-1	-0.05	19	19	0	0.00
Number of children	1,102	6,845			1,102	6,845		

Source: Arizona Department of Child Safety (DCS) administrative data.

Note: Due to rounding, percentages may not sum to 100 and the reported mean difference may not equal the difference in means between the NPP and comparison groups. Other race and ethnicity includes American Indian, African American, and Asian or Pacific Islander. Investigations for neglect and physical abuse are not mutually exclusive. Standard errors are clustered at the case level.

[a] Number of children included on the child's case.

[b] Average age of all perpetrators included on the child's case.

[c] DCS involvement in the three years before the family's service referral.

**/* Differences are statistically significant at the .01/.05 levels, respectively, two-tailed test.

IV. Patterns of NPP Dosage and Outcomes

Before estimating the impacts of NPP, we conducted two descriptive analyses to provide context for the findings. This chapter presents patterns of program dosage and changes in outcomes for children in the NPP group.

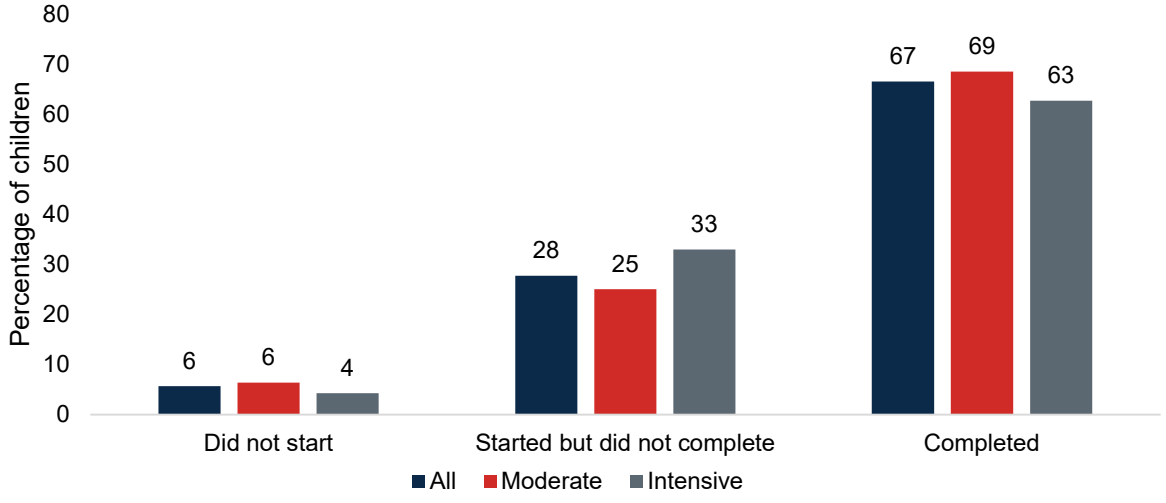
A. NPP dosage

We examined patterns in program participation to understand the extent to which children and families received the intended dose of NPP. We grouped dosage into three categories based on the reason for service discharge in the NPP provider data: (1) did not start, (2) started but did not complete, and (3) completed. Children did not start NPP if the service provider was unable to make contact with the family or the family refused services. Started but did not complete meant the family participated in NPP to some extent and had a discharge reason other than program completion. Completed meant that providers marked the child's family as having completed NPP. Participation data were only available for the NPP group and not for the comparison group.

The largest category included children whose families completed NPP (67 percent) (Figure IV.1). Another 6 percent of children belonged to families who were referred to NPP but did not start it. The middle group consisted of families who started but did not complete NPP (28 percent). Within this 28 percent, most had discharge reasons that did not fully explain the reasons why services ended, such as partially completed, not completed, or withdrew (75 percent). A smaller percentage of children's families ended NPP because of child removal (13 percent). Rates of program completion were slightly higher for moderate services (one session per week for three months) than for intensive services (two sessions per week for four months). Completion rates were similar for the two NPP providers included in the study (69 percent for Arizona's Children Association and 64 percent for Casa de los Niños).

Because of the varying levels of program dosage, the next chapter presents both the impacts of referrals to NPP and of actually completing NPP.

Figure IV.1. NPP dosage, overall and by level of services

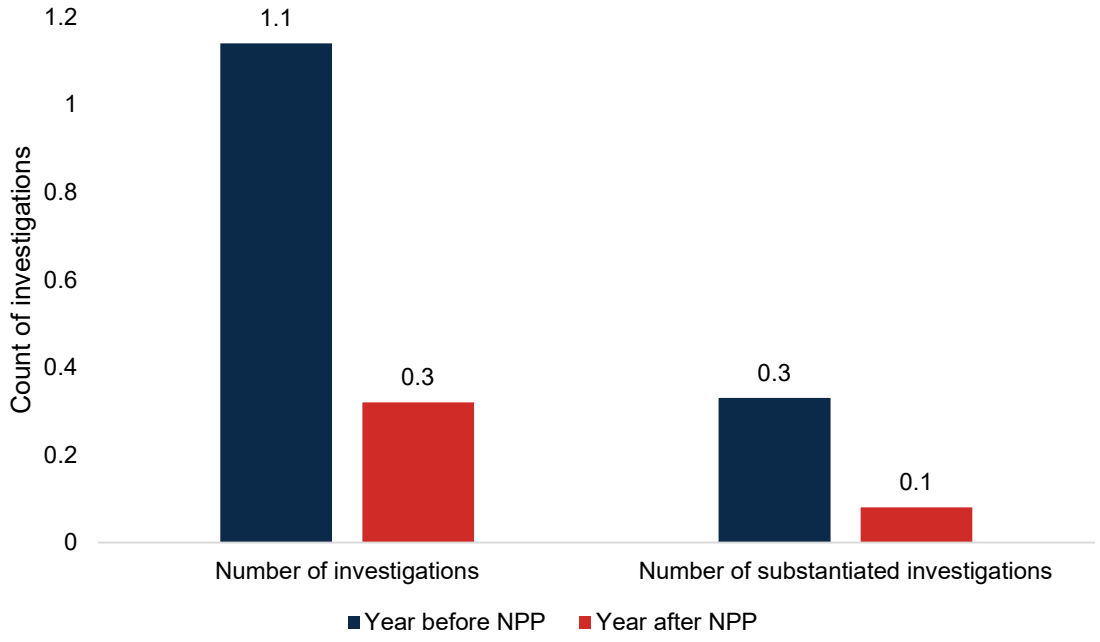


Source: Data from NPP service providers.

B. Outcome trends

To understand the effects of NPP, we conducted descriptive analyses to assess the changes in outcomes over time for all children in the NPP group. We found that children in the NPP group had more favorable outcomes in the year after NPP than in the year before it, including all children regardless of their level of NPP dosage. We compared the number of investigations and substantiated investigations in the year before and after the child's family was referred to NPP. We used counts of investigations because, by design, every child was the subject of at least one investigation before NPP. Children experienced about 1.1 investigations in the year before NPP, compared to just 0.3 investigations in the year after the program (Figure IV.2). Likewise, children in the NPP group experienced fewer substantiated investigations after NPP than before it (0.3 compared to 0.1). Knowing that children's outcomes were more favorable after NPP, we examined whether the NPP group improved more than the comparison group (see Chapter V).

Figure IV.2. Trends in outcomes for the NPP group



Source: Arizona Department of Child Safety administrative data.

Note: This figure shows the average number of investigations and substantiated investigations in the year before and after the referral to NPP for all children in the NPP group, regardless of the level of NPP dosage. The number of investigations in the year before NPP is different from the count of investigations presented in Table III.1, which used a three-year window before the referral.

V. Program Impacts

This chapter reports findings on NPP's success in achieving its goals of promoting child safety and permanency. We first present the impacts of referrals to NPP (Research Question 1), which includes the full sample of children, regardless of the level of program dosage. We then present the impacts of completing NPP (Research Question 2), which includes only the sample of children from the NPP group whose families completed NPP. Both types of impacts are policy-relevant, and use rigorous QEDs eligible for PSC review, including demonstrations of baseline equivalence for each sample. We also summarize findings from the secondary analyses, sensitivity tests, and subgroup analyses (Research Question 3). Supplemental information for each analysis is provided in Appendix B.

Overall, we found no evidence of impacts for children whose families were referred to NPP (regardless of the level of program participation); and large, favorable, and statistically significant impacts of completing NPP on outcomes in both the safety and permanency domains. Notably, there were impacts of completing NPP on removals up to one year after the end of the program. This evidence meets PSC criteria to assign NPP a program rating of "supported."

A. Impacts of referrals to NPP on main outcomes

Compared with other in-home family services, being referred to NPP did not change children's safety or permanency outcomes (Table V.1). Children whose families were referred to NPP had similar safety outcomes at each follow-up time point as those whose families were referred to other in-home family preservation services. For example, 10 percent of children in the NPP group and 11 percent of children in the comparison group were subject to an investigation immediately after the end of the program. Likewise, 3 percent of children in both study groups were subject to a substantiated investigation at the end of the program. The two study groups also had similar rates of investigations and substantiated investigations at the 6- and 12-month follow-up time points. Across all safety outcomes at each time point, the differences between study groups were fairly small, with an average absolute effect size of just 0.08 standard deviations, and none were statistically significant.

Children in both study groups also experienced removals at similar rates. About 7 percent of children in both groups experienced a removal immediately after the program and 15 percent of children in both groups experienced a removal 12 months after the program. The percentage of children who experienced a removal was larger than the percentage who were subject to a substantiated investigation because families in Arizona can appeal a substantiated finding, which takes time. In some instances, children may be removed before the appeals process has concluded. Regardless, the differences in removals between the two study groups at each follow-up time point were near zero and statistically insignificant.

Table V.1. Impacts of referrals to NPP on children's safety and permanency

Outcome	Follow-up time point	NPP group	Comparison group	Impact	Effect size
Child safety					
Investigation	0 mos	10	11	-1	-0.05
	6 mos	23	22	1	0.03
	12 mos	33	29	4	0.12

Outcome	Follow-up time point	NPP group	Comparison group	Impact	Effect size
Substantiated investigation	0 mos	3	3	0	0.00
	6 mos	6	5	1	0.13
	12 mos	9	8	1	0.12
Child permanency					
Removal	0 mos	7	7	0	0.04
	6 mos	14	12	2	0.09
	12 mos	15	15	0	0.00
Number of children	0 mos	1,102	6,845		
	6 mos	858	5,684		
	12 mos	603	4,046		

Source: Arizona Department of Child Safety administrative data.

Note: This table shows the impacts of referrals to NPP at three follow-up time points: immediately after NPP (0 mos), 6 months after (6 mos), and 12 months after (12 mos). The numbers in the NPP Group and Comparison Group columns are weighted, regression-adjusted predicted values of the outcomes. The weights to estimate impacts are the same as those in the baseline equivalence assessment. Standard errors are clustered at the case level.

**/* Impact estimates are statistically significant at the .01/.05 levels, respectively, two-tailed test.

B. Impacts of completing NPP on main outcomes

Children whose families completed NPP had lower rates of investigations and removals. Completing NPP reduced the likelihood of an investigation by 4 percentage points: approximately 7 percent of children whose families completed NPP experienced an investigation, relative to 11 percent of the comparison group (Table V.2). These children also had lower rates of substantiated investigations after the end of the program. Just 1 percent of children whose families completed NPP were the subject of a substantiated investigation after the program relative to 3 percent of the comparison group, a difference of 2 percentage points. The impacts of completing NPP on investigations and substantiated investigations immediately after the program were large, with effect sizes of -0.28 and -0.66 respectively, and statistically significant at the $p < 0.05$ level. We did not find evidence of impacts on these outcomes at 6 or 12 months after NPP.

Children whose families completed NPP had removal rates from the home that were significantly lower than those in the comparison group, at all three follow-up time points. About 8 percent of children in the comparison group experienced a removal after the end of the program, while just 3 percent of children whose families completed NPP were removed from the home, a difference of 5 percentage points. The reduction in removals became more pronounced as time passed. We found a difference between groups in removals of 6 percentage points after 6 months and 9 percentage points after 12 months. The decline in removals was large (average effect size of -0.63 across follow-up time points) and statistically significant at the $p < 0.01$ level at each follow-up time point, indicating a sustained reduction in removals for children whose families completed NPP. Consequently, NPP meets PSC criteria to receive a program rating of supported.

Table V.2. Impacts of completing NPP on children's safety and permanency

Outcome	Follow-up time point	NPP completers	Comparison group	Impact	Effect size
Child safety					
Investigation	0 mos	7	11	-4*	-0.28
	6 mos	20	22	-2	-0.09
	12 mos	31	29	2	0.06
Substantiated investigation	0 mos	1	3	-2*	-0.66
	6 mos	5	5	0	-0.05
	12 mos	6	7	-1	-0.14
Child permanency					
Removal	0 mos	3	8	-5**	-0.78
	6 mos	6	12	-6**	-0.42
	12 mos	6	15	-9**	-0.70
Number of children	0 mos	732	6,845		
	6 mos	582	5,684		
	12 mos	401	4,046		

Source: Arizona Department of Child Safety administrative data.

Note: This table shows the impacts of completing NPP at three follow-up time points: immediately after NPP (0 mos), 6 months after (6 mos), and 12 months after (12 mos). The numbers in the NPP Completers and Comparison Group columns are weighted, regression-adjusted predicted values of the outcomes. The weights to estimate impacts are the same as those in the baseline equivalence assessment. Standard errors are clustered at the case level.

**/* Impact estimates are statistically significant at the .01/.05 levels, respectively, two-tailed test.

C. Impacts of NPP on secondary outcomes

We examined the impacts of referrals to NPP and of completing NPP on the study's secondary outcomes: investigations of physical abuse versus neglect (see detailed results in Appendix B). The study data included the most severe allegation associated with each investigation. Physical abuse allegations were coded by DCS as more severe than neglect. As a result, an investigation for both types appears in the study data as physical abuse. We first explored the impacts of referrals to NPP on each type of allegation. Consistent with the overall findings for NPP referrals, we found no impacts on either physical abuse or neglect.

For children whose families completed NPP, we found that the improvement in children's safety was driven by a reduction in investigations for physical abuse. Children whose families completed NPP were 2 percentage points less likely than those in the comparison group to have been subject to an investigation where the most severe allegation was physical abuse at the completion of the program and 4 percentage points less likely 6 months after the program. Both differences between groups were large (effect sizes of -0.51 and -0.43, respectively) and statistically significant at the $p < 0.05$ level. We found no impacts of completing NPP on neglect.

D. Sensitivity tests

We conducted four sensitivity tests, including using alternative outcome measures and alternative statistical methods such as not applying weights, and found that the impact findings were robust to different analytic decisions. Referrals to NPP had no effect on children's safety or permanency, whereas completing NPP had favorable and statistically significant impacts on both outcome domains. Details of these sensitivity tests are presented in Appendix B.

E. Subgroup findings

We examined whether NPP was particularly effective for certain subgroups of the program's intended population. We studied impacts separately by the child's age, gender, and race and ethnicity, and by the family's NPP provider and level of services.

We found that the impacts of referrals to and completing NPP were generally similar across subgroups and consistent with the full-sample results. We found no evidence that referrals to NPP influenced children's outcomes for any subgroup. The results were a bit more nuanced for completing NPP. For all but one subgroup examined, there were favorable and statistically significant effects of completing NPP on reducing rates of removals. The only exception was for race and ethnicity: the impacts on removals were in the favorable direction for all groups but only statistically significant for Hispanic and White children (the groups with the largest sample sizes). There were favorable and statistically significant impacts of completing NPP on investigations among older, female children, as well as those served by Casa de los Niños and referred for moderate services. In addition, there were favorable and statistically significant impacts of completing NPP on substantiated investigations among older, male children and those served by Arizona's Children Association and referred for intensive services. Completing NPP reduced investigations among Hispanic and White children and reduced substantiated investigations among Hispanic children. In most cases, however, these impact estimates were imprecise, and we found limited evidence of differences in the impacts between subgroups.

VI. Discussion

This study examined the effectiveness of NPP on several policy-relevant outcomes. Using a QED, we compared the child welfare outcomes of children whose families were referred to NPP relative to a well-matched sample of children whose families were referred to other in-home services. We used best evaluation practices in our study design and analytic approaches, with the goal of finding credible evidence of NPP's effectiveness that might enable NPP to receive a "supported" rating by the PSC.

In this evaluation, promising effects of NPP were observed across outcomes and research questions. Although the study found that there no impacts of referrals to NPP (regardless of the level of program participation), there were several favorable and statistically significant impacts of receiving the intended dose of NPP (that is, completing the program). For children whose families completed the program, there were large, sustained impacts on rates of removals up to one year after the program. In addition, for children whose families completed the program, there were favorable short-term impacts observed on investigations and substantiated investigations of maltreatment, particularly for investigations of abuse. NPP could influence abuse more than neglect because the program contains lessons on preventing physical abuse (for example, alternatives to spanking). Although the program also includes lessons to teach parents skills that could influence neglect, it does not provide material resources or behavioral-health related supports that may be required to influence investigations of neglect (Palmer et al. 2022). For this reason, observing impacts on physical abuse conveys credibility to the overall package of findings. The credibility of these findings is also enhanced by virtue of being robust across a variety of sensitivity analyses.

There are several explanations for why NPP had more favorable effects for program completers than for those who were simply referred to NPP. One logical explanation is that NPP's theory of change requires families to get the intended dose of the program for outcomes to be achieved. If families that don't attend any sessions or only attend a small number of sessions are included in the overall test of NPP's effectiveness, it makes sense for the study to find attenuated estimates of program effectiveness. A second explanation is that the reason why some families did not complete the program is because they experienced an adverse child welfare outcome, meaning there is a mechanical relationship between NPP completion and favorable outcomes. However, the data suggest that this second explanation is unlikely to be the main driver of the observed impact. Only 13 percent of children whose families started but did not complete NPP ended programming due to child removal according to data from program providers. This means that close to 90 percent of children's families ended NPP for some other reason that would not create this mechanical relationship.[3] Data on the experiences of the comparison group and the extent to which families in the comparison group completed other in-home services would have been useful to unpack this question. However, these data were not available for the current study.

As with nearly any impact evaluation, there are limitations to this study. First, data on outcomes that are closely aligned with the content of the program (for example, parenting beliefs and behaviors from the AAPI-2) would have allowed for a more comprehensive test of the effect of NPP. Although referrals to NPP did not show impacts on subsequent child welfare outcomes, it is possible that referrals to NPP had effects on more proximal outcomes better aligned with the NPP program. A second limitation of this study is that it uses a non-experimental design. While the study has shown the equivalence of the samples

[3] It is possible that providers do not always know which families did not complete NPP due to child removal. However, the discharge reason from provider data closely tracks with DCS administrative data, which shows that 17 percent of children whose families did not complete NPP experienced a removal in the four months after the service referral.

at baseline (including the analyses that compare NPP completers against the comparison group), all QEDs suffer the same internal validity threat: there is always a potential omitted variable (like parental motivation to change behavior) that may differ across conditions, and this variable (or a collection of omitted variables) may obscure the program's true effect. A third limitation is that this study lacks detailed program data for the comparison group that were referred to providers who did not use NPP. We do not have information on the specific services provided to the comparison group, and thus do not know the extent to which services between the two groups differed.

Despite these limitations, this study provides foundational evidence about the effectiveness of NPP. The study addresses a known gap and continues the evidence-building path that has been paved by previous research on the Nurturing Parenting Programs (Brock et al. 2013; Greeno et al. 2021; Maher et al. 2011; Illinois Department of Children and Family Services 2018; Weikert et al. 2007). It highlights the promise of a program that is widely used in Arizona and was recently expanded as part of the state's redesign of its prevention services array.

Given the limitations above, additional research into the effectiveness of NPP is warranted. A more rigorous impact evaluation could address the limitations of this study and show whether the findings can be replicated in a prospective design. Importantly, given the findings from this evaluation show that NPP is effective when families receive it as intended, it will be critical to create supports to enable families to complete the program. Additional evidence of NPP's effectiveness could also make the program eligible for a "well-supported" PSC rating.

Overall, when NPP is experienced as it is intended (that is, families receive a full dose of the program and complete it), there is evidence of favorable effects on child welfare outcomes among children and families in Arizona.

References

Arizona Department of Child Safety. "DCS Report on Progress Implementing Key Components of the Family First Prevention Services Act." January 2021a. Available at https://dcs.az.gov/content/dcs-progress-implementing-ffpsa-2021. Accessed May 18, 2021.

Arizona Department of Child Safety. "Monthly Operational Outcomes Report_Jun 2022." June 2022. Available at https://dcs.az.gov/content/monthly-operational-outcomes-reportjun-2022. Accessed July 26, 2022.

Arizona Department of Child Safety. "Service Array Re-Design." May 2021b. Available at https://dcs.az.gov/service-array. Accessed May 18, 2021.

Barth, R. P., J. Landsverk, P. Chamberlain, J. B. Reid, J. A. Rolls, M. S. Hurlburt, E. M. Z. Farmer, 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*, vol. 15, no. 5, 2005, pp. 353–371.

Barth, Richard P., and Kyla Liggett-Creel. "Common Components of Parenting Programs for Children Birth to Eight Years of Age Involved with Child Welfare Services." *Children and Youth Services Review,* vol. 40, 2014, pp. 6–12.

Brock, D. J. P., L. I. Marek, C. Matteo-Kerney, and T. Bagby. "Open Groups: Adaptations in Implementing a Parent Training Program." *Health Promotion Perspectives*, vol. 3, no. 2, 2013, pp. 230–241.

Burnson, C., S. Covington, B. Arvizo, J. Qiao, and E. Harris. "The Impact of Parents Anonymous on Child Safety and Permanency." *Children and Youth Services Review,* vol. 124, 2021, article 105973. https://doi.org/10.1016/j.childyouth.2021.105973.

Centers for Disease Control and Prevention. "Parent Training Programs: Insight for Practitioners." 2009. Available at https://www.cdc.gov/violenceprevention/pdf/parent_training_brief-a.pdf.

Chaiyachati, B. H., J. R. Gaither, M. Hughes, K. Foley-Schain, and J. M. Leventhal. "Preventing Child Maltreatment: Examination of an Established Statewide Home-Visiting Program." *Child Abuse and Neglect,* vol. 79, 2018, pp. 476–484.

Child Welfare Information Gateway. "Child Welfare Practice to Address Racial Disproportionality and Disparity." Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Children's Bureau, 2021. Available at https://www.childwelfare.gov/pubpdfs/racial_disproportionality.pdf.

Child Welfare Information Gateway. "Parent Education to Strengthen Families and Reduce the Risk of Maltreatment." Washington, DC: Children's Bureau, U.S. Department of Health and Human Services, 2019. Available at https://www.childwelfare.gov/pubPDFs/parented.pdf.

Cox, D. R. *Analysis of Binary Data*. London: Chapman and Hall/CRC, 1970.

Family Development Resources Inc. "Nurturing Skills for Families—Lesson Guide for Parents (NSF-LGP)." 2007. Available at https://www.nurturingparenting.com/files/nsf_lgp_intro__toc.pdf.

Palmer, L., S. Font, A. L. Eastman, L. Guo, and E. Putnam-Hornstein. "What Does Child Protective Services Investigate as Neglect? A Population-Based Study." *Child Maltreatment*, Online ahead of print, 2022. https://doi.org/10.1177/10775595221114144.

Greeno, E. J., J. A. Cosgrove, and B. R. Lee. "The Evaluation of a Nurturing Parenting Program Implemented by Child Welfare Workers." *Children and Youth Services Review*, vol. 127, 2021, article 106118.

Hainmueller, J. "Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies." *Political Analysis*, vol. 20, no. 1, 2012, pp. 25–46.

Hedges, L. "Distribution Theory for Glass's Estimator of Effect Size and Related Estimators." *Journal of Educational Statistics,* vol. 6, no. 2, summer 1981, pp. 107–128.

Illinois Department of Children and Family Services. "Illinois Birth Through Three Waiver: Developmentally Informed Child and Family Interventions: Final Evaluation Report Reporting Period: 7/1/2013 – 9/30/2018." December 2018. Available at https://www2.illinois.gov/dcfs/aboutus/newsandreports/Documents/IL_IB3_Final_Evaluation_Report_Dec_2018.pdf.

Kids Count Data Center. "Child Population by Race in Arizona." 2021. Available at https://datacenter.kidscount.org/data/tables/103-child-population-by-race?loc=1&loct=2#detailed/2/4/false/574/68,69,67,12,70,66,71,72/423,424.

Kids Count Data Center. "Children 0 to 17 in Foster Care in Arizona." 2022. Available at https://datacenter.kidscount.org/data/tables/6242-children-0-to-17-in-foster-care?loc=1&loct=2#detailed/2/4/false/574,1729,37,871,870,573,869,36,868,867/any/12985,20455.

Maher, E. J., L. A. Marcynyszyn, T. W. Corwin, and R. Hodnett. "Dosage Matters: The Relationship Between Participation in the Nurturing Parenting Program for Infants, Toddlers, and Preschoolers *and* Subsequent Child Maltreatment." *Children and Youth Services Review,* vol. *33*, no. 8, 2011, pp. 1426–1434.

National Center on Parent, Family, and Community Engagement. "Compendium of Parenting Interventions." Washington, DC: Office of Head Start, U.S. Department of Health and Human Services, 2015. Available at https://www.acf.hhs.gov/sites/default/files/documents/ecd/compendium_of_parenting_interventions_911_508.pdf.

Orr, Larry L. *Social Experiments: Evaluating Public Programs with Experimental Methods*. Thousand Oaks, CA: Sage, 1999.

Quick-Beachy, K., C. Lee, L. McConnell, R. Orsi, Z. Timpe, and M. Winokur. "SafeCare Colorado Program Evaluation Report 2014–2017." Denver, CO: Colorado Office of Early Childhood, 2018.

Weikert, P., R. Keene, and S. J. Bavolek. "The Florida Study: A Comparative Examination of the Effectiveness of the Nurturing Parenting Programs." Park City, UT: Nurturing Parenting, 2007.

Wilson, S. J., C. S. Price, S. E. U. Kerns, S. D. Dastrup, and S. R. Brown. "Title IV-E Prevention Services Clearinghouse Handbook of Standards and Procedures, Version 1.0." OPRE Report # 2019-56. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, 2019.

Technical Appendix

This technical appendix supplements the report on the impacts of NPP. Appendix A details the study's analytic approach, including additional information on baseline equivalence, weighting, and the models used to estimate impacts. Appendix B provides supplemental information on the impacts of NPP. All data preparation and analyses for the study were conducted in Stata 16.1.

Appendix A
Analytic Approach

A. Baseline equivalence

For quasi-experimental designs (QEDs), producing a credible estimate of program impacts requires establishing baseline equivalence—that is, ensuring that children in the NPP and comparison groups were similar before the program. Establishing baseline equivalence supports that any differences in outcomes between study groups were due to NPP and not to other factors, such as compositional differences.

Consistent with the PSC guidelines, we assessed baseline equivalence on pre-program measures of the outcomes and on sociodemographic characteristics. We used pre-program measures of the study's three main outcomes, as measured in the three years before the service referral: (1) investigations, (2) substantiated investigations, and (3) removals. For investigations, by design, every child in the sample was subject to a maltreatment investigation before the referral, so we used the number of investigations as an alternative (Wilson 2019). We used direct pre-tests for substantiated investigations and removals; that is, binary measures of whether the child was subject to a substantiated investigation or was removed from the home in the three years before the service referral. We also examined sociodemographic characteristics, such as the child's age at the time of the referral, gender, race and ethnicity, and family size.[4]

For each characteristic, we examined the mean and standard deviation separately for the NPP and comparison groups as well as the difference in means. We also tested whether the difference in means was statistically significant. For all inferential analyses in the study, we clustered standard errors at the case level because the analysis was conducted at the child level, yet cases are referred to NPP and can consist of multiple children. Lastly, we calculated the difference in means in terms of effect size (standard deviation) units. For continuous variables, we calculated the effect size as the standardized mean difference effect size multiplied by Hedges' *g,* the small-sample correction factor (Hedges 1981). For binary variables, we calculated the effect size by dividing the log odds ratio of the two study groups by 1.65 (Cox 1970).

The PSC considers differences in background characteristics between study groups of less than 0.05 standard deviations to be acceptable and not require a statistical adjustment. Differences from 0.05 to 0.25 standard deviations require a statistical adjustment. As described in further detail later in this appendix, we included characteristics in this range as covariates in the models used to estimate impacts. Differences larger than 0.25 standard deviations do not meet the PSC's baseline equivalence standards.

As described in the report, children in the NPP and comparison groups were similar on demographics, family characteristics, and histories of child welfare involvement, except for race and ethnicity (Table A.1). The results were similar for the 6- and 12-month follow-up samples (Tables A.2 and A.3).

[4] The Arizona Department of Child Safety uses six race and ethnicity categories: African American, American Indian, Asian or Pacific Islander, Hispanic, White, and Other. No children in our sample were in the Other category and the two smallest categories in our sample (Asian or Pacific Islander and American Indian) combined to make up only 8 percent of the sample. Therefore, for the baseline equivalence assessment, we combined these two categories with the next smallest group (African American) so as not to overinterpret effect size differences as Cox indices tend to be (artificially) magnified when calculated in the tails of the distribution.

Table A.1. Background characteristics of the full sample, by study group

Characteristic	NPP group		Comparison group		Difference		
	Mean	(SD)	Mean	(SD)	Mean	*p*-value	Effect size
Child demographics							
Age (years)	7.4	(5.1)	7.1	(5.2)	0.3	0.196	0.05
Female	51	(50.0)	49	(50.0)	1	0.356	0.04
Race and ethnicity							
Hispanic	40	(48.9)	38	(48.5)	2	0.427	0.05
White, non-Hispanic	42	(49.3)	37	(48.2)	5*	0.028	0.13
Other	19	(39.0)	26	(43.7)	-7**	0.000	-0.25
Family characteristics							
Family size[a]	3.0	(1.7)	3.2	(1.9)	-0.2	0.107	-0.10
Caregiver age (years)[b]	35.2	(8.8)	34.9	(8.6)	0.4	0.390	0.04
Child welfare history[c]							
Average # of investigations	1.8	(1.1)	1.7	(1.2)	0.0	0.566	0.03
Investigation for							
Neglect	79	(40.9)	79	(40.9)	0	0.989	0.00
Physical abuse	36	(48.1)	35	(47.8)	1	0.716	0.02
Substantiated investigation	39	(48.8)	38	(48.6)	1	0.758	0.02
Removal	19	(39.6)	21	(40.6)	-1	0.456	-0.05
Number of children	1,102		6,845				

Source: Arizona Department of Child Safety (DCS) administrative data.

Note: Due to rounding, percentages may not sum to 100 and the reported mean difference may not equal the difference in means between the NPP and comparison groups. Other race and ethnicity includes American Indian, African American, and Asian or Pacific Islander. Investigations for neglect and physical abuse are not mutually exclusive. Standard errors are clustered at the case level.

[a] Number of children on the child's case.

[b] Average age of all perpetrators on the child's case.

[c] DCS involvement in the three years before the family's service referral.

**/* Differences are statistically significant at the .01/.05 levels, respectively, two-tailed test.

Table A.2. Background characteristics of the 6-month follow-up sample, by study group

Characteristic	NPP group		Comparison group		Difference		
	Mean	(SD)	Mean	(SD)	Mean	*p*-value	Effect size
Child demographics							
Age (years)	7.3	(5.2)	7.1	(5.2)	0.2	0.367	0.04
Female	51	(50.0)	49	(50.0)	2	0.371	0.04
Race and ethnicity							
Hispanic	39	(48.8)	38	(48.5)	1	0.673	0.03
White, non-Hispanic	42	(49.4)	37	(48.3)	5*	0.048	0.13
Other	19	(39.3)	25	(43.5)	-6**	0.005	-0.23
Family characteristics							
Family size[a]	3.1	(1.8)	3.2	(1.9)	-0.2	0.163	-0.10
Caregiver age (years)[b]	35.3	(9.0)	34.9	(8.6)	0.4	0.422	0.05
Child welfare history[c]							
Average # of investigations	1.7	(1.1)	1.8	(1.2)	0.0	0.751	-0.02
Investigation for							
Neglect	80	(39.8)	80	(40.0)	0	0.886	0.01
Physical abuse	33	(46.9)	34	(47.4)	-2	0.517	-0.04
Substantiated investigation	39	(48.8)	39	(48.8)	0	0.928	0.01
Removal	19	(39.2)	21	(40.5)	-2	0.393	-0.07
Number of children	858		5,684				

Source: Arizona Department of Child Safety (DCS) administrative data.

Note: The 6-month follow-up sample is comprised of children with outcome data available 6 months after the end of NPP. Due to rounding, percentages may not sum to 100 and the reported mean difference may not equal the difference in means between the NPP and comparison groups. Other race and ethnicity includes American Indian, African American, and Asian or Pacific Islander. Investigations for neglect and physical abuse are not mutually exclusive. Standard errors are clustered at the case level.

[a] Number of children on the child's case.

[b] Average age of all perpetrators on the child's case.

[c] DCS involvement in the three years before the family's service referral.

**/* Differences are statistically significant at the .01/.05 levels, respectively, two-tailed test.

Table A.3. Background characteristics of the 12-month follow-up sample, by study group

Characteristic	NPP group		Comparison group		Difference		
	Mean	(SD)	Mean	(SD)	Mean	*p*-value	Effect size
Child demographics							
Age (years)	7.3	(5.2)	7.2	(5.2)	0.1	0.618	0.03
Female	53	(50.0)	49	(50.0)	4	0.076	0.09
Race and ethnicity							
Hispanic	40	(49.1)	38	(48.5)	3	0.437	0.06
White, non-Hispanic	40	(49.1)	37	(48.4)	3	0.314	0.08
Other	19	(39.3)	25	(43.2)	-6*	0.039	-0.20
Family characteristics							
Family size[a]	3.0	(1.7)	3.3	(1.9)	-0.3	0.077	-0.15
Caregiver age (years)[b]	35.5	(9.6)	34.9	(8.6)	0.6	0.361	0.07
Child welfare history[c]							
Average # of investigations	1.8	(1.2)	1.8	(1.2)	0.0	0.712	0.03
Investigation for							
Neglect	82	(38.1)	81	(39.3)	2	0.485	0.06
Physical abuse	31	(46.4)	33	(47.1)	-2	0.497	-0.06
Substantiated investigation	42	(49.4)	40	(49.1)	2	0.543	0.05
Removal	19	(39.6)	21	(40.8)	-2	0.512	-0.06
Number of children	603		4,046				

Source: Arizona Department of Child Safety (DCS) administrative data.

Note: The 12-month follow-up sample is comprised of children with outcome data available 12 months after the end of NPP. Due to rounding, percentages may not sum to 100 and the reported mean difference may not equal the difference in means between the NPP and comparison groups. Other race and ethnicity includes American Indian, African American, and Asian or Pacific Islander. Investigations for neglect and physical abuse are not mutually exclusive. Standard errors are clustered at the case level.

[a] Number of children on the child's case.

[b] Average age of all perpetrators on the child's case.

[c] DCS involvement in the three years before the family's service referral.

**/* Differences are statistically significant at the .01/.05 levels, respectively, two-tailed test.

B. Weighting

Because of the differences in race and ethnicity between the NPP and comparison group, we used a weighting approach to form research groups that were more similar before NPP. We used entropy balancing to construct a reweighted comparison group that had similar background characteristics to the NPP group. Entropy balancing has several advantages relative to other matching or weighting procedures used to establish baseline equivalence, such as propensity score methods. Namely, unlike other methods, entropy balancing constructs weights by imposing that specified background characteristics be similar across research groups—thus, ensuring equivalence among those characteristics (Hainmueller 2012). In this sense, it is a more direct approach to achieve baseline equivalence than propensity score methods, which often require iterating on the propensity score model to achieve equivalence.

We obtained weights by conducting entropy balancing separately for each sample. We specified that both study groups have similar pre-program measures (or pre-program alternatives) of the three main outcomes: (1) number of investigations, (2) had a substantiated investigation, and (3) was removed from the home. We also specified that the groups be similar on child age, race and ethnicity, and family size. For each study group, the sum of the weights was equal to the number of children in that group. We confirmed that the distribution of the weights for each group was relatively evenly distributed. For example, the weights for the comparison group ranged from 0.40 to 1.65, the median was 1.04, and the interquartile range was 0.84 to 1.15.

As expected, the weighting approach was successful in creating study groups that had nearly identical pre-program characteristics. There were no statistically significant differences between the study groups and all differences were smaller than 0.25 standard deviations (Table A.4). The results were similar for the 6- and 12-month follow-up samples (Tables A.5 and A.6).

Table A.4. Background characteristics of the full sample after weighting, by study group

Characteristic	NPP group		Comparison group		Difference		
	Mean	(SD)	Mean	(SD)	Mean	*p*-value	Effect size
Child demographics							
Age (years)	7.4	(5.1)	7.4	(5.3)	0.0	1.000	0.00
Female	51	(50.0)	49	(50.0)	1	0.435	0.03
Race and ethnicity							
Hispanic	40	(48.9)	40	(48.9)	0	1.000	0.00
White, non-Hispanic	42	(49.3)	42	(49.3)	0	1.000	0.00
Other	19	(39.0)	19	(39.0)	0	1.000	0.00
Family characteristics							
Family size[a]	3.0	(1.7)	3.0	(1.7)	0.0	1.000	0.00
Caregiver age (years)[b]	35.2	(8.8)	35.1	(8.7)	0.2	0.682	0.02
Child welfare history[c]							
Average # of investigations	1.8	(1.1)	1.8	(1.2)	0.0	1.000	0.00
Investigation for							
Neglect	79	(40.9)	79	(40.9)	0	0.989	0.00
Physical abuse	36	(48.1)	36	(47.9)	1	0.821	0.01
Substantiated investigation	39	(48.8)	39	(48.8)	0	1.000	0.00
Removal	19	(39.6)	19	(39.6)	0	1.000	0.00
Number of children	1,102		6,845				

Source: Arizona Department of Child Safety (DCS) administrative data.

Note: Due to rounding, percentages may not sum to 100 and the reported mean difference may not equal the difference in means between the NPP and comparison groups. Other race and ethnicity includes American Indian, African American, and Asian or Pacific Islander. Investigations for neglect and physical abuse are not mutually exclusive. Standard errors are clustered at the case level.

[a] Number of children on the child's case.

[b] Average age of all perpetrators on the child's case.

[c] DCS involvement in the three years before the family's service referral.

**/* Differences are statistically significant at the .01/.05 levels, respectively, two-tailed test.

Table A.5. Background characteristics of the 6-month follow-up sample after weighting, by study group

Characteristic	NPP group		Comparison group		Difference		
	Mean	(SD)	Mean	(SD)	Mean	*p*-value	Effect size
Child demographics							
Age (years)	7.3	(5.2)	7.3	(5.3)	0.0	0.996	0.00
Female	51	(50.0)	49	(50.0)	1	0.418	0.04
Race and ethnicity							
Hispanic	39	(48.8)	39	(48.7)	0	0.995	0.00
White, non-Hispanic	42	(49.4)	42	(49.4)	0	0.989	0.00
Other	19	(39.3)	19	(39.3)	0	0.980	0.00
Family characteristics							
Family size[a]	3.1	(1.8)	3.1	(1.8)	0.0	0.991	0.00
Caregiver age (years)[b]	35.3	(9.0)	35.0	(8.7)	0.2	0.615	0.03
Child welfare history[c]							
Average # of investigations	1.7	(1.1)	1.7	(1.1)	0.0	0.998	0.00
Investigation for							
Neglect	80	(39.8)	80	(40.2)	1	0.782	0.02
Physical abuse	33	(46.9)	34	(47.4)	-1	0.567	-0.04
Substantiated investigation	39	(48.8)	39	(48.8)	0	1.000	0.00
Removal	19	(39.2)	19	(39.1)	0	0.995	0.00
Number of children	858		5,684				

Source: Arizona Department of Child Safety (DCS) administrative data.

Note: The 6-month follow-up sample is comprised of children with outcome data available 6 months after the end of NPP. Due to rounding, percentages may not sum to 100 and the reported mean difference may not equal the difference in means between the NPP and comparison groups. Other race and ethnicity includes American Indian, African American, and Asian or Pacific Islander. Investigations for neglect and physical abuse are not mutually exclusive. Standard errors are clustered at the case level.

[a] Number of children on the child's case.

[b] Average age of all perpetrators on the child's case.

[c] DCS involvement in the three years before the family's service referral.

**/* Differences are statistically significant at the .01/.05 levels, respectively, two-tailed test.

Table A.6. Background characteristics of the 12-month follow-up sample after weighting, by study group

Characteristic	NPP group		Comparison group		Difference		
	Mean	(SD)	Mean	(SD)	Mean	*p*-value	Effect size
Child demographics							
Age (years)	7.3	(5.2)	7.3	(5.3)	0.0	1.000	0.00
Female	53	(50.0)	49	(50.0)	4	0.080	0.09
Race and ethnicity							
Hispanic	40	(49.1)	40	(49.1)	0	0.999	0.00
White, non-Hispanic	40	(49.1)	40	(49.1)	0	0.998	0.00
Other	19	(39.3)	19	(39.3)	0	0.997	0.00
Family characteristics							
Family size[a]	3.0	(1.7)	3.0	(1.8)	0.0	0.997	0.00
Caregiver age (years)[b]	35.5	(9.6)	35.0	(8.8)	0.5	0.400	0.06
Child welfare history[c]							
Average # of investigations	1.8	(1.2)	1.8	(1.2)	0.0	1.000	0.00
Investigation for							
Neglect	82	(38.1)	81	(39.2)	1	0.530	0.06
Physical abuse	31	(46.4)	33	(47.0)	-2	0.540	-0.05
Substantiated investigation	42	(49.4)	42	(49.4)	0	0.999	0.00
Removal	19	(39.6)	19	(39.6)	0	0.999	0.00
Number of children	603		4,046				

Source: Arizona Department of Child Safety (DCS) administrative data.

Note: The 12-month follow-up sample is comprised of children with outcome data available 12 months after the end of NPP. Due to rounding, percentages may not sum to 100 and the reported mean difference may not equal the difference in means between the NPP and comparison groups. Other race and ethnicity includes American Indian, African American, and Asian or Pacific Islander. Investigations for neglect and physical abuse are not mutually exclusive. Standard errors are clustered at the case level.

[a] Number of children on the child's case.

[b] Average age of all perpetrators on the child's case.

[c] DCS involvement in the three years before the family's service referral.

**/* Differences are statistically significant at the .01/.05 levels, respectively, two-tailed test.

C. Impact estimation strategy

To provide a comprehensive overview of the impacts of NPP, we estimated two types of policy-relevant effects: the impacts of referrals to NPP (Research Question 1) and the impacts of completing NPP (Research Question 2). The first represented the average impact for all children whose families were referred to NPP, which incorporated the effects for those who received a high level of program dosage as well as those with lower levels of program participation. We also examined the impacts of NPP for children whose families completed NPP. To produce the second estimate, we excluded children from the NPP group whose families were marked in the provider data as not having completed the program.

We reassessed baseline equivalence for the sample used to estimate the impacts of program completion, which excluded the 370 children whose families did not complete the program (about one-third of the NPP group). Completers in the NPP group were similar to those in the comparison group on demographics (except for race and ethnicity), family characteristics, and histories of child welfare involvement (Table A.7). The results were similar for the samples used to estimate impacts of completion at the 6- and 12-month follow-up time points (Tables A.8 and A.9). Because some pre-program differences between NPP completers and the comparison group were larger than 0.25 standard deviations, we used the same entropy balancing approach as previously described to form more similar groups. This weighting approach was again successful in creating similar study groups. For the samples used to estimate the impacts of completing NPP, no differences between groups after weighting were statistically significant nor were any larger than 0.25 standard deviations (Tables A.10 to A.12).

For each sample, outcome, and follow-up time point, we used a multivariate weighted least squares regression model to estimate program impacts. We regressed the outcome on a binary indicator of NPP status. The regression models applied the weights from entropy balancing to account for any pre-program differences between study groups. We clustered standard errors at the case level to correct for the level of NPP assignment.

The models also included three types of covariates to adjust for any remaining pre-program differences between study groups. Including covariates may also improve the precision of the impact estimates by reducing the residual variation in the outcome measure (Orr 1999). First, all models included pre-program characteristics that showed effect size differences in the PSC's adjustment range of 0.05 to 0.25 standard deviations for any sample: child gender, caregiver age, whether the child was ever subject to an investigation for neglect in the three years before the referral, and whether the child was ever subject to an investigation for physical abuse in the three years before the referral. Second, for precision, the models also controlled for pre-program measures of the three main outcomes and the child's age, race and ethnicity, and family size. Third, the models included indicators for the year of service referral, to account for time trends, and indicators for the county where the children lived at the start of the investigation, to account for local factors that might influence their outcomes.

Table A.7. Background characteristics of NPP completers and the comparison group for the full sample

Characteristic	NPP completers		Comparison group		Difference		
	Mean	(SD)	Mean	(SD)	Mean	*p*-value	Effect size
Child demographics							
Age (years)	7.4	(5.1)	7.1	(5.2)	0.2	0.330	0.05
Female	49	(50.0)	49	(50.0)	0	0.988	0.00
Race and ethnicity							
Hispanic	41	(49.2)	38	(48.5)	3	0.285	0.08
White, non-Hispanic	41	(49.2)	37	(48.2)	4	0.138	0.11
Other	18	(38.8)	26	(43.7)	-7**	0.003	-0.26
Family characteristics							
Family size[a]	3.0	(1.8)	3.2	(1.9)	-0.1	0.299	-0.08
Caregiver age (years)[b]	35.3	(8.5)	34.9	(8.6)	0.4	0.433	0.05
Child welfare history[c]							
Average # of investigations	1.8	(1.1)	1.7	(1.2)	0.0	0.876	0.01
Investigation for							
Neglect	78	(41.4)	79	(40.9)	-1	0.721	-0.03
Physical abuse	34	(47.4)	35	(47.8)	-1	0.651	-0.03
Substantiated investigation	41	(49.3)	38	(48.6)	3	0.279	0.08
Removal	18	(38.4)	21	(40.6)	-3	0.189	-0.11
Number of children	732		6,845				

Source: Arizona Department of Child Safety (DCS) administrative data.

Note: Due to rounding, percentages may not sum to 100 and the reported mean difference may not equal the difference in means between the NPP and comparison groups. Other race and ethnicity includes American Indian, African American, and Asian or Pacific Islander. Investigations for neglect and physical abuse are not mutually exclusive. Standard errors are clustered at the case level.

[a] Number of children on the child's case.

[b] Average age of all perpetrators on the child's case.

[c] DCS involvement in the three years before the family's service referral.

**/* Differences are statistically significant at the .01/.05 levels, respectively, two-tailed test.

Table A.8. Background characteristics of NPP completers and the comparison group for the 6-month follow-up sample

Characteristic	NPP completers		Comparison group		Difference		
	Mean	(SD)	Mean	(SD)	Mean	*p*-value	Effect size
Child demographics							
Age (years)	7.4	(5.2)	7.1	(5.2)	0.3	0.342	0.05
Female	49	(50.0)	49	(50.0)	0	0.919	0.01
Race and ethnicity							
Hispanic	41	(49.3)	38	(48.5)	4	0.272	0.09
White, non-Hispanic	40	(49.1)	37	(48.3)	3	0.280	0.09
Other	18	(38.8)	25	(43.5)	-7*	0.012	-0.25
Family characteristics							
Family size[a]	3.1	(1.9)	3.2	(1.9)	-0.1	0.386	-0.08
Caregiver age (years)[b]	35.6	(8.7)	34.9	(8.6)	0.7	0.235	0.08
Child welfare history[c]							
Average # of investigations	1.8	(1.2)	1.8	(1.2)	0.0	0.972	0.00
Investigation for							
Neglect	80	(40.0)	80	(40.0)	0	0.982	0.00
Physical abuse	30	(45.7)	34	(47.4)	-5	0.125	-0.13
Substantiated investigation	41	(49.1)	39	(48.8)	2	0.622	0.04
Removal	17	(37.6)	21	(40.5)	-4	0.133	-0.15
Number of children	582		5,684				

Source: Arizona Department of Child Safety (DCS) administrative data.

Note: The 6-month follow-up sample is comprised of children with outcome data available 6 months after the end of NPP. Due to rounding, percentages may not sum to 100 and the reported mean difference may not equal the difference in means between the NPP and comparison groups. Other race and ethnicity includes American Indian, African American, and Asian or Pacific Islander. Investigations for neglect and physical abuse are not mutually exclusive. Standard errors are clustered at the case level.

[a] Number of children on the child's case.

[b] Average age of all perpetrators on the child's case.

[c] DCS involvement in the three years before the family's service referral.

**/* Differences are statistically significant at the .01/.05 levels, respectively, two-tailed test.

Table A.9. Background characteristics of NPP completers and the comparison group for the 12-month follow-up sample

Characteristic	NPP completers		Comparison group		Difference		
	Mean	(SD)	Mean	(SD)	Mean	*p*-value	Effect size
Child demographics							
Age (years)	7.4	(5.2)	7.2	(5.2)	0.2	0.518	0.04
Female	51	(50.1)	49	(50.0)	2	0.472	0.04
Race and ethnicity							
Hispanic	45	(49.8)	38	(48.5)	7	0.095	0.17
White, non-Hispanic	37	(48.4)	37	(48.4)	0	0.976	0.00
Other	18	(38.6)	25	(43.2)	-7	0.052	-0.24
Family characteristics							
Family size[a]	3.1	(1.9)	3.3	(1.9)	-0.2	0.344	-0.10
Caregiver age (years)[b]	35.9	(9.2)	34.9	(8.6)	1.0	0.189	0.11
Child welfare history[c]							
Average # of investigations	1.8	(1.3)	1.8	(1.2)	0.1	0.568	0.06
Investigation for							
Neglect	83	(38.0)	81	(39.3)	2	0.537	0.07
Physical abuse	28	(45.0)	33	(47.1)	-5	0.158	-0.14
Substantiated investigation	42	(49.5)	40	(49.1)	2	0.606	0.05
Removal	18	(38.6)	21	(40.8)	-3	0.352	-0.11
Number of children	401		4,046				

Source: Arizona Department of Child Safety (DCS) administrative data.

Note: The 12-month follow-up sample is comprised of children with outcome data available 12 months after the end of NPP. Due to rounding, percentages may not sum to 100 and the reported mean difference may not equal the difference in means between the NPP and comparison groups. Other race and ethnicity includes American Indian, African American, and Asian or Pacific Islander. Investigations for neglect and physical abuse are not mutually exclusive. Standard errors are clustered at the case level.

[a] Number of children on the child's case.

[b] Average age of all perpetrators on the child's case.

[c] DCS involvement in the three years before the family's service referral.

**/* Differences are statistically significant at the .01/.05 levels, respectively, two-tailed test.

Table A.10. Background characteristics of NPP completers and the comparison group for the full sample after weighting

Characteristic	NPP completers		Comparison group		Difference		
	Mean	(SD)	Mean	(SD)	Mean	*p*-value	Effect size
Child demographics							
Age (years)	7.4	(5.1)	7.4	(5.3)	0.0	1.000	0.00
Female	49	(50.0)	49	(50.0)	0	0.920	0.00
Race and ethnicity							
Hispanic	41	(49.2)	41	(49.2)	0	1.000	0.00
White, non-Hispanic	41	(49.2)	41	(49.1)	0	1.000	0.00
Other	18	(38.8)	18	(38.8)	0	1.000	0.00
Family characteristics							
Family size[a]	3.0	(1.8)	3.0	(1.8)	0.0	1.000	0.00
Caregiver age (years)[b]	35.3	(8.5)	35.0	(8.6)	0.3	0.547	0.03
Child welfare history[c]							
Average # of investigations	1.8	(1.1)	1.8	(1.2)	0.0	1.000	0.00
Investigation for							
Neglect	78	(41.4)	79	(40.6)	-1	0.628	-0.04
Physical abuse	34	(47.4)	35	(47.7)	-1	0.746	-0.02
Substantiated investigation	41	(49.3)	41	(49.3)	0	1.000	0.00
Removal	18	(38.4)	18	(38.3)	0	1.000	0.00
Number of children	732		6,845				

Source: Arizona Department of Child Safety (DCS) administrative data.

Note: Due to rounding, percentages may not sum to 100 and the reported mean difference may not equal the difference in means between the NPP and comparison groups. Other race and ethnicity includes American Indian, African American, and Asian or Pacific Islander. Investigations for neglect and physical abuse are not mutually exclusive. Standard errors are clustered at the case level.

[a] Number of children on the child's case.

[b] Average age of all perpetrators on the child's case.

[c] DCS involvement in the three years before the family's service referral.

**/* Differences are statistically significant at the .01/.05 levels, respectively, two-tailed test.

Table A.11. Background characteristics of NPP completers and the comparison group for the 6-month follow-up sample after weighting

Characteristic	NPP completers		Comparison group		Difference		
	Mean	(SD)	Mean	(SD)	Mean	*p*-value	Effect size
Child demographics							
Age (years)	7.4	(5.2)	7.4	(5.3)	0.0	0.995	0.00
Female	49	(50.0)	49	(50.0)	0	0.976	0.00
Race and ethnicity							
Hispanic	41	(49.3)	41	(49.2)	0	0.992	0.00
White, non-Hispanic	40	(49.1)	40	(49.1)	0	0.992	0.00
Other	18	(38.8)	18	(38.8)	0	0.981	0.00
Family characteristics							
Family size[a]	3.1	(1.9)	3.1	(1.8)	0.0	0.995	0.00
Caregiver age (years)[b]	35.6	(8.7)	35.0	(8.6)	0.6	0.321	0.07
Child welfare history[c]							
Average # of investigations	1.8	(1.2)	1.8	(1.2)	0.0	1.000	0.00
Investigation for							
Neglect	80	(40.0)	80	(40.0)	0	0.967	0.00
Physical abuse	30	(45.7)	34	(47.4)	-5	0.131	-0.13
Substantiated investigation	41	(49.1)	41	(49.1)	0	0.998	0.00
Removal	17	(37.6)	17	(37.6)	0	0.990	0.00
Number of children	582		5,684				

Source: Arizona Department of Child Safety (DCS) administrative data.

Note: The 6-month follow-up sample is comprised of children with outcome data available 6 months after the end of NPP. Due to rounding, percentages may not sum to 100 and the reported mean difference may not equal the difference in means between the NPP and comparison groups. Other race and ethnicity includes American Indian, African American, and Asian or Pacific Islander. Investigations for neglect and physical abuse are not mutually exclusive. Standard errors are clustered at the case level.

[a] Number of children on the child's case.

[b] Average age of all perpetrators on the child's case.

[c] DCS involvement in the three years before the family's service referral.

**/* Differences are statistically significant at the .01/.05 levels, respectively, two-tailed test.

Table A.12. Background characteristics of NPP completers and the comparison group for the 12-month follow-up sample after weighting

Characteristic	NPP completers		Comparison group		Difference		
	Mean	(SD)	Mean	(SD)	Mean	*p*-value	Effect size
Child demographics							
Age (years)	7.4	(5.2)	7.4	(5.3)	0.0	0.999	0.00
Female	51	(50.1)	49	(50.0)	2	0.475	0.04
Race and ethnicity							
Hispanic	45	(49.8)	45	(49.7)	0	0.998	0.00
White, non-Hispanic	37	(48.4)	37	(48.3)	0	0.999	0.00
Other	18	(38.6)	18	(38.6)	0	0.997	0.00
Family characteristics							
Family size[a]	3.1	(1.9)	3.1	(1.8)	0.0	0.999	0.00
Caregiver age (years)[b]	35.9	(9.2)	35.0	(8.7)	0.9	0.214	0.11
Child welfare history[c]							
Average # of investigations	1.8	(1.3)	1.8	(1.2)	0.0	1.000	0.00
Investigation for							
Neglect	83	(38.0)	81	(39.2)	1	0.582	0.06
Physical abuse	28	(45.0)	34	(47.2)	-5	0.132	-0.15
Substantiated investigation	42	(49.5)	42	(49.4)	0	1.000	0.00
Removal	18	(38.6)	18	(38.6)	0	0.999	0.00
Number of children	401		4,046				

Source: Arizona Department of Child Safety (DCS) administrative data.

Note: The 12-month follow-up sample is comprised of children with outcome data available 12 months after the end of NPP. Due to rounding, percentages may not sum to 100 and the reported mean difference may not equal the difference in means between the NPP and comparison groups. Other race and ethnicity includes American Indian, African American, and Asian or Pacific Islander. Investigations for neglect and physical abuse are not mutually exclusive. Standard errors are clustered at the case level.

[a] Number of children on the child's case.

[b] Average age of all perpetrators on the child's case.

[c] DCS involvement in the three years before the family's service referral.

**/* Differences are statistically significant at the .01/.05 levels, respectively, two-tailed test.

D. Subgroup analysis

We used a similar approach as described above to examine whether NPP was particularly effective for certain subgroups (Research Question 3). To examine impacts by child age, gender, and race and ethnicity, we tailored the regression models by including an interaction term between the indicator of NPP status and an indicator for the subgroup. For example, to estimate impacts separately for female and male children, we included the interaction between the indicator of NPP status and a female indicator. To examine impacts by NPP provider and service level, we tailored the regression model by including separate treatment indicators for each provider or service level instead of a single treatment indicator. For each subgroup, we tested whether the impact of NPP was statistically significant. We also tested whether the difference between the impacts of NPP among the subgroups was statistically significant, to understand if the effect of NPP varied across the subgroups.

E. Approach to missing data

Children whose families were referred to in-home services in late 2019 or early 2020 may be missing 6- and 12-month follow-up data because outcomes were only available through December 31, 2020. To be in the study sample, children had to have outcome data available immediately at the end of NPP. We defined the end of NPP as 4 months (121 days) after the service referral, because that was when the majority of services were stated to have been delivered. Thus, children's families had to be referred on or before September 1, 2020, to be in the study sample. To have outcome data available 6 months after the end of NPP, the referral date had to be on or before March 3, 2020 (that is, 10 months or 303 days before the end of 2020). Likewise, to have outcome data available 12 months after the end of NPP, the referral date had to be on or before August 30, 2019 (that is, 16 months or 489 days before the end of 2020.) This means that the samples used to estimate impacts were different at each follow-up time point.

For missing background characteristics, we excluded children with missing characteristics from the analysis, referred to as complete case analysis. About 9 percent of the children were missing race and ethnicity data and another 5 percent were missing county information. Race and ethnicity data were critical for the analysis because the PSC reviews baseline equivalence on those characteristics. County information was also important because we included county indicators as covariates in the impact models, to account for local factors. Fewer than 2 percent of children were missing other background characteristics, such as age and gender. Overall, we excluded 17 percent of the sample due to missing background characteristics.

This page has been left blank for double-sided copying.

Appendix B:

Supplemental Information on Program Impacts

This appendix provides supplemental information on the program impacts discussed in Chapter V of the report: impacts on the study's main and secondary outcomes, sensitivity tests, and subgroup findings.

A. Impacts of NPP on main outcomes

We presented the impacts of referrals to NPP on the study's main outcomes in Table V.1 of the report and the impacts of completing NPP in Table V.2. Tables B.1 and B.2 provide the same results, but include additional information: unadjusted group means, standard deviations, and exact *p*-values.

Table B.1. Impacts of referrals to NPP on main outcomes, by follow-up time point

Outcome measure	Follow-up time point	NPP group		Comparison group		Estimated effect		
		Unadjusted mean	Adjusted mean	Unadjusted mean	Adjusted mean	Impact	*p*-value	Effect size
Child safety								
Investigation	0 mos	10	10	11	11	-1	0.635	-0.05
	6 mos	22	23	22	22	1	0.702	0.03
	12 mos	33	33	29	29	4	0.197	0.12
Substantiated investigation	0 mos	3	3	3	3	0	0.987	0.00
	6 mos	7	6	5	5	1	0.460	0.13
	12 mos	11	9	7	8	1	0.514	0.12
Child permanency								
Removal	0 mos	7	7	8	7	0	0.784	0.04
	6 mos	13	14	12	12	2	0.422	0.09
	12 mos	17	15	15	15	0	0.988	0.00
Number of children	0 mos	1,102		6,845				
	6 mos	858		5,684				
	12 mos	603		4,046				

Source: Arizona Department of Child Safety administrative data.

Note: This table shows the impacts of referrals to NPP at three follow-up time points: immediately after NPP (0 mos), 6 months after (6 mos), and 12 months after (12 mos). The numbers in the NPP Group and Comparison Group columns apply the same weights as those used in the baseline equivalence assessment. Standard errors are clustered at the case level.

**/* Impact estimates are statistically significant at the .01/.05 levels, respectively, two-tailed test.

Table B.2. Impacts of completing NPP on main outcomes, by follow-up time point

Outcome measure	Follow-up time point	NPP completers		Comparison group		Estimated effect		
		Unadjusted mean	Adjusted mean	Unadjusted mean	Adjusted mean	Impact	*p*-value	Effect size
Child safety								
Investigation	0 mos	8	7	11	11	-4*	0.030	-0.28
	6 mos	20	20	22	22	-2	0.384	-0.09
	12 mos	32	31	29	29	2	0.605	0.06
Substantiated investigation	0 mos	1	1	3	3	-2*	0.029	-0.66
	6 mos	5	5	5	5	0	0.825	-0.05
	12 mos	8	6	7	7	-1	0.515	-0.14
Child permanency								
Removal	0 mos	2	3	8	8	-5**	0.000	-0.78
	6 mos	7	6	12	12	-6**	0.003	-0.42
	12 mos	8	6	15	15	-9**	0.000	-0.70
Number of children	0 mos	732		6,845				
	6 mos	582		5,684				
	12 mos	401		4,046				

Source: Arizona Department of Child Safety administrative data.

Note: This table shows the impacts of completing NPP at three follow-up time points: immediately after NPP (0 mos), 6 months after (6 mos), and 12 months after (12 mos). The numbers in the NPP Completers and Comparison Group columns apply the same weights as those used in the baseline equivalence assessment. Standard errors are clustered at the case level.

**/* Impact estimates are statistically significant at the .01/.05 levels, respectively, two-tailed test.

B. Impacts of NPP on secondary outcomes

This section presents the impacts of referrals to and of completing NPP on the study's secondary outcomes: (1) investigations for neglect and (2) investigations for physical abuse. As discussed in the report, we found no impacts of referrals to NPP on either allegation type, and large and statistically significant decreases in investigations of physical abuse for NPP completers (Table B.3).

Table B.3. Impacts of NPP on secondary outcomes, by follow-up time point

Outcome measure	Follow-up time point	NPP group		Comparison group		Estimated effect		
		Unadjusted mean	Adjusted mean	Unadjusted mean	Adjusted mean	Impact	*p*-value	Effect size
Impacts of referrals to NPP								
Investigation for neglect	0 mos	7	6	7	7	-1	0.685	-0.05
	6 mos	16	15	15	15	0	0.976	0.00
	12 mos	24	24	21	21	3	0.336	0.10
Investigation for physical abuse	0 mos	3	3	4	4	-1	0.508	-0.12
	6 mos	7	8	8	8	0	0.832	-0.03
	12 mos	13	12	10	10	2	0.495	0.11
Number of children	0 mos	1,102		6,845				
	6 mos	858		5,684				
	12 mos	603		4,046				
Impacts of completing NPP								
Investigation for neglect	0 mos	6	5	7	7	-2	0.182	-0.22
	6 mos	16	15	15	15	0	0.905	-0.01
	12 mos	23	22	21	21	1	0.724	0.04
Investigation for physical abuse	0 mos	2	2	4	4	-2*	0.023	-0.51
	6 mos	4	4	8	8	-4*	0.014	-0.43
	12 mos	11	11	11	11	0	0.897	-0.02
Number of children	0 mos	732		6,845				
	6 mos	582		5,684				
	12 mos	401		4,046				

Source: Arizona Department of Child Safety administrative data.

Note: This table shows the impacts of referrals to and completing NPP at three follow-up time points: immediately after NPP (0 mos), 6 months after (6 mos), and 12 months after (12 mos). The study data include the most severe allegation associated with each investigation, with physical abuse allegations coded by DCS as more severe than neglect, meaning that investigations for both types are coded as physical abuse. The numbers in the NPP Group and Comparison Group columns apply the same weights as those used in the baseline equivalence assessment. Standard errors are clustered at the case level.

**/* Impact estimates are statistically significant at the .01/.05 levels, respectively, two-tailed test.

C. Sensitivity tests

We conducted several sensitivity tests to verify that the impacts of both referrals to and completion of NPP were robust to the specific, different analytic decisions that we made.

1. Used alternative methods to estimate impacts

We compared the findings from the main approach used to estimate impacts, which included covariates and weights to adjust for pre-program differences between study groups, with findings from two alternative models: one that did not include covariates and another that did not apply weights. The model without covariates was a parsimonious bivariate regression of the outcome on a binary indicator of NPP status. The model without weights included all of the covariates described in Appendix A but did not apply the weights obtained from entropy balancing. Standard errors were clustered at the case level for all models. The findings from these alternative models were nearly identical to those from the study's main approach (Table B.4).

Table B.4. Impacts of NPP on main outcomes, using alternative methods

Outcome measure	Follow-up time point	Impacts of referrals to NPP			Impacts of completing NPP		
		Main approach	No covariates	No weights	Main approach	No covariates	No weights
Child safety							
Investigation	0 mos	-1	-1	-1	-4*	-4*	-4*
	6 mos	1	1	1	-2	-2	-2
	12 mos	4	4	4	2	3	2
Substantiated investigation	0 mos	0	0	0	-2*	-2*	-2*
	6 mos	1	2	1	0	0	0
	12 mos	1	3	2	-1	0	-1
Child permanency							
Removal	0 mos	0	0	1	-5**	-6**	-5**
	6 mos	2	1	2	-6**	-5**	-5**
	12 mos	0	2	0	-9**	-7**	-9**
Number of children	0 mos	7,947	7,947	7,947	7,577	7,577	7,577
	6 mos	6,542	6,542	6,542	6,266	6,266	6,266
	12 mos	4,649	4,649	4,649	4,447	4,447	4,447

Source: Arizona Department of Child Safety administrative data.

Note: This table shows the impacts of referrals to and completing NPP at three follow-up time points: immediately after NPP (0 mos), 6 months after (6 mos), and 12 months after (12 mos), using different methods. The numbers in the Main Approach and No Covariates columns apply the same weights as those used in the baseline equivalence assessment. Standard errors are clustered at the case level.

**/* Impact estimates are statistically significant at the .01/.05 levels, respectively, two-tailed test.

2. Used alternative measures of the main outcomes

The three main outcomes were coded as binary in the analysis presented thus far because the first-order questions were whether the outcome occurred. As sensitivity tests, we examined count versions of each main outcome: (1) number of investigations, (2) number of substantiated investigations, and (3) number of out-of-home placements. The findings using these count outcomes were consistent with the results on the study's main outcomes in terms of sign and statistical significance (Tables B.5 and B.6).

Table B.5. Impacts of referrals to NPP on alternative versions of main outcomes, by follow-up time point

Outcome measure	Follow-up time point	NPP group			Comparison group			Estimated effect		
		Unadjusted mean	Adjusted mean	Unadjusted (SD)	Unadjusted mean	Adjusted mean	Unadjusted (SD)	Impact	*p*-value	Effect size
Child safety										
Number of investigations	0 mos	0.1	0.1	(0.3)	0.1	0.1	(0.4)	0.0	0.244	-0.05
	6 mos	0.3	0.3	(0.6)	0.3	0.3	(0.6)	0.0	0.622	-0.03
	12 mos	0.5	0.5	(0.8)	0.4	0.4	(0.8)	0.1	0.354	0.07
Number of substantiated investigations	0 mos	0.0	0.0	(0.2)	0.0	0.0	(0.2)	0.0	0.910	0.01
	6 mos	0.1	0.1	(0.3)	0.1	0.1	(0.3)	0.0	0.373	0.06
	12 mos	0.1	0.1	(0.4)	0.1	0.1	(0.3)	0.0	0.455	0.07
Child permanency										
Number of placements	0 mos	0.2	0.2	(0.9)	0.2	0.2	(0.9)	0.0	0.957	0.00
	6 mos	0.3	0.3	(1.1)	0.3	0.3	(1.2)	0.0	0.803	0.01
	12 mos	0.5	0.4	(1.3)	0.4	0.4	(1.3)	0.0	0.951	0.00
Number of children	0 mos	1,102			6,845					
	6 mos	858			5,684					
	12 mos	603			4,046					

Source: Arizona Department of Child Safety administrative data.

Note: This table shows the impacts of referrals to NPP at three follow-up time points: immediately after NPP (0 mos), 6 months after (6 mos), and 12 months after (12 mos). The numbers in the NPP Group and Comparison Group columns apply the same weights as those used in the baseline equivalence assessment. Standard errors are clustered at the case level.

**/* Impact estimates are statistically significant at the .01/.05 levels, respectively, two-tailed test.

Table B.6. Impacts of completing NPP on alternative versions of main outcomes, by follow-up time point

Outcome measure	Follow-up time point	NPP completers			Comparison group			Estimated effect		
		Unadjusted mean	Adjusted mean	Unadjusted (SD)	Unadjusted mean	Adjusted mean	Unadjusted (SD)	Impact	*p*-value	Effect size
Child safety										
Number of investigations	0 mos	0.1	0.0	(0.3)	0.1	0.1	(0.4)	-0.1**	0.003	-0.15
	6 mos	0.2	0.2	(0.6)	0.3	0.3	(0.6)	-0.1	0.162	-0.09
	12 mos	0.5	0.4	(0.8)	0.4	0.4	(0.8)	0.0	0.617	0.04
Number of substantiated investigations	0 mos	0.0	0.0	(0.1)	0.0	0.0	(0.2)	0.0*	0.017	-0.11
	6 mos	0.1	0.1	(0.3)	0.1	0.1	(0.3)	0.0	0.884	0.01
	12 mos	0.1	0.1	(0.3)	0.1	0.1	(0.3)	0.0	0.790	-0.03
Child permanency										
Number of placements	0 mos	0.0	0.0	(0.2)	0.2	0.2	(0.9)	-0.2**	0.000	-0.19
	6 mos	0.1	0.1	(0.6)	0.3	0.3	(1.2)	-0.2**	0.000	-0.17
	12 mos	0.2	0.1	(0.8)	0.4	0.4	(1.4)	-0.3**	0.000	-0.22
Number of children	0 mos	732			6,845					
	6 mos	582			5,684					
	12 mos	401			4,046					

Source: Arizona Department of Child Safety administrative data.

Note: This table shows the impacts of completing NPP at three follow-up time points: immediately after NPP (0 mos), 6 months after (6 mos), and 12 months after (12 mos). The numbers in the NPP Completers and Comparison Group columns apply the same weights as those used in the baseline equivalence assessment. Standard errors are clustered at the case level.

**/* Impact estimates are statistically significant at the .01/.05 levels, respectively, two-tailed test.

3. Used a consistent sample to estimate impacts at each time point

As described in Appendix A, the study's main approach used different samples to estimate impacts at each follow-up time point. The samples were slightly different because families referred to services after March 2020 did not have available 6-month follow-up data and those referred after August 2019 did not have available 12-month follow-up data. Using different samples had the benefit of maximizing the sample size at each follow-up time point. At the same time, doing so meant that differences in impacts across time points may reflect compositional changes to the sample rather than actual changes in impacts over time. For this reason, we reassessed the impacts of NPP using a consistent sample at each time point: children with available 12-month follow-up data. The estimated impacts from using a consistent sample were similar in sign, magnitude, and statistical significance to those from the main approach (Table B.7).

Table B.7. Impacts of NPP on main outcomes, using a consistent sample

Outcome measure	Follow-up time point	Impacts of referrals to NPP		Impacts of completing NPP	
		Main approach	Consistent sample	Main approach	Consistent sample
Child safety					
Investigation	0 mos	-1	0	-4*	-4
	6 mos	1	2	-2	-1
	12 mos	4	4	2	2
Substantiated investigation	0 mos	0	1	-2*	-3**
	6 mos	1	1	-0	-1
	12 mos	2	1	-1	-1
Child permanency					
Removal	0 mos	0	0	-5**	-9**
	6 mos	2	2	-6**	-7**
	12 mos	0	0	-9**	-9**
Number of children	0 mos	7,947	4,649	7,577	4,447
	6 mos	6,542	4,649	6,266	4,447
	12 mos	4,649	4,649	4,447	4,447

Source: Arizona Department of Child Safety administrative data.

Note: This table shows the impacts of referrals to and completing NPP at three follow-up time points: immediately after NPP (0 mos), 6 months after (6 mos), and 12 months after (12 mos), using different samples. These analyses apply the same weights as those in the baseline equivalence assessment. Standard errors are clustered at the case level.

**/* Impact estimates are statistically significant at the .01/.05 levels, respectively, two-tailed test.

D. Subgroup findings

We examined whether NPP was particularly effective for certain subgroups, by age, gender, race and ethnicity, program provider, and service level. For child age, we divided the sample into children who were 6 years or younger versus 7 years or older at the time of the service referral because 7 was the average age for children in the sample (Table III.1). For gender, we examined impacts separately for female and male children, which were the categories available in the study data. We focused our analysis of impacts by race and ethnicity on three mutually exclusive categories to ensure sufficiently large sample sizes: Hispanic, White, and Other. The Other race and ethnicity category included American Indian, African American, and Asian or Pacific Islander children. For program provider, we looked at impacts separately for children whose families were referred to Arizona's Children Association and Casa de los Niños, and for service level, we examined impacts for moderate (one session per week for three months) versus intensive services (two sessions per week for four months). To maximize sample size and statistical power, we focused the subgroup analyses on impacts immediately after the end of the program.

Consistent with the results for the full sample, we found no evidence of impacts of referrals to NPP for each subgroup for all outcomes examined (Tables B.8 to B.12). Although the impact estimates for both age groups were not statistically different from zero, there was some evidence that referrals to NPP were more effective in reducing investigations and substantiated investigations for older children than for younger children. For the other subgroups, the impacts of referrals to NPP were small and statistically insignificant.

The results were more nuanced for completing NPP. The program appeared to be similarly effective in reducing removals regardless of children's age, gender, program provider, or service level. For children's race and ethnicity, the impacts on removals were in the favorable direction for all groups but only statistically significant for Hispanic and White children (the groups with the largest sample sizes).

The impacts of completing NPP on investigations and substantiated investigations varied somewhat across groups. The improvements in both outcomes were driven by older rather than younger children. The reduction in investigations was concentrated among female children, while the reduction in substantiated investigations was concentrated among male children. Completing NPP reduced both safety outcomes for Hispanic children. For White children, only the impact on investigations was statistically significant. We found no evidence of impacts on either safety outcome for children in the Other race and ethnicity group. In terms of NPP provider and service level, the reduction in investigations was driven by those referred to Casa de los Niños and to moderate services, and the reduction in substantiated investigations was driven by Arizona's Children Association and intensive services.

In most cases, these impact estimates were imprecise, and we did not find significant differences in the impacts between subgroups.

Table B.8. Impacts of NPP immediately after the end of the program, by child age

Outcome measure	Ages 6 years and younger		Ages 7 years and older		*p*-value for subgroup difference
	Comparison group	Impact	Comparison group	Impact	
Impacts of referrals to NPP					
Investigation	10	2	12	-3	0.010†
Substantiated investigation	3	1	3	-1	0.029†
Removal	8	1	7	-1	0.320
Number of children	3,902		4,045		
Impacts of completing NPP					
Investigation	10	-1	12	-6**	0.043†
Substantiated investigation	3	-2	3	-2*	0.904
Removal	9	-5**	7	-5**	0.872
Number of children	3,724		3,853		

Source: Arizona Department of Child Safety administrative data.

Note: The numbers in the Comparison Group columns are weighted, regression-adjusted predicted values of the outcomes. The weights to estimate impacts are the same as those in the baseline equivalence assessment. Standard errors are clustered at the case level.

**/* Impact estimates are statistically significant at the .01/.05 levels, respectively, two-tailed test.

††/† Statistically significant differences among the subgroup impact estimates at the .01/.05 level, respectively.

Table B.9. Impacts of NPP immediately after the end of the program, by child gender

Outcome measure	Female		Male		*p*-value for subgroup difference
	Comparison group	Impact	Comparison group	Impact	
Impacts of referrals to NPP					
Investigation	12	-2	11	0	0.329
Substantiated investigation	3	0	3	0	0.684
Removal	7	1	7	0	0.398
Number of children	3,916		4,031		
Impacts of completing NPP					
Investigation	12	-5*	11	-3	0.366
Substantiated investigation	3	-2	3	-2*	0.907
Removal	8	-6**	8	-5**	0.715
Number of children	3,718		3,859		

Source: Arizona Department of Child Safety administrative data.

Note: The numbers in the Comparison Group columns are weighted, regression-adjusted predicted values of the outcomes. The weights to estimate impacts are the same as those in the baseline equivalence assessment. Standard errors are clustered at the case level.

**/* Impact estimates are statistically significant at the .01/.05 levels, respectively, two-tailed test.

††/† Statistically significant differences among the subgroup impact estimates at the .01/.05 level, respectively.

Table B.10. Impacts of NPP immediately after the end of the program, by child race and ethnicity

Outcome measure	Hispanic		White		Other		*p*-value for Hispanic vs. White	*p*-value for Hispanic vs. Other	*p*-value for White vs. Other
	Comparison group	Impact	Comparison group	Impact	Comparison group	Impact			
Impacts of referrals to NPP									
Investigation	10	-2	13	-1	11	2	0.562	0.302	0.591
Substantiated investigation	3	-1	3	1	3	0	0.159	0.488	0.525
Removal	8	-2	8	3	7	0	0.120	0.704	0.377
Number of children	3,021		2,959		1,967				
Impacts of completing NPP									
Investigation	10	-5*	13	-5*	11	1	0.930	0.187	0.191
Substantiated investigation	3	-3**	3	-1	3	-1	0.221	0.308	0.933
Removal	8	-8**	8	-4*	7	-3	0.075	0.084	0.682
Number of children	2,883		2,798		1,896				

Source: Arizona Department of Child Safety administrative data.

Note: Other race and ethnicity includes American Indian, African American, and Asian or Pacific The numbers in the Comparison Group columns are weighted, regression-adjusted predicted values of the outcomes. The weights to estimate impacts are the same as those in the baseline equivalence assessment. Standard errors are clustered at the case level.

**/* Impact estimates are statistically significant at the .01/.05 levels, respectively, two-tailed test.

††/† Statistically significant differences among the subgroup impact estimates at the .01/.05 level, respectively.

Table B.11. Impacts of NPP immediately after the end of the program, by NPP provider

Outcome measure	Comparison group	Impact for Arizona's Children Association	Impact for Casa de los Niños	*p*-value for subgroup difference
Impacts of referrals to NPP				
Investigation	11	1	-2	0.393
Substantiated investigation	3	-1	1	0.565
Removal	7	0	1	0.663
Number of children	7,947			
Impacts of completing NPP				
Investigation	11	-2	-5*	0.367
Substantiated investigation	3	-3*	-1	0.221
Removal	8	-5*	-6**	0.653
Number of children	7,577			

Source: Arizona Department of Child Safety administrative data.

Note: The numbers in the Comparison Group column are weighted, regression-adjusted predicted values of the outcomes. The weights to estimate impacts are the same as those in the baseline equivalence assessment. Standard errors are clustered at the case level.

**/* Impact estimates are statistically significant at the .01/.05 levels, respectively, two-tailed test.

††/† Statistically significant differences among the subgroup impact estimates at the .01/.05 level, respectively.

Table B.12. Impacts of NPP immediately after the end of the program, by service level

Outcome measure	Comparison group	Impact for moderate services	Impact for intensive services	*p*-value for subgroup difference
Impacts of referrals to NPP				
Investigation	11	-2	2	0.154
Substantiated investigation	3	-1	1	0.390
Removal	7	-1	3	0.193
Number of children	7,947			
Impacts of completing NPP				
Investigation	11	-4*	-3	0.660
Substantiated investigation	3	-1	-3**	0.172
Removal	8	-6**	-5*	0.608
Number of children	7,577			

Source: Arizona Department of Child Safety administrative data.

Note: Intensive services last up to four months and are offered when children are at significant risk of removal from the home. Moderate services last up to three months and are offered when children face some level of risk of subsequent abuse or neglect. The numbers in the Comparison Group column are weighted, regression-adjusted predicted values of the outcomes. The weights to estimate impacts are the same as those in the baseline equivalence assessment. Standard errors are clustered at the case level.

**/* Impact estimates are statistically significant at the .01/.05 levels, respectively, two-tailed test.

††/† Statistically significant differences among the subgroup impact estimates at the .01/.05 level, respectively.

Mathematica Inc.

Princeton, NJ • Ann Arbor, MI • Cambridge, MA
Chicago, IL • Oakland, CA • Seattle, WA
Tucson, AZ • Woodlawn, MD • Washington, DC

mathematica.org

EDI Global, a Mathematica Company

Bukoba, Tanzania • High Wycombe, United Kingdom