



SCHOOL of  
PUBLIC POLICY

# 22

## Nurse Family Partnership (Memphis)

**Douglas J. Besharov**  
**Peter Germanis**  
**Caeli A. Higney**  
**and**  
**Douglas M. Call**

September 2011



Maryland School of Public Policy  
Welfare Reform Academy  
[www.welfareacademy.org](http://www.welfareacademy.org)

Part of a forthcoming volume  
*Assessments of Twenty-Six Early Childhood Evaluations*  
by Douglas J. Besharov, Peter Germans, Caeli A. Higney, and Douglas M. Call

*Note:* This report is open to public comments, subject to review by the forum moderator. To leave a comment, please send an email to [welfareacademy@umd.edu](mailto:welfareacademy@umd.edu) or fill out the comment form at [http://www.welfareacademy.org/pubs/early\\_education/chapter22.html](http://www.welfareacademy.org/pubs/early_education/chapter22.html).

## 22

# Nurse Family Partnership (Memphis)

The Nurse Family Partnership (NFP) operated in Memphis, Tennessee, from 1990 to 1993. It was designed to “help low-income, first-time parents start their lives with their children on a sound course and prevent the health and parenting problems that can contribute to the early development of antisocial behavior.”<sup>1</sup> The program had three main objectives: (1) to improve women’s health-related behaviors during pregnancy; (2) to aid parents in the attainment of parenting skills; and (3) to enhance the maternal life-course development of participating women by encouraging family planning, educational development, and self-sufficiency. The Memphis trial was undertaken to determine whether the positive findings of the Elmira trial (see chapter 18) could be replicated in a major urban area with a sample of primarily low-income black women.

David Olds, now professor of pediatrics at the University of Colorado Health Sciences Center, Harriet Kitzman, associate professor of nursing at the University of Rochester, and their colleagues (the “NFP team”) evaluated the Nurse Family Partnership program in Memphis. Olds earlier developed and evaluated the precursor program in Elmira, New York, which ran from 1978 to 1982. The NFP in Memphis was carefully evaluated using random assignment. For the children, positive results were concentrated among children with mothers with low psychological resources on cognitive and school readiness/performance measures through the age-twelve follow-up. Mothers in the program group initially were less likely to have subsequent pregnancies, to be married, to be living with the child’s father, and less likely to receive government benefits. However, by the age-twelve follow-up, many of the differences had faded out. As with the Elmira evaluation, the only major concern is that the evaluation was conducted by the same group that designed the intervention and has yet to be independently evaluated.

### Program Design

**Program group.** Given that the beneficial effects of the Elmira program were concentrated in the subgroup of women who were unmarried and from low SES families at registration, the Memphis trial targeted women who were less than twenty-nine weeks pregnant, had no previous live births, and had at least two of the following sociodemographic risk

---

<sup>1</sup>David Olds, Peggy Hill, and Elissa Rumsey, “Prenatal and Early Childhood Nurse Home Visitation,” *Juvenile Justice Bulletin* (November 1998): 1.

conditions: unmarried, less than twelve years of education, and unemployed. At the time of enrollment, 64 percent of the women were age eighteen or younger, 98 percent were unmarried, 92 percent were black, and 85 percent had household incomes at or below the poverty level.<sup>2</sup>

**Services.** The program consisted of home visits made by trained, experienced nurses, beginning during pregnancy and continuing until the child reached two years of age. Making an average of seven visits during pregnancy, nurses followed detailed program guidelines that covered the personal health and development of the mothers, as well as the development of the child. Services included parent education, social support, and referrals to other health and social services.

The nurses completed an average of twenty-six visits from the child's birth through his or her second birthday. The parent education component focused on informing parents about fetal and infant development. The nurses also encouraged the mothers to improve the health behaviors that affect child well-being. Under the program, the nurse-visitors coached the mothers in making better lives for themselves and their children. They encouraged mothers to finish school and to find employment, and also discouraged additional childbearing, but they "did so in the context of helping women set goals for themselves at a crucial stage in their own personal development."<sup>3</sup> The nurses also worked to involve other family members and friends during pregnancy, birth, and while their children were young. Finally, the nurses referred participating women to health and human services agencies for family planning, mental-health counseling, and legal aid services as needed.

**The Evaluation.** A random assignment evaluation was conducted by the NFP team. A total of 1,290 women were recruited and 1,139 (88 percent) agreed to participate in the evaluation. Black women were somewhat more likely to enroll than non-black women (89 percent vs. 74 percent), as were younger women and non-high school graduates, but the differences were relatively small.

Participants were stratified by maternal race, maternal age, gestational age at enrollment, employment status of the head of household, and geographic region (with a total of four regions). They were then randomly assigned to one of four groups. The first group received free

---

<sup>2</sup>David L. Olds, Harriet Kitzman, Carole Hanks, Robert Cole, Elizabeth Anson, Kimberly Sidora-Arcoleo, Dennis W. Luckey, Charles R. Henderson, Jr., John Holmberg, Robin A. Tutt, Amanda J. Stevenson, and Jessica Bondy, "Effects of Nurse Home Visiting on Maternal and Child Functioning: Age-9 Follow-up of a Randomized Trial," *Pediatrics* 120, no. 4 (October 2007): 832-845.

<sup>3</sup>Harriet Kitzman, David L. Olds, Charles R. Henderson, Jr., Carole Hanks, Robert Cole, Robert Tatalbaum, Kenneth M. McConnochie, Kimberly Sidora, Dennis W. Luckey, David Shaver, Kay Engelhardt, David James, and Kathryn Barnard, "Effect of Prenatal and Infancy Home Visitation by Nurses on Pregnancy Outcomes, Childhood Injuries, and Repeated Childbearing: A Randomized Controlled Trial," *Journal of the American Medical Association* 278, no. 8 (August 27, 1997): 646.

transportation for scheduled prenatal care visits. The second group received the same free transportation, but also received developmental screening and referral services when their child was six, twelve, and twenty-four months of age. The third group received home visits during pregnancy, while the fourth group received such visits during pregnancy and for the child's first two years. For the prenatal phase of the program, the NFP team compared the latter two groups to the first two groups. For the findings reported here, the comparison is based on the outcomes for group four ("the program group") vs. group two (the "control group").

The findings also single out one major subgroup—mothers with low psychological resources. This variable is based on a composite of the women's intelligence, mental health, and sense of mastery, and was dichotomized at the 50<sup>th</sup> percentile of the raw scores. The procedure split the women into two groups: a low-functioning group (40 percent of the sample) and a high-functioning group (60 percent of the sample).

## Major Findings

During the first two years of the participating children's lives, there were no program effects on their mental development, behavioral problems, or immunization rates. There were, however, important reductions in children's health-care encounters in which injuries or ingestions were detected and modest reductions in welfare use among participating mothers. At the age four-and-a-half follow-up, women who received home visits by nurses had fewer subsequent births (averaging 1.15 vs. 1.34 births) within four and one-half years following the birth of their first child.<sup>4</sup>

---

<sup>4</sup>Unless otherwise indicated, all findings are from the following studies: Harriett Kitzman, David L. Olds, Charles R. Henderson, Jr., Carole Hanks, Robert Cole, Robert Tatelbaum, Kenneth M. McConnochie, Kimberly Sidora, Dennis W. Luckey, David Shaver, Kay Engelhardt, David James, and Kathryn Barnard, "Effect of Prenatal and Infancy Home Visitation by Nurses on Pregnancy Outcomes, Childhood Injuries, and Repeated Childbearing: A Randomized Controlled Trial," *Journal of the American Medical Association* 278 (8) (August 27, 1997): 644-652; David Olds, Harriet Kitzman, Robert Cole, JoAnn Robinson, Kimberly Sidora, Dennis W. Luckey, Charles R. Henderson, Carole Hanks, Jessica Bondy, and John Holmberg, "Effects of Nurse Home-Visiting on Maternal Life Course and Child Development: Age 6 Follow-Up Results of a Randomized Trial," *Pediatrics* vol. 114, no. 6 (December 2004): 1550-1559; David L. Olds, Harriet Kitzman, Carole Hanks, Robert Cole, Elizabeth Anson, Kimberly Sidora-Arcoleo, Dennis W. Luckey, Charles R. Henderson, Jr., John Holmberg, Robin A. Tutt, Amanda J. Stevenson, and Jessica Bondy, "Effects of Nurse Home Visiting on Maternal and Child Functioning: Age-9 Follow-up of a Randomized Trial," *Pediatrics* 120, no. 4 (October 2007): 832-845; Harriet J. Kitzman, David L. Olds, Robert E. Cole, Carole A. Hanks, Elizabeth A. Anson, Kimberly J. Arcoleo, Dennis W. Luckey, Kchael D. Knudson, Charles R. Henderson, and John R. Holmberg, "Enduring Effects of Prenatal and Infancy Home Visiting by Nurses on Children: Follow-up of a Randomized Trial Among Children at Age 12 Years," *Archives of Pediatrics and Adolescent Medicine* 164, no. 5 (May 2010): 412-418; and David L. Olds, Harriet J. Kitzman, Robert E. Cole, Carole A. Hanks, Kimberly J. Arcoleo, Elizabeth A. Anson, Dennis W. Luckey, Michael D. Knudtson, Charles R. Henderson, Jessica Bondy, and Amanda J. Stevenson, "Enduring Effects of Prenatal and Infancy Home Visiting by Nurses on Maternal Life Course and Government Spending: Follow-up of a Randomized Trial Among Children at Age 12 Years," *Archives of Pediatrics and Adolescent Medicine* 164, no. 5 (May 2010):

At the age-six follow-up, there were positive effects on some indicators of children's vocabulary and mental health. There were no differences in reading or behavioral indicators. Women who received home visits from nurses continued to have fewer subsequent births and subsequent children, and had fewer months of receiving welfare assistance. Similar findings were reported at the age-nine and age-twelve follow-ups.

**Cognitive.** There were no effects on mental development, based on the Bayley Scales of Infant Development when the children were tested at the age four-and-a-half follow-up. At the age-six follow-up, nurse-visited children had higher PPVT-III scores than the control group (84.32 vs. 82.13, an effect size of 0.17 SD), Kaufman Assessment Battery for Children (KABC) scores (92.34 vs. 90.24, an effect size of 0.18 SD), and better indicators of mental health (an effect size of 0.16 SD). All of these were significant at the .05 level. (See Appendix 1 for a further discussion of effect sizes and their interpretation.)

At the age twelve follow-up, there were no statistically significant differences in Peabody Individual Achievement Test (PIAT) reading and math scores between the program group and the control group. For those children born to mothers with low psychological resources, children in the program group had statistically significantly higher PIAT scores in math and reading (88.78 vs. 85.70, an effect size of 0.25 SD).<sup>5</sup>

**School readiness/performance.** At the age-nine follow-up, the NFP team found no statistically significant differences between the program and control groups in reading and math GPA or achievement tests scores for the groups as a whole. However, for those children born to mothers with low-psychological resources, the program group had statistically significant higher GPAs in math and reading (2.68 vs. 2.44, an effect size of 0.22 SD) and group achievement test scores (44.89 vs. 35.72, an effect size of 0.33 SD).

At the age-twelve follow-up, these trends continued. The NFP team found no statistically significant differences between the program group and the control groups for math and reading GPA, and group achievement test scores. For children born to mothers with low psychological resources, the program group had significantly higher average GPAs in math and reading from grades 1–6 (2.46 vs. 2.27, an effect size of 0.18 SD) and group achievement test scores (40.52 vs. 34.85, an effect size of 0.22 SD).

---

419-424.

<sup>5</sup>Harriet J. Kitzman, David L. Olds, Robert E. Cole, Carole A. Hanks, Elizabeth A. Anson, Kimberly J. Arcoletto, Dennis W. Luckey, Kchael D. Knudson, Charles R. Henderson, and John R. Holmberg, "Enduring Effects of Prenatal and Infancy Home Visiting by Nurses on Children: Follow-up of a Randomized Trial Among Children at Age 12 Years," *Archives of Pediatrics and Adolescent Medicine* 164, no. 5 (May 2010): 412–418.

In addition, there were no statistically significant differences on measures of grade retention or special education placement for the group as a whole and for children born to mothers with low psychological resources.

**Socioemotional development.** There was no difference between the program and control groups at the age-six follow-up on internalizing or externalizing problems.

At the age-twelve follow-up, the NFP team found that children in the program group were significantly less likely to self-report “internalizing disorders” (which included anxiety or depression) than the control group (22.1 percent vs. 30.9 percent).

**Health.** During the first two years of their lives, children in the nurse-visiting group had fewer health “encounters” involving injuries and ingestions and were hospitalized fewer days for such conditions. These impacts were greatest for women with low psychological resources. The NFP concludes, based on an examination of the diagnoses associated with children’s hospitalizations and independent measures of mothers’ childrearing behaviors, that this impact may have been due to improved caregiving on the part of mothers in the program groups. These indicators were not measured in the additional follow-ups.

There were no reported differences between the program and control groups in the number of depressive or anxiety disorders at the age-nine follow-up.

**Behavior.** The Achenbach Child Behavior Checklist measured no effects on behavior when the children were tested at the age-four-and-a-half follow-up or the age-six follow-up. There were also no differences between the program and control groups in reported behavioral problems at school or at home at the age-nine follow-up.

At the age-twelve follow-up, there were no statistically significant differences on teacher-reported, parent reported, and self-reported externalizing behavioral problems; school-reported conduct grades; or on one measure of sustained attention.

**Crime/delinquency.** At the age-twelve follow-up, children in the program group were significantly less likely to have used cigarettes, alcohol, or marijuana in the past thirty days (1.7 percent vs. 5.1 percent), to have used fewer of these substances (0.02 vs. 0.08), and used those substances for significantly fewer days (0.03 days vs. 0.18 days) and over the past thirty days. Despite being statistically significant, the actual differences are so small as to be practically meaningless. There was no statistically significant difference in the percent ever arrested.

**Early/nonmarital births.** Data apparently either not collected or not reported.

**Economic outcomes.** Data apparently either not collected or not reported.

**Effects on parents.** The NFP team measured a number of different outcomes for mothers in the program and control groups.

*Pregnancies.* At the age-four-and-a-half follow-up, the NFP team found a reduction in subsequent pregnancies among women who received nurse home visits (1.15 vs. 1.34) and a longer period between the birth of the first and second child (30.25 vs. 26.60 months). Results were similar at the age-six and age-nine follow-ups, where there was a reduction in subsequent pregnancies among women who received nurse home visits (at age-six follow-up: 1.16 vs. 1.38, an effect size of 0.22 SD; no data for age-nine) and a longer period between the birth of the first and second child (at age-six follow-up: 34.38 vs. 30.23, an effect size of 0.26 SD; at age-nine follow-up: 40.73 vs. 34.09, an effect size of 0.29 SD). In addition, at the age-six follow up, women who received nurse home visits had fewer children (1.08 vs. 1.28, an effect size of 0.22 SD), although there was no statistically significant difference at the age-nine or age-twelve follow-ups.<sup>6</sup>

At the age four-and-a-half follow-up, almost all the reduction in rates of subsequent pregnancy and birth were among women who were unmarried. ~~unmarried—especially given that the rate of marriage by child age five was higher (at the 0.10 level) among nurse-visited women.~~ The NFP team reports that “the effects of the program on the rates, timing, and spacing of subsequent pregnancies was concentrated on women who initially had higher levels of psychological resources, while the effect of the program on other outcomes was essentially equivalent for both high- and low-resource mothers.”<sup>7</sup>

*Marriage and cohabitation.* The women in the program group had higher rates of cohabitation (43 percent vs. 32 percent) and of living with the father of the child (18 percent vs. 13 percent) at the age four-and-a-half follow-up but there were no statistically significant differences at the age-six follow-up. At the age-nine follow-up, women in the program group had been living with their current partner longer than women in the control group (51.89 months vs. 44.48 months, an effect size of 0.23 SD), and had lived with an employed partner longer over the course of the nine years (54.95 months vs. 46.01 months, an effect size of 0.30 SD). However, there was no statistically significant differences in whether they lived with the father of their child or if they were married.

At the age-twelve follow-up, women in the program group had been living with their

---

<sup>6</sup>David L. Olds, Harriet J. Kitzman, Robert E. Cole, Carole A. Hanks, Kimberly J. Arcoleo, Elizabeth A. Anson, Dennis W. Luckey, Michael D. Knudtson, Charles R. Henderson, Jessica Bondy, and Amanda J. Stevenson, “Enduring Effects of Prenatal and Infancy Home Visiting by Nurses on Maternal Life Course and Government Spending: Follow-up of a Randomized Trial Among Children at Age 12 Years,” *Archives of Pediatrics and Adolescent Medicine* 164, no. 5 (May 2010): 419–424.

<sup>7</sup>Kitzman et al. 2000, 1987.

partner longer than women in the control group (59.6 months vs. 52.7 months), but there was no statistically significant difference in whether the mother was partnered, cohabitating or married to the child's biological father.

*Government benefits.* At the age four-and-a-half follow-up, nurse-visited mothers spent less time receiving Aid to Families with Dependent Children (AFDC) (26.6 months vs. 32.6 months) and food stamps (41.6 months vs. 45.0 months) between the time of enrollment and three years following the end of the program. This was also true in the age-six follow-up; nurse-visited mothers spent less time receiving AFDC between the first and second follow-up (7.21 vs. 8.96, a 0.22 SD effect size) and food stamps (9.67 vs. 11.50, a 0.24 effect size). Between the age-six and the age-nine follow-up, there was no statistically significant difference in the months spent on TANF, but the difference remained for food stamp receipt (4.89 vs. 5.92, a -0.21 SD effect size).

At the age-twelve follow-up, for the cumulative twelve-year period, women in the program group received food stamps in significantly fewer months per year than the control group (6.27 months vs. 6.86 months, an effect size of -0.13 SD) and received significantly less in food stamps (\$3,222 vs. \$2,870, no effect size reported). There was no statistically significant difference in Medicaid or AFDC/TANF use, but, in combination, women in the program group received less money in government benefits on average than did women in the control group (\$8,772 vs. \$9,797, no effect size reported.)

The NFP team noted that the reduction in AFDC and food stamps use was comparable to the effect observed in the Elmira study for low-income, unmarried women, where the largest declines did not appear until after the children entered school. There were no statistically significant impacts on the use of Medicaid over the nine-year period.

*Other.* There were no statistically significant effects on maternal education achievement or employment at all follow-ups.

At the age-twelve follow-up, the program group was significantly less likely to have their roles as a mother impaired due to alcohol or drugs (0.0 percent vs 2.5 percent, no effect size reported), and more likely to have a sense of "maternal mastery" ( 101.04 vs. 99.6, an effect size of 0.14 SD). There were no significant differences on partner violence, alcohol or drug use, maternal arrests, if the mother had been jailed, or foster care placements.

**Benefit-cost findings.** Apparently a benefit-cost analysis was not performed specifically for the Memphis study. The approximate annual cost of the program was about \$4,600 per family per year (in 2006 dollars).<sup>8</sup>

---

<sup>8</sup>David L. Olds, Harriet J. Kitzman, Robert E. Cole, Carole A. Hanks, Kimberly J. Arcoleo, Elizabeth A. Anson, Dennis W. Luckey, Michael D. Knudtson, Charles R. Henderson, Jessica Bondy, and Amanda J. Stevenson, "Enduring Effects of Prenatal and Infancy Home Visiting by Nurses on Maternal Life Course and

## Overall Assessment

The NFP was carefully evaluated using random assignment by David Olds, now professor of pediatrics at the University of Colorado Health Sciences Center, Harriet Kitzman, associate professor of nursing at the University of Rochester, and their colleagues.

**Program theory.** According to Olds, the conceptual framework guiding the design of the nurse-visitation program and its evaluation was based on evidence that “suggests that parental behavior is the most immediate, powerful, and potentially alterable influence on child health during pregnancy and the early years of the child’s life.”<sup>9</sup> Through nurse visits, the program sought to improve women’s health-related behaviors, parenting skills, and personal development. The evaluation, designed to assess improvements in child health and development, parental health-related behaviors, and qualities of parenting skills, was appropriate within this context.

**Program implementation.** The nurse home visitation program was operated by the Memphis/Shelby County Health Department. Nurses completed an average of seven visits (range of zero to eighteen) during pregnancy and twenty-six (range of zero to seventy-one) from birth to the child’s second birthday. This is about the same level of service that was provided in the Elmira trial. There was, however, considerably more staff turnover in Memphis, due to a nurse shortage, so that 37 percent of participating families experienced a disrupted relationship with the nurse originally assigned to them.

Pregnant women in the control group were offered free screenings for their children for sensory and developmental problems and those with possible problems were referred to other specialists. Some also received free transportation by taxi to regular prenatal and well-child care visits. As The NFP team pointed out with respect to the Elmira trial, “The transportation and screening may have elevated the functioning of the comparison group and thus reduced the range over which the program might produce positive effects.”<sup>10</sup>

**Assessing the randomization.** There were no reported problems in the randomization process. A comparison of various program group families on a wide range of baseline characteristics indicated that the groups were comparable, with two exceptions: the program

---

Government Spending: Follow-up of a Randomized Trial Among Children at Age 12 Years,” *Archives of Pediatrics and Adolescent Medicine* 164, no. 5 (May 2010): 419–424. This cost figure represents all program costs, including administrative and indirect costs. The program lasted between 2.4 and 2.5 years, depending on when during pregnancy the mother enrolled.

<sup>9</sup>Olds, 1992, 705.

<sup>10</sup>David L. Olds, Charles R. Henderson, Charles Phelps, Harriet Kitzman, and Carole Hanks, “Effect of Prenatal and Infancy Nurse Home Visitation on Government Spending,” *Medical Care* 31, no. 2 (February 1993): 165.

group was slightly more likely to have an unemployed household head and a slightly lower household income. The range of baseline characteristics collected was considerably broader than those found in most projects and included information about the women's socioeconomic characteristics, mental health, personality characteristics, obstetrical histories, health-related behaviors (such as smoking and drug use), social support network, and intellectual functioning.

**Assessing statistical controls in experimental and nonexperimental evaluations.** The evaluation was based on random assignment, so selection bias should not be a serious problem. The NFP team used an extensive array of covariates to refine their estimates.

**Sample size.** The overall sample of the Memphis intervention consisted of 1,139 families.

**Attrition.** Thirty-one women in control group two, plus eleven in program group four, were excluded from the analysis of the postnatal phase of the program after they experienced a fetal or infant death. Including just those cases where infants survived, the NFP team achieved a response rate of 91 percent of mothers (90 percent for the control group and 93 percent for the program group) at the age four-and-a-half follow-up, the age-six follow-up,<sup>11</sup> the age-nine follow-up (92 percent for the control group and 89 percent for the program group),<sup>12</sup> and a response rate of 86 percent at the age-twelve follow-up (86 percent for the control group and 87 percent for the program group).<sup>13</sup> By the age-nine follow-up, attrition for the children was slightly higher, as achievement test scores were available for only 83 percent of the children and teacher report forms were completed for only 81 percent of the children. By the age twelve follow-up, the completion rate for child assessments was 84 percent; for teacher reports it was 79 percent and for school records it was 92 percent. The NFP team achieved similarly high rates of completion for office-based assessments, as well as from pediatric and AFDC records. They reported that attrition did not affect the comparability of the groups.

**Data collection.** The data collection relied on a various standardized tests, administrative data, and parent surveys.

**Measurement issues.** The main source of information for both follow-ups was a telephone interview. Some of the outcomes measured in the nurse home visiting project, however,

---

<sup>11</sup>Olds, et al., 2004, 1551.

<sup>12</sup>David L. Olds, Harriet Kitzman, Carole Hanks, Robert Cole, Elizabeth Anson, et al., "Effects of Nurse Home Visiting on Maternal and Child Functioning: Age-9 Follow-up of a Randomized Trial," *Pediatrics* 120, no. 4 (October 2007): 832–845.

<sup>13</sup>Harriet J. Kitzman, David L. Olds, Robert E. Cole, Carole A. Hanks, Elizabeth A. Anson, Kimberly J. Arcoleo, Dennis W. Luckey, Kchael D. Knudson, Charles R. Henderson, and John R. Holmberg, "Enduring Effects of Prenatal and Infancy Home Visiting by Nurses on Children: Follow-up of a Randomized Trial Among Children at Age 12 Years," *Archives of Pediatrics and Adolescent Medicine* 164, no. 5 (May 2010): 413.

were validated by comparisons across two or more data sources. For example, telephone interviews with the mother at twenty-four and fifty-four months postpartum were used to determine the number of months that they participated in AFDC, food stamps, Medicaid, and the Special Supplemental Nutrition Program for Women, Infants, and Children (WIC). The NFP team also accessed administrative records from the Tennessee Department of Human Services (recording a match in 93 percent of the cases) to verify AFDC or food stamp receipt. Comparing the results from both sources, the NFP team found “small differences between these sources of data and no significant treatment differences in mothers’ accuracy of report.”<sup>14</sup> They further add:

Some of the outcomes were based on maternal self-report and they often covered long intervals for recall. Nevertheless, it is important to note that analyses of the AFDC and food stamp data produced virtually identical results irrespective of whether maternal report or administrative data were used as outcomes. This increases our confidence in the mother’s ability to report other outcomes for which we did not have administrative data to check the accuracy of self-report.<sup>15</sup>

**Generalizability.** The positive impacts of the Memphis trial were concentrated among children born to women with low psychological resources having their first child. The Memphis trial suggests that some of the positive findings of the Elmira trial could be replicated with a sample of primarily black, low-income women in a major urban area. Longer-term follow-up is needed to determine whether these effects are lasting and whether the program yields other benefits, such as reduced delinquency, when the study children become adolescents.

**Replication.** The Memphis nurse visiting program was a replication of the Elmira program. A second replication is underway in Denver (see chapter 20), also by the NFP team. But, further replication and evaluation by an independent evaluator following the identical intervention protocol would be desirable. A group of independent evaluators has recently completed a randomized controlled trial of the NFP as it is replicated in Louisiana parishes.

Home visiting using paraprofessionals or other protocols has been widely tested. One comprehensive review of home visiting programs found that such programs have been less effective than the nurse home visiting model applied in Elmira.<sup>16</sup> Thus, it is important to bear in mind that the program model and program content are very important. As Richard Berk, professor of Criminology and Statistics at the University of Pennsylvania, and Peter Rossi, former

---

<sup>14</sup>Kitzman et al., 2000.

<sup>15</sup>Kitzman et al., 2000.

<sup>16</sup>See David L. Olds and Harriet Kitman, “Review of Research on Home Visiting for Pregnant Women and Parents of Young Children,” *The Future of Children* 3, no. 3 (Winter 1993): 53–92, <http://www.futureofchildren.org> (accessed July 11, 2002).

professor at the University of Massachusetts (Amherst), note, “Replications of a given evaluation may be used to incrementally define the boundaries within which generalization is possible.”<sup>17</sup>

**Evaluator’s description of findings.** The reduction in subsequent pregnancies was smaller than the corresponding effect in the Elmira study. However, the reduction in closely spaced births (less than two years apart) was about the same. The NFP team notes that this “is important because of its implications for the improved outcomes of subsequent pregnancies and for parents’ abilities to become economically self-sufficient.”<sup>18</sup> The NFP team was careful in describing their findings, cautioning readers that the positive impacts were limited to low-income, unmarried mothers and provide little benefit to the “broader population.”<sup>19</sup>

**Evaluator’s independence.** The NFP team both developed and evaluated the nurse visiting model. They, however, support independent evaluations of the model assuming it is implemented with fidelity. Moreover, the evaluation findings have been published in high-quality peer-reviewed journals.

**Statistical significance.** Statistical significance was reported for variables with p values at 5 percent or lower.

**Effect sizes.** Apparently effect sizes were either not calculated or not reported for the age four-and-a-half follow-up but were reported for the age-six, age-nine, and age-twelve follow-ups. Effect sizes generally ranged between 0.15 SD and 0.30 SD. (See Appendix 1 for a further discussion of effect sizes and their interpretation.)

**Sustained effects.** The evaluation examined the impacts through age twelve.

**Benefit-cost analysis.** Apparently not performed.

**Cost-effectiveness analysis.** Apparently not performed.

---

<sup>17</sup>Richard A. Berk and Peter H. Rossi, *Thinking About Program Evaluation 2* (Thousand Oaks, CA: Sage Publications, 1999), 25.

<sup>18</sup>Kitzman et al., 2000.

<sup>19</sup>David Olds, Peggy Hill, JoAnn Robinson, Nancy Song, and Christina Little, “Update on Home Visiting for Pregnant Women and Parents of Young Children,” *Pediatrics* 30 (4) (April 2000): 117.

## Commentary

David Olds\*

Over the past decade, we have witnessed growing interest in the potential of early intervention to alter the life-course trajectories of vulnerable children and families, with the promise of these interventions' improving children's lives and societies' economic productivity. The federal government, states, and major philanthropies are shifting public and private investments in accordance with these promises. In light of this, it is crucial that we have a solid understanding of the evidentiary foundations for these claims. The children and families we seek to help deserve interventions that work and tax payers deserve accountability in public investments.

One of the challenges facing this field is that until recently there have been few broadly accepted standards for determining whether interventions actually work. This has made it possible for some advocates to claim that programs are effective on the basis of limited or flawed evidence. Common examples of flawed designs include quasi-experimental studies that cannot rule out significant selection factors that may account for apparent program effects; flawed randomization procedures in studies designed to be randomized trials; reliance on self-reported outcomes rather than objectively measured outcomes that have clear public health and social significance; and failure to replicate randomized trials to ensure the validity and generalizability of findings.

Moreover, even if one has solid evidence that meets high scientific standards, it is not always clear that interventions tested in trials can be effectively replicated in community settings. As investigators, we have a responsibility to ensure that program models tested in scientifically controlled studies can be reproduced in complex community settings in ways that reliably reproduce program effects found in our scientifically controlled studies. One crucial element in community replication work involves integrating information systems into evidence-based programs in ways that ensure on-going accountability and a foundation for continuous quality improvement and refinement of effective program models. If we are to create an evidentiary foundation that can be relied upon to guide policy and practice, we will need to improve our science and methods of translating findings into practice.

---

\*David Olds is a professor of Pediatrics and the director of the Prevention Research Center for Family and Child Health at the University of Colorado Health Sciences Center.

In recent years, these messages have begun to filter into the halls of Congress and state legislatures. Part of this is due to the gradual accumulation of well-conducted randomized trials of early interventions for children and families and to concerted efforts by some groups to base policy on good research, and in particular on randomized controlled trials. The Coalition for Evidence-Based Policy (a branch of the bi-partisan Council on Excellence in Government) has taken the lead in advocating for founding state and federal policy on the results of randomized controlled trials, promoting the same standards employed by the FDA in approving drugs and medical devices (<http://www.evidencebasedprograms.org/>). Partly as a result of the efforts of the Coalition, policy makers are beginning to ask the right questions about the evidence base for proposed policies and practices in education and human services. The bar is being raised.

It is with these challenges in mind that our team has designed and implemented the trials of the Nurse Family Partnership. We have been deeply concerned about the need for replication of findings in separate trials with different populations living in different contexts and points in our country's history; ensuring valid implementation of randomization; minimizing attrition bias; valid measurement of socially and clinically important outcomes; and examining long-term functional and economic effects of the program. We have taken the position that we ought not to promote the Nurse Family Partnership unless there is replicated evidence that it can produce meaningful effects on clinically important outcomes.

Our team thus welcomes the application of these standards to early intervention research as a whole, as Besharov and his colleagues have done in this book. Doing so will raise the evidentiary bar and increase the likelihood that programs with the greatest chances of success will be made available to children and families who need them. In the long run, raising this bar will help families, because scientists and advocates will base estimates of program impact on even more solid evidence, thus reducing the cynicism and lost hope that can result from faulty claims, and increasing the likelihood that policies and practices will achieve the goals for vulnerable children and families that have drawn all of us together in this common commitment.

Now, I offer a few comments on their review of our studies.

The reviews of our trials offered in this volume are appropriately critical, balanced, and consistent with our own concerns about the limitations of the evidence derived from these trials. As Besharov and his colleagues indicate, we have attempted to address most of our concerns about the limitations of our findings by replicating the original trial with different populations, living in different contexts, and at different points in our country's history.

One of Besharov and his colleagues' concerns with the three trials of the NFP has to do with the lack of independent evaluation. I acknowledge the importance of independent replication. Until recently, I have been concerned that this program model is relatively complex to deliver and that it is expensive to conduct well designed trials of it. Unless other investigators are highly motivated, there may be temptations to dilute the program or cut corners on the research in order

to get by with less money, undermining their ability to ascertain its real impact. To put this in perspective, in the mid-1980's, it took me four years and nine funding sources to raise the \$7 million required to test the first phase of the first replication of the NFP in Memphis. Less committed investigators would have not undertaken this effort or may have settled for a smaller sample size, a shorter program, or a less extensive measurement design. Moreover, until we began the national replication of this program in 1996, I was not convinced that we had articulated the essential elements of the program well enough to replicate it reliably in new community settings without investigator input to ensure quality program implementation. The first three trials of the NFP required varying levels of investigator oversight to ensure program integrity.

For these reasons, to date we have not promoted independent replication. But as the program is now replicated in hundreds of communities across the United States with support from the Nurse Family Partnership National Service Office ([www.nursefamilypartnership.org](http://www.nursefamilypartnership.org)), I support the eventual conduct of independent replication of our trials of the program as it is delivered in existing community settings. When summing across all program sites currently conducting the program, I suspect that the program on average is not producing the same level of impact as we found in our original trials, where we had more complete control over the delivery of the program. For this reason, I think we should hold off on testing the program in community settings until the system of national replication being built by the NFP National Service Office is fully operational. As the NFP national office refines its training and technical assistance to sites and as the basic program model is augmented by our research team to address inevitable shortcomings in its design (e.g., augmentations to address parental mental illness, intimate partner violence), I expect that local NFP programs eventually will exceed the impacts found in the original trials.

Finally, Besharov and his colleagues note that we have no economic analyses of the Memphis and Denver trials. The Washington State Institute for Public Policy has conducted an economic evaluation of this program that integrates data from the Elmira, Memphis, and Denver trials of the program to derive an overall estimate of program costs and savings, which estimates that on a per-family basis government and society realize about \$17,000 in savings for every family served.<sup>1</sup>

We owe a great deal to Besharov and his colleagues for helping to raise the evidentiary standards for early intervention research in this book, as this eventually will contribute to better intervention program models and policies.

---

<sup>1</sup>Steve Aos, Roxanne Lieb, Jim Mayfield, Marna Miller, and Annie Pennucci, *Benefits and Costs of Prevention and Early Intervention Programs for Children and Youth* (Olympia, WA: Washington State Institute for Public Policy, September 2004), <http://www.wsipp.wa.gov/rptfiles/04-07-3901.pdf> (Accessed July 12, 2010).

**Note:** This report is open to public comments, subject to review by the forum moderator. To leave a comment, please send an email to [welfareacademy@umd.edu](mailto:welfareacademy@umd.edu) or fill out the comment form at [http://www.welfareacademy.org/pubs/early\\_education/chapter22.html](http://www.welfareacademy.org/pubs/early_education/chapter22.html).