

1995

## Evidence or Assertions? The Outcomes of Family Preservation Services

Julia H. Littell

*Bryn Mawr College*, [jlittell@brynmawr.edu](mailto:jlittell@brynmawr.edu)

[Let us know how access to this document benefits you.](#)

Follow this and additional works at: [http://repository.brynmawr.edu/gsswsr\\_pubs](http://repository.brynmawr.edu/gsswsr_pubs)

 Part of the [Social Work Commons](#)

---

### Custom Citation

Littell, Julia H. "Evidence or Assertions? The Outcomes of Family Preservation Services." *Social Service Review* 69 (1995): 338-351.

This paper is posted at Scholarship, Research, and Creative Work at Bryn Mawr College. [http://repository.brynmawr.edu/gsswsr\\_pubs/7](http://repository.brynmawr.edu/gsswsr_pubs/7)

For more information, please contact [repository@brynmawr.edu](mailto:repository@brynmawr.edu).

# Debates with Authors

## Evidence or Assertions? The Outcomes of Family Preservation Services

Julia H. Littell  
*Bryn Mawr College*

In their review of research on the effects of family preservation services in child welfare, Howard Bath and David Haapala (*Social Service Review* 68 [September 1994]: 386–404) raise questions about the validity of three recent controlled studies in this area.<sup>1</sup> The results of at least two of these studies challenge the widely held notion that family preservation services (FPS) are effective in preventing the out-of-home placement of children. Careful examination of the studies' methods, thoughtful interpretations of findings, and efforts to integrate this information into existing knowledge about the effects of FPS are clearly warranted.

When faced with empirical evidence that contradicts theory, one can dismiss the new evidence, attempt to modify the theory, or ignore the discrepancy. The article by Bath and Haapala presents one approach to the reconciliation of apparent contradictions between theory and research on the effects of FPS.<sup>2</sup> After an attempt to dismiss the null findings of several randomized experiments on methodological grounds, the authors cite studies that employed much weaker designs as support for their hypothesis that FPS prevent out-of-home placements.

The authors' criticisms of the randomized experiments are incomplete and often inaccurate. They use assertions rather than evidence to dismiss the findings of these studies. Their failure to take into account issues of validity in studies that appear to support their own beliefs constitutes a serious lapse, one that results in misrepresentation of the weight of the evidence in this field. Their review also misses important lessons that can be drawn from the research in this area.

### Statistical Power

The claim that "design weaknesses leading to low statistical power are largely responsible for the negative conclusions drawn from the state-initiated experi-

Permission to reprint a comment printed in this section may be obtained only from the author.

mental FPS evaluations” (p. 389) deserves close examination. Bath and Haapala assert that the statistical power in these studies was low as a result of the heterogeneity of respondents, treatment inconsistency, and instability of outcome measures. The authors never mention the size of effects the studies should have been able to detect, the level of power desired, or the sample sizes required to meet these goals—nor do they present convincing evidence that the experiments fell short of these goals.

Although it has been claimed that FPS result in dramatic reductions in the proportion of cases that require placement,<sup>3</sup> evaluations should be designed to detect more modest effects. For the purposes of this discussion, I assume that a reduction in the risk of placement of 20 percent has some practical significance. Although we may want to detect smaller differences (say, 10%), these are unlikely to be meaningful for policy purposes. Assuming the use of a one-tailed test and an alpha level of .05, the power to detect between-group differences of 20 percent (centered on 50%) was approximately .88 in New Jersey, .97 in California, and over .99 in Illinois.<sup>4</sup> This means that there was an 88–99 percent chance that the experiments would have detected program effects of 20 percent or more *if the programs had such effects*. Levels of power of this magnitude are more than adequate for research on treatment effects.<sup>5</sup>

The Illinois study had the statistical power to detect even smaller effects within 1 year after random assignment.<sup>6</sup> In fact, a very small but statistically significant difference between groups *was* found in Illinois: FPS appeared to result in an overall *increase* in the risk of placement for children—an increase of approximately 5 percent at 1 year after referral.<sup>7</sup> This effect could be attributed to the “case finding” phenomenon Bath and Haapala describe, although the gap between the experimental and control groups in placement rates disappeared once variations in case characteristics were taken into account.<sup>8</sup>

#### *Heterogeneity of Respondents*

This brings us to the issue of heterogeneity of respondents and the problems this poses for the design and interpretation of FPS evaluations. It is true that family preservation programs serve a very diverse clientele, and, as Bath and Haapala argue, such heterogeneity can reduce statistical power.

We are concerned here with client heterogeneity as it affects the main dependent variable, that is, the likelihood of placement. In this respect, Bath and Haapala cite studies in which there were differences among subgroups in the rates of placement after FPS treatment. The authors refer to this as evidence of the “differential effectiveness” of FPS (p. 390). Because there are also differences among subgroups in the likelihood of placement *in the absence of FPS*, a comparison of subgroup placement rates after FPS is not sufficient to determine whether the program has differential effects. Program effects (and differential effects) can be gauged only by comparing “base rates” of placement in the absence of FPS with rates after FPS intervention. Information on subgroup “base rates” can be obtained from control groups in randomized experiments.

As Bath and Haapala noted, most studies have not had the sample sizes needed to examine program effects for specific subgroups of clients. But the authors failed to present the evidence that does exist in this area. The Illinois experiment included examination of program effects for 16 subgroups.<sup>9</sup> Contrary to Bath and Haapala’s claims,<sup>10</sup> all but three of these subgroups had sample sizes sufficient to detect program effects of 20 percent at a level of power of .8.<sup>11</sup> Results showed that FPS *increased* the risk of placement for households headed by single adults but had no effect for any of the other

subgroups.<sup>12</sup> The program also had no effect on the risk of subsequent maltreatment for any of the subgroups.<sup>13</sup> Hence, at least in the Illinois study, there is no reason to believe that heterogeneity is an explanation for negative findings.

#### *Treatment Inconsistency*

Bath and Haapala suggest that “uncertainty about what the interventions . . . actually consisted of . . . suggests poor construct validity” (p. 392). On the contrary, reports on each of the three experiments include extensive analysis and discussion of the nature of services provided to clients.<sup>14</sup>

Lack of consistency in the treatment of clients both within and across sites is a reality—in FPS and in virtually any social program. Because the nature of family problems varies from one case to the next and because the community and social settings in which FPS operate are quite diverse, these programs are never “single undifferentiated entities.”<sup>15</sup> Attention to variations in the treatment provided to families within these studies indicates that the experiments were not based on the assumption that the treatment was undifferentiated. Although the authors point to the New Jersey experiment as “the study that had the most consistent intervention” (p. 398), Leonard Feldman’s analysis showed significant between-site differences in the duration and intensity of services provided to clients in that state.<sup>16</sup>

The approach taken in Illinois was that it made sense to accept variation within the FPS program because there is no evidence of the superiority of any one model over another. The Illinois study can be viewed as a natural experiment in which variations among programs can be examined. This experiment included 18 programs in six sites.<sup>17</sup> Three of the six experimental sites (two in Chicago and one in Peoria) had sample sizes sufficient to detect program effects of 20 percent (centered on 50%) at a level of power of .8. Site-specific analyses showed that children in the FPS groups in both of the Chicago sites were significantly *more* likely to be placed than children in the control groups in those sites.<sup>18</sup> Differences in placement rates between the program and control groups in the Peoria site (which had a single FPS provider) were not significant.<sup>19</sup>

There is no foundation for the statement that “had positive effects been demonstrated in these studies, it would have been impossible to specify what exactly was responsible for obtaining the results” (p. 392). The Illinois study was certainly large enough to support analyses of this kind—and, in fact, my colleagues and I have reported results of analyses of relationships between FPS service characteristics and outcomes.<sup>20</sup>

#### *Placement as an Outcome Measure*

As Bath and Haapala suggest, there are a number of problems with the use of placement as an outcome.<sup>21</sup> However, contrary to their suggestion, these are not related to the stability (reliability) or validity of placement as an outcome measure.<sup>22</sup> The authors have confused issues in the reliability and validity of measures with questions about the appropriateness and interpretation of an outcome variable.

The reliability of placement as an outcome measure appears to be quite straightforward: we want to know whether placement has occurred for each case at various points in time after referral to FPS. Data on placement events can be derived from state foster care records. In many states, these records are computerized and updated continuously; they are used to track children in the foster care system and for payment purposes. When linked to payments

for foster care, the reliability of data on placements is supported by the fact that there are negative fiscal consequences (to either payor or payee) if data are entered incorrectly.<sup>23</sup> In Illinois these records appear to be fairly reliable (although, to my knowledge, there have been no empirical studies of the reliability of administrative data on placements). In some studies, information on placements has also been obtained directly from caseworkers and parents, who are presumably knowledgeable about the placement status of certain children. Variations in definitions of placement across studies and changes over time in placement policies and rates within jurisdictions raise issues of external validity (i.e., whether the findings apply to other settings and periods of time), not reliability. There is no support for the statement that “placement status is an unstable measure with an inherently large error variance, which, in turn, contributes to small effect sizes” (p. 393).

With respect to construct validity, Bath and Haapala suggest that placement is “only an indirect and distal measure of the family functioning changes that FPS interventions claim to be able to make *in order to avert placements*” (p. 392, emphasis added). Because placement prevention has been the primary objective of FPS, placement is a direct measure of whether a program does what it was intended to do, not a proxy for something else. As long as there are claims that FPS “are able to prevent out-of-home placements” (p. 399), such claims must be tested through the use of placement as an outcome variable. However, this is not the only relevant outcome. The effects of FPS on various aspects of child and family functioning are important—and these effects have been examined in several studies.<sup>24</sup>

The view that many child welfare placements are “unnecessary” and that placements are “failures” of FPS is a social construction of reality. The use of placement as an outcome measure in evaluation research does not mean that placement is necessarily a bad thing. Counting placements is not the same thing as counting “failures.” Any positive or negative connotation is in the eyes of the beholder, and, as Bath and Haapala suggest, the meaning of out-of-home placement depends to a great extent on the circumstances of the individual case. The recognition that placements *are* necessary and “good” in some cases raises questions about the wisdom of placement prevention as an *overriding objective* of FPS.

### Internal Validity

Bath and Haapala suggest that the findings of little or no difference between program and control groups may have occurred as a result of diffusion, imitation, or compensatory equalization of treatments.<sup>25</sup> If cases in the control groups received services that were similar to FPS, comparable outcomes would be expected. Comparisons of the amounts and types of services provided to program and control groups are required to determine whether some equalization of treatments occurred.

Detailed comparisons of services provided to program and control cases were conducted in the Illinois experiment.<sup>26</sup> In that study, one-fifth of the cases in the control group were never opened for services in the state child welfare agency, and 51 percent of those that were opened received no services of any kind during the first 90 days after random assignment.<sup>27</sup> Overall, the control group received far less intensive services than the FPS group (e.g., control cases received a median of 2.5 hours of face-to-face contact with workers in the first 90 days of service, compared with 70.3 for FPS cases).<sup>28</sup> Cases in the FPS groups were much more likely to receive counseling (93% vs. 37% of controls) and concrete services (89% vs. 31%).<sup>29</sup> The FPS cases

received a wider array of concrete services (median of four different types of concrete services vs. a median of zero for the control group) and were more likely to receive case assistance than controls (74% vs. 7%).<sup>30</sup> Interviews with a subsample of clients in the experiment support the conclusion that FPS cases received much more extensive help than cases in the control group.<sup>31</sup> Interviews with public child welfare workers helped to confirm and explain these differences: workers who served cases in the control group usually wanted to provide better services for their clients but were simply unable to do so given much higher caseloads (an average of 50, compared with five for FPS workers).<sup>32</sup> So we can conclude that, in Illinois at least, internal validity was not threatened by diffusion or imitation treatments or by compensatory equalization.

There are other issues of internal validity that Bath and Haapala did not address. In the New Jersey experiment, 22 percent (33) of the 150 cases assigned to the FPS group were not provided with FPS because they did not meet the eligibility criteria, the caregiver refused to participate in the program, or a child was considered at risk of harm and removed from the home.<sup>33</sup> These cases were not tracked and were excluded from the analysis. In contrast, all of the cases assigned to the control group were retained in that group. The loss of a substantial proportion of FPS cases clearly could have affected the comparability of the experimental and control groups in New Jersey.<sup>34</sup> The finding of higher placement rates in the control group than the FPS group in New Jersey may be due largely to the fact that some "difficult" cases—including cases in which placement was truly imminent—were excluded from the FPS group but not from the control group.

### Summary

In the California and Illinois studies, findings that FPS did not prevent placements cannot be explained by insufficient statistical power or threats to internal validity. Although client heterogeneity and treatment inconsistency probably had some effect on statistical power, the overall level of power in these studies was so high to begin with that it is unlikely that the results can be explained by these factors. Further, analyses that took into account client heterogeneity and between-site differences in treatments in Illinois confirm the findings that FPS did not prevent placements. Claims that outcome measures were unstable and that there may have been some equalization of treatments among experimental and control cases are not credible.

In New Jersey, it is possible that the exclusion of certain kinds of cases from the FPS group (but not from the control group) was responsible for observed differences between groups.

The authors did not explain the fact that a number of statistically significant differences in the largest experiment were in the "wrong" direction. (That is, the Illinois FPS program appeared to increase the risk of placement for children in the Chicago area and for those in single-parent families.) Although Bath and Haapala acknowledge the possibility that FPS may have a "case finding" effect, it must be understood that this effect is exactly the opposite of what one would expect to find if the placement prevention hypothesis were true.

Failure to reject the null hypothesis can be due to several conditions: the null hypothesis may be true (i.e., the program may have no effect), the underlying theory may be faulty, the reliability or validity of measures may be weak, or power may be insufficient to detect real differences between groups.<sup>35</sup> Bath and Haapala's article deals only with issues of validity and statistical power. Issues of truth and theory remain.

## The Weight of the Evidence

Early evaluations of family preservation services were not “unequivocally positive,” nor did they all report “high placement prevention success” (p. 387). The results of early controlled studies were mixed: four randomized experiments and one comparison group study found that the programs did not produce significant overall reductions in placement.<sup>36</sup> Three randomized experiments found statistically significant—but small—reductions in placement in favor of the experimental groups.<sup>37</sup>

Of the studies reviewed by Bath and Haapala, only those with the weakest controls for threats to internal validity show program effects on placement. The authors’ emphasis on possible methodological limitations in the three focal experiments stands in stark contrast to a complete lack of attention to design problems inherent in the quasi experiments they reviewed.

The quasi experiments employed various overflow designs. Referring workers can often manipulate these designs to ensure that certain families receive intensive services. Some cases that are turned away initially from the program may be referred again when there is an opening for services, whereas other options (including placement) are pursued in the remaining overflow cases.<sup>38</sup> Without a careful record of the course of each case that enters an overflow group, the comparability of the overflow group with FPS cases cannot be assured. Thus, overflow studies may suffer from the threats to internal validity that Thomas Cook and Donald Campbell termed “selection” and “mortality.”<sup>39</sup>

The overflow studies cited by Bath and Haapala were quite small. The study reported by Sally Wood, Keith Barton, and Carroll Schroeder included 26 families in the FPS group and 24 who did not receive FPS.<sup>40</sup> The study by Ira Schwartz, Philip AuClaire, and Linda Harris involved an initial sample of children who were approved for placement; 58 of these cases were diverted to a FPS program, and a random sample of 58 of the cases that remained was taken as a comparison group.<sup>41</sup> The evaluation of the Utah Homebuilders program (reported by Mark Fraser, Peter Pecora, and Haapala) had an overflow comparison group of 38 cases, out of which nearly one-third (12 cases) were lost to follow-up.<sup>42</sup> This loss of a substantial proportion of comparison cases is sufficient to invalidate the study, because the comparability of treatment and comparison groups cannot be assured. Fraser and his colleagues attempted to deal with this problem by matching a subsample of FPS and comparison cases on several variables. However, the variables used in the matching process were only weakly related to placement, and the number of matched pairs was quite small. Further, the researchers in the Utah study relied on different sources of information on placements in the program and comparison group.<sup>43</sup>

Bath and Haapala cite data from the state of Michigan on changes in placement rates over time as further evidence of the effects of FPS. When these data were examined in terms of rates of foster care placement per thousand children, there was no clear relationship between FPS and placement rates.<sup>44</sup> In any event, fluctuations in placement rates over a few years do not provide evidence for the effects of FPS because placement rates are affected by many factors outside of these programs.<sup>45</sup>

There is also no convincing support for the cost-savings arguments put forth in Michigan. Family preservation programs clearly are more costly than traditional child welfare services for intact families and less costly than out-of-home placements. But, without convincing evidence that these programs prevent placements, there is no reason to expect substantial reductions in

costs. The California study is the only randomized experiment that provided data on service costs. Although the costs of FPS were said to “compare favorably” with the costs of out-of-home care, when the total costs of services provided and placements required by cases in the FPS and control group are compared, it is clear that FPS resulted in *increased* costs in California.<sup>46</sup>

Bath and Haapala’s conclusion that “there is strong evidence that FPS *are* able to significantly reduce the placement rates of served, at-risk children” (p. 389) appears to be based on three overflow comparison studies, the New Jersey experiment, and data on placement rates from the state of Michigan that, as we have seen, all suffer from threats to internal validity.<sup>47</sup> The recommendation that “until convincing evidence is forthcoming, it would be wisest for program developers to base their initiatives on model program packages that have some proven effectiveness” (p. 399) is confusing. Without convincing evidence, there is no proof of effectiveness.

In sum, the empirical evidence on the effects of FPS on the out-of-home placement of children is mixed. The findings of recent controlled studies call into question earlier claims about the success of FPS in preventing placement. The lack of effects on placement appears to be due, at least in part, to difficulties in targeting FPS to families that are truly “at risk” of placement.

### Targeting

Bath and Haapala suggest that it is not “fair” to evaluate the outcomes of FPS when relatively few of the referred cases are at “imminent risk” of placement (as was the case in the three recent experiments). Yet, if FPS are to have a substantial effect on placements, they must first solve this targeting problem. Given widespread claims of the success of FPS in preventing placements and the rapid proliferation of state-supported FPS initiatives, it is certainly reasonable to assess both the targeting accuracy and placement prevention effects of FPS projects.

Bath and Haapala imply that difficulties in the targeting of FPS are relatively new,<sup>48</sup> but there is evidence to the contrary. Seven randomized experiments conducted in the 1970s and 1980s on programs that provided intensive in-home services designed to prevent placement showed that relatively few of the cases referred were at risk of placement.<sup>49</sup>

The authors underestimate the complexity of issues in the targeting of FPS. This is more than a matter of workers’ reluctance to refer the “right” cases. The need for placement is sometimes difficult to predict because family circumstances can change rapidly and in ways that cannot always be foreseen by caseworkers. Although referring workers may not be certain about whether or when placement will be necessary, they sometimes state that placement is imminent in order to obtain thorough case assessments or services for families through FPS.<sup>50</sup> Many child protective services (CPS) workers believe that “unnecessary placements” are very rare; thus, in their view, there are few placements to be prevented. They often have other objectives for cases that they pursue through FPS.<sup>51</sup> Workers knowingly refer cases that are not at risk of placement because there are few alternatives for these families.<sup>52</sup> The fact that FPS provide immediate, intensive, and concrete help for families in a service system in which such resources are scarce has led to an accretion and shifting of goals for FPS.<sup>53</sup> To CPS workers, FPS may seem most useful in addressing previously unmet service needs of intact families,<sup>54</sup> which constitute the largest proportion of cases in the child welfare system. Thus, although most CPS workers place a high value on the services provided by FPS programs, they do not tend to view them as an alternative to placement.<sup>55</sup>



## Contextual Issues

Bath and Haapala provide a useful discussion of contextual factors that affect FPS and their evaluations. As they suggest, there are substantial variations across settings in client and community characteristics, service options, decision-making practices, and the policies and politics of child welfare. These variations can limit the generalizability of FPS evaluations.

Contrary to the authors' assertions, contextual factors have little to do with the internal validity of evaluations or the strength of conclusions one can draw about the effects of specific FPS programs.<sup>56</sup> Rather, information about the service context can help us understand targeting problems, limitations of FPS, and why some approaches may be more successful in certain locations or with specific types of clients than others.

Recognition of the unique contexts in which FPS operate also points to the need to tailor these services to fit local needs, even though this tailoring may contribute to "treatment inconsistency" and "model drift." Although it might be desirable to find program models that are so robust that they can withstand the various adaptations needed to be effective in different sites, it does not appear that this has been achieved in FPS.<sup>57</sup>

All of this suggests that the question, Does FPS work?—which is the central concern of Bath and Haapala's review—does not adequately reflect real-world constraints and variations in the ways in which FPS operate.

## Toward More Realistic Expectations

Is it reasonable to expect FPS (or any short-term intervention) to have dramatic, lasting effects on family functioning, leading to substantial reductions in rates of out-of-home placement and the costs of child welfare services? The findings of recent experiments suggest that the answer to this question is no.

This answer makes sense in light of the serious problems faced by families in the child welfare system, the lack of supports for these families outside of FPS, and the impact of larger social, economic, and environmental influences on families. We have suggested that the child welfare system alone cannot be expected to solve problems of substance abuse, poverty, mental illness, and family violence that can lead to child maltreatment and out-of-home placement.<sup>58</sup>

What should we expect? The FPS experiments in New Jersey and Illinois found a few significant differences in favor of the FPS groups in measures of child and family functioning.<sup>59</sup> In Illinois, it was shown that these effects were short-term and that they dissipated over time.<sup>60</sup> Thus, it may be reasonable to hope for some positive changes in families as a result of FPS, although these benefits are likely to be modest and (in the absence of adequate follow-up services) short-term.

Through intensive, in-home services we can learn a great deal about the nature of problems faced by the families in child welfare. In some cases, this will lead to the discovery of conditions that pose serious risks to children, new reports of child maltreatment, and out-of-home placements. The value of FPS is that they provide immediate, intensive, and flexible services where they are needed. But it is not realistic to expect these programs, remarkable as they are, to produce any quick fix.

## Conclusions

What does the research on FPS *really* tell us? First, it tells us that there are widespread problems in the targeting of FPS.<sup>61</sup> These programs are not serving the group of families for whom they were intended. It appears that FPS are

being used to serve aims other than placement prevention. Second, we have learned that the expectation that FPS will result in dramatic reductions in placements and in the costs of child welfare services is unrealistic.

The debate over the effects of FPS still centers on very few rigorous studies. More solid evidence is needed, particularly on the effects of the Homebuilders model.<sup>62</sup> There is also need for more refined analyses of effects of various types of FPS for different kinds of families. The question, Does FPS work? promotes an overly simplistic view of families, program processes, and the intricate "service context" in which they interact.

Through the development of the Homebuilders program, Haapala and his colleagues have made extraordinary contributions to the field of child welfare. Homebuilders and other family preservation programs have resulted in tangible improvements in the quality and quantity of services available to families in child welfare. These programs offer many examples of "best practice" in the field. Now, faced with mounting evidence that initial expectations for these programs have not been met, it is time to move beyond the rhetorical arguments that seem to engulf family preservation services toward new understandings of what these programs can and cannot accomplish.

## Notes

1. The three focal experiments are described in Leonard H. Feldman, "Assessing the Effectiveness of Family Preservation Services in New Jersey within an Ecological Context" (New Jersey Division of Youth and Family Services, Bureau of Research, Evaluation, and Quality Assurance, Trenton, N.J., 1991); Walter R. McDonald and associates, "Evaluation of AB 1562 In-Home Care Demonstration Projects: Final Report" (Walter R. McDonald and associates, Sacramento, Calif., 1990); and John R. Schuerman, Tina L. Rzepnicki, and Julia H. Littell, *Putting Families First: An Experiment in Family Preservation* (Hawthorne, N.Y.: Aldine de Gruyter, 1994).

2. Here, "theory" is used in its informal sense, meaning an explanation or conjecture. There is no single theory behind FPS; rather, these programs rely on various combinations of crisis intervention theory, family systems theory, social learning theory, and ecological theory (Richard P. Barth, "Theories Guiding Home-Based Intensive Family Preservation Services," in *Reaching High Risk Families: Intensive Family Preservation in Human Services*, ed. James K. Whittaker, Jill Kinney, Elizabeth M. Tracy, and Charlotte Booth [Hawthorne, N.Y.: Aldine de Gruyter, 1990]).

3. For example, Jill Kinney, David Haapala, and Charlotte Booth stated that "for families with at least one child at imminent risk of placement, Homebuilders['] success in avoiding placement at 12 months following the initiation of service has varied from 73 to 91% of families served" (*Keeping Families Together: The Homebuilders Model* [Hawthorne, N.Y.: Aldine de Gruyter, 1991], p. 185). The assumption here is that the risk of placement has been reduced from 100 percent to 27 percent or less.

4. Differences of 20 percent centered on 50 percent would result in placement rates of 60 percent in one group and 40 percent in the other. This difference is the same as a "generic effect size" of .4 (Mark W. Lipsey, *Design Sensitivity: Statistical Power for Experimental Research* [Newbury Park, Calif.: Sage, 1990]). Power estimates were derived from tables provided by Jacob Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 2d ed. (Hillsdale, N.J.: Lawrence Erlbaum, 1987). Estimates of the level of statistical power are based on the number of cases in the treatment and control groups. In the New Jersey study, there were 97 families in the control group and 117 in the FPS group—not including 33 cases assigned to FPS that were subsequently turned back and lost to follow-up (Feldman, "Assessing the Effectiveness of Family Preservation Services" [n. 1 above]). The California study had 152 families in both groups (McDonald and associates [n. 1 above]). In the Illinois experiment, there were 995 families in the FPS group and 569 controls; valid data on placements were available on 984 FPS cases and 552 control cases (Schuerman, Rzepnicki, and Littell, *Putting Families First* [n. 1 above]). Because the treatment and control groups in the New Jersey and Illinois studies

were of unequal size, harmonic means were used to estimate power (the harmonic means are 106 for New Jersey and 707 for Illinois) (see Cohen [above]). The argument that effective sample sizes were actually much smaller, because of bad targeting, is specious. Targeting accuracy merely shifts the percentages. For example, in a well-targeted program, between-group differences may be centered on values over 50 percent (e.g., this would be the case if placement rates were 90% in one group and 70% in the other), whereas differences of the same magnitude will be centered on lower values in a poorly targeted program (e.g., placement rates of 30% and 10%). Differences centered on values above or below 50 percent yield relatively large effect sizes—that is, they are *easier* (require less statistical power) to detect than differences that center on 50 percent. However, there may be a “floor effect” when placement rates in the control group are so low that there are few placements to be prevented. In this case, any differences between groups are bound to be slight. This may have occurred in the first few months after random assignment in the Illinois study, although placement rates in the control group rose to over 20 percent within the first year (see Schuerman, Rzepnicki, and Littell, *Putting Families First* [n. 1 above]).

5. Cohen (n. 4 above) recommends .8 as a minimum standard for statistical power.

6. In the Illinois experiment, the level of power to detect a 10 percent difference between groups centered on 50 percent was approximately .98; the power to detect differences of 10 percent centered on 20 percent was over .99 (see Cohen [n. 4 above]).

7. Schuerman, Rzepnicki, and Littell, *Putting Families First* (n. 1 above).

8. *Ibid.*

9. See John R. Schuerman, Tina L. Rzepnicki, Julia H. Littell, and Amy Chak, “Evaluation of the Illinois Family First Placement Prevention Program: Final Report” (Chapin Hall Center for Children at the University of Chicago, Chicago, 1993); and Schuerman, Rzepnicki, and Littell, *Putting Families First* (n. 1 above). The subgroups examined were families that were new to the child welfare system; families in which a child had been physically injured prior to referral; cases of chronic neglect; families with housing problems; families with severe income or resource deficits; families with cocaine problems; families with alcohol problems; cases with other substance abuse problems; families in which a parent had a chronic mental illness or serious emotional problem; cases with marital or adult relationship problems; cases with serious child-care skill deficits; families in which a child had a health, developmental, or learning problem; cases in which the primary caregiver was a teenager; households headed by a single adult; households with extended family members; and cases in which protective custody of a child had been taken within 1 year prior to referral. These categories are not mutually exclusive.

10. Bath and Haapala stated that “in the cited studies, overall sample sizes are not large enough to simultaneously allow for the examination of important client subgroups while retaining adequate statistical power to detect treatment effects” (p. 390).

11. Samples of 64 cases or more in the control group and 96 or more in the FPS group would be needed to detect differences of 20 percent (centered on 50%) with a power level of .8 (see Cohen [n. 4 above]). The only subgroups that did not meet these size requirements were cases of chronic child neglect, families with substance abuse problems other than cocaine or alcohol abuse, and families in which the primary caregiver was a teenager (see Schuerman, Rzepnicki, and Littell, *Putting Families First* [n. 1 above]).

12. Schuerman, Rzepnicki, and Littell, *Putting Families First* (n. 1 above). For single-parent families, the risk of placement at 1 year after random assignment was .20 for control cases, compared with .29 for those in the FPS group. Earlier findings that FPS reduced the risk of placement for families with marital problems and for cases in which a child had been seriously injured prior to referral (Schuerman et al. [n. 9 above]) were not borne out in the final analysis.

13. Schuerman, Rzepnicki, and Littell, *Putting Families First* (n. 1 above). Earlier evidence that FPS reduced the risk of maltreatment for families in which there had been chronic child neglect (Schuerman et al. [n. 9 above]) was not upheld in the final analysis.

14. See Leonard Feldman, “Evaluating the Impact of Family Preservation Services in New Jersey” (New Jersey Division of Youth and Family Services, Bureau of Research, Evaluation, and Quality Assurance, Trenton, N.J., 1990), and “Assessing the Effective-

ness of Family Preservation Services" (n. 1 above); McDonald and associates (n. 1 above); Julia H. Littell and Jeanne Howard, with Stephen Budde, Cindy Chrisman, Brenda Eckhardt, and Diane Pellowe, "The Nature of the Work with Families in Family First" (Chapin Hall Center for Children at the University of Chicago, Chicago, 1990); Julia H. Littell, Jeanne Howard, Tina L. Rzepnicki, Stephen Budde, and Diane Pellowe, "Intervention with Families in the Illinois Family Preservation Program" (Chapin Hall Center for Children at the University of Chicago, Chicago, 1992); John R. Schuerman, Tina L. Rzepnicki, Julia H. Littell, and associates, "Evaluation of the Illinois Family First Placement Prevention Programs: Progress Report" (Chapin Hall Center for Children at the University of Chicago, Chicago, 1989-92). Schuerman et al. (n. 9 above); and Schuerman, Rzepnicki, and Littell, *Putting Families First* (n. 1 above).

15. Peter Rossi used this term to refer to studies that did not delve into the "black box" of treatment (Peter H. Rossi, *Evaluating Family Preservation Programs: A Report to the Edna McConnell Clark Foundation* [New York: Edna McConnell Clark Foundation, 1991]). Regarding the three focal experiments in their review, Bath and Haapala stated that "the assumption behind this evaluation approach [is] that the intervention is a 'single undifferentiated entity'" (p. 392).

16. Feldman, "Assessing the Effectiveness of Family Preservation Services" (n. 1 above).

17. One geographic site with two FPS programs was dropped from the Illinois experiment (Schuerman, Rzepnicki, and Littell, *Putting Families First* [n. 1 above]).

18. Ibid.

19. Ibid.

20. Ibid.

21. For a discussion of these issues, see John R. Schuerman, Tina L. Rzepnicki, and Julia H. Littell, "From Chicago to Little Egypt: Lessons from an Evaluation of a Large Scale Child Welfare Family Preservation Program," in *Family Preservation Services: Research and Evaluation*, ed. Kathleen Wells and David E. Biegel (Newbury Park, Calif.: Sage, 1991).

22. Bath and Haapala stated that "the placement measure . . . has dubious construct validity in that it is only a proxy for the more immediate or proximal changes a family is presumed to make during the intervention, and it fails to account positively for desirable placement outcomes. Its reliability is also suspect, as it is dependent on so many fluid macro or contextual factors, to say nothing of the numerous definitions of placement that have actually been used" (p. 393).

23. Mark E. Courtney and Raymond C. Collins, "New Challenges and Opportunities in Child Welfare Outcomes and Information Technologies," *Children and Youth Services Review* 73 (September-October 1994): 359-78.

24. For example, see Feldman, "Assessing the Effectiveness of Family Preservation Services" (n. 1 above); and Schuerman, Rzepnicki, and Littell, *Putting Families First* (n. 1 above).

25. For a description of these threats to internal validity, see Thomas D. Cook and Donald T. Campbell, *Quasi-Experimentation: Design and Analysis Issues for Field Settings* (Boston: Houghton Mifflin, 1979).

26. See Schuerman, Rzepnicki, and Littell, *Putting Families First* (n. 1 above). Data on the amounts and types of services provided to control cases were not obtained in the California and New Jersey experiments (McDonald and associates [n. 1 above]; Feldman, "Assessing the Effectiveness of Family Preservation Services" [n. 1 above]).

27. Schuerman, Rzepnicki, and Littell, *Putting Families First* (n. 1 above).

28. Ibid.

29. Ibid.

30. Ibid.

31. Ibid.

32. Ibid.

33. Feldman, "Assessing the Effectiveness of Family Preservation Services" (n. 1 above).

34. Cook and Campbell (n. 25 above) refer to this threat to internal validity as "mortality." It has also been termed "differential attrition."

35. Elazar J. Pedhazur and Liora Pedhazur Schmelkin, *Measurement, Design, and Analysis: An Integrated Approach* (Hillsdale, N.J.: Lawrence Erlbaum, 1991).

36. See D. N. Willems and R. DeRubeis, "The Effectiveness of Intensive Preventive Services for Families with Abused, Neglected, or Disturbed Children: Hudson County Project Final Report" (New Jersey Division of Youth and Family Services, Bureau of Research, Trenton, N.J., 1981); Stephen J. Leeds, "Evaluation of Nebraska's Intensive Services Project" (National Resource Center on Family Based Services, Cedar Rapids, Ia., 1984); and Hennepin County Community Services Department, "Family Study Project: Demonstration and Research in Intensive Services to Families" (Hennepin County Community Services Department, Minneapolis, 1980). Steven Szykula and Matthew Fleischman reported on two studies ("Reducing Out-of-Home Placements of Abused Children: Two Controlled Field Studies," *Child Abuse and Neglect* 9 [1985]: 277–83). One, a randomized experiment, found no significant overall program effects on placement. Although the number of cases was extremely small, the program appeared to reduce the risk of placement among less difficult cases but not among more difficult ones. The second study examined the numbers of placements made before, during, and after program implementation; this design does not control for threats to internal validity because the number of placements was small and many factors outside of the program could have accounted for change over time. A comparison group study was reported by Christina Mitchell, Patricia Tovar, and Jane Knitzer ("The Bronx Homebuilders Program: An Evaluation of the First 45 Families" [Bank Street College of Education, New York, 1989]).

37. Mary Ann Jones, Renee Neuman, and Ann W. Shyne, *A Second Chance for Families: Evaluation of a Program to Reduce Foster Care* (New York: Child Welfare League of America, 1976); G. Halper and M. A. Jones, "Serving Families at Risk of Dissolution: Public Preventive Services in New York City" (Human Resources Administration, Special Services for Children, New York, 1981); and Charles G. Lyle and John Nelson, "Home Based vs. Traditional Child Protection Services: A Study of the Home Based Services Demonstration Project in the Ramsey County Community Human Services Department" (Ramsey County Community Human Services Department, St. Paul, 1983).

38. This pattern of repeated referrals of selected cases to FPS was observed in Illinois prior to the implementation of a randomized experiment in that state.

39. Selection is "a threat when an effect may be due to the difference between the kinds of people in one experimental group as opposed to another. Selection is therefore pervasive in quasi-experimental research, which is defined in terms of different groups receiving different treatments" (Cook and Campbell [n. 25 above], p. 53). Differential attrition or mortality is a threat to internal validity when subjects drop out of a particular treatment group during the course of an experiment. "This results in a selection artifact, since the experimental groups are then composed of different kinds of persons" (p. 53).

40. Sally Wood, Keith Barton, and Carroll Schroeder, "In-Home Treatment of Abusive Families: Cost and Placement at One Year," *Psychotherapy* 25, no. 3 (1988): 409–14.

41. Three of the FPS cases experienced placement during the entire follow-up period and were excluded from analysis. Ira M. Schwartz, Philip AuClaire, and Linda J. Harris, "Family Preservation Services as an Alternative to the Out-of-Home Placement of Adolescents: The Hennepin County Experience," in Wells and Biegel, eds. (n. 21 above).

42. Mark W. Fraser, Peter J. Pecora, and David A. Haapala, *Families in Crisis: The Impact of Intensive Family Preservation Services* (Hawthorne, N.Y.: Aldine de Gruyter, 1991).

43. Data on placements were obtained from both clients and workers in the family preservation group, whereas information on placements in the comparison group was obtained only from workers (*ibid.*).

44. Ira M. Schwartz, "The Systemic Impact of Family Preservation Services: A Case Study," in *Home-Based Services for Troubled Children*, ed. Ira M. Schwartz and Philip AuClaire (Lincoln: University of Nebraska Press, 1994).

45. The threats to internal validity that Cook and Campbell (n. 25 above) term "history," "selection," and "the interaction of history and selection" are relevant here. In this context, "history" refers to events outside the program that may affect placement rates. For example, Schwartz (n. 44 above) noted that a recent drop in the substantiation rate (i.e., the proportion of abuse and neglect reports that were confirmed), brought about by policy changes in the definitions of abuse and neglect and the nature of evidence required for substantiation, may have contributed to a decline in the rate of foster care case openings in Michigan. The selection of sites for initial FPS implementa-

tion in Michigan was not random, and initial differences between selected sites and other sites are not fully understood. Program planners often implement pilot projects in settings in which they are most likely to succeed. In Michigan, it is possible that the commitment of child protective services (CPS) workers to the ideal of family preservation was greater in the initial FPS sites than in other sites. Regarding the interaction of history and selection, the implementation of FPS may have systemic effects on local child welfare decision-making practices. In selected sites, workers may become less likely to seek out-of-home placements for children as a result of attention paid to the value of preserving families, apart from any improvements in family functioning that may be brought about by FPS services.

46. See McDonald and associates (n. 1 above). Because the California FPS program did not prevent placements, the costs of placements in the two groups were roughly equivalent (\$141,375 for 152 FPS cases, and \$145,388 for 152 control cases). However, the FPS program itself cost an average of \$4,767 per family, or approximately \$725,000 for the 152 families in the experimental group (see Schuerman, Rzepnicki, and Littell, *Putting Families First* [n. 1 above]).

47. The authors also note that results of smaller studies of FPS are "encouraging" (p. 399) but provide no references to support this claim.

48. Bath and Haapala suggest that the "pool of preventable placements" has become smaller in recent years (p. 395). To support this notion, they cite data from California that show a decrease over a 15-year period in the percentage of child maltreatment investigations that resulted in placement. However, it should be understood that there was a dramatic increase in the denominator (the number of investigations) during this time. Like most states, California experienced an increase in foster care admissions between 1985 and 1990. See Richard P. Barth, Mark Courtney, Jill Duerr Berrick, and Vicky Albert, *From Child Abuse to Permanency Planning: Child Welfare Services Pathways and Placements* (Hawthorne, N.Y.: Aldine de Gruyter, 1994).

49. See Mitchell, Tovar, and Knitzer (n. 36 above); Szykula and Fleischman (n. 36 above); Willems and DeRubeis (n. 36 above); Halper and Jones (n. 37 above); Leeds (n. 37 above); Lyle and Nelson (n. 37 above); Jones, Neuman, and Shyne (n. 37 above); and a follow-up report by Mary Ann Jones (*A Second Chance for Families, Five Years Later: Follow-Up of a Program to Prevent Foster Care* [New York: Child Welfare League of America, 1985]).

50. Evidence of this practice comes from the state of Washington and from our interviews with CPS workers in Illinois. In Washington, where referring workers must state that a child is at imminent risk of placement in order to refer families to the Homebuilders program (see Kinney, Haapala, and Booth [n. 3 above]), a CPS worker was quoted as saying, "The truth is that I lie about imminent risk. We all do if we think the families need the service" (Duff Wilson, "Big No-Bid Contract with DSHS Draws Fire for Cost," *Seattle Times* [March 8, 1994], p. A1). Another CPS worker said, "You do what you have to do to get the service. We lied" (Wilson [above]). This practice was also recounted in Illinois, where CPS workers described it as "advocacy" on behalf of their clients.

51. See Julia H. Littell, John R. Schuerman, Tina L. Rzepnicki, Jeanne Howard, and Stephen Budde, "Shifting Objectives in Family Preservation Programs," in *Advancing Family Preservation Practice*, ed. E. Susan Morton and R. Kevin Grigsby (Newbury Park, Calif.: Sage, 1993).

52. This idea was suggested by Stephen Budde on the basis of his dissertation research in progress (personal communication, September 1994).

53. *Ibid.*; Littell et al. (n. 51 above).

54. Stephen Budde, "Placement and Family Preservation Referral Decisions by Child Protection Investigators: A Qualitative Study" (paper presented at the National Association for Family-Based Services Sixth Annual Empowering Families Conference, Seattle, December 11, 1992).

55. *Ibid.*

56. Bath and Haapala stated that "any failure to demonstrate the superiority of an FPS program over normal service controls may . . . be due to contextual factors that can affect both the internal and external validity of any evaluation design" (p. 397).

57. For example, a small quasi experiment showed that Homebuilders—which is arguably the premier FPS model—did not appear to be effective in preventing placements in the Bronx (see Mitchell, Tovar, and Knitzer [n. 36 above]). This project included an overflow comparison group of 11 families and a treatment group of 22

families referred from the same source. Although the study clearly lacked the statistical power to detect program effects of practical significance, it is important to note that placement rates in the treatment group were *higher* than those in the comparison group.

58. Schuerman, Rzepnicki, and Littell, *Putting Families First* (n. 1 above).

59. Feldman, "Assessing the Effectiveness of Family Preservation Services" (n. 1 above); Schuerman, Rzepnicki, and Littell, *Putting Families First* (n. 1 above).

60. Schuerman, Rzepnicki, and Littell, *Putting Families First* (n. 1 above).

61. In the states of Michigan and Washington, Homebuilders and others have undertaken special projects aimed at solving the targeting problems.

62. Several promising efforts to augment the research in this area are now underway. For example, Betty Blythe and her colleagues have mounted a randomized experiment in Detroit that includes efforts to improve the targeting of FPS. Further evidence of the effect of FPS will also be developed in a national evaluation funded by the U.S. Department of Health and Human Services assistant secretary for planning and evaluation.

## Evaluation Outcomes of Family Preservation Services and the Way Ahead: A Reply to Littell

Howard I. Bath

*Marymead Child and Family Centre, Canberra, Australia*

David A. Haapala

*Bold Solutions, Tacoma, Washington*

In her response to our article "Family Preservation Services: What Does the Outcome Research *Really* Tell Us?" (*Social Service Review* 68 [September 1994]: 386–404), Julia Littell questions some of our arguments with particular reference to the research undertaken by the Chapin Hall group.<sup>1</sup> The research from this group has contributed a great deal to our understanding of family preservation services (FPS) and the complex issues that surround their implementation; however, the experimental component of their research and the conclusions that were reached were, we believe, fundamentally flawed. Conceptually, the experimental design was sound enough, but, as is often the case in such research, the researchers had little control over powerful factors in the service delivery context that led to serious problems with internal validity. It was the premise of our article that the no-effect conclusions from two major experimental, and widely cited, FPS evaluations "need to be interpreted with great caution" (p. 399)—in our view, this remains the case.

Ironically, we find ourselves agreeing with both of the conclusions to Littell's paper, in particular, her assertion that "the question, Does FPS work? promotes an overly simplistic view of families, program processes, and the intricate 'service context' in which they interact" (p. 346). It was precisely this concern that prompted our paper in the light of the simplistic interpretations of the Illinois and California<sup>2</sup> evaluations that were gaining currency. We acknowledge that the authors of both studies were more cautious in interpreting their results than external commentators and the media have been, but these