

APPENDIX A

Low-Cost Randomized Controlled Trial of Safe Families for Children

Final Report

Report to the Laura and John Arnold Foundation

February 15, 2021

by

Dr. Mark F. Testa

Distinguished Professor Emeritus
School of Social Work, University of North Carolina at Chapel Hill

and

Dr. Stephen Budde

Executive Vice President
Juvenile Protective Association

On behalf of

Children's Home Society of America

ACKNOWLEDGEMENTS

The authors would like to acknowledge the support and contributions of those persons and organizations that made this evaluation possible. First and foremost, we wish to acknowledge the support of Richard H. Calica, past director of the Illinois Department of Children and Family Services (IDCFS), who requested the evaluation of the Safe Families for Children (SFFC) program. We are also grateful to the DCFS directors who succeeded him for their continued support of the project. They include: Denise Gonzales (Acting Director: November 15, 2013 to January 24, 2014); Bobbie Gregg (Acting Director: February 28, 2014 to January 19, 2015); Cynthia Tate (Interim Director: January 20, 2015 to February 16, 2015); George H. Sheldon (Appointed Acting Director on February 17, 2015 and confirmed by the Senate on January 13, 2016 and served to June 15, 2017); Lise T. Spacapan (Acting Director: June 16, 2017 to June 26, 2017); Beverly (BJ) Walker (Acting Director: June 27, 2017 to February 15, 2019); Debra Dyer-Webster (Interim Director: February 16, 2019 to April 14, 2019); and Marc D. Smith (Acting Director: April 15, 2019 to present). There were numerous individuals within the IDCFS who had direct responsibility for the implementation of the project. They include: Peter Digre, Anne Gold, Nora Harms-Pavelski, Dr. Kimberly Mann, Mary Rogers, and Michael Ruppe.

Very special thanks go to David Anderson, Executive Director of Lydia Home and developer of SFFC, and his staff: Ryan Mobley, Joyce Moffitt, Abel Ortiz, Carlos Sanchez, and Aarika Yarger. It takes courage to participate in a rigorous evaluation to learn the truth about your program's effectiveness. Their willingness to act on the lessons learned is an inspiring example of the program's commitment to continuous quality improvement.

An earlier draft of this paper was reviewed by Andrew Barclay, Statistician, Fostering Court Improvement; Joseph P. Doyle, Erwin H. Schell Professor of Management at the MIT Sloan School of Management, and Allan R. Ellis, Associate Professor, North Carolina State University. The recommendations that we were able to address resulted in significant improvements to the manuscript.

We are thankful for the consultation and assistance of Dr. Ding-Geng Chen, Wallace H. Kuralt Distinguished Professor, University of North Carolina at Chapel Hill, in providing us with the computer code for estimating the frequentist and Bayesian p-values and probabilities of superiority of SFFC compared to services as usual.

We are also indebted to Children's Home Society of America (CHSA) for serving as the fiscal agent for the project. Special appreciation goes to David Bundy, Brian Manes, Frank Poremski, Amy Ramirez, Nancy Ronquillo, and Michael Shaver for their role in project management.

The SFFC evaluation was supported through a grant from the Laura and John Arnold Foundation.

TABLE OF CONTENTS

EXECUTIVE SUMMARY	1
INTRODUCTION	11
Results-Oriented Accountability.....	12
Framework for Evidence Building.....	13
Current Report.....	14
SECTION ONE: IDENTIFY AND EXPLORE	15
Target Population and Primary Problem.....	15
Theory of Change.....	16
Informal and Formal Dichotomy	18
Multidimensional Construct of Formality.....	22
Research Question.....	28
Comparable Treatment Groups	29
Falsifiable Logic Model	34
SECTION TWO: DEVELOP AND TEST	39
Usability Testing	40
Power Analysis.....	43
Referral Challenges.....	45
Goodness of Fit with Theory of Change	47
Steps Taken to Boost Referral Volume.....	50
Initial Implementation and Formative Evaluation	53
Characteristics at Baseline	56
Protocol Deviations and Treatment Crossovers	57
SECTION THREE: COMPARE AND LEARN	61
Deflection from Foster Care (Primary Outcome)	67
Protective Custody (Proximal Outcome)	76
Permanence of Care at One Year (Distal Outcome 1).....	78
Recurrence of Maltreatment (Distal Outcome 2).....	81
Cost Savings.....	83
Effects of the Treatment on the Treated.....	85

SECTION FOUR: REPLICATE AND ADAPT	89
SECTION FIVE: APPLY AND IMPROVE	94
DISCUSSION	96
NEXT STEPS	104
CONCLUSION	110
REFERENCES.....	113
APPENDIX.....	124

EXECUTIVE SUMMARY

Safe Families for Children (SFFC) is a promising program to prevent foster care removal and the recurrence of maltreatment of children whose parents are investigated by Child Protective Services (CFS) for alleged child neglect and abuse. To achieve these purposes, SFFC recruits and oversees a network of unpaid host families with whom parents can voluntarily place their children in times of need. On the basis of promising findings from usability and formative evaluations conducted at the University of North Carolina (UNC), the study's principal investigators pre-registered a summative evaluation plan with the Open Science Framework (<https://osf.io/35ygk>) in August of 2017.¹

The summative evaluation was recently completed with 99 families and their 216 children who were enrolled in the study for the two-year period ending December 31, 2018. It tracked primary and secondary child welfare outcomes using public administrative data available to the investigators through June 30, 2020. The report also includes data from the usability testing and formative evaluation, which were completed on a total of 276 families and their 597 children who were enrolled in the study prior to January 1, 2017. The study employs the Bayesian paradigm, which is consistent with the scientific principle of phase-based evidence building in intervention research. It combines the prior information accumulated during the formative phase with the new data collected during the summative phase to provide a comprehensive analysis of SFFC's "intent-to-treat" impact.

¹ The registration of clinical trials prior to data analysis and publication serves both ethical and scientific purposes (Zarin, & Keselman, 2007). Most notably, it helps guard against "fishing" for significant statistical associations between the intervention and a variety of potential outcomes, which can arise by chance and give the erroneous impression of a program's effectiveness.

The SFFC evaluation is one of several initiatives that the Illinois Department of Children and Family Services (IDCFS) rolled out and evaluated under a supplemental implementation plan that the United States District Court for the Northern District of Illinois Eastern Division approved pursuant to the *B.H. v. Smith* Consent Decree (*B.H. Decree*). The *B.H. Decree* is one of over 30 settlements currently or previously in effect across the U.S (Kosanovich & Joseph, 2005). These decrees bind a CPS agency and the attorneys acting on behalf of the “plaintiff” class to the implementation of a plan to address entrenched injustices in the protection and care of abused, neglected, and vulnerable children.

Compliance management under the *B.H.* consent decree adheres to guidelines outlined in the *Framework to Design, Test, Spread, and Sustain Effective Practice in Child Welfare* (the Framework; Framework Workgroup, 2014). The U.S. Children’s Bureau funded the Framework to help CPS agencies build evidence for effective child welfare interventions and spread their implementation. By embedding an unbiased assignment mechanism in routine agency operations and then tracking results with existing administrative data, it is possible to guide evidence building at low cost through successive phases of increasingly generalizable validity (Shadish, Cook & Campbell, 2002). The SFFC evaluation demonstrates both the merits and challenges of this strategy. By adhering to a phase-based approach to evidence building, this study demonstrates the feasibility of systematically winnowing and ultimately enlarging the supply of promising, supported, and well-supported interventions in a cost-effective manner that satisfies the demanding evidence standards of results-oriented accountability (Testa & Poertner, 2010) and new federal legislation (Wilson, Price, Kerns, Dastrup & Brown, 2019).

The report is subdivided into five sections that align with the Framework’s five phases of evidence building. The first section describes the *Identify and Explore* phase, which defines the

scope and nature of the problem and outlines the theory of change which the program implements to attain the desired outcomes. The second section presents the findings from the *Develop and Test* phase, which confirms program usability during initial implementation and conducts a formative evaluation that tests statistically whether program outputs and improvements are trending in the desired direction. The third section presents findings from the *Compare and Learn* phase, which supports full implementation and a pre-registered, summative evaluation to assess whether the intervention created practical improvements that are causally attributable to the intervention. The fourth section describes the lessons learned from the *Replicate and Adapt* phase, which assesses the extent to which similar outcomes can be reproduced through the spread of the supported intervention to different populations and across varied settings. The last section outlines future steps that SFFC can take at the *Apply and Improve* phase to sustain positive results and adapt to changing circumstances through continuous quality improvement (CQI) and quality service reviews (QSRs).

The first section, Identify and Explore, describes the research question that the principal investigators pre-registered on the Open Science Framework. It has been modified slightly to define more precisely the target population and shorten the follow-up period to conform to the outcome definitions used in the federal Child and Family Services Review.² Framed in **PICOTS** terms (population, intervention, comparator, outcomes, timeframe, and setting), the research question is as follows with modifications in italics:

² One of the reviewers of a draft version of this report correctly pointed out that the study population should have been registered as Illinois children whose parents are investigated by CPS for alleged abuse and neglect and investigators deem appropriate for SFFC. In addition, the plan originally called for tracking outcomes for a period of 24 months. In this report, the follow-up period is shortened to 12 months to conform to the definitions used in the federal Child and Family Services Review (CFSR). The setting for the study now includes both CPS and SFFC in the wider community, which was previously omitted.

Are Illinois children whose parents are investigated by CPS for alleged abuse and neglect *and investigators deem appropriate for SFFC (P)*:

- less likely to enter state protective custody and formal foster care (primary **Outcome**) *during the 12 months following enrollment in the study (T)*,
- just as likely to avert subsequent episodes of *indicated* abuse/neglect (secondary **Outcome 1**) *during the same follow-up period*,
- and more likely to be maintained in or reunified with their birth families *at one year after allocation to the study* (secondary **Outcome 2**),

if they are referred to SFFC's network of voluntary host families (**I**) as compared to children from statistically equivalent families who are served by CPS as usual (**C**) *in the wider community (S)*?

The three directional hypotheses that the plan proposed to test are as follows:

- Among child subjects investigated for maltreatment, the percentage taken into protective custody or later removed into foster care after allocation to the treatment groups will be lower for the SFFC intervention groups compared to families who receive child protective services as usual (primary hypothesis).
- Among child subjects investigated for maltreatment, the percentage who had a recurrence of indicated maltreatment after the date of allocation to the treatment groups will be statistically equivalent for both intervention and comparison groups.
- Among child subjects investigated for child maltreatment, the percentage who are maintained in the custody of their parents or returned to their physical custody

within one year after allocation to the treatment groups will be higher in the intervention group than the comparison group.

The study was approved by Institutional Review Boards (IRBs) at the Illinois Department of Children and Family Services (IDCFS), the University of North Carolina at Chapel Hill (UNC), and the University of Illinois at Urbana-Champaign (UIUC). Because IDCFS is able to track primary and secondary outcomes through its administrative data systems, the study operates at low cost by forgoing primary data collection. The Department also agreed to automate an unbiased allocation mechanism (alternation) and a continuous data monitoring dashboard (proximal and distal outcomes) directly into its data system. Both automations further reduce costs and help bolster confidence in the (internal) validity of the findings on program effectiveness. In 2015, the Laura and John Arnold Foundation (LJAF) awarded a grant to the Children's Home Society of America (CHSA) to serve as the fiscal agent for the evaluation of the program by independent evaluators from UNC and the Juvenile Protective Association (JPA) in Chicago.

During the formative phase of the evaluation, the version of the program that operated in Cook County showed unacceptably high levels of non-compliance with study protocols. These deviations resulted in a high proportion (> 10%) of cross-overs from the comparison group to family hosting. Because of continuing high levels of non-compliance in Cook County, IDCFS leadership and the independent evaluators agreed to confine the pre-registration of the summative evaluation to the downstate regions of the state. Conversely, an intent-to-treat analysis of administrative data indicated a positive impact of referrals to the SFFC network on the primary outcome of deflecting children from protective custody and formal foster care in jurisdictions outside of Cook County, Illinois. The ratio of the transition (hazard) rate of removal

in these “downstate” counties for children referred to SFFC was one-third (0.335) of the rate for children assigned to services as usual (SAU). Group differences of this magnitude are considered to be between a small and medium effect size (Azuero, 2016).

The downstate Illinois formative evaluation also showed a promising impact of program referrals on the secondary outcome of the permanence of children’s living situation at one year after their enrollment in the study. The odds of permanence were 3.778 times greater for the intervention than the comparison group. Again the positive effect was confined to downstate regions. An accompanying secondary outcome, prevention of the recurrence of indicated maltreatment, showed no differences between the intervention and comparison groups. Even though the expectation going into formative evaluation had been that SFFC referrals would reduce repeat maltreatment, by the time we pre-registered the summative evaluation we thought a more realistic expectation was that SFFC exposes children to no greater danger than SAU. Our reasoning was that deflection of children from formal foster care could expose them to higher risks of repeat victimization than SAU by maintaining them in potentially abusive situations or reunifying them too quickly before their families had been adequately helped. Fortunately, the formative evaluation showed that there was no difference in repeat rates of indicated child maltreatment between the two groups in the twelve months following enrollment.

Downstate enrollment of families in the summative evaluation concluded six-months earlier than the planned ending date of June 30, 2019. The reasons for the shortened enrollment period included the appointment of a new IDCFS director, who accepted the validity of the intervention in the absence of well-supported evidence of its effectiveness, and the eagerness of the SFFC provider to dispense with automated alternation to the comparison group, which halved the number of referrals that could be hosted. The downside of the abbreviated enrollment period

was that only 567 families were allocated to the study, which fell below the targeted sample size of 750 determined to optimal for rejecting the null hypothesis of no improvement with 90% power. The smaller than desired sample size elevates the chances that practically important improvements in outcomes could erroneously be rejected as statistically indistinguishable from no difference (type II errors) because of inadequate statistical power.

An intent-to-treat analysis of the primary outcome of deflection from protective custody and foster care showed a smaller but still practically important effect size during summative evaluation than during formative evaluation. The ratio of the transition (hazard) rate of children's experiencing removal in downstate counties for children referred to SFFC was a little more than two-thirds (0.602) of the rate for children assigned to SAU. The null hypothesis of no difference was rejected at the .10 level (1-tailed test). Even though the chances of making a type I error (i.e. rejecting the null hypothesis that the program is ineffective) doesn't meet conventional significance thresholds, a persuasive argument can be made for supplementing this "frequentist" approach with a Bayesian paradigm that incorporates prior formative evidence to strengthen the statistical conclusion validity of the summative findings. Applying the Bayesian paradigm improved the *p*-value to a highly significant .004 level.

The summative evaluation also showed a weaker impact of program referrals on the secondary outcome of the permanence of children's living situation than the formative findings. The odds of permanence at one year after enrollment in the summative evaluation were only 1.491 times greater for the intervention than the comparison group compared to 3.778 times greater during formative evaluation. Nonetheless, applying the Bayesian paradigm indicated that the cumulative effect was statistically significant at the .009 level (1-tailed test). Finally, the summative findings on repeat rates of victimization remained consistent with the formative

results. The transition (hazard) ratio of indicated maltreatment in the intervention group was statistically indistinguishable from the comparison group for both phases of the study. Applying the Bayesian paradigm yielded a statistically negligible p -value of .834 when both the formative and summative evidence were taken into account.

The findings from the summative evaluation of SFFC replicate the promising results from the formative evaluation with respect to the primary outcome of deflection from protective custody and formal foster care in downstate Illinois. In addition, the data showed that few children languished in indefinite hosting arrangements, which critics had feared could occur when parents are deprived of the legal representation that formal foster care affords. The Illinois data showed that most children (70%) stayed in SFFC homes for fewer than two months and none stayed longer than one year. In contrast, 70% of children placed in formal foster care stayed for longer than one-year with one-half experiencing extended stays in state custody for longer than two years. Only 20% of children taken into formal foster care were reunified with their families within a year of assignment compared to nearly 80% of children deflected to the SFFC program. Finally, the higher rate of deflection and family permanence in the intervention group was not accompanied by any greater risk of repeat maltreatment.

The rigorous design used to evaluate the SFFC program can be mined *post hoc* for other specific details that can help clarify issues currently being debated over the comparative risks and benefits of the widespread practice of diverting abused and neglected children to what critics now call a “hidden” or “shadow” foster care system. These terms are used pejoratively to draw attention to the millions of children in the U.S. who are deflected from the formal foster care system primarily to the informal care of blood relatives and fictive kin. Evaluation of the SFFC program allows for a more refined analysis by narrowing the comparison to the informal

“hidden” care of children by unrelated families. As defined by Gupta-Kagan (2020): “Hidden foster care occurs when CPS agencies cause a change in a child’s physical custody without any family court action, without placing the child in the agency’s own custody, and without reporting the child’s removal to the federal government” (Gupta-Kagan, 2020: 848). Given the paucity of information about the scope and outcomes of hidden foster care, the arguments over its comparative risks and benefits are based largely on value judgments, cherry-picked anecdotes, and worse/best case scenarios. The results from this evaluation offer a fresh perspective on these issues.

Many of the envisioned advantages of SFFC are variations on themes commonly invoked to justify the use of informal kinship care as an alternative to formal foster care. Even though host families may not be motivated by the same sense of family duty and altruism that can help sustain informal kinship care arrangements, the originators of the program believe that host families draw inspiration from faith-based norms of charity and hospitality, which can function as close equivalents to the intrinsic motivations that spring from extended family bonds.

An important feature of SFFC is its unpaid quality. In addition to saving taxpayer dollars, the absence of monetary compensation removes the financial temptation to enter a host-family role out of self-interest rather than beneficence. Placing children voluntarily with host families without wresting legal custody from the parents potentially spares the parents a stressful and often drawn-out, legal-adversarial process. The end-result of the formal process can be more invasive and destructive of family integrity than the implicit “comply or else” threats, which critics allege the CPS system uses to pressure parents into turning their children over to strangers. The evidence presented in this report suggests that sharing child custody informally with a host family in times of need, even if the choice is not entirely under conditions of parents’

own choosing, may ultimately preserve family integrity better in the long run than usual judicial and CPS bureaucratic procedures.

INTRODUCTION

Safe Families for Children (SFFC) is a promising program to prevent foster care removal and the recurrence of maltreatment of children whose parents are investigated by Child Protective Services (CPS) for alleged child neglect and abuse. In order to achieve these purposes, SFFC recruits and oversees a network of unpaid host families with whom parents can voluntarily place their children in times of need. Founded in 2002 by Lydia Home Association (LYDIA), a Chicago based Christian social service agency, SFFC partners with churches, ministries, and local community organizations to offer temporary 24-hour care to children whose families are being investigate for alleged child abuse or neglect.

SFFC is currently operating in over 40 sites across the U.S. In spite of its broad appeal, the efficacy of the program has never been rigorously tested in the field under controlled experimental conditions. In 2013, the Illinois Department of Children and Families Services (IDCFS) approached the lead investigator about designing an evaluation of the program in usual practice settings, which could provide credible evidence of the program's effectiveness (or lack thereof) in preventing placement into formal foster care and the recurrence of child maltreatment.

The lead investigator submitted the design proposal to a competition, which the Coalition for Evidence-Based Policy (CEBP) sponsored, to fund low-cost, randomized controlled trials (RCTs) that seek to build actionable evidence about "what works" in U.S. social spending. Because IDCFS was able to track primary outcomes through its administrative data systems, the study could operate at low cost by forgoing primary data collection. The Department also agreed to automate an unbiased allocation mechanism (alternation) and a continuous monitoring

dashboard (proximal and distal outcomes) directly into its data system, which further reduced costs.

The CEBP competition selected the SFFC proposal as one of five finalists. Even though the study was not funded under the competition, the Laura and John Arnold Foundation (LJAF) later invited the lead investigator to resubmit the proposal through its regular funding channels. In 2015, the LJAF awarded a grant to CHSA to serve as fiscal agent for the evaluation of the program.

Results-Oriented Accountability

The SFFC project is one of several initiatives that IDCFS rolled out and evaluated under a supplemental implementation plan developed in connection with the *B.H. v. Smith* Consent Decree (*B.H. Decree*). The United States District Court for the Northern District of Illinois Eastern Division approved the implementation plan in 2016. The *B.H. Decree* is one of over 30 settlements previously or currently in effect across the U.S (Kosanovich & Joseph, 2005). These consent decrees, which a federal or state court oversees, bind the CPS agency and the attorneys acting on behalf of the “plaintiff” class to the implementation of a plan to address entrenched injustices in the protection and care of abused and neglected children in state legal custody.

The *B.H. Implementation Plan* adopts an approach to consent decree management, which Testa and Poertner (2010) refer to as “results-oriented accountability.” Other scholars have labelled analogous approaches “double-loop learning” (Argyris & Schön, 1995) and the “experimentalist approach” (Noonan, Sabel & Simon, 2009). These methods share a common orientation to consent decree management. Instead of ordering a public agency to comply with a fixed regime of solutions, the parties agree to an evidence-building and accountability process, which recognizes initiatives as provisional, reviews agency progress periodically, and conducts

rigorous experimental or quasi-experimental evaluations that either support continued rollout of the programs or call for their replacement based on the best scientific evidence (Noonan, Sabel & Simon, 2009). In this respect, the process echoes the results-oriented approach to social problem-solving, which the psychologist, Donald T. Campbell, enunciated decades ago in his treatise, *The Experimenting Society* (Campbell, 1969) as follows:

Administrators and parties must advocate the importance of the problem rather than the importance of the answer. They must advocate experimental sequences of reforms, rather than one certain cure-all, advocating Reform A with Alternative B available to try next should an honest evaluation of A prove it worthless or harmful (Campbell, 1969: 427).

Framework for Evidence Building

The *B.H.* Supplemental Implementation Plan conceives of compliance management as progressing through successive phases of “increasingly generalizable studies” (Shadish, Cook,



and Campbell, 2002) as outlined in the *Framework to Design, Test, Spread, and Sustain Effective Practice in Child Welfare* (Framework Workgroup, 2012). The U.S. Children’s Bureau funded the Framework to help guide child welfare agencies in building evidence for and spreading effective child welfare interventions.

Figure 1— The Framework

The Framework conceives of evidence-building and information accumulation as cycling through five phases or “tollgates” prior to

scaling up the program for full dissemination³ (see Figure 1). The first phase is *Identify and Explore*, which defines the scope and nature of the problem and selects promising innovations for installation based on the best available evidence of past success. The second phase is *Develop and Test*, which confirms program usability during initial implementation and conducts a formative evaluation that tests statistically whether program outputs and improvements are trending in the desired direction. The third phase is *Compare and Learn*, which supports full implementation and a summative evaluation to assess whether the intervention created practical improvements that are causally attributable to the intervention. The fourth phase, assuming evidence of effectiveness, is *Replicate and Adapt*, which spreads evidence-supported interventions (ESIs) and evaluates whether similar outcomes are reproducible with different populations across varied settings. The last phase is *Apply and Improve*, which monitors whether positive changes are sustainable and improvable over time through continuous quality improvement (CQI) and quality service reviews (QSRs).

Current Report

This report presents summary results for the SFFC initiative under *B.H*, which successfully transitioned through the first and second phases of *Identify and Explore* and *Develop and Test*. The third phase of *Compare and Learn* was recently completed, which involves full implementation and summative evaluation. In accordance with the conditions of LJAF award, the lead evaluators pre-registered the SFFC summative evaluation on the Open Science Framework (<https://osf.io/35ygk>) for the two-year enrollment period that began on January 1,

³ The phases closely parallel the four-stage model of public health intervention research (Mercy, Rosenberg, Powell, Broome & Roper 1993) and the IOM five step model of preventive intervention research (Institute of Medicine, 1994).

2017 and ended December 30, 2018 (Testa & Budde, 2017). This report present results of the summative evaluation that tracked primary and secondary child welfare outcomes using public administrative data available to the investigators through June 30, 2020. It employs the Bayesian paradigm, which is consistent with the scientific principle of phase-based evidence building in intervention research, to combine the prior information accumulated during the formative phase with the new data collected during the summative phase to provide a comprehensive analysis of SFFC's intent-to-treat impact.

SECTION ONE: IDENTIFY AND EXPLORE

The goals for the first phase of the Framework are as follows: 1) define the target population to be served and the primary problem to be solved; 2) explicate a theory of change and assess the best available research evidence in support of an intervention to effect the change; 3) develop a research question that specifies an appropriate comparator for drawing valid causal inferences; and 4) summarize the aforesaid elements into a “falsifiable logic model” that abstracts the implementation activities, program outputs, and proximal outcomes that are hypothesized as contributing to the achievement of the desired distal outcomes.

Target Population and Primary Problem

The target population for the SFFC intervention are children who are the subjects of a report of child maltreatment that has been accepted for formal investigation by CPS. The primary problem that SFFC seeks to remedy is the removal of children into state protective custody and formal foster care. The key difference between foster care removal and informal hosting is that the former requires court action that transfers children to the temporary legal custody of the CPS

agency. In contrast, informal hosting leaves legal custody with the parents on the assumption that they will place the children voluntarily with a host family until the program or CPS agency deems it safe to restore them to the physical custody of their parents.

Theory of Change

The decision to refer a family to SFFC is a form of “safety-organized practice” (Meitner & Albers, 2012). The aim is to *preserve* children in the legal custody of their parents by developing a safety plan that insures children’s safekeeping while the parents remediate the conditions that brought them to attention of CPS or the agency clears them of the allegations of maltreatment. The plan is based on a formal safety assessment conducted by CPS, which assesses whether children are at serious risk of harm and in need of an immediate change in physical custody to insure their safety. In Illinois, CPS uses the Child Endangerment Risk Assessment Protocol (CERAP; Fluke, Edwards, Bussey, Wells, & Johnson, 2001). Based on an itemized assessment of sixteen safety threats, the investigator decides whether there is “clear evidence or other cause for concern” regarding the safety of any or all of the children. If the investigators in consultation with their supervisors deem one or more of the children to be “unsafe,” protocol requires the development of a safety plan. When the plan calls for a change in the children’s living situation, CPS may elect to delegate to a third-party, such as SFFC, the primary responsibility of arranging the voluntary alternative care of the children. If the parents do not comply with the plan, the understanding is that one or more of the children may be removed from the home and placed into formal foster care.

Even though IDCFS rule and procedure declare that family cooperation with a safety plan is voluntary, some legal advocates allege that the threatened restrictions on a family’s living arrangements are inherently coercive (Redleaf, 2018; Gupta, Kagan, 2020). Parents who agree to

leave the home or place the children informally with relatives or SFFC host families are told they cannot modify the plan without risking removal of the children until they are cleared of the maltreatment allegations. The Family Defense Center (2016) argues that parents' inability to make modifications or withdraw from the agreement demonstrates the fallacy of calling safety plans "voluntary." By deflecting children from formal foster care to the informal care of relatives and unrelated host families, critics allege that CPS in effect is creating a "shadow" or "hidden" foster care system that infringes on parents' and children's fundamental right to family integrity with few meaningful due process checks on the asymmetrical power relationship between parents and the state (Redleaf, 2018). Nonetheless, the opinion of the Seventh Circuit U.S. Court of Appeals in *Dupuy v. Samuels* (2006) was that compliance with a safety plan was the result of "voluntary choices" by parents to temporarily relinquish physical custody of their child that gave them an alternative to going to court which they would not otherwise have (Gupta-Kagan, 2020).

The proliferation of asymmetrical power relationships in modern society between individual persons and collective agents, such as hospitals, police, and CPS authorities, creates new challenges for the regulation of potentially intimidating relationships through legal and policy reforms (Coleman, 1982). One approach to lessening the potentially coercive aspects is to introduce greater formality into the relationship, for example, through contracts, due process, equal rights, and the rule of law. An alternative approach is to delegate discretion to persons linked together through informal solidarities of kin relations, neighborhood organizations, faith-based communities, and voluntary associations. The crucial test of superiority is whether one alternative (informal competencies or formal organization) demonstrates greater effectiveness than the other in accomplishing the broader social purposes that both intend to fulfill.

Informal and Formal Dichotomy

The voluntary hosting of children pulls caregiving relationships in the informal direction, whereas the placement of children into licensed foster care pushes it in the formal direction. The social science literature defines informal competencies as behaviors and value patterns that evolve spontaneously and naturally from the diffuse interactions of ordinary people living and working together (Mouselis, 1968). Formal organization is governed by fixed rules and policies devised by expert and other authorities to achieve a specific social purpose.

The sociologist Talcott Parsons (1951) identified six pattern variables that mirror the informal-formal dichotomy. Of particular relevance to child welfare is the specificity versus diffuseness of the scope of responsibility that a proxy agent or collective agent exercises on behalf of the well-being of another person or class of persons (principals). For example, licensed foster care is a less diffuse and more specific principal-agent (agency) relationship than informal kinship care in the sense that the scope of foster parents' responsibilities are specific to the goals, services, and conditions set by the CPS agency. In contrast, the responsibilities felt by grandparents and other relatives span an almost limitless range of family duties, personal commitments, support obligations, and cultural traditions. SFFC falls somewhere between these two poles.

Another pattern variable that Parsons (1951) identified to distinguish formal from informal organization is universalism versus particularism. This pair, which correlates with specificity versus diffuseness, refers to the degree to which agents are expected to treat principals impartially according to a universal norm or preferentially on the basis of their particular relationship to the other person. Licensed foster parents are expected to lean in the direction of universalism, whereas relatives are allowed and even encouraged to tilt in the direction of

particularism and favoritism. The embedding of the informal norm of nepotism in the foster home licensing process, for example, permits relatives to be approved under less stringent standards for the care of a specific child. In contrast, unrelated caregivers are licensed to care for any and all children that the CPS agency deems appropriate for placement in their home. Recently laws have changed which permit CPS agencies to treat as “fictive kin” a SFFC family that previously hosted a particular child so that the child can be placed in the home under similar rules and policies that govern the placement of children with genetic relatives.⁴

It is possible to continue moving down Parson’s list of pattern variables and ordering each pair according to the informal-formal dichotomy. For example, the amount of affectivity versus affective neutrality that is appropriate or expected in a given situation is another pattern variable that distinguishes informal from formal organization. On the informal side people are permitted to express their emotions freely, whereas on the formal side they are expected to be restrained and affectively neutral toward their role partners. SFFC host families, again, fall somewhere in the middle of these two poles inasmuch as host families are expected to form a compassionate bond with the placing parents and their children. Likewise, regarding achievement versus ascription, host families choose to welcome children into their home whereas children are born into their own family. The same ordering can be done with the remaining pattern variables regarding self-interest versus collective interest and instrumental versus expressive action. For most practical applications, it is sufficient to focus attention on the two pattern variables of specificity-diffuseness and universalism-particularism. Their cross-

⁴ The Illinois law that became effective in June of 2015 also stipulated that a fictive kin with whom a child is placed shall apply for licensure as a foster family home within 6 months of the child's placement with the fictive kin. However, the law also stated that the CPS agency shall not remove a child from the home of a fictive kin on the basis that the fictive kin fails to apply for licensure within 6 months of the child's placement with the fictive kin, or fails to meet the standard for licensure. As a consequence, few homes of fictive kin become licensed.

classification forms a typology that has been used to categorize the major tensions and organizational responses that give rise to periodic cycles of child welfare reform in the United States and elsewhere (Testa, 2008). The other pattern variables tend to cluster in predictable ways in each of the four quadrants of the typology.

The more specific is the scope of public responsibility for child welfare, the more constraints there are on external interventions into autonomous family life. Removal of a child is justified in these circumstances only if the physical safety, health, and sustenance needs of the child are threatened, e.g., there are bruises, burns, malnourishment, or other visible signs of harm. Since the 1960s, public child welfare has pulled away from a specific focus on the “battered child” (Kempe et al., 1962) to a more diffuse scope of public interest in the overall well-being of the child. As the scope of public interest widened, children acquired special rights in the fiduciary relationship, which proxy and collective agents could enforce by exercising personal or class-action claims against the state.

The organizational shape that the remedies can take will vary depending on which end of formal-informal continuum ruling authorities deem appropriate for correcting the identified deficiencies. Sometimes a ruling requires the embedding of the informal competencies of everyday family life regarding care, commitment, and trust into the formal system, as the U.S. District Court required of Illinois in 1976 to favor the placement of children in the homes of relatives as the preferred placement of choice (Testa, 2020). Other times it involves rulings on the incorporation of formalities and expert systems certified elsewhere, such as child abuse pediatrics, which has been the focus of litigation in cases involving abusive head trauma (Redleaf, 2018). Whether the disruption of an existing formal system stems from embedding

informal competencies or incorporating external formalities, it can be expected that disputes will arise over which cluster of pattern variables should take precedence.

The disputes that have arisen over the hidden foster care system and related deflection programs like SFFC raise issues that can be informed by a systematic examination of the outcomes associated with the promising innovation compared to business as usual. Since the 1960s, advances in the scientific understanding of the importance of secure emotional attachments for a child's healthy social and emotional development have helped shift policy preferences away from the specificity, universalism, and affective neutrality that governed foster care in the past toward a greater acceptance of diffuseness, particularism, and affectivity as reflected in statutes that favor kinship placement, encourage diversion to informal alternative care, and promote legal permanence through adoption and guardianship over long-term foster care.⁵

The formal-informal dichotomy and its related typology of scope of public interest by locus of organization have yielded important insights when applied to the study of kinship care (Testa, 2013, 2017). However, its cognitive adequacy suffers when it is generalized beyond kinship to other sorts of informal solidarity that are rooted in friendship, neighborliness, and hospitality. Critics of deflection programs argue that when the model shifts from family-to-family agreements that are at the foundation of private alternative care to CPS-mediated, voluntary alternative care where host families become agents of CPS in implementing a safety

⁵ It is interesting to note in this regard that licensed foster parents were once counseled to be affectively neutral toward the children placed in their homes. Grandparents and other relative caregivers were of course exempt from this rule. Nowadays the appropriateness of affective neutrality, especially in the care of young children, is being reconsidered in light of new scientific knowledge. Best practice in foster care now allows for greater affectivity irrespective of kinship.

plan, the parents, who have no counsel appointed to assist, are left worse off than if formal foster care had been used at the outset (Redleaf, 2020). In subsequent sections, we evaluate the empirical adequacy of the claim that SFFC is more damaging to family integrity than SAU. Before presenting these results, we elaborate a revised theory of change that avoids the either-or solution of substituting formality or restoring informality.

Multidimensional Construct of Formality

Social theorists, such as Arthur Stinchcombe (2001), argue for setting aside the formal-informal dichotomy in favor of a multidimensional construct:

Formalities of the most various kinds can be described by the degree to which they are cognitively adequate to the situations they govern, are communicable to the people who must act in those situations, and are improvable and in fact improving (Stinchcombe, 2001: 54).

Stinchcombe's proposition can be restated as follows: the greater the degree to which a theory of change satisfies the three criteria of adequacy, communicability, and trajectory of improvement, the more likely that the action governed by the formality will work better in accomplishing its intended purposes than diffuseness, affectivity, particularism, ascription, and other pattern variables that are associated with informal everyday life.

The criteria of adequacy, communicability, and continuous quality improvement align with the phases of evidence-building depicted in Figure 1 above. Even though the phases are ordered sequentially, the progression is typically non-linear. For example, after checking the statistical validity of a logic model at the *Develop and Test* phase or after confirming its internal validity at the *Compare and Learn* phase, it is commonplace for investigators to return to the

outer loop of *Identify and Explore* to improve the model's construct validity. This step may involve some recalibration of the measurement instruments, elimination of extraneous variables, inclusion of any essential influences omitted from the model, and a re-specification of the theory of change to enhance the communicability of the model to both the providers and recipients of the intervention.

According to its operational policies and procedures manual, SFFC “hosts vulnerable children and creates extended family-like support for socially isolated families through a community of devoted volunteers motivated by compassion to keep children safe and ultimately together with their parents” (Safe Families for Children, 2018a:4). SFFC founder, David Anderson, calls SFFC a movement of the Church to care for the most vulnerable. He notes that it revitalizes the ancient practice of *philoxenia*, i.e., hospitality or “love of strangers”, which he notes remains one of the distinguishing features of Catholic teaching (Anderson, 2014). In this respect, SFFC is more closely aligned with the pattern variable of universalism than particularism. Further, SFFC views its movement as an effort both to “prevent child abuse and return the church back to the forefront of caring for children” (Safe Families for Children, 2018a: 1).

The construct validity of SFFC appeals to biblical traditions and Catholic teachings, which the Host Family Handbook (Safe Families for Children, 2018b) summarizes as follows:

- Our aim is to mobilize a movement of people who will reach out to families in a crisis.
- Our passion is to connect people willing to be a ‘spiritual extended family’ or a kind of “godparent” to those who lack such support at their point of most need.

- Our understanding is that biblical hospitality has become a lost practice but at one point was a foundational ministry of the early church. This type of hospitality is risky, demanding, and must be deliberate.

Phrasing the program’s mission in religious terms enhances its communicability to the congregations of believers who are the targets of recruitment as host families. At the same time, the religious references can sow seeds of suspicion and doubts in the minds of non-believers. Considering the adverse publicity that has dogged faith-based communities, such as the sexual abuse scandals in the Catholic Church and the Boys Scouts of America, SFFC has attracted its share of suspicions about its hidden purposes and interests. The thought of turning children over to the unlicensed care of religiously motivated caregivers raises concerns in some minds about whether children may be exposed to higher than usual risks of corporal punishment (“spare the rod, spoil the child”), home schooling, the withholding of medical care, or unconventional worshiping rituals (Melton & Anderson, 2008). While authoritative parenting can be beneficial to the healthy social and emotional development of the child (Baumarind, 2013), there is also the risk that it could veer off into unsafe directions, such as condoning harsh beatings to prevent “spoiling the child,” substituting prayer for medical treatment, or overlooking abuse by persons who occupy positions of religious authority, such as priests and ministers (Bottoms, Shaver, Goodman & Qin, 1995). Lastly, parents of different faiths and perspectives may also be concerned about the impact on their children’s own religious upbringing. Given these possible harms, many child advocates may question if it is worth the risks that potentially accompany the deflection of alleged victims from the legal protections and public accountability that state legal custody and formal foster care can afford children and families.

In spite of the above misgivings, there are many facets to the voluntary placement of children in faith-based alternative care, which are appealing to CPS frontline staff, administrators, and policymakers. SFFC alleviates some of the time pressures and uncertainty that CPS investigators typically encounter when attempting to strike a balance between the “impossible imperatives” (Duerr Berrick, 2017) of minimizing government intrusion while simultaneously protecting children from harm. Because placement into formal foster care is a surer (albeit not fail-safe) way of insulating children from repeat maltreatment, the availability of a voluntary safe alternative can inhibit the “when in doubt, yank ‘em out” reflex that can trigger removal at the early phases of child protective investigation. SFFC also helps alleviate the anxiety over making the wrong choice in maintaining the children in the family’s custody rather than taking them into foster care. Even though SFFC involves the temporary separation of children from their parents, the agreement typically lasts less than two months and 75% of parents regain full custody of their children within a year of referral to SFFC.

In the case of private (non-CPS involved) and voluntary (CPS-involved) kinship care, much of the apprehension that SFFC elicits over the risks posed by relocating children to unlicensed homes is alleviated by the belief that the children are safely nestled in an “affine agency relationship” of care, commitment, and trust (Testa 2013). An affine agency relationship is one in which a child is placed with a member of a child’s extended family, clan, or tribe, whom the family and wider community trust will act in the child’s best interest as if those interests were the member’s own. SFFC extends the range of informal placement options by mobilizing the care and support of children by voluntary host-families when kinfolk are unavailable, unable, or unwilling to lend assistance. Whether the same trust assurances should extend beyond kinship to faith-based communities most likely will vary with the degree to which

the birth family shares the same religious and other cultural traditions of the host family. It should not be forgotten that it was the distrust between the Catholic Church and the Protestant organizations that transported predominantly Catholic children on so-called “orphan trains” in the late 19th century, which prompted the construction of large Catholic institutions as the preferred “community-based” alternative to the out-of-state placement of children in foster family care.

Restated in the jargon of social science, SFFC’s contributions to the overarching outcomes of child safety, family permanence, and child and adolescent wellbeing rests on the theory that voluntary hosting by families substitutes “bridging social capital” for the “bonding social capital” that is lacking in the existing network of support (Testa, Bruhn & Helton, 2010). By involving diverse families in regular interactions of reciprocal exchange and mutual trust, the ties that develop between placing parents and host families can potentially continue well after family reunification and help lessen the hardships that many socially isolated families experience when trying to meet their family’s needs. The Host Family Handbook (SFFC, 2018) phrases this expectation as follows:

The power and uniqueness of Safe Families is the connection between the placing parent and Host Family. Because Host Families care for children without compensation, they have a unique opportunity to befriend and connect with the placing parent, like becoming part of their extended family. Our hope is that this relationship will continue after the children are returned home. This doesn’t happen all the time but it certainly is our desire (SFFC, 2018:3).

The absence of monetary compensation for SFFC host families translates into potentially huge public dollar savings. Paid foster family care and associated administrative expenses is

approximately \$85 per day. An unpublished study prepared by the Chicago-based Juvenile Protection Association (JPA; Budde, Thompson, Go, Douthitt & Pryor, 2009) reported an average length of stay of 53 days for children placed with SFFC host families in Illinois. Most of the children returned home to their birth families, but 8 percent were taken into public custody. This compares favorably to the 16 percent of indicated victims in Illinois who were taken into state custody by IDCFS during a similar time period (Fostering Court Improvement, 2013). A reduction of just one week in the average time children stay in foster care recoups the entire costs of the program if the shortened time can truly be attributed to the SFFC intervention. It was primarily the interest of the IDCFS director in obtaining credible evidence of the causal connection to potential savings, which prompted his outreach to the lead investigator. Before committing to an expansion of the state contract to support the SFFC program, he wanted a study of the promised cost savings, which adequately approximated the “counterfactual” of what might have happened to those same children who instead of being hosted by SFFC had experienced CPS and the accompanying risk of foster care placement as usual.

Lastly, the SFFC theory of change sets an expectation that the movement’s collective impact will extend beyond the personal succor offered to individual families. The hope is that the spread of good deeds and hospitality norms will help strengthen the motivational investment of religious and secular institutions in cooperative efforts to improve the general conditions of poor and vulnerable populations regardless of their particular religious traditions (Melton & Anderson, 2008). A definitive answer to this question is beyond the scope of the present study and deserves future attention. The following subsections detail the specific research question that both the formative and summative evaluations intended to answer.

Research Question

The research question that the principal investigators pre-registered on the Open Science Framework has been modified slightly to define more precisely the target population and shorten the follow-up period to conform to the outcome definitions used in the federal Child and Family Services Review.⁶ The research question is framed in **PICOTS** terms (population, intervention, comparator, outcomes, timeframe, and setting) and modifications italicized as follows:

Are Illinois children whose parents are investigated by CPS for alleged abuse and neglect *and investigators deem appropriate for SFFC (P)*:

- less likely to enter state protective custody and formal foster care (**primary Outcome**) *during the 12 months following enrollment in the study (T)*,
- just as likely to avert subsequent episodes of *indicated* abuse/neglect (**secondary Outcome 1**) *during the same follow-up period*,
- and more likely to be maintained in or reunified with their birth families *at one year after allocation to the study* (**secondary Outcome 2**),

if they are referred to SFFC's network of voluntary host families (**I**) as compared to children from statistically equivalent families who are served by CPS as usual (**C**) *in the wider community (S)*?

Even though a definitive answer to this research question is impossible because the exact same families cannot be exposed simultaneously to both the intervention and comparison conditions, it is possible to provide a high-quality approximation by ensuring that the comparison

⁶ A reviewer of a draft of this report correctly noted that the target population is actually only a subset of child subjects investigated for maltreatment. Only children in families deemed appropriate for SFFC were allocated to the comparison and intervention groups.

group is statistically equivalent, on average, to the intervention families within the bounds of chance error. This is best accomplished by using an unbiased allocation mechanism, which generates comparable treatment groups that are alike with respect to many observable and unobservable risk and protective factors except for the treatment each group receives.

Comparable Treatment Groups

The lead investigator agreed to design and implement a rigorous SFFC evaluation provided that IDCFS would agree to automate an unbiased allocation mechanism as part of its information system. In this way, a statistically equivalent group of families allocated to the intervention and comparison groups could be tracked in real time to measure differences in rates of foster care removal and repeat maltreatment. IDCFS already had experience with an in-house, automated allocation routine it programmed for an evaluation of Illinois's differential response (DR) program, which compared usual and alternative responses to selected, screened-in reports of child maltreatment (Fuller, Nieto & Zhang, 2013). The alternation routine that the Department programmed for the current study implemented the same post-randomization consent design (Adamson, Cockayne, Puffer & Torgerson, 2006; Zelen, 1979) that the DR evaluation utilized. IDCFS leadership maintained that it would be too onerous (and ethically questionable) to follow a conventional consent process of first asking alleged perpetrators whether they would participate in a research study and then randomizing them to SFFC (intervention) or CPS as usual (comparison).

Compliance with safety plans is a factor that investigators take into consideration when recommending and making removal decisions. To pledge no penalty against parents in an environment that delegates downstream decision-making to follow-up workers, attorneys, and judges, who were not parties to the original consent agreement, would be difficult if not

impossible to honor. Moreover, investigators allocate to intervention and comparison groups only those families they consider appropriate for SFFC. In this respect, the study is less like “research” and more like a “quality service review” of practice, which the Belmont Report defines as interventions that are designed solely to enhance the well-being of an individual patient or client and are believed to have a reasonable expectation of success. On the recipient side, parents are likely to agree to an intervention only if they believe their chances of retaining legal custody and regaining physical custody are better than they would be if their children were taken into formal foster care (Greenberg, Shroder & Onstott, 1999). Resentful demoralization is likely to be a greater threat to internal validity in child welfare studies than in drug trials or other medical studies where assignment to treatment is more easily concealed. Because the CPS agency cannot guarantee participants that they would suffer no penalty or loss of benefits to which they were otherwise entitled if the participant discontinued participation in the safety plan, they opted for the post-randomization consent design over the conventional, pre-randomization consent design. The lead investigator submitted the study proposal to the IRBs at IDCFS and the University of North Carolina at Chapel Hill and later joined the application to the University of Illinois at Urbana-Champaign. All three IRBs approved the study.

A post-randomization consent design typically results in a lower rate of participation in the intended treatment than a pre-randomization consent design. Holland (1988) refers to this type of study as a “randomized encouragement design.” It is a fully valid type of RCT which offers persons who are pre-allocated to the intervention the option of participating or not participating in the assigned treatment. Families pre-allocated to the comparison group receive treatment as usual, which could result in any of the following actions: dismissal of the allegation or, if indicated for maltreatment, closure of the case without further services, protective removal

of the child for 48 hours, the option of in-home family services, or formal placement of the child into foster care for an unspecified duration. The SFFC option was not offered to the comparison group. In this way, they were spared the disappointment of being denied services that developers believed improved their chances of holding on to their children.

Families pre-allocated to the intervention group were exposed to all of the usual treatment options plus the choice of participating in a SFFC host-family arrangement. The offer of a promising innovation of unknown efficacy to families without their first consenting to participation in the program more closely approximates how innovations are rolled-out in the real world, which doesn't require informed consent from families before offering or providing services. The downside of pre-allocation, however, is that incomplete compliance is more likely than with a conventional randomized design that limits participation to persons who signal beforehand their willingness to comply prior to randomization (Zelen, 1979). At the same time, full compliance cannot be guaranteed even when participants are randomized after giving their informed consent to participate in the study. Full compliance with assigned treatments cannot ethically be compelled in either design. To preserve the statistical advantages of randomization, therefore, it is necessary to conduct a so-called "intent-to-treat" (ITT) analysis in which the outcomes for all subjects allocated to intervention and comparison groups are tracked and compared regardless of whether the families fully comply with their assigned treatments. Even though ITT analysis results in a "diluted" estimate of the treatment effect by pooling the outcomes of the fully compliant with the outcomes of those who drop out of treatment or refuse participation in the study, it does provide an externally valid estimate of the net benefit the public can expect when rolling-out a promising innovation in usual care or practice settings. The Department's primary interest was learning whether SFFC was working in general and not

specifically for a possibly unrepresentative and compliant subsample of families under artificial experimental conditions. Instead, its goal was to learn of the extent of child protective need for safe alternatives to removal, measure the take-up of the offer, and ascertain the effect of the offer of participation in SFFC on the deflection of children from formal foster care and the prevention of the recurrence of child maltreatment.

In the interests of concealing the allocation process and discouraging staff from tampering with assignments, the IDCFS computer programmers installed a “behind the scenes,” computerized routine for assigning families. Even though the routine was billed within the agency as a “randomizer,” it was actually a binary, yes-no alternation routine that assigned every other referral to the comparison or intervention arm of the experiment. The process of forming study groups through alternation is a subset of an array of unbiased assignment methods, which include allocating subjects according to the day of the week, date of birth, medical record number, or the order in which subjects show up for services. One of the earliest trials that historians register as an RCT was actually not a randomized experiment but an unbiased assignment routine implemented in the late 19th century by the Danish Nobel laureate, Johannes Fibiger (Hróbjartsson, Gøtzsche & Gluud, 1998). Subject allocation depended on the day of admittance. New patients admitted on alternate days received either the standard or experimental treatment.

Alternation continued to be the gold standard for medical research until the late 1940s, when it was superseded by allocation using random numbers (Chalmers, 2001). The reason that randomization superseded alternation is not because of its statistical superiority but because a randomized sequence is more easily concealed from tampering than alternation. Properly implemented, both alternation and randomization satisfy the two essential conditions of unbiased

allocation: 1) generation of an unbiased sequence, and 2) unbiased implementation of the sequence. Even though alternation poses a greater threat to the second condition than randomization (i.e., knowledge of the prior assignment potentially biases the selection of subsequent allocations), the alternation sequence was effectively concealed in the SFFC study. Only supervisors had the security clearance to activate the “randomizer button,” and they were geographically dispersed across the state. Referrals to SFFC were relatively rare occurrences (less than two in any single office per week). Any concerted effort to manipulate the selection process by stacking up potential SFFC referrals to favor assignment to the intervention group would have required much forethought and coordination.⁷

Disposition

Unqualified Report Other Written Reports Harassment

Investigation Retention Code: 5 years

Level Of Intervention:

Case(s) Created: Intact Family Services - Tier 1

Family Recommended For Safe Family Program

Does family agree to participate in Safe Family Program?: Yes No

Handoff Meeting Date/Time: [Generate Handoff Document](#)

Figure 2 displays a snapshot of the randomizer screen, which indicated whether the investigator was authorized to proceed with asking the family whether they want to participate in the SFFC program. After speaking with the family, the investigator checked whether

Figure 2—IDCFS Automated Alternation Routine

the parents agreed or did not agree to participate. An answer to this question had to be provided in order for the investigator to close out the investigation. If the family was assigned to the comparison group, the second question about the family’s agreement to participate in SFFC did not pop up.

⁷ It quickly became evident during initial implementation that concealment must be working because IDCFS staff routinely complained that all of their families were being assigned to the “control” group even though the weekly runs showed that exactly half of the referrals had been assigned to the intervention group and the other half to the comparison group.

Falsifiable Logic Model

After clearly articulating the problem to be solved, identifying the appropriate target population, explicating a theory of change based on the best available practice and research evidence, and developing a PICOTS question that specifies an appropriate comparator for drawing valid causal inferences, the last step in the *Identify and Explore* phase is to assemble the pieces into a “falsifiable logic model” (Epstein & Klerman, 2012).

A logic model is a planning document that translates purposes – proximal, intermediate, and distal outcomes—and abstracts key implementation activities and program outputs into a formal document for reaching those goals. Much like a blueprint, a logic model is a graphical presentation of overall plans, preferably one-page, which describes some important features of the program to be implemented and specifies the hypothesized causal linkages among program components. While there are many templates available in the literature, not surprisingly we used the one developed by Testa (2010) as described in his book, *Fostering Accountability: Using Evidence to Guide and Improve Child Welfare Policy*.

At a minimum, a logic model should satisfy the three criteria of adequacy, communicability, and continuous quality improvement, which Stinchcombe (2001) identified as crucial for proper implementation of the actions governed by the logic model. Pictorially, the model should summarize the connections among the following components: 1) population that is the target of the intervention; 2) resources or interventions for achieving the desired outputs and outcomes, including the comparator that will be used to draw causal inferences; 3) program elements including the manual, staffing and training requirements, coaching and supervisory activities, and support for ensuring adherence to the program model (fidelity) and best practice standards; 4) outputs and deliverables of program activities; 5) proximal, intermediate, side-

effects, and distal outcomes indexing the desirable and undesirable changes in people and systems; 6) exogenous historical, political, cultural, and social factors that influence the readiness and capacity of the organization to undertake program activities; 7) theory of change based on research, best practices, and experience, which describes the underlying causal mechanisms and assumptions about why the program is expected to work; and 8) the general end-values under which specific outcomes can be classified, such as equity, efficiency, economic benefit, freedom, voice, subjective well-being, group solidarity and social integration.

The inclusion of “falsifiable” in front of logic model, as Epstein and Klein (2012) explain it, is meant to draw attention to the importance of taking continuous quality readings of the goodness of fit between what is supposed to happen according to the logic model and what is actually happening on the ground. This is especially important at the *Develop and Test* phase of initial implementation and formative evaluation. Early warnings about the difficulties a program may be having in passing its own logic model should give pause about proceeding to the *Compare and Learn* phase of full implementation and summative evaluation. Whereas conventional logic models also specify a sequence of steps, a falsifiable logic model includes considerably more implementation detail with specific quantitative and qualitative benchmarks for monitoring the achievement of program outputs and proximal outcomes. It is important for a promising innovation to pass usability testing and hit the desired benchmarks at initial implementation before progressing to full implementation and summative evaluation.

Figure 3 elaborates the PICOTS question into a SFFC logic model. It illustrates the hypothesized causal pathways by which SFFC is expected to translate into measurable outputs that effect the desired outcomes for a target population of interest. Two of the major steps for constructing a falsifiable logic model are: 1) defining the scope of activities and mechanisms for

Figure 3 -- Logic Model for Safe Families for Children

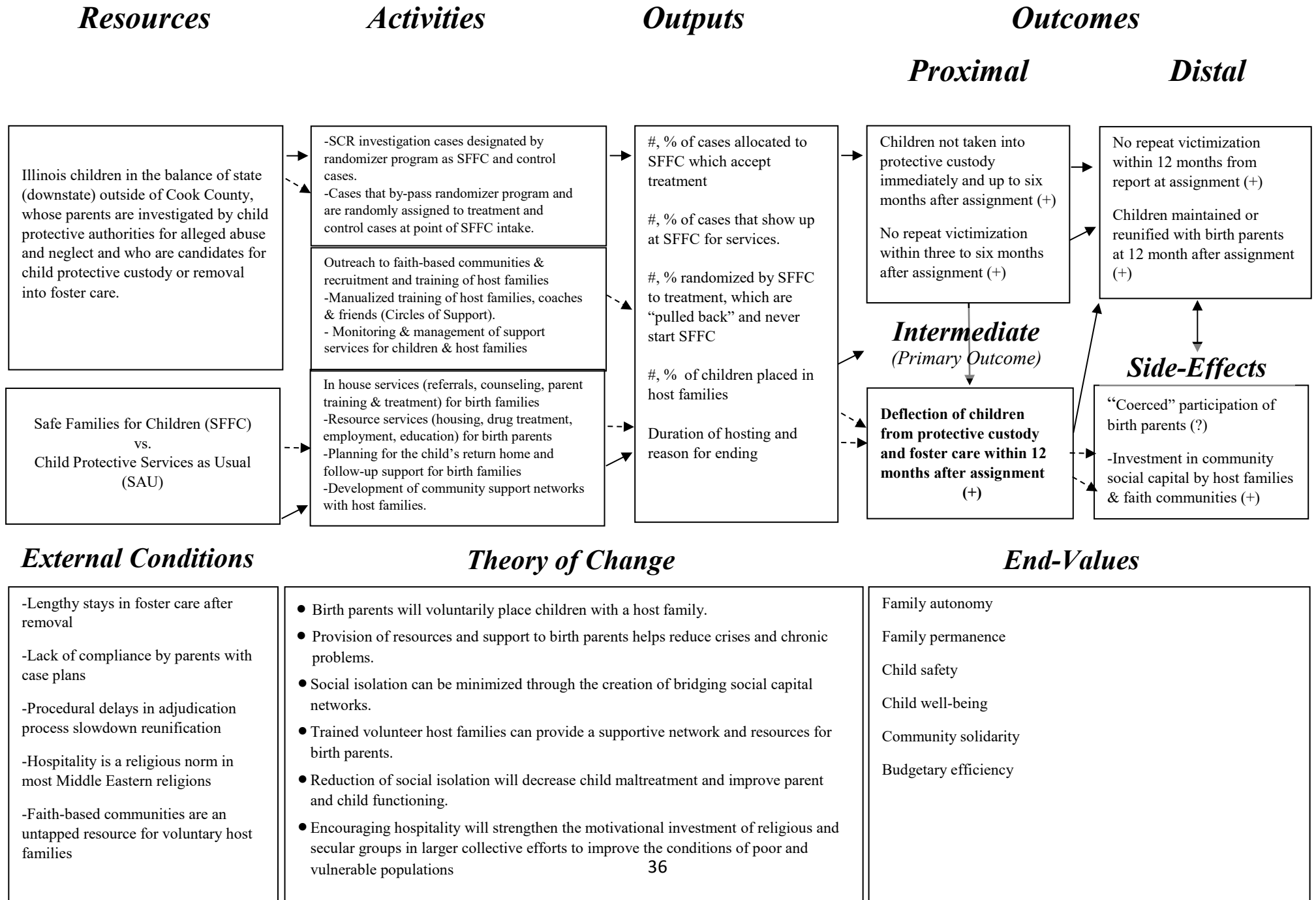


Table 1. —Core SFFC program assumptions, outcomes, and measures

Core Program Assumptions	Outcome	Measure
<i>Foster care deflection:</i> SFFC provides a safe alternative to child welfare custody, which can significantly reduce the number of children entering the formal foster care system.	<i>Deflection from formal foster care (primary):</i> Prevention of removal of child from the home and placement into formal foster care.	Among child subjects investigated for maltreatment, % taken into protective custody or later removed into foster care from day 1 to 12 months after randomization.
<i>Child abuse prevention:</i> Providing resource-poor parents with a safe, temporary place for children without threat of losing legal custody helps avert subsequent abuse/neglect episodes.	<i>Repeat victims within 12 months of prior indicated report (secondary):</i> Re-victimization of children within 12 months from the date of the initial report at assignment.	Of all children who were victims of an indicated maltreatment report, the percentage who were victims of another indicated maltreatment report within 12 months of the date of the index report at assignment.
<i>Family support and stabilization:</i> Many SFC host families become the “fictive” extended family support that a parent needs, which helps birth parents maintain full legal custody or quickly regain physical custody of their children.	<i>Children maintained or reunified with birth parents at 12 months after assignment (secondary):</i> Maintenance of a child or reunification with the birth or extended family at 12 months after assignment.	Among child subjects investigated for child maltreatment, % who were maintained in the custody of their parents or returned to their physical custody at 12 months after assignment.

achieving change; and 2) enhancing the communicability of the standards, protocols, outputs and proximal outcomes to downstream agents that share responsibility for achieving the desired change.

Table 1 lists the primary and secondary outcomes identified in the logic model, including the operational measures and core program assumptions made about the superiority of the SFFC intervention compared to treatment as usual. Even though the pre-registered analysis plan called for tracking outcomes for a period of 24 months, the follow-up period was shortened to 12

months to conform to the definitions used in the federal Child and Family Services Review (CFSR).

Foster Care Deflection: The primary outcome, which was registered on *OSF.io*, is the time to the event of protective custody or court-sanctioned removal of the child into public foster care. In an emergency, IDCFS investigators have the authority to remove children from the physical custody of their parents without a court order or warrant. However, protective custody lapses within 48 hours of removal unless a State’s Attorney approves the filing of a petition with the juvenile court to retain the child in state custody.

Child maltreatment prevention: Each new episode of alleged child maltreatment receives a State Central Registry (SCR) number and sequence letter (A thru ZZ), which uniquely identify each new household configuration and the history of prior reports (starting with A) for that specific household configuration. In addition to the SCR number each family member receives a person-specific identifier, which enables the tracking of subsequent reports of maltreatment and findings of indicated maltreatment for each child regardless of household changes.

Family support and stabilization: The theory of change, which underlies the SFFC intervention, is that host families will become the “functional equivalent” of a supportive social network that is currently absent from a parent’s life. By offering the prospect of an ongoing supportive relationship beyond the date that SFFC returns the child to parental custody, it is hoped that birth parents will retain full custody of their children, which reduces the overall time children spend in foster care. Because the families allocated to the different treatment groups are nearly identically balanced in terms of the date the families were allocated to the intervention and comparison groups, the outcome can be measured by ascertaining the whereabouts of the

children (e.g. never removed, reunified back with their parents, or still in foster care) 12 months after their enrollment in the study.⁸

SECTION TWO: DEVELOP AND TEST

This section of the report presents the findings from the *Develop and Test* phase of evidence building, which confirms program usability during initial implementation and conducts a formative evaluation that tests statistically whether program outputs and improvements are trending in the desired direction. The goal for usability testing is to assess whether the unbiased allocation mechanism is reliable and the implementation and data collection activities are stable enough for initial implementation and formative evaluation. Usability testing of the installation of the new referral procedures was originally scheduled to last only three quarters to give time for SFFC and CPS leadership to train new investigators in the philosophy of SFFC and to learn how to access the program through the use of the randomizer. Lower than desired quarterly referral volume, however, required the extension of usability testing for an additional two quarters.

The purposes of usability testing are to check the goodness of fit (construct validity) between what is supposed to happen according to the falsifiable logic model and what is actually happening on the ground and to correct for any deviations before initiating the next phase of

⁸ As noted in the introduction, SFFC is currently operating in over 40 sites across the U.S. Despite its broad appeal, the efficacy of the program has never been rigorously tested. Although several descriptive studies have been done, which support the potential of the program as promising, the level of evidence mustered in these studies doesn't meet the federal standard required to receive a designation of "supported by research evidence." In order to receive this designation, at least one rigorous RCT in a usual care or practice setting must show the practice to be superior to an appropriate comparison practice. In addition, one of the RCTs must show that the practice had a sustained effect of at least six months beyond the end of treatment when compared to the comparison group.

study. After correcting any problems that a program may be experiencing in passing its own logic model, the next phase involves testing the extent to which observed improvements during initial implementation are associated with assignment to the intervention and receipt of the intended treatment. The next several sections present findings in the phased order in which the

Table 2.—Allocation of Families to SFFC Intervention and Comparison Groups for All Three Phases of the Study, 2014-2018

Quarter	Allocation		Total
	Comparison	Intervention	
Usability (Cook & Northern Region)			
2014 3	1	1	2
2014 4	3	4	7
2015 1	4	3	7
2015 2	3	4	7
2015 3	2	1	3
Subtotal	13	13	26
Formative (Cook & Downstate Regions)			
2015 4	2	2	4
2016 1	14	20	34
2016 2	52	55	107
2016 3	48	51	99
2016 4	35	44	79
Subtotal	151	172	323
Summative (Downstate Regions Only)			
2017 1	5	11	16
2017 2	14	9	23
2017 3	7	6	13
2017 4	1	5	6
2018 1	1	2	3
2018 2	10	10	20
2018 3	6	10	16
2018 4	6	4	10
Subtotal	50	57	107
TOTAL	214	242	456

analyses were conducted: *usability testing* of referral and allocation methods for cases assigned during the five quarters prior to October 1, 2015 (n = 26 families); *formative evaluation* of program outputs and proximal outcomes for the five quarters prior to January 1, 2017 (n = 323 families); and *summative evaluation* of primary and distal outcomes during the *Compare and Learn* phase for the eight quarters ending December 31, 2018 (n = 107). Table 2 presents the quarterly allocation of cases by assignment group for all three phases of the study.

Usability Testing

During initial implementation, the lead evaluator tested the usability of the automated alternation mechanism and several

key reporting procedures. During usability testing, the number of times that CPS supervisors

activated the assignment link fell well below the desired target goal of 120 allocations per quarter. CPS leadership took affirmative steps to encourage investigators to make greater use of the SFFC hosting option. In February of 2016, IDCFS agreed to include SFFC as one of the initiatives in the *B.H.* Supplemental Implementation Plan that the United States District Court for the Northern District of Illinois Eastern Division approved in October of that year.

One of the reasons for the low volume of referrals, it was later learned, was that many CPS investigators by-passed the automated allocation process and continued to make referrals directly to SFFC as had been the practice in the past. Even though the change in procedure was widely publicized, the evaluators in consultation with LYDIA and IDCFS decided against forcing all traffic through the automated process. Instead, they instituted a back-up randomization routine at the point of direct contact between families, investigators, and the SFFC intake staff. The changes made to accommodate the addition of another allocation mechanism are detailed in the sub-section below on the findings from the formative evaluation.

The next test focused on the integrity of record linkages across administrative data systems. The combination of SCR number and sequence letter uniquely identifies each new episode of alleged child maltreatment at the household level. This unique identifier links to other administrative data, which enables tracking the history of the case from the date of the initial report to the dates of subsequent events, such as protective custody, removal into foster care, and subsequent reports of maltreatment.

Usability testing demonstrated that it was a simple and low-cost task to link the families allocated to the SFFC study to IDCFS administrative data. This permitted making linkages to baseline characteristics to check the adequacy of the alternation method in creating statistically equivalent intervention and comparison groups. The alternation routine evenly divided the 26

families referred during usability testing to the intervention (n = 13) and the comparison (n= 13) groups. Even though a sample of 26 is much too small for conducting valid statistical tests, there were no practically important differences between the two groups with regard to available baseline variables, with the exception of geographical location. Date of report, number of prior reports, investigation finding, and the percentage of unfounded reports that were expunged from the abuse registry were all evenly balanced between the comparison and intervention groups. The geographical imbalance (Comparison: 23% Cook vs. Intervention: 54% Cook) was chalked up to a statistical fluke that could be expected to disappear with a larger sample size.

The second purpose of usability testing is to check the assumptions made for power analysis and the algorithms developed for longitudinally tracking the primary and secondary outcomes that were to be used to evaluate the efficacy of the SFFC program. During the formative evaluation period, approximately 3% of all CPS investigations initiated in Illinois began with police, physicians, or CPS investigators' taking children immediately into protective custody whom they assessed to be at imminent risk of harm. IDCFs administrative data showed that the CPS agencies allowed 6% of protective custodies to lapse without petitioning the court for adjudication of wardship. Another 6% of protective custodies lapsed because the State's Attorney rejected the filing of a petition or the Judge denied the State's petition to grant temporary custody to the CPS guardianship administrator. For the remaining 88% of protective custodies, the Court's granted temporary custody to the CPS agency (84%) or continued the hearing to a later date (4%).

In the majority of cases, however, CPS investigators refrained from taking temporary custody. Instead, the alleged victims of maltreatment stayed in the custody of their parents so long as the results of the CERAP indicated it was safe for the children to stay while the

investigator collected additional information to determine whether there was credible evidence that maltreatment had occurred. If the investigator instead assessed the children to be unsafe and successfully petitioned for their removal into state custody, the children and family receive a unique case identifier that facilitates tracking each child's subsequent episodes in foster care and whether the children are eventually reunified with their parents, are adopted, discharged to the legal guardianship of a relative, or retained in long-term foster care.

Power Analysis

During proposal development, the lead investigator conducted a power analysis, which posited a comparison removal rate of 16% versus an intervention removal rate of 8%. The expected removal rate for the comparison group was based on 2012 data from UNC's Fostering Court Improvement (FCI) website (http://fosteringcourtimprovement.org/state_websites.php). The anticipated removal rate for the intervention group was based on a previous study conducted for SFFC by the Juvenile Protective Association (Budde, Thompson, G, Douthitt & Pryor, 2009).

Over the five quarters of usability testing, 15% of the comparison group (4 of 26 children) were removed and placed in protective custody or foster care compared to 11% of the intervention group (2 of 19 children). Even though the usability sample size of 45 is too small to draw statistically valid conclusions about expected differences, the estimated removal rates were close enough to the 16% and 8% estimates for the comparison and intervention groups, respectively, that the lead investigator decided to retain the original estimates based on the FCI data for the power analysis.

Assuming a 16% removal rate for the comparison group and projecting an 8% for the intervention group, the odds ratio (effect size) for the intervention effect is 0.46. For the analysis

plan registered on the OSF, we rounded the effect size estimate up to 0.5. That is, the intervention is projected to reduce the odds of removal by one-half. An effect of this magnitude is considered between a small and medium *effect size* (Chen, Cohen & Chen, 2010). Expressed in positive terms as deflection from foster care, the inverse of the expected odds ratio is 2.0. To reject a null hypothesis that the true odds ratio is 1 (i.e., 50:50) rather than 2.00 at the 0.05 (1-tailed) alpha level with 90% power, requires the enrollment of 375 families in each arm of the study (intervention and comparison groups). Because outcomes are measured at the individual child level and children are clustered within households, this study requires the same type of power analysis as a cluster-randomized trial (Boutron, Moher, Altman, Schulz, Ravaud, 2008). A ratio of 1.7 children per report was assumed based on the ratio of 45 children who were linked as subjects of the maltreatment report in the 26 usability cases.

Reaching a sample size of 375 treated families and 638 children would require assigning 470 family units to each arm of the study on the assumption that 80% of families in the intervention group would fully comply with their assigned treatment of SFFC hosting. Whereas a total sample size of 940 families is optimal for detecting a statistically significant effect, strictly speaking, formative evaluation does not require reaching this target goal. A smaller sized sample could suffice even though the thresholds of type I errors (i.e., concluding that the intervention is effective when it is not: *false-positives*) and type II (i.e. concluding that it is not effective when it is: *false-negatives*) would be higher than desired using conventional significance levels. The goal of formative evaluation is *not* to render a summary judgment of the statistical significance of an observed effect size but to get a mid-course reading on whether it is wise to proceed to summative evaluation where statistical conclusion validity becomes more important. For purposes of formative evaluation, therefore, the study set the thresholds of type I and type II

errors at 0.15 and 0.80, respectively. Failure to show evidence of expected differences at these weaker thresholds of significance should give pause about proceeding to the next phase of full implementation and summative evaluation.

Referral Challenges

Nearly all efforts at social innovation underestimate the difficulties in reaching their intended target populations. The same was true for this study. IDCFS and LYDIA administrators did not foresee that it would be much of a challenge in hitting the desired sample targets. SFFC data showed that an average of 60 families and about 100 children per quarter had previously been served by SFFC host families between 2013 and 2014. Doubling the level of participation to 120 families per quarter would yield the desired number of family referrals within 24 months. Before proceeding to full implementation and summative evaluation, however, it was important first to test the willingness of workers and supervisors to press the randomizer button. The early returns during usability testing were far from encouraging.

During the first three quarters from October 1, 2014 to June 30, 2015, IDCFS supervisors activated the SFFC allocation routine 16 times. Even though the small number of allocations was adequate for usability testing (i.e., there was an even split of 8 cases to the intervention group and 8 to the comparison group), it was an early warning sign that doubling the alternation volume to 120 families per quarter was going to be much more of a challenge than IDCFS and LYDIA had originally anticipated. Therefore, usability testing was extended for another two quarters to give time to test various approaches to boosting the use of the SFFC allocation routine.

At the end of usability testing, only 13 cases in total had been assigned to the SFFC intervention and 13 to the comparison condition of CPS as usual. A much larger than usual fraction of usability cases that CPS supervisors deemed appropriate for SFFC were subsequently indicated (substantiated) for child maltreatment, 73%, versus only 28% for all investigations conducted in Cook County and the northern counties of Illinois during this same time period. The especially high proportion of anticipated indicated reports, which correlates positively with multiple subsequent reports of maltreatment, suggested that CPS workers and supervisors were accurately honing in on the selective population of children most in need of hosting by SFFC. On the other hand, a larger proportion of hosted cases in Cook County exceeded the 60-day limit for making a final determination of maltreatment than other cases investigated in Cook during the same study period (28% vs. 18%). This disparity suggests another less laudable reason for the higher than usual indication rates. From discussions with workers, it appeared that some IDCFS investigators, especially in Cook County, were using SFFC as a way-station for separating children from their parents while they gathered more conclusive evidence to justify removing the children into formal foster care. As noted later in the Discussion section, this alternative use of SFFC as a temporary shelter rather than a deflection program deviated from the underlying theory of change as to how the program was supposed to reduce removal rates. If left uncorrected, SFFC could end up having the opposite effect to the intended one of preserving family integrity.

The referral problem notwithstanding, usability testing demonstrated that the automated allocation mechanism and reporting procedures were functioning as intended. The 26 usability cases linked seamlessly to 45 children who were involved in the maltreatment report. The demographic characteristics of the children in families deemed appropriate for SFFC matched

the expectations of program developers. The average age of children was 5.6 years old. Aggregating child characteristics up to the family level showed that most of the referrals involved only one child (54%) and 4% involved larger families of four or more children. The age profile of one-child families and sibling groups showed most were no older than seven years of age (64%). A small percentage of cases included children in their teenaged years (11%). In addition, 56% were male, and 60% were African-American. The child-level distributions for report sequence, county of referral, and indicated for maltreatment were similar to the distributions for allocated family cases

The 45 children linked to a total of 78 caretakers, of which 62 (79%) were birth parents, 7 (9%) were kinship caregivers, and the remaining 9 (12%) were other adults in the home. Nearly one-half (47%) of the children resided in single-parent homes, and 27% resided in two-parent families. The remaining 26% lived in a variety of extended family and multiple-adult households. In order to learn more about the reasons children were reported for maltreatment, we linked caretaker data to the specific allegations that instigated maltreatment reports on each involved child. The children linked to 75 separate allegations of maltreatment: 41% of the allegations involved inadequate shelter or environmental neglect, 31% involved substantial risk of abuse, 19% inadequate supervision, and 9% involved a variety of miscellaneous harms, including burns, bone fractures, and sexual abuse.

Goodness of Fit with Theory of Change

An in-depth qualitative review of 20 non-expunged cases (35 children) conducted with the usability sample offered an opportunity to check how well case narratives lined-up with some of the key assumptions specified in the logic model (see Figure 3 above). One of the foundational assumptions is that chronic stress experienced by families can be mitigated at the early stages of

CPS intervention through the creation of “bridging social capital” that links socially isolated families to a wider supportive network of volunteer host families, coaches, and friends. The key assumptions that are outlined in the logic model are listed here as follows:

- Birth parents will voluntarily place children with a host family.
- Provision of resources and support to birth parents helps reduce crises and chronic problems.
- Social isolation can be minimized through the creation of bridging social capital networks.
- Trained volunteer host families can provide a supportive network and resources for birth parents.
- Reduction of social isolation will decrease child maltreatment and improve parent and child functioning.
- Encouraging hospitality will strengthen the motivational investment of religious and secular groups in larger collective efforts to improve the conditions of poor and vulnerable populations

To assess the plausibility of these assumptions, the investigators mined electronic case notes for instances where case workers mentioned the Safe Families program. During usability testing, IDCFS staff recorded that five (39%) out of the 13 families assigned to the intervention group accepted the offer of temporary hosting of children. SFFC staff later recorded all five as participating in the program in their case notes. Even though the observed compliance rate of 39% was only one-half of the expected 80% rate used in the power analysis, it was agreed that no special efforts should be undertaken to boost compliance rates in order to insure that participation was voluntary and not “coerced” (see Side-Effects in Figure 3).

Validating the assumptions about bridging social capital’s supplementing the dearth of bonding social capital in the family’s current support network involved a deeper dive into the electronic case notes. Out of the five intervention cases that accepted the hosting offer, the electronic case notes for two of the cases were expunged from the accessible database and hence concealed from re-analysis. In all three of the remaining cases, there was accessible evidence in the case notes for gaging the construct validity of the theory of change. The case notes demonstrate that prior to or concurrent with the referral to SFFC, there were regular inquiries

from the CPS investigators about the nearby availability of kin to lend a hand. Below are the relevant passage from the three sets of accessible case notes:

- CASE 1: The worker asked the boyfriend and his mother if the children's mother had any family that could keep her son, while she works on her mental health issues. They stated that all of her family are in Florida or New York.
- CASE 2: A relative of the family agreed to call additional family members and report back with the results. She called back stating all of the family members she knew were contacted and none were willing to help the mother with her children.
- CASE 3: Investigator talked to the maternal grandfather of child. He informed this worker that he and his wife cannot take another child into their home. He informed this worker that he and his daughter do not get along, even though he and his wife are the legal guardians of her 4 year-old son.

The average hosting duration of the five cases that approved the referral to SFFC (including the two case with concealed notes) was 68 days. The shortest stay was 6 days and the longest lasted 228 days. Even among the families with brief associations, many of the families continued to stay in touch after the hosting arrangements ended. Below are some relevant passages from the three cases with accessible case notes:

- CASE 1: In September, the mother confided to this worker that she was feeling overwhelmed and stressed over the weekend. So she called the SFFC host family who had looked after her son last winter and asked if they could watch him for a couple days. The family picked him up on Saturday morning and the mother picked him back up on Monday morning. She said she felt comfortable reaching out to them.
- CASE 2: At the end of February, the CPS worker met with the child in his new SFFC home. The parents requested that their child stay in the SFFC home for about two months due to the fact that they do not have a stable home for the child. In March, the birth parents visited with the child and host family at McDonald's. The CPS worker wrote that the boy appeared bonded to the host mother. As the SFFC hosting arrangement was nearing its end, the child was transferred with his mother to a parent and child wing at a drug recovery center for a 90-day treatment. The mother stated her SFFC partners were in regular contact and were willing to help in any possible way. Postscript: CPS eventually took legal custody of the child in November of the following year and placed him in the home of his former SFFC family as a formal fictive kin (HFK) placement.
- CASE 3: The CPS Investigator arrives at home to help mom prepare for the SFFC host to come and get the children. Mother stated that she understood she needed to seek housing for her and the children. Mom was informed that if she could not locate a resource by Monday that the SFFC agency could help extend the hosting arrangement for the kids. Hosting lasted for 6 days. NOTE: No further mention of SFFC is recorded after hosting ended.

The above passages are unprompted observations made by CPS workers in administrative records of how families maintained or did not maintain social ties after the hosting period ended.

The recordings are consistent with the developers' theory of change: SFFC has the potential to mobilize a self-sustaining, voluntary network of support and resources which fragile families can draw upon in times of need. The lack of ongoing contact for a hosting arrangement that lasted only six days is not surprising. The fact that the bridging social capital created through SFFC hosting can later transmute into the bonding social capital of fictive kinship, as illustrated in the CASE 2 above, suggests another positive side-effect of the program.

Steps Taken to Boost Referral Volume

Even though usability testing indicated that the alternation method of allocation was producing statistically equivalent groups and the program's theory of change appeared to have sufficient construct validity to warrant proceeding to initial implementation, the low frequency of supervisors' use of the randomizer hyperlink continued to be a major worry as the usability testing period wound down. The Department had taken a number of deliberate steps to increase the use of the hyperlink. These included: personal conversations with each area administrator in Cook and Northern regions to discuss the demonstration and explore possible barriers or alternative plans to reach the desired levels of referrals. The Department sent e-mails to all child protection supervisors in Cook and Northern regions to remind them of the requirement to consider SFFC families as a resource for children and to utilize the hyperlink. In spite of these steps, the hyperlink was activated only a handful of times as the study prepared to enter the *Develop and Test* phase of initial implementation and formative evaluation.

It had been anticipated that more than half of the expected SFFC referrals would originate from the state's largest population hub of Cook County (which includes the city of Chicago). In the end, however, only 35% of referrals came from Cook. In order to increase the volume, SFFC and IDCFS opened up participation to all regions of the state during formative evaluation. It was

also during this time, that SFFC staff informed the evaluators of several protocol deviations that they were observing across the state but mainly in Cook County. The first deviation was the direct referrals from intact family workers, which were not part of the original study protocol. Intact family workers served families in which the assessment of risk to the children were not severe enough to warrant removal of the children at the time of case opening. Because these cases were not subject to the IDCFS alternation routine, they should not have been randomized by SFFC. The next deviation concerned the sizable uptick in direct referrals from CPS investigators, which by-passed the automated allocation routine. Rather than clamping down on direct referrals, which risked further dampening practitioners' enthusiasm for assigning families to the program, the leadership team decided to install a back-up randomization procedure at SFFC intake. The team briefly explored the possibility of granting SFFC staff direct access to the IDCFS "randomizer button," but privacy concerns precluded giving blanket privileges to non-state employees. As an alternative, the investigators developed a back-up randomization procedure for referrals when workers phoned in directly to SFFC. The choice of going with a true randomization procedure rather than mimicking the alternation procedure used by IDCFS was motivated by the greater difficulty with concealing an alternated sequence of cases from SFFC staff compared to a randomized sequence. This was a particular concern in light of the fact that the organization responsible for randomization was also the same agency delivering the intervention.

The back-up randomization routine functioned as follows: SFFC intake staff manually entered referrals that by-passed the IDCFS alternation procedure into a spreadsheet, which ran a macro that assigned families to the intervention and comparison groups based on a random sequence of assignment categories. After assignment, SFFC management added additional case

information to the spreadsheet, which, among other fields, listed referral date and the IDCFS state central registry (SCR) number and the unique sequence letter from A (initial report) to UX (the highest sequence of subsequent reports registered). This required manual look-ups in the IDCFS information system, which increased the chances of data-entry errors that the IDCFS automated allocation process avoided by linking the allocation record to all administrative information immediately upon assignment. Because the record linking information had to be entered manually by SFFC staff, the evaluators lost the ability to track proximal outcomes in real time for direct SFFC referrals until they could download the assignment spreadsheet, which the Department posted to the Department's SharePoint site at the beginning of each month.

In addition to the back-up randomization procedure, IDCFS reaffirmed its commitment to the use of the SFFC program by incorporating the following provision into the Supplementary Implementation Plan that IDCFS and plaintiffs filed in February of 2016 with the United States District Court for the Northern District of Illinois Eastern Division under the *B.H. vs. Sheldon* Consent Decree:

Under SAFE Families for Children (SFFC), IDCFS assists families in need with services to protect children and support keeping families together. SFFC places children at risk of removal in vetted volunteer families to avoid their placement into foster care. SFFC strives to meet three objectives: child welfare deflection, child abuse prevention, and family support and stabilization. SFFC has been in operation in Cook and Northern regions of Illinois for thirteen years. Due to a grant from the Arnold Foundation, SFFC was recently expanded state-wide to provide services to children and to evaluate the program. Challenges with the roll-out of the program evaluation include:

- Lack of anticipated participation by workers and identified candidates given limited education about the benefits of SAFE Families as well as various case issues related to the SAFE Families model.
- Reluctance of workers to refer children to SAFE Families out of concern that a child would be assigned to the control group and not to SAFE Families.

The Strategic Planning team will drive the continued progress of this initiative by breaking down barriers to success.

To give SFFC and IDCFS sufficient time to address these limitations, LJAF agreed to the postponement of formative evaluation for an additional two quarters. It also agreed to cut the target goal of 120 families in half to 60 families per quarter. The lead investigator put SFFC and IDCFS on notice that if substantial progress towards increasing the volume of referrals was not made by the end of the 3rd quarter of 2016, proceeding to summative evaluation might be in jeopardy.

Initial Implementation and Formative Evaluation

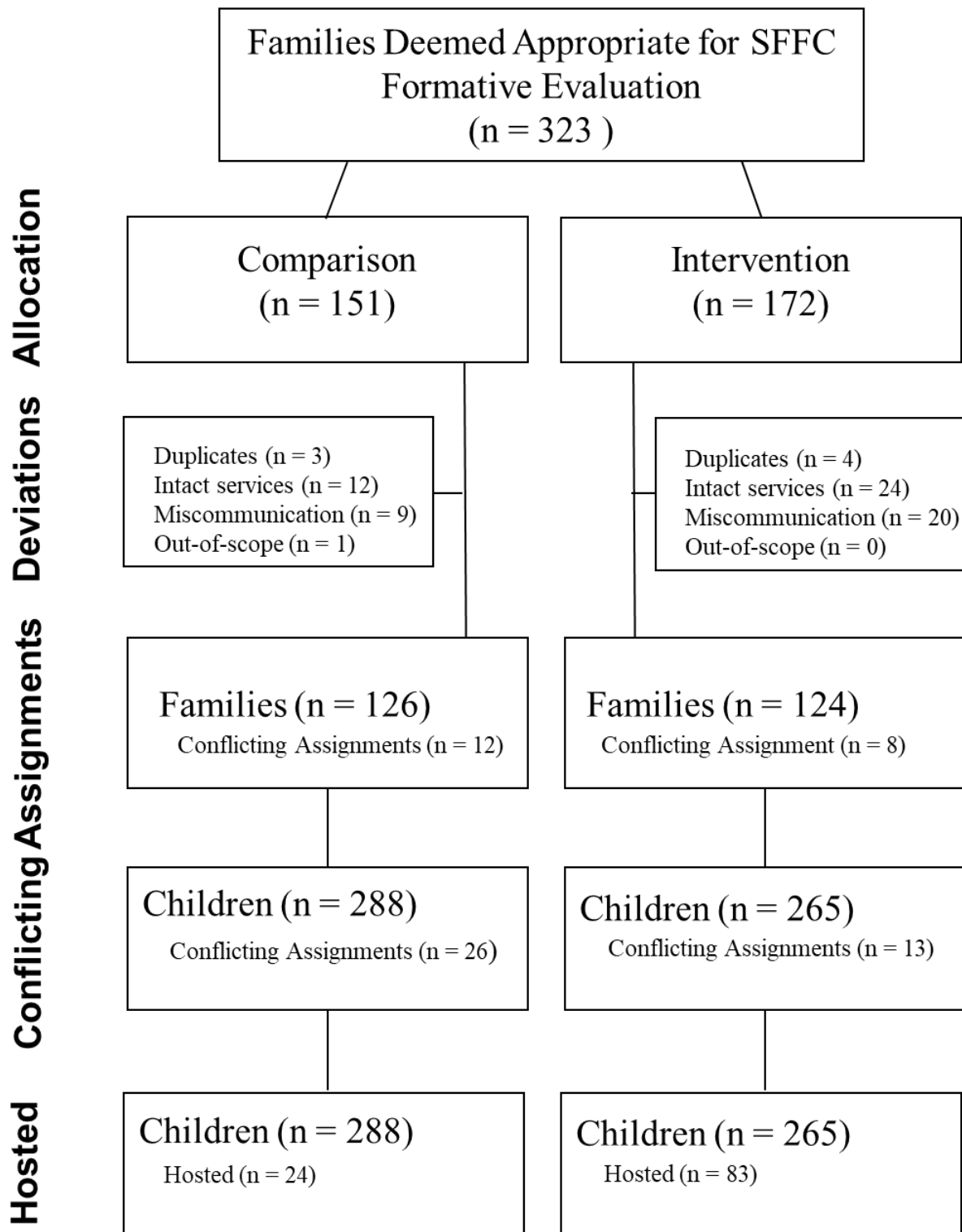
The Department's highly visible reaffirmation of support for the SFFC initiative and the investigator's installation of a back-up randomization mechanism had an immediate impact. The project was able to hit its revised targets of 60 families per quarter. During formative evaluation, IDCFS and SFFC together assigned families 323 times to the program (see Table 2 above). Given the increased sample size, the investigators were able to ascertain whether the differences in proximal outcomes were trending in the expected direction so that full implementation and summative evaluation could proceed as planned.

During the course of formative evaluation, investigators uncovered 37 protocol deviations (see CONSORT diagram in Figure 4 below) after dropping the 36 intact families. On the IDCFS side (not shown in diagram), this occurred four times because the automated allocation procedure barred re-assignment only within report sequence and not for the entire state central registry (SCR) case. Families could potentially cross-over from the comparison to the intervention condition if they were reported multiple times for different maltreatment allegations. Another two families were assigned twice by IDCFS because the household configuration of alleged perpetrators had changed, which generated a separate SCR number. IDCFS had promised to fix both glitches, but both problems persisted throughout summative evaluation.

The more common reason for protocol deviations was that some families were assigned separately by both IDCFS and SFFC. This occurred four times on the IDCFS side because the Department was unaware that SFFC had already assigned the case. It occurred 26 times on the SFFC side because intake staff relied on self-reports from investigators or families for determining whether or not the families had already been assigned by IDCFS. Sometimes this occurred between the report date and date the family was approached by IDCFS investigators to determine their willingness to participate in the SFC program. Other times it occurred after the IDCFS assignment was made and there was miscommunication between the IDCFS office and SFFC. To correct this problem, IDCFS granted read-only access so SFFC staff could check the IDCFS assignment screen prior to randomizing referrals at SFFC intake. This helped minimize the frequency of multiple assignments made during summative evaluation. The remaining case was dropped because the last maltreatment report was filed prior to the start of usability testing.

Figure 4 displays the flow of the 323 allocated cases as filtered through the exclusion criteria and de-duplication procedures to arrive at a final sample of 250 families that were

Figure 4- Consolidated Standards of Reporting Trials (CONSORT) diagram showing the flow of families from allocation to final sample, formative evaluation, State of Illinois



Source: Step0_8_CONSORT_2_4_2021_ArnoldReport.sps

referred to SFFC during formative evaluation: 126 allocated to the comparison group and 124 to the intervention group. These families linked to 553 children, five of whom appear in multiple household configurations. Retaining only the first record of multiple assignments yielded a final formative sample of 288 children in the comparison group and 265 children in the intervention group.

Characteristics at Baseline

All but two of the 323 cases allocated to the study during formative evaluation linked to IDCFS administrative data. The two unlinked cases received their assignments at SFFC and the investigation information taken at intake was missing or incorrect. Even though usability testing showed no obvious baseline differences between treatment groups, the much larger formative sample allowed us to take another look at whether the allocation procedures at IDCFS and SFFC were sufficiently unbiased to support causal inferences about the program's effectiveness.

Testing for baseline differences in randomized controlled trials is no longer recommended (de Boer, Waterlander, Lothar, Kuijper, Steenhuis & Twisk, 2015). According to CONSORT guidelines, a table showing baseline characteristics for each treatment group can help practitioners and policymakers judge the relevance of the intervention for their particular target populations. However, significance tests of baseline differences are discouraged. Not only are they unnecessary, they are potentially misleading (Moher, Hopewell, Schulz, Montori, Gøtzsche & Devereaux, 2010). As mentioned above, an unbiased allocation mechanism, such as alternation, a lottery, or a table of random numbers, guards against selection biases that can threaten the validity of the assumption that participants differ only with respect to the offered treatment (e.g., SFFC vs. SAU). It does not guarantee, however, that the groups are equivalent at

baseline. It only means that any differences in baseline characteristics are the result of chance rather than systematic bias.

Table 3 presents descriptive information about the baseline characteristics of investigation referrals made during formative evaluation. The geographical spread and timing of allocations are well balanced between treatment groups for both the sample of original allocations and the final sample. The same is true for report sequence and case status. The case characteristic that trends toward imbalance is the percentage of reports that were subsequently indicated for maltreatment (55.3% within the comparison group versus 66.1% within the intervention group). Because a report that is indicated for maltreatment is an important predictor of removal from parental custody, it will be treated as a potential confounder of the effect of SFFC on the primary outcome of deflection from foster care. Other than this difference, there were no other sizable imbalances identified at baseline. Therefore, we felt reasonably assured that both alternation and the back-up randomization routine at SFFC were yielding sufficiently balanced distributions to support causal inference at the summative evaluation phase.

Protocol Deviations and Treatment Crossovers

Figure 4 shows that compared to the intervention group a larger number of families allocated by IDCFS to the comparison group erroneously received a second assignment at SFFC intake, which conflicted with the original one. Further inspection showed that protocol deviations were significant only in Cook County and negligible in downstate regions. In Cook County, 12% of the comparison group received a conflicting assignment compared to only 3% in the intervention group. The percentage of conflicting assignments in the downstate comparison groups averaged 5%. The differences by geographical area hinted at the possibility of some deliberate

Table 3.-- Baseline Characteristics of SFFC Referrals at Formative Evaluation by Sample and Treatment Group, State of Illinois

Characteristic	Allocated Sample (N = 323)				Final Sample (N = 250)			
	Number		Percentage		Number		Percentage	
	Comparison	Intervention	Comparison	Intervention	Comparison	Intervention	Comparison	Intervention
Geography								
Central	24	25	15.9%	14.5%	21	20	16.7%	16.1%
Cook	91	106	60.3%	61.6%	76	78	60.3%	62.9%
Northern	33	39	21.9%	22.7%	26	24	20.6%	19.4%
Southern	3	2	2.0%	1.2%	3	2	2.4%	1.6%
Allocation quarter								
2015 4	2	2	1.3%	1.2%	2	2	1.6%	1.6%
2016 1	14	20	9.3%	11.6%	14	16	11.1%	12.9%
2016 2	52	55	34.4%	32.0%	41	38	32.5%	30.6%
2016 3	48	51	31.8%	29.7%	37	38	29.4%	30.6%
2016 4	35	44	23.2%	25.6%	32	30	25.4%	24.2%
Report sequence								
A	91	101	60.7%	59.1%	79	77	62.7%	62.1%
B	32	24	21.3%	14.0%	24	17	19.0%	13.7%
C	7	14	4.7%	8.2%	7	9	5.6%	7.3%
D or higher	20	32	13.3%	18.7%	16	21	12.7%	16.9%
Finding								
Indicated	83	113	55.3%	66.1%	71	79	56.3%	63.7%
Unfounded	67	58	44.7%	33.9%	55	45	43.7%	36.3%
Case Status								
Closed	95	121	63.3%	70.8%	83	85	65.9%	68.5%
Expunged	55	50	36.7%	29.2%	43	39	34.1%	31.5%

manipulation or deception by staff and families for the purpose of reversing the assignment of cases to the comparison group. The greater number and proportion of treatment crossovers from the comparison group to SFFC hosting in Cook County (n = 20, 12%) compared to downstate counties (n = 4, 3%) reinforced these suspicions. Whether or not there was deliberate manipulation, the higher than expected crossover rate indicated that implementation in Cook County wasn't adhering with sufficient fidelity to the SFFC logic model.

In spite of the much lower rate of treatment crossovers downstate, we recognize that even low levels pose thorny challenges to detecting an intervention effect in controlled experiments. Treatment crossovers that occur when families that are assigned to services as usual but nonetheless manage on their own or with the intercession of CPS workers to engage their children in treatment should be minimized as much as possible in order to estimate accurately the average causal effect of the intervention on the families who are agreeable to accepting this form of assistance. Another type of non-compliance occurs when families assigned to the intervention fail to show up for treatment as intended. Even though “no shows” weaken the chances of detecting a significant intervention effect, this type of incomplete compliance at least gives some feel for the net impact of the intervention as implemented in practice. As explained by Shadish, Cook & Campbell (2002), it is often of great policy interest to learn about the impact of an available treatment where participation in the treatment cannot be mandated or coerced but only encouraged. This is usually the situation with most forms of safety-organized practices, like SFFC. Voluntary participation invariably entails incomplete compliance to some degree.

A recommended statistical fix that we use to address the problem of no-shows involves analyzing outcomes as though the families and children had completed the treatment to which they were originally assigned. This so-called “intent-to-treat” analysis preserves the benefits of

unbiased assignment for causal inference but yields an unbiased estimate *only of the effect of being assigned to treatment* and not of actually receiving the treatment (Gerber & Green, 2012).

It was precisely this interest in learning about the practical impact of SFFC expansion on removal rates, regardless of family participation rates, which motivated the IDCFS director at the time to solicit a rigorous evaluation of the program. His interest in also learning about the efficacy of SFFC for the small proportion of families that actually received treatment was secondary to his primary interest in learning whether the net benefits were worth the costs, not just in monetary terms, but also in terms of the practical consequences for family integrity and the recurrence of maltreatment. The developers of the program, while also interested in collective impact were not content with just a rigorous “intent-to-treat” evaluation. They wanted to learn whether their theory of change was valid and if participation in the program truly improved the lives of the individual families and children as a result of receiving the support of host families.

In order to realize the dual objectives of estimating net impact at the collective level and clinical efficacy at the individual child and family level, the investigators decided it was necessary for treatment crossovers during summative evaluation to rise no higher than 5%. This was slightly higher than the upper threshold observed for the downstate intervention group during formative evaluation, but lower than the 10% to 20% crossover range sometimes found for field experiments in the literature (Gerber & Green, 2012). By the end of the formative evaluation period, the percentage of crossovers to treatment in Cook County had already surpassed the 10% threshold.

Our discussions with practitioners about the crossover problem in Cook County made it abundantly clear that maintaining fidelity to the SFFC guidelines would not be easy. Because

crossovers from the comparison to the intervention remained unacceptably high in spite of the urgings from top DCFS administration, further efforts to engage Cook County offices in the summative evaluation ceased after September of 2017. The lead investigators agreed to pre-register the evaluation only for the downstate regions of the state.

SECTION THREE: COMPARE AND LEARN

The purpose of *Compare and Learn* is to render a summary judgement of a program's effectiveness. As explained in their book by Rossi, Lipsey & Freeman (2004):

The findings of summative evaluations are usually intended for decisionmakers with major roles in program oversight.... Such evaluations may influence significant decisions about the continuation of the program, allocations of resources, restructuring, or legal action. For this reason, they require information that is sufficiently credible under scientific standards to provide a confident basis for action and to withstand criticism aimed at discrediting the results (p. 36).

On June 9, 2017, the lead evaluators created a project on Open Science Foundation (*osf.io*) to conduct a low-cost summative evaluation of the effectiveness of the SFFC program in preventing removals of children into foster care. The registration of clinical trials prior to data analysis and prior to the publication of results serves both ethical and scientific purposes (Zarin, & Keselman, 2007). On the basis of the formative findings, we pre-registered the summative evaluation on August 20, 2017 for referrals made in downstate Illinois. Even though enrollment in the summative evaluation had begun earlier in the year, the evaluators were “blind” to the outcomes because no IDCFS data had been downloaded since December of 2016 in order to be

consistent with the “no tweaking, no peeking” rule for summative evaluations

(<http://www.socialinnovationcenter.org/archives/769>). No “tweaking” means that the structure of the intervention is locked down and no further modifications are made after formative evaluation is concluded. No “peeking” means that data analysis occurs only at the conclusion of reporting periods, which helps guard against “fishing” for significant statistical associations in the interim, which can arise by chance.

The three key hypotheses registered on *OSF.io* are listed as follows:

- Among child subjects investigated for maltreatment,⁹ the percentage taken into protective custody or later removed into foster care from day one to 24 months after allocation will be lower for the SFFC intervention groups compared to families who receive services and usual (primary hypothesis).
- Among child subjects investigated for maltreatment, the percentage who had a subsequent report of substantiated maltreatment within the first quarter and each subsequent quarter from the date of allocation to 24 months after the investigation will be equivalent for both intervention and comparison groups.
- Among child subjects investigated for child maltreatment, the percentage who were maintained in the custody of their parents or returned to their physical custody within one year to 24 months after allocation will be higher in the intervention group than the comparison group.

Enrollment in the summative evaluation of SFFC lasted from January 1, 2017 to December 30, 2018. The enrollment period was originally scheduled to end in June of the

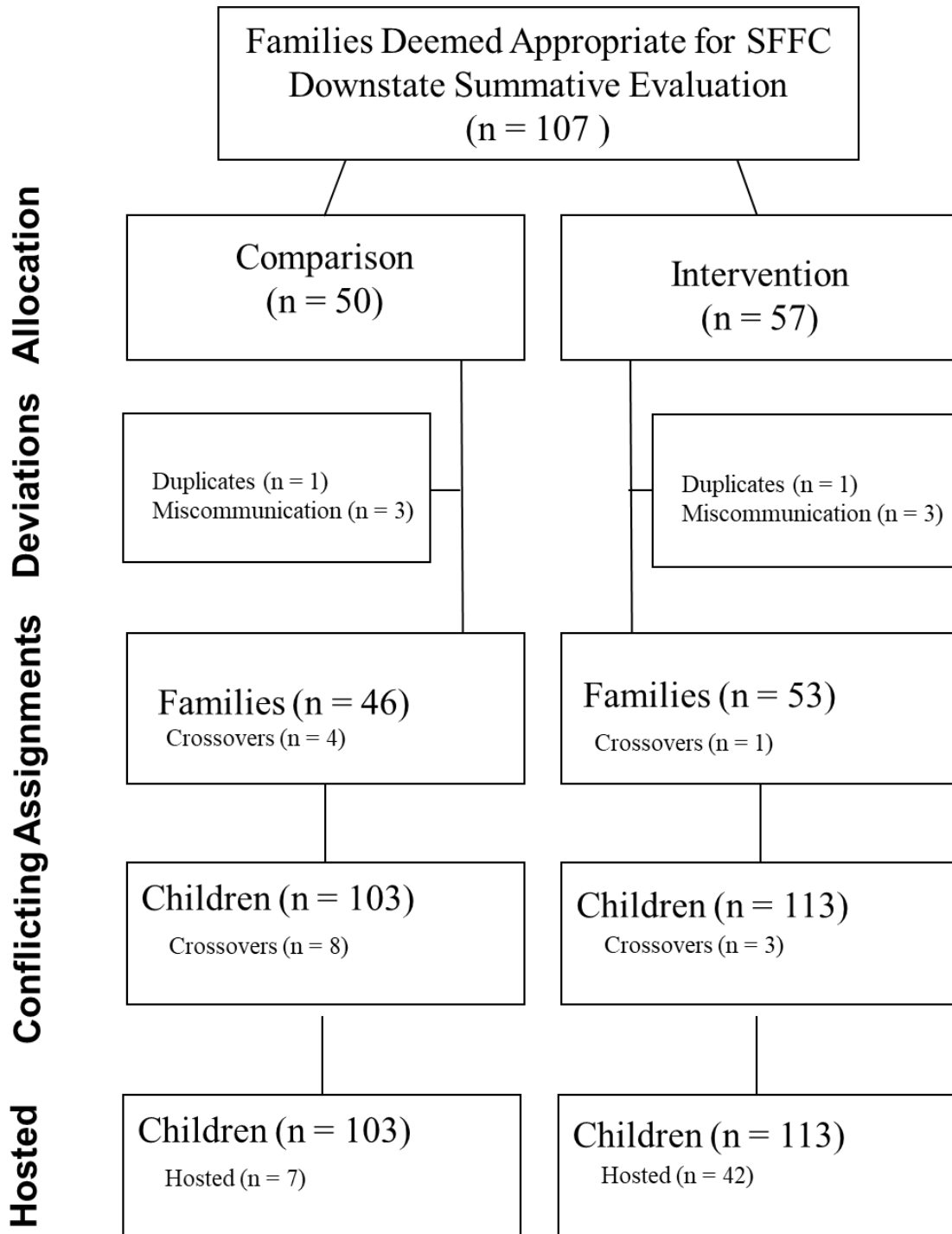
⁹ A reviewer of a draft of this report correctly noted that the target population is actually only a subset of child subjects investigated for maltreatment. Only children in families deemed appropriate for SFFC were allocated to the comparison and intervention groups.

following year. However, the appointment of a new IDCFS director in June of 2017, who accepted the validity of the intervention in the absence of well-supported evidence of its effectiveness, and the eagerness of the SFFC provider to dispense with automated alternation, which halved the number of referrals that could be hosted, drained any remaining enthusiasm that originally existed in the Department for prolonging the experiment. As a result, the follow-up period was shortened to 12 months instead of 24 to accommodate the needs of the developers to have publishable results by the end of 2020. In addition, shortening the follow-up period to 12 months enables use of outcome definitions utilized in the federal CFSR, which facilitates replications of the study's findings.

In addition to the shortened enrollment period, the dropping of Cook County from the pre-registered trial greatly reduced the size of the study sample. The CONSORT diagram in Figure 5 tracks the flows of family and child cases available for analysis. After accounting for protocol deviations, there were 99 families and 216 children retained in the summative sample. Even though this sample size doesn't support statistical inference at the power thresholds originally intended, the integrity of the experiment improved (fewer duplicate assignments and crossovers), which puts the summative evaluation on less shaky grounds for drawing valid causal inferences about the net impact of the program on key child welfare outcomes.

Table 4 displays the baseline characteristics of the allocated and finals samples for the summative evaluation in downstate Illinois. The one difference that displayed an imbalance during formative evaluation – indicated findings of maltreatment – no longer shows an imbalance in the summative evaluation sample. An imbalance that arose by chance during summative evaluation – report sequence – is another important predictor of deflection from foster care. The intervention group has a much lower percentage of sequence A reports than the comparison

Figure 5—Consolidated Standards of Reporting Trials (CONSORT) diagram showing the flow of families from allocation to final sample, summative evaluation, downstate Illinois



Source: Step0_8_CONSORT_2_4_2021_ArnoldReport.sps.

Table 4. -- Baseline Characteristics of SFFC Referrals at Summative by Sample and Treatment Group. Downstate Illinois

Characteristic	Allocated Sample (N = 107)				Final Sample (N = 99)			
	Number		Percentage		Number		Percentage	
	Comparison	Intervention	Comparison	Intervention	Comparison	Intervention	Comparison	Intervention
Geography								
Central	21	24	42.0%	42.1%	19	23	41.3%	43.4%
Northern	25	28	50.0%	49.1%	23	25	50.0%	47.2%
Southern	4	5	8.0%	8.8%	4	5	8.7%	9.4%
Allocation quarter								
2017 1	5	11	10.0%	19.3%	5	10	10.9%	18.9%
2017 2	14	9	28.0%	15.8%	13	9	28.3%	17.0%
2017 3	7	6	14.0%	10.5%	7	6	15.2%	11.3%
2017 4	1	5	2.0%	8.8%	1	5	2.2%	9.4%
2018_1	1	2	2.0%	3.5%	1	1	2.2%	1.9%
2018 2	10	10	20.0%	17.5%	8	9	17.4%	17.0%
2018 3	6	10	12.0%	17.5%	6	9	13.0%	17.0%
2018 4	6	4	12.0%	7.0%	5	4	10.9%	7.5%
Report sequence								
A	31	28	62.0%	49.1%	30	25	65.2%	47.2%
B	5	14	10.0%	24.6%	5	13	10.9%	24.5%
C	2	6	4.0%	10.5%	2	6	4.3%	11.3%
D or higher	12	9	24.0%	15.8%	9	9	19.6%	17.0%
Finding								
Indicated	32	36	64.0%	63.2%	31	34	67.4%	64.2%
Unfounded	18	21	36.0%	36.8%	15	19	32.6%	35.8%
Status								
Closed	42	48	84.0%	84.2%	39	45	84.8%	84.9%
Expunged	8	9	16.0%	15.8%	7	8	15.2%	15.1%

group (47.2% vs. 65.2%). Because sequence A reports are associated with a lower risk of removal, the imbalance could obscure the true effect of SFFC on the primary outcome of deflection from foster care. Therefore, it will be treated as a potential confounder in the summative analysis.

The integrity of the summative experiment is bolstered by fewer number of protocol deviations as illustrated in the CONSORT diagram (see Figure 5). The number of protocol deviations were only 8 cases (7%) compared to 23% during formative evaluation. Even though the percentage of conflicting assignments didn't change much, the difference between the comparison and intervention groups with respect to the proportions of families that received hosting support widened to 31 percentage points.

In spite of the wider spread, the higher than hoped-for percentage of treatment crossovers (9%) in the comparison group still makes it difficult to draw valid inferences about the efficacy of SFFC at the individual child and family level. Likewise, the high failure-to-treat proportion (63%) makes it challenging to detect a statistically significant, net improvement at the collective level among families assigned to SFFC regardless of whether they actually participated in the program. On the brighter side, the findings from the formative evaluation buoyed program developers' hopes that the intervention effect on the primary outcome of deflection from foster care would be large enough to be statistically distinguishable from no difference even with a much smaller sample size.

As a supplementary option, the investigators began looking into the Bayesian paradigm as an alternative to the frequentist approach for testing statistical significance. As an alternative to discarding sample data gathered at the formative evaluation phase, the Bayesian paradigm elegantly incorporates formative findings with summative evidence in order to render a more

comprehensive summary of the program’s merits. The Bayesian perspective on statistical significance is more in keeping with the scientific principle of phase-based evidence building advocated in this report than the frequentist approach (Chen, Testa, Anson & Brevard, 2020). Because the surface and deep structures of the program remained unchanged between formative and summative evaluations, use of prior information seems appropriate. Nonetheless, given that the Bayesian paradigm is less widely applied in child welfare intervention research, we report statistical results using both approaches.

In the following sections, we report results for the primary and secondary outcomes pre-registered on *OSF.io*. For the most part, we employ operational definitions of child welfare outcomes, which the U.S. Children’s Bureau uses to assess conformity with federal child welfare requirements related to safety and permanence. Relying on the same definitions facilitates the replication of results when drawing information from different data systems. It also helps guard against the temptation to cherry-pick definitions that present the program in the best possible light. Exceptions to standard definitions are duly noted.

Deflection from Foster Care (Primary Outcome)

As discussed in earlier sections, the primary outcome that SFFC aims to impact is the deflection of alleged and indicated victims of child maltreatment from formal foster care by facilitating their temporary hosting in the homes of volunteers who offer assistance and support to vulnerable families in crisis. The U.S. Code of Federal Regulations (CFR) defines foster care as 24-hour alternative care for children placed away from their parents or guardians and for whom a CPS agency has placement and care responsibility (45 CFR §1355.20). Placement and care responsibility has two components: physical custody and legal custody. Both concepts are separable to some degree. A CPS agency can temporarily seize physical custody of a child

(usually restricted by statute to 48 hours) without a court's transferring to it legal custody of the child; or the agency can retain legal custody of the child and return physical custody to the parents, as with overnight visits and trial reunifications. Only when a court fully restores both physical and legal custody to birth parents and transfers legal custody to guardians or adoptive parents is the child considered to be discharged from foster care.

Deflection from foster care encompasses alternative care arrangements wherein birth parents retain legal custody of their children regardless of whether they maintain physical custody or share it with another family. When families make these arrangements privately, the placement is called *private alternative care*. This phrasing borrows from terminology used in the study of kinship care (Testa, 2013). When the arrangement is facilitated through the mediating help or encouragement of a CPS agency, without the agency's petitioning for legal custody, the placement is called *voluntary alternative care*. Placements made by the CPS agency after a court has awarded it temporary custody or guardianship of the children is called *public alternative care*, or simply foster care. These interchangeable terms also encompass temporary out-of-home placements made by the agency (commonly called protective custody), while it decides whether to petition the court for legal custody. Lastly, *permanent alternative care*, refers to living arrangements that the court authorizes by granting full legal and physical custody to adoptive parents, permanent guardians, or legal custodians. In keeping with these naming conventions, we shall refer to the hosting of children by SFFC families as voluntary alternative care even though we acknowledge that the appropriateness of this designation is currently a subject of dispute (Redleaf, 2018). We take up this issue later on in the Discussion section.

Table 5 reports the main results from the analysis of IDCFS administrative data. The first set of results examines the differences in proportions and odds ratios for all of the outcomes

Table 5. -- Results from Tests of Differences in Proportions and Odds Ratios: Is the SFFC Intervention Superior to Services as Usual (SAU) within a Year of Assignment in Downstate Illinois?

Phase		Deflection from Foster Care			No Protective Custody		
		Proportions			Proportions		
Formative	SFFC(I)	y_{1I}/n_{1I}	= 99/115	= 0.861	y_{1I}/n_{1I}	= 108/115	= 0.939
	SAU(C)	y_{1C}/n_{1C}	= 77/120	= 0.642	y_{1C}/n_{1C}	= 99/120	= 0.825
Summative	SFFC(I)	y_{2I}/n_{2I}	= 81/113	= 0.717	y_{2I}/n_{2I}	= 104/113	= 0.920
	SAU(C)	y_{2C}/n_{2C}	= 63/103	= 0.612	y_{2C}/n_{2C}	= 78/103	= 0.757
		Odds Ratios			Odds Ratios		
Formative	OR, CI*	3.455	[1.480, ∞]		3.273	[0.981, ∞]	
Summative	OR, CI*	1.607	[0.776, ∞]		3.704	[1.456, ∞]	
Formative Frequentist <i>p</i> -value		0.008			0.053		
Summative Frequentist <i>p</i> -value		0.141			0.011		
Bayesian <i>p</i> -value		0.009			0.003		
Formative Freq. Prob(I > C)		0.992			0.937		
Summative Freq. Prob(I > C)		0.860			0.990		
Bayesian Prob(I > C)		0.992			0.997		
Phase		Permanence at One Year			No Recurrence of Maltreatment		
		Proportions			Proportions		
Formative	SFFC(I)	y_{1I}/n_{1I}	= 102/115	= 0.887	y_{1I}/n_{1I}	= 104/115	= 0.904
	SAU(C)	y_{1C}/n_{1C}	= 81/120	= 0.675	y_{1C}/n_{1C}	= 109/120	= 0.908
Summative	SFFC(I)	y_{2I}/n_{2I}	= 84/113	= 0.743	y_{2I}/n_{2I}	= 105/113	= 0.929
	SAU(C)	y_{2C}/n_{2C}	= 68/103	= 0.660	y_{2C}/n_{2C}	= 94/103	= 0.913
		Odds Ratios			Odds Ratios		
Formative	OR, CI*	3.778	[1.619, ∞]		0.954	[0.225, 4.048]	
Summative	OR, CI*	1.491	[0.710, ∞]		1.257	[0.337, 4.688]	
Formative Frequentist <i>p</i> -value		0.005			0.949		
Summative Frequentist <i>p</i> -value		0.188			0.733		
Bayesian <i>p</i> -value		0.009			0.834		
Formative Freq. Prob(I > C)		0.995			0.525		
Summative Freq. Prob(I > C)		0.814			0.634		
Bayesian Prob(I > C)		0.991			0.583		

*1-sided 95% CIs except for No Recurrence of Maltreatment, which is a 2-sided 95% CI

Note: SFFC = Safe Families for Children; SAU = services as usual. The frequentist *p*-values for chi-square tests of the difference in proportions and the proportional hazard rate are adjusted for the clustering of children within family units.

Source: SFFC_Children451_Clusters195.do; SFFC_Children451_Clusters195.R

specified in the pre-registered analysis plan. The deflection of alleged and indicated victims from foster care is the primary outcome of interest. Assignment to the intervention group authorized investigators to ask families, whom they deemed appropriate for hosting, whether they agreed or not to their making a referral to SFFC. Assignment to the comparison group did not pop-up an agreement box, which was alerted investigators to refrain from referring the family to SFFC (see Figure 2 above). Instead, they were to offer services as usual, which could entail leaving the child in parental custody if the CERAP deemed the child “safe,” or if deemed “unsafe,” encouraging voluntary kinship care under the governance of a safety plan or taking the child into protective custody and placing them in formal foster care with relatives or licensed foster families. The guiding hypothesis was that assignment to the intervention reduced the likelihood of CPS’s taking protective custody and placing the child into foster care compared to services as usual.¹⁰

As noted below, the pre-registered analysis plan specified the use of transition (hazards) ratios to reach a summary judgment about intervention superiority. The advantage is that transition ratios make full use of the date and time information that are stored in most administrative data systems. For example, not only can it be determined whether a child was removed into foster care but also how quickly it occurred after allocation. The downside is that transition models are harder to communicate to a general audience than more familiar descriptive statistics such as the observed proportions or odds of children deflected from foster care. To

¹⁰ One of the reviewers raised the issue of whether non-blinded assignment to treatment might encourage investigators to give the SFFC arm more leeway before taking protective custody of the children. The answer is yes, and as the reviewer notes, the discouragement to take legal custody constitutes the bulk of the intervention effect. It is the lesser likelihood of CPS’s taking legal custody of the children that is the primary outcome and reason for parents’ agreeing to accept the offer of SFFC hosting as an alternative to removal. Whether there are additional spillovers from SFFC to other areas of child and family wellbeing are topics for future research.

facilitate communication, therefore, we start with the observed differences in the proportions and odds of deflections from foster care during the formative and summative evaluation periods.

Table 5 displays the proportions and odds of deflection from the date of allocation to treatment groups to a year later. It shows that the proportions of children deflected from foster care were consistently higher among children assigned to the intervention group compared to the comparison group. The odds ratio converts the different proportions into a single measure which indicates a greater likelihood of experiencing the outcome when the odds ratio is greater than 1 (i.e., > 50:50) and a lesser likelihood when the ratio is less than 1. Odds ratios close to 1.0 imply no difference between treatment groups in the occurrence of an event.

Focusing first on the odds ratio for deflection (the primary outcome), children assigned to the intervention group were 3.455 times as large as the odds of deflection from foster care for the comparison group during the formative evaluation period. The ratio fell to 1.607 times as large during the summative evaluation period. Unlike the frequentist p -value for the larger odds ratio estimated during formative evaluation, the p -value for the smaller odds ratio estimated during summative evaluation indicates that the odds of the SFFC's superiority is not distinguishable statistically from 1 (no difference) at the conventional .05 level. From the frequentist perspective, the summary judgment is that SFFC is no more effective than SAU in deflecting children from foster care. However, it seems inefficient to ignore totally the strong effect observed during formative evaluation. Because the program registered for summative evaluation is fundamentally the same as the program tested during formative evaluation, the Bayesian paradigm offers a more comprehensive summary of the program's effectiveness. When the formative results are updated in light of the new information from the summative evaluation, the p -value changes to a highly

significant value of 0.009.¹¹ We also make use of a Monte-Carlo, simulation-based statistical model (Chen & Fraser, 2017) to assess the extent to which the program is superior to SAU.¹² The results indicate that the likelihood that SFFC is truly superior, Bayesian Prob($I > C$), in deflecting children from foster care is 99 percent certain (see Table 5).

The analysis plan registered with *OSF.io* in 2017 specified the use of a more precise statistic for assessing superiority, which is reported in Table 6. The transition ratio, also referred to as a hazard rate in the literature, takes into account the amount of time elapsed before the outcome or event of interest happens. Events that never happen during an evaluation period, i.e. the child is never removed, are said to be “censored” observations. Follow-up stops at the date of last observation. The transition ratio indicates the proportionate amount that the risk of experiencing a specific outcome or event is expected to vary at any particular time. In this study, a transition ratio less than 1.0 implies that assignment to SFFC lowers the chances of an event’s occurring at any particular time, whereas greater than 1.0 implies that SFFC raises the chances. Estimates close to 1.0 imply no difference between SFFC and SAU in the chances of the event.

The *OSF* analysis plan specified that a transition ratio below 0.5 is considered practically important for the primary outcome of deflection. A transition ratio below 0.5 signifies that

¹¹ All three peer reviewers of the draft report questioned the necessity of conducting separate statistical analyses of the formative and summative samples given that the program remained fundamentally unchanged between the two phases. The odds ratio for the pooled sample is 2.223, which is statistically significant at the .009 level based on the frequentist approach. Thus both perspectives support the conclusion that SFFC is superior to SAU with regard to deflection from foster care. The primary reason for favoring the Bayesian approach is that the analysis plan was registered after “cherry-picking” from the demonstration sites the most promising finding for replication.

¹² We are indebted to Dr. Ding-Geng Chen for assisting us in estimating the frequentist and Bayesian p -values and probabilities of superiority, Prob($I > C$). The calculations involved simulating 1 million separate intervention studies in which the formative and summative samples for each of the four outcomes were used to construct the data likelihood distributions. The Bayesian posterior distribution incorporates the formative data into the data likelihood from the summative sample. A p value was obtained from this posterior distribution to test the null hypothesis of non-superiority. The Monte Carlo estimate of Prob($I > C$) is the proportion of the total number of simulations with $I > C$ among the 1 million simulations. All calculations were implemented in R software, which is posted on the *OSF*.

assignment to SFFC reduces the estimated risk of removal by more than 50% on any particular day after assignment. The reason the transition ratio is favored over the odds ratio is that it facilitates updating prior information in light of newer information from replication studies without needing to take into account different durations of follow-up intervals (DePanfilis & Zuravin, 1998). While as a practical matter the ratio of daily transition probabilities or relative risks of removal can rise or fall as time elapses, the model we use makes the simplifying assumption that the relative risk ratio stays proportionate over time.¹³

Table 6 displays the transition ratios observed during both formative and summative evaluation periods for the three outcomes that could occur at any moment after assignment. In this study, protective custody could be taken immediately upon assignment or up to a year later after which the observation is censored at 366 days. Because the follow-up period for assessing the secondary outcome of family permanence is fixed at one year after assignment, it is omitted from Table 6. All of the available information for assessing the superiority of SFFC over SAU for maintaining or restoring family permanence is conveyed in Table 5 above. The data in Table 6 show that the ratio of the daily probability of removal relative to the comparison group fell below the threshold of 0.5 that the analysis plan set for assessing practical significance but

¹³ A reviewer of the draft report recommended first checking the proportionality assumption using graphs. Both the survival and log-log plots visually suggest that the hazards functions for the comparison and intervention are parallel. An appendix displays these two graphs for the primary outcome of removal using the formative sample data. A global test based on Schoenfeld residuals suggests that the proportional-hazards assumption cannot be rejected at conventional significant levels. As Allison (2000) notes, it is unlikely that the proportional-hazards assumption is ever exactly satisfied, but that's true of nearly all statistical assumptions. Violation of the proportionality assumption is akin to suppressing interactions in ordinary least squares regression. It's done all the time. As Allison (2000) notes, the coefficient that is estimated is a sort of average effect over the range of times observed in the sample.

Table 6. -- Results from Tests of Proportional Transition Ratios: Is SFFC Intervention Superior to Services as Usual?

Phase	Observed Outcomes within a Year of Assignment								
	Foster Care Removal Transition Ratios			Protective Custody Transition Ratios			Recurrence of Maltreatment Transition Ratios		
	Exp($\lambda_{(t)}$)	1-sided CI		Exp($\lambda_{(t)}$)	1-sided CI		Exp($\lambda_{(t)}$)	2-sided CI	
Formative	0.335		[- ∞ , .691]	0.338		[- ∞ , 1.032]	1.057		[0.273, 4.093]
Summative	0.602		[- ∞ , 1.081]	0.297		[- ∞ , .667]	0.797		[0.229, 2.768]
	Coefficients			Coefficients			Coefficients		
Phase	$\lambda_{(t)}$	SE	1-sided CI	$\lambda_{(t)}$	SE	1-sided CI	$\lambda_{(t)}$	SE	2-sided CI
Formative	-1.093	0.437	[- ∞ , -.370]	-1.086	0.676	[- ∞ , .031]	0.056	0.686	[-1.298, 1.409]
Summative	-0.508	0.354	[- ∞ , .078]	-1.212	0.489	[- ∞ , -.405]	-0.227	0.631	[-1.472, 1.018]
Formative Frequentist <i>p</i> -value		0.007			0.055			0.935	
Summative Frequentist <i>p</i> -value		0.077			0.007			0.720	
Bayesian <i>p</i> -value		0.004			0.002			0.834	
Formative Freq. Prob(I > C)		0.993			0.945			0.532	
Summative Freq. Prob(I > C)		0.924			0.993			0.641	
Bayesian Prob(I > C)		0.996			0.998			0.583	

Note: The frequentist *p*-values for t tests of the difference in proportional transition ratios are adjusted for the clustering of children within family units.

Source: SFFC_Children451_Clusters195.do; SFFC_Children451_Clusters195.R

averaged a little higher during summative evaluation. Controlling for the imbalances in the distribution of indicated and sequence A reports doesn't alter the results to any appreciable degree. The p -values calculated from both the frequentist and Bayesian perspectives indicate that transition ratio observed during both evaluation periods are statistically distinguishable from one (no difference) at a significance level of $p < .08$ for a one-tailed test. The second set of coefficients listed in the table are the logs of the transition ratio. They convey the same results as the transition ratios. Their coefficients and associated standard errors (SE) and confidence intervals (CI) are reported here to satisfy the interests of fellow researchers.

Delving a bit more into these statistical matters, a one-tailed test of statistical significance as opposed to a two-tailed test allows for a thumbs up-or down confirmation of whether the data support the hypothesis that the intervention is superior to services as usual. In this instance, the thumbs point up. Even though the pre-registered analysis plan called for a two-tailed test, we now believe that a one-tailed test is equally appropriate (Chen, Testa, Ansong & Brevard, 2020). Because we set an effect size of 0.5 based on prior information from the formative evaluation, we weren't operating entirely in the dark when we developed the analysis plan. Our ability to formulate a directional hypothesis of intervention superiority favors the use of a one-tailed statistical test. Given that the summative evaluation was designed to decide whether to maintain or possibly expand funding for the project, a directional hypothesis that the outcome for SFFC(I) is superior to SAU(C), i.e., $H_a: I > C$, seems appropriate. A one tailed test of significance also leads to greater statistical power for deciding whether or not to continue investing public funds in the program.

As shown in Table 5 above, the observed proportion of children deflected from foster care during formative evaluation was 0.861 for SFC and 0.642 for SAU. The difference in

proportions indicates an intervention effect of 0.219. The difference in proportions during summative evaluation, however, narrowed to 0.105 ($= .717 - .612$) and is no longer statistically significant. As formulated by Scriven (1997), the results from a formative evaluation are used primarily for program improvement purposes and are usually not recorded. The results from summative evaluation are supposed to stand on their own and hence do not explicitly incorporate the prior information from the formative evaluation. The Bayesian paradigm does not follow this practice of separately analyzing formative and summative data. Instead it uses the data from the formative evaluation as prior evidence that is updated with the findings from the summative evaluation to build cumulatively toward a summary judgement about program effectiveness. As shown in Table 5, applying the Bayesian paradigm results in a p -value $< .009$ and 99.2% certainty that SFFC is superior to SAU in deflecting children from foster care.

Protective Custody (Proximal Outcome)

Mining administrative data for additional clues about intervention integrity and intervention validity offers another pathway for exploring the underlying causal mechanisms and deep structure, which are contributing to the program's effectiveness. As shown in the SFFC logic model, a key proximal outcome that can be easily tracked before the full extent of the risks to the child is known is the proportion taken into state protective custody (PC). Even though this component of state intervention is folded into our broader measure of foster care removal, it is worth tracking protective custody as a separate component in order to get an early reading of success or troubles ahead. In a related publication (Chen, Testa, Ansong & Brevard, 2020), we ignored so-called "lapsed" protective custodies which lasted less than 48 hours. This choice resulted in our *not* rejecting the null hypothesis of no difference. However, upon further investigation we saw that much of the impact of SFFC occurs within 48 hours of when the

investigation begins. In this final report, we apply the definition of deflection that we pre-registered on *OSF*, which imposes no time-limit on protective custodies.

CPS workers, local law enforcement officers, and physicians who are treating a child may take physical custody of the child without the consent of their parents if they believe the health and welfare of the child to be in imminent danger. During formative evaluation, over 90% of protective custodies in Illinois were taken by CPS workers. The remainder were divided evenly between physicians and police. In order to retain physical custody of a child beyond 48 hours, IDCFS must obtain court approval or else release the child back to the custody of their parents or legal guardians. Approximately 35% of all of the protective custodies that were taken immediately at the start of investigation were rejected by the State's Attorney or lapsed by IDCFS without court action. The proportion drops to 23% for protective custodies taken a week or more after receipt of the report.

Table 6 above showed that the intervention effect on the transition ratio for protective custodies was highly significant in both practical and statistical terms. The transition ratio was slower than the practical importance threshold of 0.5 and the Bayesian p -value was statistically significant at the .002 level. The same conclusion was supported by the proportions and odds ratios reported in Table 5 above. These findings reinforce the summary judgement that SFFC is superior to SAU in deflecting children from foster care. However, two questions linger: 1) how lasting is the effect of SFFC on preserving family integrity, and 2) how safe are the children in the "hidden" system of voluntary alternative care?

Permanence of Care at One Year (Distal Outcome I)

To recap, the evidence presented in Tables 5 and 6 above is supportive of the developers' claim that SFFC successfully deflects alleged and indicated victims of maltreatment from the formal foster care system to the voluntary alternative care of host families. As displayed in the logic model (see Figure 3 above), deflection is the primary outcome that the analysis plan preregistered for assessing program effectiveness. Confidence in the validity of the claim is further reinforced by the secondary finding that SFFC reduces the relative risks of protective custody, which the logic model identifies as a key mediating influence on the primary outcome. Having successfully passed through these two "falsifiable tollgates" on the path to accumulating credible evidence of effectiveness, the summative analysis could justifiably stop here. However, in order to qualify as a "supported" (as opposed to just "promising) intervention under the standards promulgated by the Prevention Services Clearinghouse (Wilson, Price, Kerns, Dastrup & Brown, 2019), it is also necessary to demonstrate a favorable effect of the intervention for at least six months beyond the end of treatment. Because SFFC does not have a definite end-point but can continue for as long as families need support, we selected an end-point for treatment that "corresponds to when the majority of a clearly defined set of services were slated to have been delivered" (Wilson, Price, Kerns, Dastrup & Brown, 2019: 44). We calculate that approximately 90% of the 215 children to whom SFFC delivered hosting resources finished the program within six months of their allocation to treatment groups. Therefore, setting 12 months after the allocation date seems an appropriate end-point for assessing whether SFFC had a superior effect on family permanence of at least six months beyond the time that most children stopped receiving voluntary alternative care from host families.

Even though the net effect of alternating the assignment of families to SFFC resulted in a larger proportion of children being successfully deflected than SAU, the program did not completely shield hosted children from subsequently being taken into protective custody or foster care. An estimated 28% (n = 9) of children hosted during formative evaluation in downstate Illinois subsequently entered foster care, with nearly half of them taken into custody within a week after leaving the host family. Most of the removed children (> 80%) were still in foster care 18 months after enrollment.

The large percentage of children who are retained in foster care reflects historical patterns and practices in Illinois which have long made the state an outlier compared to the rest of the nation. Comparative data show Illinois registers the lowest rate of family reunification and the longest median length of formal foster care among all 50 states and the District of Columbia. The most recent data released by the U.S. Children's Bureau for 2017 show that Illinois ranked lowest in the percentage of children discharged during their first year in foster care to the physical custody of birth parents, relatives, guardians and adoptive parents—10.6%. The next lowest percentage was 24.9% for DC and the weighted average for the rest of the U.S. was 39.0% with South Carolina topping the rankings at 58.3%.

The practice of delaying the discharge of children to permanent homes, especially during the first year of formal foster care, is deeply entrenched in the Illinois system. This ingrained bias is one of the systemic injustices that the *B.H.* consent decree aims to remedy. In spite of repeated pledges by successive IDCFS directors to hasten the tempo of permanence, the annual permanency rates have remained remarkably stable over the last decade. Hence, it didn't come as much of a surprise that only 9 (7%) of the 132 children taken into foster care for 8 or more days

during the formative evaluation period were back home ($n=7$) or placed in voluntary kinship care ($n = 2$) within 12 months after their enrollment in the study. Further, the proportions were the same for children whether they were removed into foster care from the intervention group or from the comparison group. The insight gleaned from the equivalence in discharge rates is that SFFC succeeds in reducing foster care not by shortening the time children stay in care but by preventing their entry into care in the first place.

Table 5 above displayed the proportions and odds ratios of children who were consistently maintained in their own home or who at one year after allocation to treatment were back living with their families. During the formative evaluation period in downstate Illinois, the odds of permanence at one year were 3.778 times as large for children assigned to SFFC than children assigned to SAU. The permanency odds ratio observed during summative evaluation declined to 1.491 times as large. While the intervention effect was in the desired direction, the p -value was no longer statistically significant. Nonetheless, when the weaker effect is conditioned on the much stronger prior effect, the probability that SFFC is superior than SAU—Bayesian $\text{Prob}(I>C)$ —exceeds 99%.

The summary judgement that SFFC serves better the purpose of maintaining the integrity of vulnerable families than SAU should come as welcomed news to folks who are distrustful of the hidden foster care system. Recent criticisms have focused on the concern that deflection from formal foster care threatens family integrity by shielding CPS discretionary practices from legal and judicial oversight (Gupta-Kagan, 2020; Redleaf, 2020). The reason for concern is not entirely baseless. As noted above, approximately one-third of protective custodies taken at the start of investigation are lapsed by the CPS agency because of insufficient evidence to justify

removing legal custody from parents. Even when protective custody is taken weeks into an investigation, nearly a quarter are lapsed, rejected by State's Attorneys, or denied by judges. No similar restraint exists on the asymmetrical power relationship between families and CPS when investigators induce parents to seek voluntary alternative care under threat of removing their children into formal foster care. The accumulated weight of the evidence suggests, however, that such misuse of state authority is infrequent. Most children hosted by SFFC return home quickly and not a single hosting episode in Illinois lasted for longer than a year before CPS reunited children with their parents, either ended intact family services, or else took the children into formal foster care.

Recurrence of Maltreatment (Distal Outcome 2)

Whereas recent criticism of deflection policies has emphasized the threats to family integrity, an opposite argument has been made that such policies tilt the power imbalance too far in the direction of "parental autonomy rights to raise children without intervention by the state" (Bartholet, 2012: 1324). The data mustered in support of this criticism rests on the claim that recurrence rates among confirmed victims of maltreatment run around 30% to as high as 50% (DePanfilis & Zuravin, 1998). Scrutiny of the studies and systematic reviews on which such claims are based, however, reveals a wide range of variability depending on the definition of recurrence, the units of analysis, the length of follow-up intervals, and selected data analysis strategies (DePanfilis & Zuravin, 1998). For purposes of this evaluation, we chose the definition of recurrence that the U.S. Children's Bureau uses for its Child and Family Services Reviews (CFSRs) to assess state conformity with federal child welfare requirements.

The federal indicator of recurrence of maltreatment measures whether a CPS agency is successful in preventing subsequent maltreatment of a child who was the subject of a prior indicated report. The follow-up interval for making this determination is 12 months after the initial report, excluding re-reports within 14 days of the previous report. This exclusion rule helps eliminate those false (positive) indicators of recurrence, which are based on allegations registered later but pertain to the original maltreatment report. Applying the federal definition, the recurrence rate among all indicated reports recorded in downstate Illinois during the period covered by the formative evaluation averaged 13.8%. The recurrence rate during this same period in Cook County was 9.8%.

The fact that Cook County exhibits a lower recurrence rate than downstate Illinois in spite of a higher concentration of poverty and social disadvantage in the central city of Chicago raises some baffling issues. On the one hand, Cook County takes into formal foster care a much smaller share of its child population than downstate counties. According to critics of deflection policies, lower removal rates should result in higher recurrence rates. On the other hand, Cook retains children in foster care much longer and returns far fewer to the custody of their birth parents. According to the same logic, longer retention rates should lower recurrence rates. How do these two countervailing influences play out in light of SFCC's success in deflecting children from foster care?

Table 5 above displays the differences in the proportions of children assigned to each treatment group who did not experience a recurrence of maltreatment within a year of their prior indicated report. Both the proportions and associated odds ratios are approximately equivalent in spite of the fact mentioned above that a significantly higher proportion of children assigned to the SFCC program were deflected from protective custody and foster care. The high *p*-values for

the difference in proportions and the odds ratios of recurrence point to statistically similar recurrence rates in both treatment groups. The absence of significant differences in the transition ratios to recurrence reinforces this conclusion (see Table 6). The low probability that the intervention's effect is superior to SAU (0.583) means that the diversion of indicated victims of maltreatment to voluntary alternative care does not put them at any greater risk of repeat maltreatment than SAU. Unlike the other measures, we report the two-sided 95% CI because in light of concerns over the safety of children diverted to the hidden foster care system, there is no *a priori* reason to believe that one response is safer than the other.

Cost Savings

Deflection from foster care to the voluntary, uncompensated care of host families promises to save potentially hundreds of thousands of dollars in public expenditures. How believable are those claims? The average daily administrative costs associated with licensed foster care was approximately \$45 per child per day in Illinois. The average daily maintenance assistance payment per child for these same children was about \$40 per day. The combined daily administrative and maintenance costs of \$85 compares to approximately \$7 per child per day in SFFC administrative fees based on an average duration of 55 days of hosting in downstate counties.

During the 12 months following enrollment, the average length of paid foster care was 52 days longer for children assigned to SAU than children assigned to SFFC. Overall, children assigned to the comparison group during the formative and summative phases of the study consumed an annual average of 116 days of paid foster care compared to 64 days in the intervention group. Multiplied by the average of \$85 per day yields an estimated gross savings of \$4,420 in paid foster care for every child assigned to SFFC. The net savings after subtracting out

the SFFC administrative fees (\$690 per family/2 children) is approximately \$4,090 per child.

Applied to all 223 children assigned downstate to the comparison group during the two phases of the evaluation would have yielded over \$900,000 in savings to the state.

From a child's point of view, however, the costs of foster care cannot be measured solely in dollar terms but also in terms of the time they must spend away from their families, friends, neighbors, and familiar surroundings. From a child's point of view, you should ask how much of a real difference is there between spending time in some stranger's home that is licensed by the state compared to staying in another stranger's home recruited through SFFC? Answers to this question cannot be found in the kinds of administrative data that are available in this study. The voices, feelings, and attitudes of the children are largely absent from administrative records and case notes. Nonetheless, it is easy to imagine that a child's experiences must be different depending on the restrictions a CPS safety plan imposes on family visitation, the degree of deference accorded to parental wishes, and the cumulative length of time children must spend away from the familiar presence of their parents, schoolmates, friends, and neighbors.

Critics and proponents of voluntary alternative care agree that the trauma of separation leaves lifelong scars whether the children are removed to public foster care or hosted in voluntary alternative care. To assess the magnitude of the harm, it is important to know the differences, if any, between SFFC and SAU in the total time children are absent from their homes. Adding the duration of hosting by SFFC families to the cumulative time children spent in formal foster care through the end of the study period increases the average number of days children spent out of home from 129 to 147 days for the intervention group. It stays approximately the same at 264 days for the comparison group after taking into account treatment crossovers. Subtracting the two averages yields a net difference of 116 fewer days spent away

from familiar surroundings. In spite of the concerns expressed by advocates that the hidden foster care system invariably leads to indefinite out-of-home stays that can be just as distressing as long-term foster care, the data show there are real differences in the time children spend away from their homes if their families are offered SFFC compared to SAU. At the same time, it must be acknowledged that these differences are not guaranteed. In Section IV, we consider a very different set of family outcomes in Cook County, Illinois, where the time spent in voluntary alternative care appears to have very different consequences for the total time spent away from home compared to downstate Illinois. Before taking up this issue, however, we first estimate the average treatment effect of SFFC on the families and children that actually received the care and support offered by host families.

Effects of the Treatment on the Treated

Consistent with the goals and ideals of SFFC, the accumulated evidence supports the claim that the program in downstate Illinois had a positive impact on the primary outcome of deflecting children from foster care. It also appeared to elevate the chances that the children will be living with their parents at 12 months after being allocated to the intervention or comparison group. These intent-to-treat findings are of great interest to policymakers and administrators because they summarize the net effects of rolling-out SFFC to a selective population of families whom investigators deem good candidates for the intervention. While practitioners can grasp the basic idea of average causal effects, many practitioners still find confusing the claim that the net improvement reflects only the effect of being assigned to treatment and not the effect of actually receiving the treatment. The question they pose is how can the mere offer of treatment possibly have any influence on the actual outcomes that individual families and children experience?

What is frequently left unsaid is that it is not the offer of treatment but what comes after the offer is made that matters. Setting aside possible placebo effects, in order to distill the actual effect of treatment on the treated (what practitioners are interested in) from the average effect of the intention to treat (what policy makers and administrators want to know), it is necessary to make the simplifying assumption that the offer of treatment has no direct bearing on potential outcomes other than their indirect effect through the encouragement of people's participation in the program. It will always be difficult to detect a non-zero summary (average) causal effect of the experimental assignment on outcomes if a truly effective program treats only a small proportion of its intended target population. In this sense, the intent-to-treat effect is a "diluted" summary of what a true average treatment effect on the treated would look like if 100% of the assigned families were to participate fully in the program. The problem that incomplete compliance poses to an evaluation is not that it biases estimation but rather that it becomes increasingly difficult to distinguish collective impact statistically from no difference (zero) as compliance rates trend smaller. Unless sample sizes run into the thousands or the magnitude of the intervention effect is quite large, it can be difficult to detect statistical evidence of superior impact given low compliance rates. If the primary question that policymakers and administrators are interested in answering is whether a promising program makes a collective impact of some practical importance, then in a sense it is irrelevant whether a large or small proportion of assigned subjects fully comply with the assigned treatment. On the other hand, if the primary interest is in how well the intervention works for families that fully participate in the program, a net-impact answer to the ITT question will not suffice.

Fortunately, statisticians and econometricians have devised novel methods for teasing out the effect of the treatment on the treated from the average differences observed in ITT studies.

These methods require making the simplifying assumption that the mechanism of assigning subjects to treatment groups influences individual-level outcomes only by increasing a subject's propensity to participate in the program. One can then conceive of the average difference in ITT outcomes between treatment groups as borne entirely by the subset of families that fully comply with their assigned treatment to the intervention group. Alternating the assignment of families to SFFC by requiring supervisors to press a "randomizer button" should no more affect child welfare outcomes, in and of itself, than flipping a coin, tossing dice, or consulting a table of random numbers. Only insofar as investigators act on the assignment to the intervention group by contacting parents, securing their consent, making a referral, having SFFC locate a suitable host family, and then confirming that the children have moved in with that family, is it reasonable to expect SFFC to have a shot at changing the outcomes for children and their families at the person level.

Excluding the automated assignment mechanism itself as a cause of the differences in outcomes seems a sensible assumption to make in the SFFC evaluation. The more troublesome corollary is that full compliance also assumes that there are no crossovers from comparison to intervention groups. This is where the SFFC evaluation tripses on shakier grounds. Ignoring crossovers might make sense if they are infrequent as they were in downstate Illinois (< 5%). However, when they begin exceeding 10%, as they did in Cook County, treatment crossovers become harder to ignore.

Even if crossovers do not fall below an ignorable threshold, the effect of the treatment on the treated, also called a complier average causal effect (CACE), can be derived from the summary difference between experimental groups irrespective of the degree of program

participation in either group.¹⁴ Compliers in this context refer only to the children of families who were induced to go into hosting because their random allocation called for them to be referred to SFFC. The 11 (5%) out of the 223 children allocated in downstate Illinois to the comparison group who crossed over to hosting during both formative and summative phases are compliers because their families opted for hosting in spite of their allocation to SAU. Excluding the assignment mechanism as a cause and separating out crossovers as a special group, the difference in the estimated number of days that compliers assigned to SFFC might otherwise spend away from their home (including hosting days) widens from an estimated 116 days (the ITT effect) to 187 days (the CACE) during the year following enrollment. This difference of over 6 months is imagined to capture what those very same children might have experienced had their only option been child protective services as usual.

The observed average time in foster care among compliers is 245 days compared to 210 days among the “no-shows” (Bloom, 2005). The children in the kinds of families that participate in SFFC appear to be at greater underlying risk of long-term foster care than other children whose families fail to participate. Electronic case notes suggest that participating families in the intervention group are more socially isolated than non-participants. Administrative data also show their children are younger on average. Given the deficits in bonding social capital among compliers due to the absence or unwillingness of extended family to step in as full-time

¹⁴ The use of the intervention indicator as an instrumental variable to estimate the CACE was suggested by one of the peer reviewers. The first-stage equation of the effect of assignment to the intervention group on the probability of hosting is: $y = 0.049 + .275\text{intervention}$. The t-statistic for the coefficient on the instrumental variable in the first-stage equation is 7.97, which is well above the t-statistic of 3 recommended by Angrist (1996) to satisfy one of the requirements of an instrumental variable. The hosting coefficient in the two-stage-least-squares equation is -187 days, which is significant at the .052 level based on a two-tailed test of significance. The other requirement is that the instrument influences the outcome only through its effect on the propensity to participate in the program. The reasonableness of this assumption is discussed above in the body of the report.

caregivers, a plausible explanation for why SFFC works is that the bridging social capital that isolated families are able to accumulate from participating in SFFC helps them forestall or avoid entirely the loss of their children to long-term foster care. No-shows probably have access to other sources of social capital that enables them to retain custody of their children or retrieve them from public custody more quickly. Extending the follow-up interval beyond one year shows a much larger impact of SFFC on the treated. The Treatment-on-Treated (TOT) difference between groups widens from 157 days to 404 days. This difference represents more than a year's less time that children assigned to SFFC end up spending in formal foster care compared to the time they might have spent if their parents' only option had been child protective services as usual.

SECTION FOUR: REPLICATE AND ADAPT

The purpose of the *Replicate and Adapt* phase is to assess the generalizability (external validity) of the spread of an evidence-supported intervention (ESI) to other populations and settings that differ from the enabling context in which the ESI was originally implemented and found to be effective (Testa, DePanfilis, Huebner, Dionne, Deakins & Baldwin, 2014). In the jargon of implementation science, an enabling context refers to the environment and capacity within a community or formal system, including policy and socioeconomic factors, which make it possible to implement a social innovation with fidelity to its supported logic model (Wandersman, Duffy, Flaspohler et al, 2008).

Throughout this report, we have emphasized the fact that the accumulated evidence in support of the superiority of SFFC over SAU applies only to CPS systems in the balance of Illinois (downstate) outside of Cook County. By the end of the first year of formative evaluation, it had become obvious that the program in Cook was diverging from the SFFC logic model in

several problematic ways. First as alluded to above, the program was failing to achieve key outputs and proximal outcomes. The failings were particularly pronounced in Cook County. Not only were referrals falling below expectations, but the number of treatment crossovers in excess of clerical mistakes was exceeding ignorable thresholds. Whereas downstate no hosting violations occurred in the comparison group through June 30, 2016, 13% of the comparison group in Cook had crossed over to treatment.¹⁵

Second, our discussions with supervisors and front-line staff about the crossover problem unearthed familiar misgivings that we've heard from practitioners in the past. The investigators and supervisors with whom we spoke viewed IDCFS's restricting SFFC only to families "randomized" to the intervention group as unethical. Even though IDCFS leadership backed the use of experimental designs to demonstrate that promising programs truly worked, the more vocal investigators and supervisors defended the violation of experimental protocols as a "badge of honor." From their viewpoint, vulnerable families were gaining access to a valuable resource that a computer algorithm had arbitrarily denied them. *Post-hoc* analysis of the formative data, however, gives a different impression of who is actually losing and gaining from the crossover violations in Cook County.

Unlike the outcomes in downstate counties, foster care removals in Cook County were actually higher in the intervention group than in the comparison group. Our discussions with practitioners brought to light that some CPS investigators in Cook County had come to depend on SFFC as a temporary "holding pen" for keeping children safe while they collected additional evidence to justify taking the children into foster care. Deflecting children to informal alternative

¹⁵ Step5_2_HostingServices_11_8_2020_ArnoldReport line 298-311.

care released investigators from the immediate time pressures that protective custody imposed. Removal; requires court approval within 48 hours of CPS's taking protective custody. It can be difficult to build a convincing case for taking legal custody when the hourglass is quickly emptying. Besides providing an alternative to protective custody, the formative data for Cook County also suggests that SFFC hosting was buying time for investigators to complete their investigations.

During the formative evaluation period, Cook County investigators were taking an average of 20 additional days to complete investigations beyond the statutory limit of 60 days. While the extra time is likely confounded with the greater complexity that hosted cases present, the pattern was opposite to the pattern observed in downstate offices. In downstate Illinois, hosting was associated with investigators' completing investigations on time. On average, hosted cases took fewer days to complete than non-hosted cases in downstate Illinois.

In Cook County, the indirect effect of hosting on removals as mediated by the extra days of investigation beyond the 60-day limit was statistically significant at the $p < .01$ level. For every two weeks beyond the statutory limit, removal odds increased by 9% in Cook County. The path diagram in Figure 6 quantifies the role that the number of days that investigations extend beyond the 60-day limit plays in mediating the relationship between SFFC hosting and foster care removal as moderated by the different enabling contexts in Cook County versus downstate Illinois. We fit this so-called "moderated mediation model" (Hayes, 2013) to formative data to isolate the indirect association between hosting and removal as mediated by the extra days investigators took to complete their investigation.

The way to read the chart is to start with the path coefficient, a , which summarizes the relationship between hosting and investigation days, and the coefficient b , which summarizes the

relationship between investigation days and removal. The product of the two coefficients ($a \times b$) is an estimate of the indirect effect of hosting on removal as mediated by the number of extra days taken to complete the investigation. Because the model includes Cook County as a

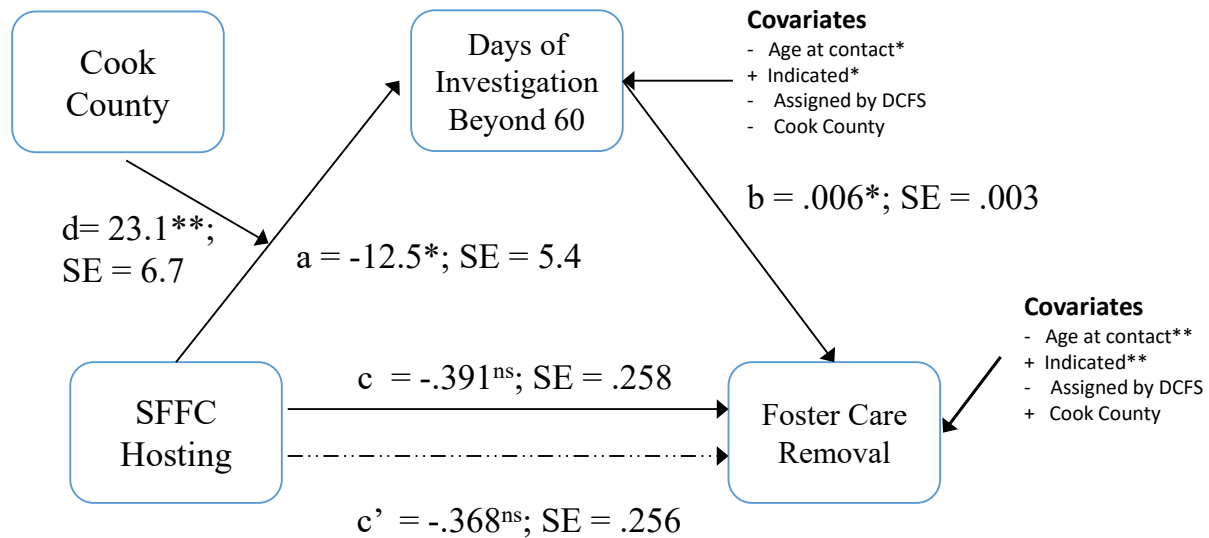


Figure 6. Moderated mediation model of the moderating effects of Cook County on the mediating role of the length of investigations beyond the statutory limit of 60 days in the relationship between SFFC hosting and foster care removal. The dotted line denotes the effect of SFFC hosting on removals when days of investigation are not included as a mediator. Paths b , c , and c' are unstandardized logistic regression coefficients; a is an unstandardized OLS regression coefficient; and d is an interaction term. ** $p < .01$, * $p < .05$.

“moderator” of the path from hosting to extra investigation days, the product of $a \times b$ ($-12.5 \times .006 = -.0075$), is an estimate of the indirect effect of hosting on removals within the enabling context of downstate Illinois.. It indicates that hosting indirectly *decreases* the (log) odds of removal in downstate Illinois. Doing the same calculation for Cook County, but this time adding in the “interaction” term, d ($23.1 - 12.5 = 10.6$) changes the sign from negative to positive for an

the indirect effect of 0.064 ($= 10.6 \times .006$).. This indicates that hosting indirectly *increases* the (log) odds of removal in Cook County. Simulating the indirect effect of hosting for the 20 extra days that it adds to investigations in Cook County increases removal odds by 12%.¹⁶ The indirect effect of hosting in downstate Illinois is the opposite. Hosting is associated with reduced odds of removal.

The change in the sign of the indirect effect of hosting from negative to positive suggests that a different enabling context may be operating in Cook County.¹⁷ The positive role that hosting in downstate Illinois plays in deflecting children from foster care is absent in Cook County. The deflection rates for children in the intervention group were no greater than the rates in the comparison group. The absence of an observed difference may be related to the higher proportion of crossovers in Cook, which makes it harder to detect a true difference between treatment groups. But it may also stem from the different roles that hosting plays in helping investigators fulfill their job expectations as defined by the local enabling context.

Critics of hidden foster care argue that CPS investigators are asked to do an impossible job that pushes them toward making “decisions under a deeply ingrained ‘better safe than sorry’ mentality” (Redleaf, 2018: 3983). In this context, the additional time that SFFC gives an investigator to complete an investigation could actually work at cross purposes with the intended goal of helping parents maintain some semblance of co-control and partnership with CPS in

¹⁶ ModeratedMediationModel_SFC_Hosted.xlsx. Tab FC_PCRemoval_M7

¹⁷ One of the reviewers found it surprising that foster care placement was reported to be higher in the intervention group than the comparison group in Cook County. However, the difference was trivial. Even if the difference were statistically distinguishable from zero, we should not interpret the change in sign as causal because hosting involves a non-random subset of families that were allocated to the intervention group. Because of the confounding potential due to selection bias, the correct inference to draw is that the spread of SFFC beyond downstate Illinois should be delayed until fidelity to program’s logic model can be improved.

shaping a trajectory of improvement that serves the needs of children and families rather than the needs of a bureaucracy to minimize agency liability. The different consequences associated with hosting in downstate Illinois compared to Cook County indicate that resolution of this tension does not inevitably tip in favor of unfettered discretion and abuse of state power that denies parents their fundamental rights. The local enabling context can insulate investigators and families from this competing pressure to facilitate co-creation and partnership. Differences in local enabling context may help explain the differences in how SFFC played out in downstate Illinois compared to Cook County. During formative evaluation, there were no hosted cases downstate where investigations lasted longer than 90 days. In Cook County, 26% of hosted cases extended 30 or more days beyond the 60-day limit. Relying on SFFC to extend the time desired to gather additional evidence without needing to petition the court for custody of the children may be a perfectly justifiable adaptation of the program in the larger scheme of things. But this particular deviation from the logic model is contrary to the SFFC theory of change and is clearly not the intended purpose program developers had in mind when they launched their movement.

SECTION FIVE: APPLY AND IMPROVE

After answering the primary research question about whether a social innovation works (accuracy) and offering a plausible interpretation for why it works (communicability) the last condition that Stinchcombe (2002) identifies as enhancing the prospects for sustainable success is whether there are discernible pathways for increasing the fidelity of implementation to its logic model (trajectory of improvement). Addressing this question involves taking account of the additional measures listed under the heading of *Outputs* in Figure 3 above).

As discussed in the *Develop and Test* section, there were two ways that families could take SFFC up on its offer to find host families. Either CPS investigators could refer them to SFFC if the automated assignment mechanism allocated them to the intervention group or the families could show up at SFFC's door and have intake staff randomize them to the intervention or comparison group. Not unsurprisingly, families who were randomized to the intervention group after showing up at SFFC's offices were far more willing to participate in the program (56%) than the families whom IDCFS allocated to the intervention group (23%). SFFC was also better able to prevent treatment crossovers. Whereas 14% of the families that IDCFS assigned to the comparison group managed to circumvent protocols and obtain hosting support, only one SFFC comparison case (1%) crossed over to treatment. This occurred because IDCFS and LYDIA made an exception to protocol and waived the family into the program.

Increasing compliance rates among the families that SFFC assigned to the intervention group suggests one possible trajectory for improvement. However, closer inspection of the reasons for incomplete compliance among parents that showed up at the SFFC offices suggests that the observed 56% compliance rate may be as good as it can get. Among the 41 no-shows, 15 (36%) had relatives step forward, 10 (25%) had their children removed by CPS within days of making the referral, 4 families (10%) withdrew their consent, and the remaining 12 (29%) were pulled back because other accommodations became available, a suitable host family couldn't be found, or the parents withdrew for other reasons.

At just 23% compliance, the low participation of families alternated to the intervention group by IDCFS suggests that there may be room for raising compliance levels among families referred directly by IDCFS. Recall that the alternation process consisted of two steps: 1) automated allocation of the families to intervention and comparison groups that investigators

deemed good candidates for SFFC; and 2) the agreement of parents to contact SFFC for help in locating host families. After filtering out duplicate assignments and other protocol deviations, there were a total of 142 families that IDCFS allocated to the intervention group across all three phases of the evaluation. Only 32 families (23%) fully complied with the intended treatment; another 54 (38%) didn't consent to CPS's making the referral; 35 (25%) of the families that did consent failed to make contact with SFFC; 13 (9%) were denied assistance because of miscommunication, clerical errors, or difficulties in locating suitable homes; and the remaining 8 (5%) pulled back for other miscellaneous reasons. The large number of families who initially agreed to the referral but failed to make contact suggests possible room for improvement. Before investing resources into elevating compliance rates, however, it is probably best to examine the different enabling contexts, as described in the prior section, in which these corrective actions are to be implemented.

DISCUSSION

The purpose of this study is to provide a summary judgement of the effectiveness of a promising social innovation, SFFC, in ensuring children's safety and ultimately keeping them together with their parents or relatives. The accumulated evidence gathered from this investigation shows that SFFC had a positive impact on the primary outcome of deflecting children from foster care and on the secondary outcome of their living with parents or relatives without state oversight at 12 months after enrollment. The chances of removal were significantly lower for children receiving SFFC compared to services as usual. There were also no concerning differences in the recurrence of maltreatment between the two treatment groups.

A secondary purpose of this study, which is why the Arnold Foundation funded the evaluation, is to demonstrate the feasibility of incorporating rigorous evaluation methods at low cost into routine CPS quality improvement operations. The proposal was that CPS agencies could improve operations by programming an unbiased allocation mechanism into their information systems and tracking outcomes with existing administrative data. Taking this so-called experimentalist approach to compliance management is a novel way of ensuring results-oriented accountability under a federal consent decree. As noted in the introduction, instead of ordering compliance with a fixed set of solutions, the experimentalist approach recognizes that court-ordered initiatives are provisional. The court periodically reviews progress using an agreed upon set of performance metrics. The CPS agency conducts rigorous experimental and quasi-experimental evaluations that can either support continued rollout of the initiatives or call for their replacement based on the best available empirical evidence.

A potentially valuable contribution of this study is its use of the Bayesian paradigm to statistically evaluate program impact (Chen et al., 2020). This paradigm aligns well with the three criteria of cognitive adequacy, communicability, and trajectory of improvement, which Stinchcombe identifies as the hallmarks of a well-built formal system for the governance of effective social action. Its cognitive adequacy is amply documented beginning with Bayes and Price (1763) and applied most recently to intervention research (Chen & Ansong, 2019; Chen & Fraser, 2017). It is more easily communicated to policymakers, judges, administrators, and practitioners who must understand and transmit the abstractions of the formal system to funders and recipients of the action. It dispenses with the awkwardness of the frequentist approach that relies on a double negative that assumes that the null hypothesis is true and studies whether the data are inconsistent with this assumption (Iverson, 1984). Lastly, it projects a trajectory of

improvement that is cumulative and more in keeping with the phased approach to evidence building than the conventional frequentist approach that separately analyzes formative and summative data. As summarized by Iverson (1984):

The fact that we can express our prior opinion of a parameter in a prior distribution means that we do not have to start at the very beginning again each time. There are many times when we do have prior information, and that makes it possible to use prior informative distributions. With informative prior distributions we get posterior distributions that are more peaked and with smaller variances and therefore get shorter Bayesian probability intervals for the parameters than we do with non-informative prior distributions.

These statistical advantages along with the use of one-tailed tests to assess intervention superiority lowers the costs and burdens of conducting rigorous evaluations as part of routine quality improvement operations.

The implementation and evaluation of SFFC provide valuable insights into the multiple challenges that a CPS agency confronts when attempting to incorporate a court-ordered, experimentalist approach into routine operations. The IDCFS already had experience with rigorous evaluations under its title IV-E waiver demonstrations and federal research grants. A few demonstrations exceeded expectations, such the subsidized guardianship experiment that helped change federal policy, but most yielded results that contradicted the beliefs and hopes of many practitioners and policymakers.¹⁸ As with social experiments in general, the agency's score card was no worse than usual: one success for every four failed attempts. Given these grim odds

¹⁸ A list of key initiatives in Illinois, which had been announced with great fanfare but didn't pan out as hoped, includes: 1) the IV-E training waiver demonstration; 2) the permanency planning TARGET initiative, the Illinois

of success, however, it can be grueling to sustain an institution's commitment to an experimentalist approach, especially when there are multiple transitions in leadership as occurred at IDCFS during the course of this evaluation. Taking five or more years before either welcomed or disappointing news can finally be unveiled drains the enthusiasm of even the most ardent champions of evidence-based policymaking.

In addition to Rossi's (1978) "iron law," which states that "the expected value for any measured effect of a social program is zero," SFFC faced its share of regular implementation challenges that threaten the chances for detecting success even when programs are truly superior to SAU. As with other innovations, program developers and sponsors underestimated the difficulty of reaching the intended target population of families and children. Even though the Department's efforts at increasing referrals eventually paid off during the third and fourth quarters of formative evaluation, enthusiasm waned just as the project entered the critical phase of summative evaluation. Even if the high volume of referrals had continued, however, the irritation and frustration of CPS staff's needing to conform to the dictates of a "randomizer button" would likely have generated continued resistance and protocol deviations. In Cook County, Illinois, the number of treatment crossovers eventually got so high that the investigators and IDCFS leadership closed down the experiment in this state's largest population center.

Admittedly it is difficult to hold back judgment of the validity of a service option when many folks believe they already can "see and feel" that it alleviates the panic and anxiety that all parents experience when they confront the prospect of losing custody of their children. It is a tough sell to persuade them that a RCT is needed to demonstrate effectiveness before public funds can be spent on the innovation. An experimentalist approach provides the strongest summary evidence of the effectiveness of a promising program if not at the individual family

level then at the aggregate level. It is the next best approximation to answering the desired but impossible “counterfactual” question of what might have happened if the exact same family that received the intervention simultaneously were not offered this option.

A rigorous field experiment of a program option like SFFC, which had already been in place for a decade, creates special challenges that threaten but not completely undermine the ability to draw meaningful causal inferences about whether a program works as intended. In spite of the lower than desired levels of compliance, the strong effect of the intervention on the primary outcome of deflecting children in downstate Illinois from protective custody and foster care still shines through. At the same time, it is sobering to contemplate that the news might have been different if we hadn’t first conducted a formative evaluation that raised doubts about whether the appropriate enabling context existed in Cook County for implementing the program with fidelity to its logic model.

The absence of a similar positive impact in Cook County is a riddle in need of further unraveling. It is also a forewarning about the potential perils of rolling out promising but untested programs to vulnerable populations of families and children. Anecdotal evidence suggests that some investigators in Cook County had come to depend on SFFC as a temporary “holding-pen” for keeping children safe while they gathered additional evidence to support taking them into foster care. For this and other reasons, Cook County was dropped from the summative evaluation. But even in downstate Illinois, where the results were more in line with expectations, we were left wondering whether the benefits could have been greater if compliance rates had been higher. As noted above, many families assigned to SFFC in downstate Illinois did not ultimately use the service, either because they withheld consent or because there was incomplete follow-thru by investigators even after parents consented to the referral. Incomplete

compliance with assigned treatments doesn't compromise the internal validity of the estimated effects of offering the SFFC option. It does weaken the ability to detect the full impact of the intervention for families who actually receive the treatment. Issues of non-compliance and partial engagement in services are endemic challenges in the implementation of child welfare, mental health, and medical interventions. While we have examined some of the reasons for non-compliance anecdotally, a deeper inquiry into the "deep structure" of problem (Resnicow, Soler, Braithwaite, Ahluwalia, & Butler, 2000) is warranted in order to inform future quality improvement strategies.

A basic rule of implementation science is that achieving outcomes requires not just valid interventions and competent implementation but also an enabling context, which enhances the infrastructure, skills, and motivation of practitioners and make possible the system's implementation of the theory of change with fidelity to the intended plan and purpose. To make a practically important impact, all three components must be operating adequately. The malfunctioning of any one component can compromise the entire enterprise. The federal judge who approved the *B.H.* implementation plan stipulated the importance of adhering to the tenets of implementation science in order to succeed at pulling off a complex and massive reform of the Illinois CPS system. The challenges in applying the experimentalist approach to the implementation and evaluation of SFFC indicate that there is still a long distance to travel before a phase-based approach to compliance management can be fully realized.

The challenges are best illustrated by the staunch resistance that the dissemination of SFFC met in the state of New York. In January of 2020, the New York Office of Children and Family Services issued draft regulations that, if approved, would have allowed nonprofits in the state to begin operating the SFFC model. The release sparked over a hundred written responses.

A diverse array of legal, judicial, voluntary associations, and family advocacy organizations lined-up on opposite sides of what became a highly polarized debate over the unknown risk and benefits of hidden foster care versus the risks and benefits of formal foster care (Fitzgerald, 2020).

A full airing of the issues is beyond the scope of this report. Gupta-Kagan (2020) provides a concise summary of differing points of view across the ideological spectrum. He notes that those distrustful of the hidden foster care system have written critically of its adverse implications for family integrity, which he correctly notes changes the person with whom children live but then embellishes by adding “often permanently” (Gupta-Kagan, 2020: 852). The accumulated evidence from this study, however, shows that a permanent change in physical custody rarely occurs outside of the formal system. A far greater risk to family integrity is the inordinately lengthy stays in foster care that typically accompany the removal of children into formal foster care in Illinois. There are other stakeholders, he acknowledges, who are concerned for the opposite reason, namely that CPS agencies defer to family integrity too much. Their criticism is that the hidden foster care system leaves children in what they see as unsafe situations without the safety precautions of formal foster care. But again the evidence available from this study shows that there is no greater recurrence of maltreatment in host homes than in licensed foster homes.

A major advantage of this study in informing the debate over the risks and benefits of informal alternative care is that it “controls” for the confounding influence of kinship. Nearly all previous studies compare formal foster care by strangers with informal alternative care by relatives. We are among the first to concur with critics who say too many CPS systems misuse the option of kinship care by withholding pertinent information about service and custody

options available to them. Especially problematic is the withholding of information about the financial support and long-term guardianship assistance that are available to formal kinship caregivers (Testa, 2020).

In the case of SFFC, the financial need of the caregiver is not at issue. Hosting is as close to an act of selfless hospitality as you can find in contemporary child welfare practice. Whereas it may be unjust to use the coercive power of the state to exploit the kinship altruism of needy relatives, it is a wholly different matter to accept the voluntary contributions of compassionate strangers. On the other hand, non-profits must always remain vigilant about the misuse of the option by both personal and collective agents. As we saw in the case of Cook County, the CPS agency may have been defecting from the true purpose of the SFFC program, which is intended to be less invasive and less threatening of family autonomy than SAU. However, as Gupta-Kagan (2020: 871) states, “the remedy is not a prohibition on the practice, but more process,” or as Stinchcombe (1985) would have put it—more formality.

Gupta-Kagan (2020) provides a helpful trajectory of improvement that is consistent with the findings of this study. As noted above, the Illinois data show that most children (70%) stayed in SFFC homes for fewer than two months. Setting a three-month maximum length of time for SFFC hosting, as suggested by Gupta-Kagan (2020), after which time the child must be reunified or the courts must become involved, seems like a reasonable place to start. Gupta-Kagan (2020) offers other recommendations for improving the procedural and substantive rationality of the CPS system.

NEXT STEPS

While the impact of SFFC on the primary outcome of deflection from formal foster care in downstate Illinois is quite promising, the limited generalizability of the findings calls for further research on the effectiveness and implementation of SFFC in different contexts. The formative evaluation showed no effects of SFFC on deflection rates in Cook County. Discussions with DCFS staff and *post-hoc* analyses that showed indirect effects of hosting in raising the likelihood of removal (as mediated by the extra days Cook County investigators took to complete their investigation) indicate that investigators sometimes utilize SFFC to keep children safe while they take more time to collect the evidence needed to justify placing children. While contrary to the SFFC theory of change, such decisions may make sense for Cook County investigators given that the county registers the lowest per capita rates of removal in the country (School of Social Work at The University of North Carolina at Chapel-Hill et al., 2018). Given that Cook County juvenile court's thresholds for removal are relatively high, it is possible that there are proportionately fewer cases in which temporary hosting by SFFC can successfully prevent foster care placement.

The limited generalizability of findings point to the broader importance of understanding how investigators make referral decisions and why that is essential to understanding SFFC implementation and outcomes. It is notable that a large majority of downstate control group cases did not result in protective custody (76% in the summative evaluation sample) or foster care placement (61%) within one year. From an outcomes evaluation perspective, this indicates that a majority of participants were not at risk of primary outcomes that were being measured, potentially inhibiting the ability of interventions to demonstrate improvements relative to the counterfactual. From a decision-making perspective, these findings strongly suggest that

placement prevention was often not the primary reason for referral. Especially, but not solely in areas with low per capita placement rates, SFFC may sometimes be viewed by investigators as preferable to existing options that are suboptimal but frequently used in practice for cases in which children are not normally placed. For example, SFFC might reasonably be viewed as a better option than leaving the children in a homeless shelter with a mother who fears domestic violence or preferable to leaving them with relatives who lack the appropriate accommodations to look after the children. Evidence for this type of decision-making strategy comes from experimental evaluations in the 1990s of another placement prevention program, family preservation services (FPS). The FPS evaluations in Illinois and elsewhere showed that, contrary to the theory of change assumption that children were at imminent risk of placement, very few control group cases were actually placed within three months of assignment (Schuerman, Rzepnicki, & Littell, 1994; Westat, Inc, Chapin Hall, James Bell & Associates, 2001). In Illinois, it appeared that investigators often supported the ideals of family preservation but used referrals for family preservation service to better address the considerable risks and needs of intact families, not to prevent imminent placement (e.g., Schuerman, Littell, Rzepnicki, & Budde, 1992).

There was some anecdotal evidence in our discussions with DCFS staff that investigators referred some families at an early point in the investigation, and that SFFC appeared to serve as a useful backup option in case it was needed later in the investigation. For example, the investigator may have made the SFFC referral prior to exploring all options for willing kin caregivers or prior to knowing whether a parent would be referred for outpatient or inpatient substance abuse or mental health services. Given that SFFC was ultimately not utilized in some

or many of the intervention group cases, this referral strategy may have contributed to lowering compliance rates and diluting the effects of SFFC on outcomes in the ITT analyses.

Other important topics related to SFFC implementation and its theory of change that were largely beyond the scope of the current evaluation include the clinical reasons for referral, the extent to which SFFC enhances parenting competences, the frequency of movements within the host-family network, safety planning and other issues related to concerns over hidden foster care, and topics related to the differences in the racial identities and religious affiliations of providers and receivers of hosting services.

Examining the types of risk and safety issues that prompt referrals, and the relationship of these characteristics to SFFC engagement (compliance) and outcomes will likely provide better information about the need for SFFC and the types of cases for which it is most and least effective. These reasons for referral to SFFC represent the central clinical challenges in child welfare, including child safety issues, intimate partner violence, homelessness, psychiatric hospitalizations and mental health crises, substance abuse, and medical problem.

Essential features of the SFFC theory of change require further inquiry. As described earlier, given the social isolation and lack of bonding social capital in many families involved in the child welfare system (Testa et al, 2010), SFFC can potentially provide bridging social capital that connects families to new and sustaining relationships and resources. Host families can potentially reduce social isolation and provide these network connections to parents and children both during hosting and after hosting ends. Given that the risk and safety issues noted above are often chronic, temporary hosting in and of itself is likely to be insufficient in many cases. We identified examples of bridging social capital becoming bonding social capital when host families became fictive kin foster parents for some children. However, the extent and quality of

host family involvement with parents and children during and following hosting has not been studied systematically. Further, the extent to which SFFC hosts provide coaching or support (e.g., during visits) related to parenting is unknown. Thus, it is not clear whether and how often SFFC actually serves as a parenting intervention. Similarly, it would be helpful to know whether SFFC support increases access to resources and engagement in effective clinical services.

We have learned that SFFC is not immune to the problem of placement instability that plagues the foster care system. Our record reviews revealed multiple examples of children changing host families or going to another host family temporarily for respite care. The early descriptive evaluation found that 12 percent of children referred by DCFS stayed with at least two hosts during their first SFFC spell (JPA: Budde et al, 2009). As with the growing body of research in the child welfare system, there is a need to better understand the timing, frequency, and reasons for (e.g., factors related to the child, family, and host) host changes, and the extent to which changes are predictive of subsequent outcomes.

Gupta-Kagan (2020) highlights the deflection of large numbers of children into informal placements with relatives and emphasizes that these arrangements are not adequately monitored by the child welfare system and often deprive parents of legal rights. He raises the concern that parents may be “coerced” into transferring custody of their children based on the threat of foster care placement and court involvement. Safety plans are often the procedural vehicle through which these informal placement arrangements are made. Given the purposeful lack of court involvement in SFFC hosting, similar concerns understandably arise. We strongly support the importance of open and transparent tracking of the utilization and outcomes of SFFC referrals by the child welfare system. In addition, researchers and child welfare professionals should carefully examine three important issues that have emerged from this dialogue.

First, as we note above, child welfare professionals should attempt to locate viable relatives and fictive kin for placements prior to referring to SFFC. The limited review of available case records we examined showed that investigators did in fact attempt to locate and engage relatives prior to SFFC referrals. This inquiry should be expanded and meeting this criterion for referral should become part of ongoing quality assurance and quality improvement efforts. Second, we should track and study whether SFFC unintentionally decreases the utilization of formal kinship and fictive kin placements, and hence, the family's access to formal supports and services available through the foster care system. If reductions in formal kinship care result from SFFC, it would be interesting to study the effects of these changes on reunification and long-term permanency outcomes. Third, and perhaps most importantly, we should examine the utilization of safety plans in SFFC cases with the aims of understanding the extent to which parental participation in SFFC results from threats of formal placement and court involvement, and of determining the relationship of safety plans to subsequent outcomes.

Finally, given the history of racism in our country and considerable racial disparities in the child welfare system, it is essential to acknowledge and study issues in SFFC related to race and religion in assessing key indicators of implementation fidelity (e.g., host family involvement with parents) and variation in outcomes. While the downstate child sample was 60% African American, only about 5% of host families are African American. As with any intervention, it will be important to examine the effects racial matching and differences on the felt experiences and outcomes of minority families, especially rates of reunification following hosting. Further, at least in Cook County, there is concern among some prominent African- American child welfare professionals about referring African-American families to SFFC given the racial makeup and religious background of SFFC staff and hosts. SFFC professionals working at the national level

note that they often face similar concerns and skepticism in other communities. While the nature, extent, and impact of such concerns about SFFC have not been systematically examined, these concerns need to be heard and addressed openly and on an ongoing basis. Failure to do so may inhibit the utilization and effectiveness of SFFC in primarily African American communities. It is possible for example, that race related concerns about SFFC contributed to the lower than expected utilization rate in Cook County.

SFFC's parent organization, LYDIA, is a Christian non-profit that extends "the truth of Scripture and God's love to people of all religious, racial, and economic backgrounds." The impressive ability of SFFC to recruit large numbers of families to host children and support families with considerable needs likely stems primarily from "religious social capital," which is "defined as the social resources available to individuals and groups through their social connections with a religious community" (Maselko, Hughes, & Cheney, 2011, p. 759). At the same time, especially in the context of frequent differences in the race and ethnicity of host families and the children and parents they serve, it is important to monitor the extent to which parent and child preferences are supported by host families. This should include tracking how SFFC staff and host families address dilemmas that can arise when parents or children have differing views of religion, or cultural or personal perspectives that conflict with religious beliefs of the host families (e.g., Schatz & Horejsi, 1992; Anderson & Mikula, 2002). Of particular concern would be cases in which the parents are atheists or Muslims, or instances, or when parents or children identify as LGBTQ.

CONCLUSION

Social spending in the United States is undergoing a major shift in accountability from expert judgement and public opinion on what is believed should work to rigorous scientific evidence of what is empirically demonstrated to work. The Family First Prevention Services Act (Family First), requires that no less than one-half of future spending on services authorized under the Act must meet the legislation's highest evidence standard for well-supported practice. This high standard requires evidence of superiority of the practice to an appropriate comparison practice in a usual care or practice setting. To meet this standard, superiority must be supported by at least two rigorous randomized controlled trials (or, if not available, studies using a rigorous, quasi-experimental research design). The trials must also demonstrate a sustained effect of at least six months beyond the end of treatment, when compared to a control group, and one of them has to demonstrate a sustained effect of at least one year.

As of December of 2020, only 39 (8%) of the 506 programs cataloged on the California Evidence-Based Clearinghouse for Child Welfare (2020) met the standard of well supported by research evidence. Only six of these well-supported interventions were specifically designed or commonly used for children and families served by the CPS system. A major reason for the dearth of well-supported interventions is that a high volume of promising innovations are later found at summative evaluation to have had no detectable impact or at best small impacts that are not commensurate with their costs (Epstein & Klerman, 2012). As a benchmark, the Laura and John Arnold Foundation (2018) and Jim Manzi (2016) report data showing that the odds of progressing through all phases of evidence building from promising innovation to replicable program for most of the RCTs conducted in business, education, criminology, political science, and economics seldom best one improvement for every four unsuccessful attempts. Given the 4:1

odds against replicable improvements, it is essential that the child welfare field enhance its capacity for building evidence rigorously and efficiently as a routine part of its continuous quality improvement (CQI) operations.

Applying A Framework to Develop, Test, Spread and Sustain Effective Practice in Child Welfare to the implementation and evaluation of the promising innovation of SFFC illustrates how the supply of evidence-supported interventions can efficiently be winnowed and eventually enlarged through rigorous CQI in child welfare. By embedding an unbiased assignment mechanism in routine CQI operations and tracking results with existing administrative data, it is possible to guide evidence building through successive phases of increasingly generalizable validity. Learning at earlier rather than later phases that a promising innovation is having difficulty passing through the validity “tollgates” of its own logic model helps conserve and redirect resources before too much time and effort are misspent rolling out an innovation that is unlikely to show positive improvements when appropriately evaluated. The SFFC evaluation demonstrates the merits of this approach. Had we followed the usual process of conducting a summative evaluation without first testing it in a formative evaluation, we would have likely misinterpreted the resulting null findings as evidence of overall program ineffectiveness. While *post-hoc* analyses might have revealed that the null findings were limited to Cook County because of implementation and cross-over problems, it is far better to identify threats to validity in advance so that only those programs that demonstrate implementation integrity and trend toward statistical significance are pre-registered for summative evaluation. Building replication into the evidence-building process makes it more likely that positive summative improvements will be reproducible in future studies.

A major dilemma that reform-minded practitioners, administrators, and policymakers face is knowing when to enlarge the scope of the corrective actions so that deficiencies are brought under the governance of the formal system and when to address them by delegating discretion to the informal competencies of kin, friends, neighbors and good Samaritans external to the agency (voluntary alternative care) or to expert systems certified elsewhere (such as child abuse pediatrics). Many of the current controversies in child welfare are derivatives of this more general dilemma of formalization. Applying *A Framework to Develop, Test, Spread and Sustain Effective Practice in Child Welfare* to the implementation and evaluation of promising innovation offers one option for resolving the dilemma. Admittedly, this route to well-supported interventions is an arduous path to climb. By following a phase-based approach to evidence building, this study demonstrates that it is possible to fill the need at low cost and to satisfy the rigorous evidence standards required by new federal legislation and results-oriented consent decrees.

REFERENCES

- Adamson, J., Cockayne, S., Puffer, S., Torgerson, D.J. (2006). Review of randomised trials using the post-randomised consent (Zelen's) design. *Contemporary Clinical Trials*, 27, 305-319.
- Allison, P. D. (2000). *Survival analysis using the SAS® system: A practical guide*. Cary, NC: SAS Institute, Inc.
- Anderson, D. (2014). The church and child protection: The safe families story: Volunteers offer hospitality to families in need. *Child Abuse & Neglect*, 38(4), 612-617.
- Anderson, G. R., & Mikula, J. (2002). Spirituality and religion in child welfare practice. *Christianity and social work*. Botsford, CT: North American Association of Christians in Social Work.
- Argyris, C., & Schön, D. A. (1996). *Organizational learning II: theory, method and practice*. Reading, Mass: Addison-Wesley.
- Azuero, A. (2016). A note on the magnitude of hazard ratios. *Cancer*, 122(8), 1298–1299.
<https://doi.org/10.1002/cncr.29924>
- Bartholet, E. (2012). Creating child-friendly child welfare system: Effective early intervention to prevent maltreatment and protect victimized children. *Buffalo Law Review*, 60(5), 1323-1374.
- Baumrind, D. (2013). Is a pejorative view of power assertion in the socialization process justified? *Review of General Psychology*, 17(4), 420-427. doi:10.1037/a0033480

Bayes, T., & Price, R. (1763). An essay towards solving a problem in the doctrine of chance. By the late Rev. Mr. Bayes, communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S. *Philosophical Transactions of the Royal Society of London*, 53, 370–418.

doi.org/10.1098/rstl.1763.0053

Bloom, H. S. (2005). *Learning more from social experiments: Evolving analytic approaches*.

New York: Russell Sage Foundation.

Bottoms, B. L., Shaver, P. R., Goodman, G. S., & Qin, J. (1995). In the name of god: A profile of religion-related child abuse. *Journal of Social Issues*, 51(2), 85-111.

doi:10.1111/j.1540-4560.1995.tb01325.x

Boutron, I., D. Moher, D. G. Altman, K. F. Schulz, and P. Ravaud. 2008. Methods and processes of the CONSORT group: Example of an extension for trials assessing nonpharmacologic treatments. *Annals of Internal Medicine*, 148, W60–W66.

Budde, S., Thompson, R., Go, L., Douthitt, C. & Pryor, S. (2009). *Safe Families for Children evaluation report*. Chicago: Juvenile Protective Association.

California Evidence-Based Clearinghouse for Child Welfare, accessed on December 30, 2020,

<http://www.cebc4cw.org/>

Campbell, D. T. (1969/1988). Reforms as experiments. In E.S. Overman (ed.) *Methodology and epistemology for social science: Selected papers, Donald T. Campbell*. Chicago: The University of Chicago Press.

- Chalmers, I. (2011). Why the 1948 MRC trial of streptomycin used treatment allocation based on random numbers. *Journal of the Royal Society of Medicine*, 104(9), 383–386.
<http://doi.org/10.1258/jrsm.2011.11k023>
- Chen, D. G., & Ansong, D. (2019). Bayesian spatial-temporal modeling of space and time dynamics: A practical demonstration in social and health research. *Journal of the Society for Social Work and Research*, 10(2), 275–299. <https://doi.org/10.1086/703444>
- Chen, H., Cohen, P., & Chen, S. (2010). How Big is a Big Odds Ratio? Interpreting the Magnitudes of Odds Ratios in Epidemiological Studies. *Communications in Statistics - Simulation and Computation*, 39(4), 860–864. <https://doi.org/10.1080/03610911003650383>
- Chen, D. G., & Fraser, M. (2017). A Bayesian approach to sample size estimation and the decision to continue program development in intervention research. *Journal of the Society for Social Work and Research*, 8(3), 457–470. <https://doi.org/10.1086/693433>
- Chen, D. G., Testa, M. F., Ansong, D., & Brevard, K. C. (2020). Evidence Building and Information Accumulation: Using the Bayesian Paradigm to Advance Child Welfare Intervention Research. *Journal of the Society for Social Work and Research*, 11(3), 483–498. <https://doi.org/10.1086/711376>
- Coleman, J. S. (1982). *The asymmetric society*. Syracuse, NY: Syracuse University Press.
- de Boer, M. R., Waterlander, W. E., Kuijper, L. D. J., Steenhuis, I. H. M., & Twisk, J. W. R. (2015). Testing for baseline differences in randomized controlled trials: An unhealthy research behavior that is hard to eradicate. *International Journal of Behavioral Nutrition and Physical Activity*, 12(1), 4. doi:10.1186/s12966-015-0162-z

- DePanfilis, D. & Zuravin, S.J. (1998). Rates, patterns, and frequency of child maltreatment recurrences among families know to CPS. *Child Maltreatment*, 3(1), 27-42.
- Duerr-Berrick, J. (2018). *The impossible imperative : Navigating the competing principles of child protection*. New York, NY: Oxford University Press.
- Dupuy v. Samuels*. (2006). 465 F.3d757 (7th Circuit).
- Epstein, D., & Klerman, J. A. (2012). When is a program ready for rigorous impact evaluation? The role of a falsifiable logic model. *Evaluation Review*, 36(5), 375–401.
<https://doi.org/10.1177/0193841X12474275>
- Fitzgerald, M. (2020). Faith-Based Movement to ‘Host’ At-Risk Kids Hits Opposition in New York. (June 17). *The Imprint*. <https://imprintnews.org/news-2/host-home-safe-families-hits-opposition-new-york/44553>
- Fluke, J., Edwards, M., Bussey, M., Wells, S., & Johnson, W. (2001). Reducing recurrence in child protective services: Impact of a targeted safety protocol. *Child Maltreatment*, 6(3), 207-218. doi:10.1177/1077559501006003002
- Fostering Court Improvement. (2013). *State websites. Illinois: Victims removed to foster care* (Archived).
- Framework Workgroup. (2014). *A framework to design, test, spread, and sustain effective practice in child welfare*. Washington, D.C.: Children’s Bureau, Administration for Children and Families, U.S. Department of Health and Human Services.

- Fuller, T., Nieto, M., & Zhang, S. (2013). *Differential Response in Illinois: Final Evaluation Report*. Urbana, IL: Children and Family Research Center, School of Social Work, University of Illinois at Urbana-Champaign.
- Gerber, A.S. & Green, D.P. (2012). *Field experiments: Design, analysis and interpretation*. New York: W.W. Norton & Company.
- Greenberg, D., Shroder, M., & Onstott, M. (1999). The Social Experiment Market. *The Journal of Economic Perspectives*, 13(3), 157-172.
- Gupta-Kagan, J. (2020). America's hidden foster care system. *Stanford Law Review*, 72(4), 841-914.
- Hayes, A.F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression based approach*. New York: The Guilford Press.
- Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, 81(396), 945-960.
- Hróbjartsson, A., Gøtzsche, P. & Gluud, C. (1998). The controlled clinical trial turns 100 years: Fibiger's trial of serum treatment of diphtheria. *British Medical Journal*, 317, 1243-1245.
- Institute of Medicine. (1994). *Reducing risks for mental disorders: Frontiers for preventive intervention research*. Washington, DC: National Academy Press.
- Iverson, G.R. (1984) *Bayesian statistical inference*. Newbury Park, CA: Sage.
- Kempe, C.H., Silverman, F.N., Steele, B.F., Droegemueller, W. & Silver, H.K. (1962). The battered-child syndrome. *Journal of the American Medical Association*, 181, 17-24.

Kosanovich, A. & Joseph, R.M. (2005). Child welfare consent decrees: Analysis of thirty-five court actions from 1995 to 2005. Washington DC: CWLA and ABA Center on Children and the Law.

Laura and John Arnold Foundation. "How to Solve U.S. Social Problems When Most Rigorous Program Evaluations find Disappointing Effects (Part One in a Series)," Straight Talk on Evidence, accessed March 27, 2018, <http://www.straighttalkonevidence.org/2018/03/21/how-to-solve-u-s-social-problems-when-most-rigorous-program-evaluations-find-disappointing-effects-part-one-in-a-series/>

Manzi, J. (2016). Luncheon Symposium- Rapid Cycle Testing: Lessons from the Private Sector, accessed March 27, 2018, <http://www.recsconference.net/2016videos.htm>

Maselko, J., Hughes, C., & Cheney, R. (2011). Religious social capital: Its measurement and utility in the study of the social determinants of health. *Social science & medicine*, 73(5), 759-767.

Melton, G. B., & Anderson, D. (2008). From safe sanctuaries to strong communities: The role of communities of faith in child protection. *Family & Community Health: The Journal of Health Promotion & Maintenance*, 31(2), 173-185.
doi:10.1097/01.FCH.0000314577.48603.b0

Meitner, H. & Albers, M. (2012). *Introducing safety-organized practice*. Madison, WI: National Council on Crime and Delinquency, Children's Research Center.

Mercy, J. A., Rosenberg, M. L., Powell, K. E., Broome, C. V., & Roper, W. L. (1993). Public health policy for prevention of violence. *Health Affairs*, 12(4), 7-29.

- Moher, D., Hopewell, S., Schulz, K. F., Montori, V., GÃtzsche, P. C., & Devereaux, P. J. (2010). CONSORT 2010 Explanation and Elaboration: Updated guidelines for reporting parallel group randomised trials. *J Clin Epidemiol*, 2010(Journal Article).
<https://doi.org/10.1016/j.jclinepi.2010.03.004>
- Mouselis, N.P. (1971). *Organization and bureaucracy: An analysis of modern theories*. Chicago: Aldine Publishing Company.
- Noonan, K.G., Sabel, C.F., & Simon, W.H. (2009). Legal accountability in the service-based welfare state: Lessons from child welfare reform. *Law & Social Inquiry*, 34 (3), 523-568.
- Parsons, T. (1951). *The social system*. New York: The Free Press.
- Resnicow, K., Soler, R., Braithwaite, R. L., Ahluwalia, J. S., & Butler, J. (2000). Cultural sensitivity in substance use prevention. *Journal of Community Psychology*, 28(3), 271–290. [https://doi.org/10.1002/\(SICI\)1520-6629\(200005\)28:3<271::AID-JCOP4>3.0.CO;2-I](https://doi.org/10.1002/(SICI)1520-6629(200005)28:3<271::AID-JCOP4>3.0.CO;2-I).
- Redleaf, D. L. (2018). *They took the kids last night: How the child protection system puts families at risk* (Davis Library). Santa Barbara, California : Praeger, an imprint of ABC-CLIO, LLC, [2018]. <https://catalog.lib.unc.edu/catalog/UNCb9406645>
- Rossi, P.H. (1987). The iron law of evaluation and other metallic rules. *Research in Social Problems and Public Policy*, 4, 3-20.
- Rossi, P.H., Lipsey, M.W. & Freeman, H.E. (2004). *Evaluation: A systematic approach*, 7th ed. Thousand Oaks, CA: Sage Publications, Inc.
- Safe Families for Children (2018a) *Operational policies and procedures*. Chicago: Author.

Safe Families for Children (2018b) *Host family handbook*. Chicago: Author.

Schatz, M.S. & Horejsi, C. (1992). *Religion and the foster home. Fostering Families. A*

specialized training program designed for foster care workers & foster care parents.

Department of Social Work Colorado State University, Fort Collins, Colorado. Designed in consultation with the Colorado Department of Social Services.

School of Social Work at The University of North Carolina at Chapel Hill, Juvenile Protective Association, Chapin Hall at The University Of Chicago, Survey Research Laboratory at The University Of Illinois, Helen Bader School of Social Welfare at The University Of Wisconsin, Erikson Institute, University of Illinois at Urbana-Champaign (Mark Testa, P.I.). (2018). *Illinois birth through three waiver demonstration: Final evaluation report*. Springfield, IL: Illinois Department of Children and Family Services.

Schuerman, J. R., Rzepnicki, T. L., Littell, J. H., & Budde, S. (1992). Implementation issues.

Children and Youth Services Review, 14(1-2), 193-206.

Schuerman, J. R., Rzepnicki, T. L., & Littell, J. H. (1994). *Putting families first: An experiment in family preservation*. Chicago: University of Chicago Press.

Scriven, M. (1997). *Evaluation thesaurus (4th ed.)*. Newbury Park, CA: Sage.

Shadish, W.R, Cook, T.D., & Campbell, D.T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin Company.

Stinchcombe, A. L. (2001). *When formality works: Authority and abstraction in law and organizations*. Chicago: The University of Chicago Press.

- Testa, M.F. (2008). New permanency strategies for children in foster care. In D. Lindsey & A. Shlonsky (Eds.), *Child welfare research: Advances for Policy and Practice* (pp. 108-124). Oxford: Oxford University Press.
- Testa, M.F. (2010). Logic of results-oriented accountability. In M.F. Testa & J. Poertner (Eds.). *Fostering accountability: Using evidence to guide and improve child welfare policy* (pp.75-100). New York: Oxford University Press.
- Testa, M.F. (2013). Systems of kinship care: Enduring challenges and emerging opportunities. *Journal of Family Social Work, 16*(5), 349-363.
- Testa, M.F., DePanfilis, D., Huebner, R., Dione, R., Deakins, B. & Baldwin, M. (2014). Bridging the gap between research and practice: The work of the steering team for the child welfare research and evaluation translational framework workgroup. U.S. Children's Bureau. *Journal of Public Child Welfare, 8*(3), 333-353.
- Testa, M. F. (2017). Introduction: Kinship care policy and practice (first issue). *Child Welfare, 95*(3), 13-39.
- Testa, M.F. (2020). A behind the scenes look at *Youakim v. Miller* and its aftermath. In M. Testa, R. Hill & C. Ingram, (Eds.) (2020). *Reflections on kinship care: Learning from the past, implications for the future* (pp. 85-110). Washington, DC: CWLA
- Testa, M. F., & Budde, S. (2017, August 20). Safe Families for Children (SFC) Evaluation. Retrieved from osf.io/j24h3
- Testa, M., Bruhn, C. & Helton, J. (2010). Comparative safety, stability, and continuity of children's placements in informal and formal substitute care arrangements. In M.B.

- Webb, K. Dowd, B.J. Harden, J. Landsverk & M.F. Testa, (Eds.). *Child welfare and child well-being: New perspectives from the national survey of child and adolescent well-being* (pp. 159-191). New York: Oxford University Press.
- Testa, M.F. & Poertner, J. (2010). *Fostering accountability: Using evidence to guide and improve child welfare policy*. Oxford: Oxford University Press.
- The Family Defense Center. (2016). *Understanding and responding to department of children and family services' abuse and neglect investigations in Illinois*. Chicago: The Family Defense Center.
- Wandersman, A., Duffy, J., Flaspohler, P., Noonan, R., Lubell, K., Stillman, L., Blachman, M., Dunville, R., & Saul, J. (2008). Bridging the gap between prevention research and practice: The interactive systems framework for dissemination and implementation. *American Journal of Community Psychology*, 41(3-4), 171-181.
<https://doi.org/10.1007/s10464-008-9174-z>
- Westat, Inc., Chapin Hall Center for Children, & James Bell Associates. (2001). *Evaluation of family preservation and reunification programs: Interim report*. Washington, D.C.: US Department of Health and Human Services.
- Wilson, S. J, Price, C., Kerns, S., Dastrup, S. & Brown, S. (2019). *Title IV-E Prevention Services Clearinghouse handbook of standards and procedures, Version 1.0*. OPRE Report # 2019-56, Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. Available from

https://preventionservices.abtsites.com/themes/ffc_theme/pdf/psc_handbook_v1_final_50_8_compliant.pdf (accessed June 1, 2020).

Zarin, Deborah A. and Alla Keselman. *Registering a Clinical Trial in ClinicalTrials.Gov*. Vol. 131 2007. doi:10.1378/chest.06-2450.

Zelen, M. (1979). A new design for randomized clinical trials. *New England Journal of Medicine*, 300, 1242-1245.

APPENDIX

Graphical Assessments of the Proportional Hazards Assumption

