Journal of Child Psychology and Psychiatry 48:3/4 (2007), pp 355–391 doi:10.1111/j.1469-7610.2006.01702.x

Programs for parents of infants and toddlers: recent evidence from randomized trials

David L. Olds,[1] Lois Sadler,[2] and Harriet Kitzman[3]

[1]University of Colorado, USA; [2]Yale University, USA; [3]University of Rochester, USA

Programs for parents of young children hold considerable promise for improving children's life-course trajectories and for reducing health and development problems and associated costs to government and society. To date, this promise has not been achieved. Fulfilling the potential of parenting interventions will require substantial improvements in current practice for developing and testing such programs. Intervention development will be improved if clinicians and investigators ground parenting interventions in theory and epidemiology; and carefully pilot them to ascertain program feasibility, participant engagement, and behavioral change prior to testing them in randomized trials. Studies of parenting interventions will be improved if they adhere to the highest standards for randomization; if they examine objectively measured outcomes with clear public health relevance; and if they minimize selection factors known to compromise the analysis of data. Policy and practice recommendations for parenting interventions will be improved if they are based upon replicated randomized controlled trials, if the interventions are tested with different populations living in different contexts, and if they are examined in dissemination studies before public investments are made in such programs. Procedures need to be developed to ensure that the essential elements of evidence-based parenting programs can be implemented reliably in a variety of practice settings so that they will produce their intended effects. To date, few programs have met these high programmatic and evidentiary standards, with the result that many large-scale policy initiatives for at-risk parents have failed. Evidence is accumulating, however, that some programs delivered by professionals, especially nurse home visiting programs for pregnant women and parents of young children, produce replicable effects on children's health and development, and that these programs can be reliably reproduced with different populations living in a variety of community settings. **Keywords:** Intervention, infancy, parenting, perinatal, prevention, research design.

In this article, we review studies of preventive interventions that have attempted to improve parental competencies early in the life of the child as a means of promoting child health, development, and behavior. We focus on interventions applied before problems with parenting have occurred, give particular attention to work that has been reported in the past decade, and synthesize and evaluate this work for the guidance it may provide in formulating the next set of scientific questions and in steering public policy and practice in ways that are likely to produce the most desirable impact.

Testing and disseminating preventive interventions that focus on parenting early in the life cycle is fraught with challenges, including choosing the right aspects of parenting to attempt to improve, understanding salient barriers to or facilitators of behavioral change, designing interventions that reliably engage parents and bring about changes in parental competencies, conducting scientifically valid trials of those interventions, and moving those that are well tested into effective practice. Even though there are many questions that need to be resolved, we think that the literature is beginning to provide a sufficiently coherent picture about what is likely to work, so that some policy and practice recommendations can be made with some confidence.

In this review, we give particular attention to the extent to which intervention trials followed detailed procedures for developing promising interventions (Mrazek & Haggerty, 1994) and conducting valid preventive intervention research (Begg et al., 1996; Moher, Shulz, & Altman, 2001; Flay et al., 2004). The recommended process for intervention development focuses on the use of epidemiology and theory for the identification of target populations, intervention focus, content, and methods; careful attention to participant engagement and behavioral change; and pre-testing and piloting of interventions in small-scale randomized controlled trials (RCTs) prior to the conduct of larger trials (Mrazek & Haggerty, 1994). The review focuses only on those interventions examined in randomized trials, and gives particular attention to features of research design that can have substantial influence on the validity of trials, including the method and timing of randomization, accounting for all study participants randomized, measurement of outcomes with clear public health value, and reliance on replication (Begg et al., 1996; Moher et al., 2001; Flay et al., 2004). It is especially important that interventions found to work in one setting be replicated in at least one community setting prior to its being offered for public investment. Given that intervention trials are conducted with the goal of possibly affecting public investment in the intervention, we have addressed

Conflict of interest statement: No conflicts declared.

© 2007 The Authors
Journal compilation © 2007 Association for Child and Adolescent Mental Health.
Published by Blackwell Publishing, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA

issues of program cost and cost savings whenever cost data are available. Improving these aspects of intervention development and research is crucial for enhancing the lives of vulnerable children and families through policy and practice.

Why parenting?

In hundreds of studies, there is overwhelming consistency in the relationship between features of early parental care and child intellectual, behavioral, and emotional outcomes (see, for example, Bornstein, 1995). Sensitive, responsive care in the early months of the child's life is especially important. When parents tune into their infants' communicative signals, interpret them accurately, and respond in ways that meet their infants' needs, children are more likely to respond in synchronous ways, display signs of secure infant attachment, and exhibit better behavioral and emotional adjustment later in life.

These findings are consistent with an evolutionary view of parent–child relationships, which emphasizes the importance of parental commitment and attending to the needs of the child (Bowlby, 1969). Mayes and colleagues have noted that the capacity for parental commitment and care must be embedded in highly conserved brain-based systems that are activated at developmentally governed periods to promote the child's protection and development (Mayes, Swain, & Leckman, 2005). From the parent's perspective, the period surrounding birth, and especially birth of a first child, is likely to require the greatest transformation of the individual's hedonic homeostasis (Clutton-Brock, 1991) and corresponding brain-based neural circuitry. In spite of powerful evolutionary pressures to protect and promote the health of infants, parents differ in the degree to which they competently manage this set of tasks. The evolutionary perspective does not account for these individual differences in parenting, and is hard to test empirically, in part because differences in species and strain can be so large that generalizations across species are problematic (Mayes et al., 2005).

In humans, there is some controversy about the role that parenting plays in determining the functioning of offspring, given shared genetic endowment and the likelihood that children evoke different patterns of behavior from their parents. Is the relationship causal or are the well-functioning offspring of sensitive and responsive parents simply revealing their genetic endowment shared with their parents? Adoption studies have suggested that genetically vulnerable children are more likely to express psychiatric problems (Tienari et al., 1994) or criminality (Bohman, 1996) when they are reared in dysfunctional families than when they are reared in well-functioning families. The recent work of Caspi et al. (2003, 2004) shows that polymorphisms in genes that encode monoamine oxidase-A and the serotonin transporter (5-HTTLPR) interact with qualities of the care-giving environment, especially child abuse and neglect, to determine the likelihood that the child will exhibit, respectively, severe antisocial behavior and violence on the one hand or depression on the other. These genetic vulnerabilities are expressed only under conditions of extreme stress or child abuse. In the extreme, at least, parenting can interact with the genotype to affect child outcomes.

In a series of ingenious cross-fostering and handling experiments with rodents, Meaney and his colleagues have shown that maternal care (licking/grooming), independently of genetic background, affects the amount of licking and grooming female pups, once they reach maturity, provide to their offspring. The female offspring of low licking and grooming dams can be turned into high licking and grooming dams through cross-fostering by high licking and grooming dams or by handling (Francis, Diorio, Liu, & Meaney, 1999). Moreover, the early care-giving experiences of rats and rhesus monkeys have been shown to affect individual differences in both the maternal behavior of offspring and their reactivity to stress (reviewed in Mayes et al., 2005). A critical question is whether corresponding improvements in early parental care-giving and child outcomes can be produced in humans.

Promoting competent care-giving in humans

The implications of these highly innovative studies for human interventions are not as straightforward as it may seem on the surface. It is one thing to have evidence that maternal care affects child outcomes. It is quite another to bring about reliable adaptive changes in maternal care in humans, especially given that contextual factors, parents' earlier experiences, their current behavioral dispositions, and genetic makeup can have profound impacts on the care they provide to their offspring. In spite of these constraints, a number of trials have been reported in the past 10 years that give us reason to believe that carefully crafted programs aimed at improving parents' early care of the young child can have significant and enduring effects on children's health and behavioral adaptation. A number of such efforts have failed to produce the desired effect, however, and it appears that the failure can be traced to insufficient development of the intervention models with clear attention to engaging parents in the program and specification of methods for reliably bringing about changes in targeted aspects of parenting or family context.

Considerable attention has been focused on issues of program implementation in intervention research (e.g., Fixsen, Naoom, Blase, Friedman, & Wallace, 2005), given that even well-designed programs delivered poorly will not produce their intended effects.

© 2007 The Authors
Journal compilation © 2007 Association for Child and Adolescent Mental Health.

Evidence is mounting, however, that the more critical issue has to do with program design. Indeed, a case can be made that a significant portion of the variation in program implementation can be traced to differences in critical features of program design, such as selecting the right target population, intervening at points in development during which individuals have a greater sense of vulnerability, and providing services that from the participants' perspectives are likely to reduce that vulnerability. Programs that are more thoroughly developed in these respects are more likely to engage parents, produce reliable reductions in dysfunctional care, and produce effects on child outcomes.

We have chosen to focus on interventions that have been tested in RCTs, as they produce the most valid results when the question has to do with program impacts. There is substantial variation in the quality of randomized trials, however. We will use this review to identify common challenges with research design, implementation, and analysis that deserve to be addressed as new trials are undertaken. We highlight these issues with the hope that doing so will spur improvement in research design and implementation.

In recent years, the field of prevention science has developed a set of standards for the development of promising interventions that relies on theory, basic research, and epidemiology to identify targeted risk and protective factors as crucial elements in this process (Mrazek & Haggerty, 1994). The combination of systematic intervention development with rigorous testing (Flay et al., 2005) has led to the identification of highly effective preventive interventions in childhood and adolescence that consistently meet high evidentiary standards and that are recommended for public funding (Aos, Lieb, Mayfield, Miller, & Pennucci, 2004; Elliott, 1998).

Intervention development

The Institute of Medicine (IOM) recommends that preventive interventions go through a set of pre-trial stages of intervention development and formative research to increase their likelihood of success (Mrazek & Haggerty, 1994), including the use of epidemiologic data to identify putative mediators that contribute to the outcome targeted for prevention, the use of theoretically and empirically grounded strategies for changing those mediators, careful pre-testing and piloting of the interventions, and the conduct of small-scale RCTs (Table 1). Only when interventions have met these earlier milestones should they be subjected to full-scale RCTs. The results of those larger trials, in turn, should be used to refine the intervention in an ongoing cycle of research and program improvement.

In the IOM framework, parenting is considered a mediator targeted for improvement because of its putative influence on a wide range of child outcomes. It is not surprising that parenting is a core mediator in many of the more successful preventive interventions examined from infancy through adolescence (Elliott, 1998).

Table 1 Standards for intervention development

Use of theory, epidemiology, and developmental research to identify:
relevant aspects of parenting to address
methods of changing behavior
barriers to and facilitators of competent parenting (putative mediators)
Construction of preliminary program model using theory, research, qualitative research, and focus groups of parents and prospective service providers to determine:
feasibility
acceptability
Conduct of pretest and pilot studies to estimate:
Parental engagement
Change in targeted mediators and aspects of parenting
Sample size required for test of intervention in large-scale RCT
Refine existing program plan based upon pretest and pilot work before testing in full-scale RCT
Use results of RCTs to conduct formative research to improve the intervention model, which then will be tested in subsequent trials in process of continuous refinement of evidence-based intervention

Source: Adapted from Mrazek & Haggerty, 1993

Following the IOM framework, epidemiologic and developmental data are reviewed to identify those aspects of parenting that are most crucial at this phase of the life cycle. It is important to understand barriers (e.g., maternal depression, intimate partner violence, compromised intellectual functioning) and facilitators of competent parenting (such as parenting self-efficacy), as addressing these ecological and personal factors is likely to affect interventions' success.

It is particularly important, in our view, for interventions to have a theory of program engagement, as many interventions that depend upon parent participation fail to reach and involve the targeted population. Programs need to be able to answer this fundamental question: 'Why would parents want to spend their time participating in this program?' The health belief model (e.g., Rosenstock, 1990; Spoth, Redmond, & Shin, 2000) is especially relevant in this regard, as it addresses parents' motivation to participate in preventive interventions and to change their behavior. The success of parenting programs will depend upon the degree to which parents' concerns and motivations are integrated into the program design and effective clinical methods for behavioral change are employed by the staff (Miller & Rollnick, 2002). We approach this review with the working hypothesis that programs that address these fundamental issues well will be more successful than those that ignore them or treat them superficially (Olds, Hill, Robinson, Song, & Little, 2000).

© 2007 The Authors
Journal compilation © 2007 Association for Child and Adolescent Mental Health.

For many interventions, the simplest indication of program implementation is the number of sessions completed. In earlier reviews of the literature on home visiting for vulnerable families, Olds and Kitzman (1990, 1993) emphasized the importance of intensive home visiting, with sufficient numbers of visits completed to accomplish the programs' goals. While completing some threshold number of sessions is needed, program success cannot be predicted with simple measures of the numbers of completed intervention sessions. Success has to be related to the clinical skill and persuasive power of the staff. Miller and Rollnick (2002) note that the most fundamental challenge is to motivate the individual to change; once motivation is activated, change can be accomplished in relatively short order. This observation, and the observation that higher risk parents sometimes receive more intervention sessions (Olds & Korfmacher, 1997), augurs against finding simple dose–response relationships between quantity of program delivered and program success.

Intervention research standards

Stages of research

Research on parenting interventions can be usefully examined within the broader framework of prevention science (Olds, 2006). Preventive intervention research has been modeled after that used in cancer control (Greenwald & Cullen, 1984). As shown in Table 2, in preventive intervention research, interventions are carefully developed and tested initially under well-controlled, randomized *efficacy* trials and then evaluated as delivered in increasingly naturalistic conditions through *effectiveness* trials. Efficacy trials are critical to learning about the effects of programs under ideal conditions, but they often have limited application to real-world settings because the researcher has exerted control over many aspects of the implementation process. Consequently, most of the interventions tested in efficacy trials require a second stage of testing in effectiveness trials in which interventions are examined again in community contexts. A third level of research, known as dissemination research, is needed to understand more completely those factors that affect the successful dissemination and implementation of evidence-based interventions (Flay et al., 2005).

Society for Prevention Research Evidentiary Standards

The Society for Prevention Research recently has published a set of evidentiary standards designed to guide research, policy, and practice regarding preventive interventions (Flay et al., 2005), some of which are included in Table 2. We have included those standards that we believe are most relevant to research design and reporting and have excluded here details on measurement and data analysis. For efficacy trials, Flay et al. require that investigators specify the intervention tested so that others can implement it; examine the public health outcomes that the intervention claims to affect; conduct follow-up assessment at least 6 months beyond the end of the intervention; use psychometrically sound outcome measures; use the most rigorous research designs possible (with RCTs being the clear preference); clearly specify the sample and how it was derived; demonstrate the practical significance of public health impact; and replicate consistent effects with at least two high-quality studies, which meet all of the standards given above. The full set of standards includes requirements for statistical analyses which are not covered here.

In order for an intervention to claim effectiveness, it must have met all of the standards for an efficacious intervention, plus the following: The program must be fully operationalized in manuals, training, and technical support; a theory of causal mechanisms needs to be stated; a clear statement must be made about for whom and under what conditions the intervention can be expected to be effective; measures of intervention exposure, integrity, and implementation should be available; the real-world

Table 2 Intervention research stages and standards

Stages of intervention research
Efficacy trials (under nearly ideal conditions)
Effectiveness trials (under real-life-conditions)
Dissemination research (studies of conditions that affect successful replication of evidence-based interventions)
Efficacy trials
Clear operationalization of intervention
Use of most rigorous research design possible
Clear specification of sample
Use of valid outcome measures
Appropriate statistical methods
Impact of practical public health value
Impacts maintained at least 6 months after end of intervention
Replication of program impact in at least two separate trials
Effectiveness trials must meet all of the standards for efficacy trials, plus:
Program operationalized in manuals, training, and technical support
Theory of causal mechanisms
Clear statement of population that benefits
Measures of intervention exposure, integrity, and implementation
Real-world target population and sampling method given
Practical value of intervention specified
Clear statement of hypotheses, including population most likely to benefit
Two high-quality trials
Dissemination standards
Evidence must meet standards for effectiveness
Evidence must be available that intervention can be delivered with fidelity to model tested
Cost information must be available
Intervention must be supported by monitoring and evaluation tools

Source: Adapted from Flay et al., 2004

© 2007 The Authors
Journal compilation © 2007 Association for Child and Adolescent Mental Health.

target population and method for sampling should be specified in order to clarify the extent to which the findings apply to the target population; the practical value of the intervention should be specified; and at least two high-quality trials should be conducted that meet all of these criteria.

Finally, for an intervention to be ready for broad-scale dissemination, it must meet all of the criteria for effectiveness; it must be supported by relevant materials and evidence that the program can be implemented with fidelity; it must have clear cost information; and it must be supported by monitoring and evaluation tools for providers so they can assess its performance as it is scaled up.

Rarely have programs supported in policy and practice met these standards. If we are going to use scarce resources to make a difference in parenting and child outcomes, however, we must bring research and policy-making into alignment with these standards.

Costs analysis

In policy discussions of early childhood programs focused on parenting, consideration of program costs and economic benefits will play a prominent role. The Washington State Institute for Public Policy (WSIPP), the research arm of the Washington State legislature, has conducted a thorough review and cost analysis of a wide range of prevention programs from the standpoint of their impact on a set of socially important outcomes (crime, substance abuse, educational outcomes such as test scores and graduation rates, teen pregnancy, suicide, child abuse and neglect, and domestic violence) (Aos et al., 2004). A number of the programs reviewed here are examined in this report. While not all of the cost savings likely to be derived from an investment in these programs are monetized, the cost analysis of programs reviewed in that report are nearly identical to those reported in a recent economic analysis of early childhood interventions produced by the RAND Corporation (Karoly, Kilburn, & Cannon, 2005). In two cases, the Washington State report examined cost savings while the Rand report did not; otherwise, their estimates of program costs and savings are nearly identical.

One of the challenges of applying consistent economic analyses across all early parenting programs is that the lengths of follow-up differ substantially among interventions, which may affect estimates of cost savings if interventions produce long-term effects. In addressing this issue, the Washington State Institute notes that the issue that must be addressed from the policy perspective is whether the evidence is strong enough to warrant dissemination and public funding when there is insufficient data to determine the degree to which a program produces long-term effects or recovers its cost.

Consort reporting standards

In 1996, the editors of high-impact medical journals agreed to a common set of standards for the reporting of randomized clinical trials known as CONSORT (Consolidated Standards of Reporting Trials) (Begg et al., 1996; Moher et al., 2001), some of which are included in Table 3. The CONSORT standards include specifications about elements that need to be included in reports (such as sample sizes, methods of randomization, masking of data gathering, numbers and reasons for participant attrition, and numbers of cases analyzed). There also is a requirement that an accounting of all cases recruited, registered in the trial, randomized, excluded, dropped, lost to follow-up and followed at subsequent phases of follow-up be fully accounted for by treatment assignment in a flow diagram or table.

We are particularly concerned about the timing of randomization in relationship to participant recruitment, as it is not uncommon for investigators

Table 3 Selected reporting standards for randomized controlled trials, based upon CONSORT

Sample specification, including eligibility criteria, settings and locations where data were collected
Intervention design and implementation
Statement of objectives and hypotheses
Outcomes clearly defined and measured, including specification of those that are primary and secondary
Sample size and how it was determined
Randomization
Method to generate random allocation sequence, including details of restriction, such as blocking or stratification
Allocation concealment
Implementation (who enrolled participants, who generated the allocation sequence, and who assigned participants to groups?)
Specification of statistical methods
Participant flow
Numbers randomized to each group
Numbers who received treatment
Numbers completing study protocol
Numbers analyzed for primary outcomes
Deviations from study design and reasons for deviations
Baseline data, including demographic and clinical characteristics of each group
Recruitment period with dates of recruitment and follow-up
Method of randomization, including its timing in relationship to recruitment
Masking of data gathering
Analyses of treatment-control equivalence on background
Numbers of cases analyzed, including whether analysis was 'intention-to-treat'
Ancillary Analyses, including subgroup analyses and adjusted analyses, indicating those pre-specified and those exploratory

Note: Does not include CONSORT standards for outcomes and estimation reporting, adverse events, and comment (interpretation, generalizability, and general interpretation of results in context of current evidence). The reader should consult the CONSORT standards for the complete set of standards for reporting the results of randomized trials (Begg et al., 1996; Moher et al., 2001).

© 2007 The Authors
Journal compilation © 2007 Association for Child and Adolescent Mental Health.

to randomly assign participants to intervention and control conditions prior to obtaining consent to participate and conducting baseline assessments. If participants are assigned before they agree to participate, it is likely that those assigned to one condition may decline participation if they are unhappy with their assignment. We are particularly concerned with the use of this procedure in studies of home visiting, in that parents who are involved in drug abuse or other illegal activities may refuse to receive a home visitor for fear of being detected; these concerns are less salient for those assigned to the control group, as visitation and surveillance of conditions in the home are reduced. While it is possible that families may drop out from the control group if they are disappointed with not receiving home visits, it is our experience that the greater concern involves the most at-risk family members not wanting their antisocial behavior to be discovered and thus selecting out of home visiting interventions.

Employing the CONSORT standards presents a challenge for some of the studies reported on early parenting interventions because the scientific reports of these interventions either have been reported in non-medical journals or are available only in the form of unpublished reports, where the CONSORT standards have not been required uniformly. To clarify this issue, to the extent possible, we reviewed unpublished reports and contacted the authors of the reports to determine the extent to which the CONSORT standards were met, but perhaps not reported. We examine these features of study design and implementation and consider these issues in the interpretation of findings.

Finally, it is common for investigators to report subgroup effects when analyzing complex interventions with heterogeneous samples. While it is likely that such programs are going to affect different segments of the population in different ways, it is risky to place too much confidence in results derived from such analyses, as significant results may emerge simply as a result of investigators conducting so many statistical tests. There are three ways in which such analyses can yield results in which we can have greater confidence: 1) the subgroup effects can be hypothesized at the start of the trial; 2) random assignment can be conducted within stratification blocks on which subgroup analyses are based; and 3) the effects can be consistent with theory and research. The CONSORT standards require that the author specify those analyses that were pre-specified and those that were exploratory to help reviewers evaluate the meaning of statistically significant effects.

Scope of review

We have focused this review on randomized trials of interventions conducted in developed countries and reported in the past decade that include significant components focused on promoting parents' care of their infants and toddlers (pregnancy through child age three) before parenting problems have emerged. Because of substantial differences in health and human services, poorer economic conditions, and differences in culture, the results of trials from developing countries cannot be generalized easily to the contexts of developed societies, where most of the studies on parenting interventions have been conducted.

We have not conducted a systematic review, but rather have examined trials that in our opinion have significant implications for deepening our understanding of the science of improving vulnerable parents' care of their children and for guiding policy decisions in this area. We have excluded two of the most widely disseminated programs of support for parents, Triple P (Sanders, Markie-Dadds, & Turner, 2003) and the Incredible Years (Webster-Stratton, Reid, & Hammond, 2001, 2004) because there are no trials of these programs that focused exclusively on children less than three years of age. Moreover, those interventions tend to be applied primarily once parenting problems have emerged. A recent trial of a home-visitor intervention for infants with failure to thrive (Raynor, Rudolf, Cottrell, Marchant, & Cooper, 1999) and MacMillan's trial of nurse home visiting for parents who have abused and neglected their children (MacMillan et al., 2005) also introduce interventions once problems have emerged and thus do not fall into the scope of this review. While we may have missed some interventions that meet these criteria (programs for parents during pregnancy or the first three years of life that have been tested in randomized trials and reported since 1996), we have covered the major interventions that meet these criteria.

Promising perinatal interventions

During the past decade, evidence has accumulated that altering the earliest experiences of low birthweight newborns can improve their early brain development (both structure and function) and cognitive functioning later in infancy; and that home visiting and educationally enriched child care can have long-lasting effects on the cognition and behavior of low birth-weight newborns in the higher end of the low-birth-weight range (2000–2500 grams).

Infant Health and Development Program

Prematurity and low birth weight, of course, are significant risks for infant mortality and subsequent health and developmental problems (McGauhey, Staffield, Alexander, & Ensminger, 1991). In 1982, with major funding from the Robert Wood Johnson

© 2007 The Authors
Journal compilation © 2007 Association for Child and Adolescent Mental Health.

Foundation, a team of investigators conducted a study of an educational intervention for low birth-weight newborns known as the Infant Health and Development Program (IHDP).

Program model. The investigators adapted a promising program for the intervention, consisting of home visiting, parenting groups, and educationally enriched daycare for children birth through age 3, which previously had been tested with young children and their families living in poverty and produced highly promising effects (Ramey et al., 1992). The first year of service consisted of home visiting; in the second and third, families were provided center-based child care and monthly parent-group meetings. The center-based infant and toddler program was designed to promote children's exposure to increasingly complex cognitive tasks and language experiences and was articulated carefully in previous research. The home visiting was conducted by bachelors-prepared child development specialists who encouraged parents to play interactive games with their infants and toddlers (Sparling & Lewis, 1979). While the original sample for which the intervention was developed (low-income) was quite different from the one for which it was applied in this trial (low birth-weight), many of the details of program design and implementation were worked out in the original study.

Research design and methods. The investigators conducted an 8-site randomized controlled trial of this intervention. Nine hundred eighty-five low-birth-weight newborns were randomly assigned either to IHDP ($n = 377$) or a control condition (608). At the most recent follow-up at child age 18, assessments were completed on 67% of those assigned to IHDP and 63% of the controls. The sample was socio-demographically mixed (Infant Health and Development Program, 1990). Fortunately, the investigators stratified the randomization by children's birth-weight ($\leqslant$2000 vs. 2001–2500 g). In all reports on this program, the investigators provided meticulous accounting of those recruited, randomized and assessed at various stages of follow-up and used a variety of psychometrically sound methods of assessing maternal, child, and family functioning.

Major findings. In earlier phases of follow-up, the investigators found large intervention effects on cognition and behavior at the end of the program, with those benefits concentrated in children born in the higher birth-weight stratum (Infant Health and Development Program, 1990). By child ages 5 and 8, those effects began to attenuate, especially for those in the lower weight subsample (Brooks-Gunn et al., 1994; McCarton et al., 1997).

The investigators recently reported the results of a follow-up of the children at age 18 (McCormick et al., 2006). As in previous reports on this trial, the investigators provided a complete accounting of the participants from the stages of randomization through the 18-year follow-up broken down by treatment assignment. The primary analysis was based upon an intention-to-treat model. The outcomes were measured with direct tests of the children's achievement and other psychometrically sound methods.

Among 18-year-old children in the higher weight group, those in the intervention had better language development and math achievement and fewer risky behaviors than their control-group counterparts. There were no discernible long-term benefits for children in the lower weight stratum. Nor were there effects for either birth-weight group on grade retention and placement in special education, outcomes that might help offset the high cost of the intervention. Nevertheless, it is remarkable that 18 years after birth the program signal could be detected at all. This trial is thus of immense significance in telling us that something important can be achieved for vulnerable infants born in the 2000–2500 g range.

The authors reason that the very low birth-weight newborns may benefit from longer interventions into the school years (McCormick et al., 2006), but evaluations of the intervention that formed the basis of the IHDP program with socially disadvantaged children who were not low birth weight found no advantage of continuing intervention into the elementary school years (Campbell & Ramey, 1995). Why then would extending the intervention with very low birthweight children be more effective?

The Washington State Institute for Public Policy and the Rand Corporation have estimated that this intervention costs $49,000 per infant in 2004 dollars, with no estimated cost savings through child age 8. Fewell and Scott (1997) estimated that the program cost about $15,146 for the last year of program operation, in what appear to be 1988 dollars. The program is likely to have cost a lot less in the first year of operation given that center-based child care and parent-groups did not begin until the second year. Fewell and Scott did not, however, include administrative costs in their estimate. The most recent follow-up data are likely to lead to cost savings but the cost of the program is unlikely to be recovered by these improvements in functioning. We need to find ways of helping very low birth-weight babies and doing so at lower costs.

Newborn Individualized Developmental Care and Assessment Program (NIDCAP)

Over the past two decades, Heideliese Als and her team have developed a promising intervention for very low birth-weight newborns known as NIDCAP (Als, 1982).

Program model. NIDCAP is an intervention for very premature newborns (<30 weeks) applied in the NICU by caregivers and parents. It consists of daily

© 2007 The Authors
Journal compilation © 2007 Association for Child and Adolescent Mental Health.

observations of newborns' developmental competencies and recommendations to parents and NICU staff for ways of supporting infants' development. The intervention is based upon the premise that premature newborns are fetuses who find themselves in an environment (the technological neonatal intensive care unit) for which they are not adapted from an evolutionary perspective. NIDCAP attempts to reduce the discrepancy between the womb and the NICU by taking into consideration the individual infant's threshold of behavioral organization, by diminishing stress, and by supporting each infant's strengths and competencies. Developmental specialists help caregivers understand infants' stress and comfort signals and make recommendations for care-giving that are designed to protect infants' regulatory balance in an effort to promote their strengths and reduce regulatory vulnerability. The NIDCAP approach has been carefully developed in over two decades of research and is operationalized thoroughly in a set of training materials.

Research designs and methods. Several trials of NIDCAP have been conducted that show consistent effects on electrophysiological and developmental outcomes, but challenges with research design and implementation have led some reviewers to question the consistency of results (Jacobs, Sokol, & Ohlsson, 2002). We have reviewed the two most recent reports (Als et al., 2003; Als et al., 2004).

The first report presents the results for 92 infants who had been randomized in a 3-site trial to NIDCAP ($n = 45$) or a control condition ($n = 47$) (Als et al., 2003). The numbers originally randomized are not given, nor is there a CONSORT diagram to account for the timing of randomization in relationship to obtaining informed consent and completing baseline assessments, or to account for losses after randomization. The investigators examined medical, neurodevelopmental, and family outcomes.

In a more recent trial of NIDCAP (Als et al., 2004), a CONSORT diagram is provided. In this trial, 33 participants were randomized to NIDCAP ($n = 18$) or the control condition ($n = 15$). Sixteen of the NIDCAP infants and 14 controls remained in the trial through 9 months of age. The second study used a wide range of sophisticated analyses of brain structures and functions in addition to measures of child health, growth, and neurodevelopment to determine program impact.

Major findings. In the first trial (Als et al., 2003), NIDCAP infants were functioning better than controls on a variety of infant outcomes, including weight, length, head circumference, and enhanced autonomic, motor, state, attention, and self-regulatory functioning (Als et al., 2003). Unfortunately, the absence of accounting of those registered in the trial reduces the significance of these highly promising findings.

In the second trial, infants assigned to NIDCAP, compared to those in the control group, were found to have superior brain structure (mature fiber structure), brain functioning (e.g., increased coherence between frontal and a wide range of mainly occipital brain regions), and neurobehavioral functioning at 2 weeks and 9 months (Als et al., 2004). The complete accounting of study participants and timing of randomization increases the validity of these findings.

It will be important to conduct follow-up assessments of the children and families enrolled in trials of NIDCAP and eventually to conduct a multi-site trial of the intervention in a wide range of sites so its clinical and economic impact can be estimated more thoroughly.

Comment on perinatal interventions

While the home visiting program examined in Project Care (the intervention upon which IHDP was based) produced virtually no benefits when delivered to low-income parents and children (Ramey et al., 1992), it is possible that properly developed home visiting may be effective for parents of low birth-weight newborns, given their heightened sense of vulnerability and receptivity to offers of help (Olds & Kitzman, 1990; 1993). Promising results from the Vermont intervention program for low birth-weight newborns (Achenbach, Howell, Aoki, & Rauh, 1993) and a trial of paraprofessional home visiting for parents of term low birth-weight newborns in Jamaica (Gardner, Walker, Powell, & Grantham-McGregor, 2003) support the need for continuing research on home visiting for parents of low birth-weight newborns.

Home visiting to promote child health and development

Home visiting has received considerable attention in the US as a means of improving child health and development by improving women's prenatal health and parents' competencies in managing their lives and the care of their children. In 1989, the National Commission to Prevent Infant Mortality (1989) advocated for the expansion of home visiting programs as a means of preventing infant mortality; in 1991, the U.S. Advisory Board on Child Abuse and Neglect (1991) advocated for the expansion of home visiting programs as a means of preventing child abuse and neglect; and in 1998, the Council on Child and Adolescent Health of the American Academy of Pediatrics (1998) recommended that primary care be augmented with home visiting services for vulnerable children and families.

Most reviews of the literature on home visiting programs, however, have produced sobering pictures of the prospects for home visiting programs in general

© 2007 The Authors
Journal compilation © 2007 Association for Child and Adolescent Mental Health.

(e.g., Olds & Kitzman, 1993; Gomby, Culross, & Behrman, 1999; Sweet & Appelbaum, 2004). A recent review of the literature by the Centers for Disease Control (Hahn et al., 2003) provides a more optimistic picture of home visiting for the prevention of child abuse and neglect, but Gomby (2005) has noted that this optimistic picture is based upon the assumption that proxy measures for maltreatment (reported injuries, parents' attitudes toward care-giving) are good substitutes for official records, an assumption that has been questioned by some (e.g., Chaffin, 2004). What accounts for the popularity of home visiting and the varying reviews?

Home visiting programs for parents during pregnancy and the early years of the child's life share many common goals, but they differ in important ways in the backgrounds of the visitors, the segments of the parent population they target, the specific content and clinical methods of the programs, and the structure provided to the visitors in delivering the services.

Hawaii Healthy Start and Healthy Families America

When the US Advisory Board on Child Abuse and Neglect recommended that home visiting programs be made universally available throughout the United States, it promoted a program of paraprofessional home visiting for vulnerable families, known as the Hawaii Healthy Start Program (HSP) which had been developed and implemented state-wide in Hawaii. The Hawaii State legislature funded the development and evaluation of a pilot program on Oahu, in which there was no control group. Quasi-experimental studies of the pilot program found that graduates of the program had much lower rates of child maltreatment than did families with similar social characteristics not enrolled in the program (see Duggan et al., 1999). These findings led to the statewide expansion of the program.

Hawaii Healthy Start/Healthy Families America Program Model

The National Committee to Prevent Child Abuse (now known as Prevent Child Abuse America) took up this charge and developed a national initiative to prevent child maltreatment based upon HSP, known as Healthy Families America (HFA). The leaders of HFA have indicated that the HFA initiative is defined by 'its commitment to a set of principles rather than to a single, monolithic approach' (Daro & Harding, 1999). The problem with such a stance is that each HFA program is then a separate initiative or program model and it becomes impossible to determine whether HFA delivered in San Diego is the same program as delivered in New York City or Buffalo. The results of trials produced in one setting thus cannot be applied to others. Nevertheless, there are commonalities among most HFA programs that we summarize here: they identify families during pregnancy or the postpartum period using the Kempe Family Stress Checklist who are at risk for child abuse and neglect and offer them voluntary home visiting services; staff visit families in their homes from 3 to 5 years following a visitation schedule that diminishes in frequency as families improve in their functioning; and the visitors focus on helping parents become more competent parents. Because state and local HFA programs do not have to follow a 'monolithic' approach, there is substantial variability in local program design and implementation.

We now have a published trial of the original Hawaii program, unpublished reports of 3 trials of HFA programs (Healthy Families New York, Healthy Families Alaska, Healthy Families San Diego), plus a published trial of an augmented version of HFA conducted in Santa Barbara that can help us shed light on the potential value of this approach to supporting vulnerable families.

Hawaii Healthy Start

(Duggan et al. 1999; 2004) conducted a randomized trial of HSP in three organizations, each with two sites on the Island of Oahu, where the program was originally developed.

Research design and methods. The study was conducted as a randomized controlled trial. One of the strengths of this study is that the intervention was tested as the program was delivered in real-life community practice, making the results of the trial applicable to the program delivered in these and perhaps similar settings. The investigators randomized 730 at-risk families to HSP ($n = 395$), a control condition (290), or a testing control group ($n = 45$) to assess the impact of repeated measurement; 684 of these families (373 in HSP, 270 in the main control group, and 41 in the testing control group) consented to participate and completed baseline assessments. There is no CONSORT diagram and the determination of eligibility, randomization and baseline assessments were separated in time in ways that increased the possibility that participants might decline participation after they gained knowledge of their treatment assignments but before the baseline assessments were conducted. Outcome assessments were completed on maternal outcomes at child age 2 on 329 HSP participants and 238 controls (83% and 82% of those randomized, respectively).

The authors examined program effects using a sophisticated measurement design consisting of maternal report of parenting behaviors and other outcomes, observations of parental behavior and the home environment, medical record reviews, and substantiated child protective service reports. The authors' use of multiple sources of information to ascertain program impact increases the study's validity.

© 2007 The Authors
Journal compilation © 2007 Association for Child and Adolescent Mental Health.

For those who completed the baseline assessments, the treatment groups were essentially equivalent on standard socio-demographic background characteristics, but no information is given on levels of risk based upon the initial screening. Nor is there information on the risk status of those who declined participation after randomization.

According to Duggan and colleagues, the major challenge with this study is that the paraprofessional visitors completed far fewer home visits than called for by the program model and the program staff shifted from a risk reduction strategy to a 'strengths-based' approach to working with families. Within one year after being referred for the service, 51% of the families had dropped from the program. Thirty-one percent of the families refused the service, with the majority of those refusing early in the program. There were differences among the three agencies in features of program implementation. Nevertheless, the investigators conducted assessments and reported results on every case that was evaluated, irrespective of program participation.

Major findings. There were few indications that the program affected the targeted outcomes: Compared to parents in the control group, HSP parents were more likely to report that they had a pediatric provider who understood their concerns, but there were no program effects on use of well-child care or immunization rates. Nor were there program effects on families' use of other health and human services in the community.

HSP mothers reported better mental health at the one-year follow-up, but there were no effects on maternal life-course (e.g., subsequent pregnancies, educational achievement), or qualities of the home environment. On the other hand, HSP mothers reported greater use of non-violent methods of disciplining their child, experienced less stress in parenting at the end of the first and second years of the program, and felt more competent in their parenting at the end of the second year.

There were no program effects on parents' reports of abusive care-giving or observations of care given in the home, although HSP families reported less use of common verbal and corporal punishment, an effect due to one agency's reduction in parents' threatening to spank (Duggan et al., 2004). HSP mothers also reported less neglectful behavior through the first three years of the child's life.

There were no program effects on children's use of the emergency department, hospitalizations, or injuries; or on children's development measured through child age three. HSP children had higher psychomotor development than their control-group counterparts at one of the three agencies (Duggan et al., 1999). Duggan and colleagues point out that there were 4 variables on which the HSP families did better than the controls and yet 4 variables on which the HSP families did worse.

These results were disappointing, in part, the authors concluded, because of high rates of participant attrition, limited participant engagement, and scant training of supervisors and visitors in addressing the multiple, complex risks for child abuse and neglect. The implication of this interpretation is that the underlying program model is effective and that the major challenge has to do with poor implementation. As we shall see below, the results of the San Diego trial of HFA, in which HFA was augmented and implemented extraordinarily well, challenge the interpretation that the weak results are due simply to poor implementation.

Alaska Healthy Families Program

Recently, Duggan et al. (2005) reported on the evaluation of the Healthy Families program in Alaska, tested in the form of a randomized trial.

Program model. The program delivered in Alaska had the same program goals as the national Healthy Families America program and was similar to the one examined in the Hawaii trial reviewed above. Leaders of the Alaska program invested considerable resources in developing a well-articulated method of building trusting relationships with families and promoting adaptive behavior change consistent with program goals, including a significant investment in reflective supervision as a means of supporting the paraprofessional visitors' work with families. The HFAK program used Individual Family Support Plans (IFSP) and promoted families' taking responsibility for achieving their goals. Workers were expected to keep track of families' goals and their progress in achieving them. HFAK program sites were expected to adhere to national HFA critical elements promoted as crucial for program success (including identification of new parents who are in need; the use of voluntary services; and caseloads that are in the 15–20 families per-visitor range and supervisor-to-staff ratios that are 1:5). HFAK sites were required to go through the HFA basic training, participate in statewide meetings, and work toward HFA credentialing. In spite of all of this clarity of goals and methods, there was no uniformly applied program curriculum. Moreover, Duggan et al. concluded that the HFAK programs deviated substantially from the implementation standards advocated by HFA national and the state office responsible for conducting the program.

Research design and methods. Duggan and colleagues conducted a randomized trial of the program with a sample of 364 families randomized to treatment ($n = 179$) and control ($n = 185$) conditions, recruited from 6 communities.

Families consented to participate after they were determined to be eligible but before the baseline assessments were conducted. Between random

© 2007 The Authors
Journal compilation © 2007 Association for Child and Adolescent Mental Health.

assignment and the conduct of baseline interviews, 39 families were lost, some due to families changing their minds about participating; others could not be located. No information is provided on the how many of the 39 had been assigned to the treatment condition versus the control condition; nor is there information on the numbers who refused and could not be located by treatment assignment. Three hundred twenty-five families completed the baseline interview, 162 in the HFAK group and 163 in the control. Eighty-five percent of those who completed the baseline assessments (75% of those randomized) were assessed at 2 years of age, but the numbers assessed in each of the treatment conditions is not given, nor is the degree to which the treatment and control groups were equivalent among those assessed at 24 months.

The investigators developed a thorough measurement design that was consistent with the goals of the program. Outcomes were measured using maternal report of parenting behavior, observations of parenting and the home environment, medical records, child protective service records, and developmental tests of the child at age 2.

Major findings. The program developers hypothesized that the intervention would impact child abuse and neglect; caregivers' high school graduation, rates of rapid repeat births, parents' financial self-sufficiency, the use of safety plans among caregivers at risk of domestic violence, parenting stress, caregivers' mental health and substance abuse problems, linkage to a Medical Home, quality of the home environment, and delays in child development.

While the program had no impact on state-verified cases of child abuse and neglect, HFAK mothers reported using mild physical and psychological disciplinary tactics significantly less than did control mothers. There were no differences in other aspects of self-reported parenting behavior, observations of mother–child interaction, or children's rates of hospitalization and ambulatory care for conditions that were potentially preventable. HFAK families were, however, less likely to have extremely poor home environments, as measured by the HOME scale.

There were no differences between HFAK mothers in educational achievement, rapid repeat birth, household income, parental employment, families' use of community resources or social support, and domestic violence. Nor were there any differences in maternal depression, self-esteem, confidence in adult relationships, or maternal substance use, although HFKA mothers were less likely to report elevated parenting stress scores.

There were no statistically significant differences between HFAK mothers and the control group in their knowledge about parenting, recognition of developmental delay in their children, or acceptance of corporal punishment as a disciplinary tactic, although KFAK mothers reported greater confidence (self-efficacy) in their parenting role. There were no HFAK-control differences in children's linkage with a Medical Home, their use of preventive care, immunizations, although there was a trend ($p = .10$) for HFAK families to have homes that were more conducive to children's development as measured on the HOME scale. While there were no program-control differences in children's hospitalizations or emergency department visits for injuries requiring medical care, HFAK children scored higher on the Bayley scales of mental development at 24 months of age, were less likely to have scores that fell below 85 on that scale, and were reported by their mothers to have fewer behavioral problems that fell into the borderline or clinical range on the Achenbach Child Behavior Checklist.

For the four outcomes on which the program had significant overall effects, the impact of the program was greater for families at lower risk at baseline, indicated by such things as initial parenting stress index scores and the absence of domestic violence at baseline.

Many of the primary outcomes targeted by this intervention were not affected, but the program did improve children's mental development by child age 2, which is a significant accomplishment. This outcome is particularly remarkable, given that the program itself deviated so much from the program model standards. Given that a corresponding effect was found in the San Diego trial of HFA (described below), which washed out over time, it will be particularly important to see whether this effect endures.

New York State HFA evaluation

The New York State Office of Children and Family Services has conducted a randomized trial of the HFA program in 3 selected counties in New York State, and produced two reports on this trial. The first covered the first-year results and the second examined program impacts on parenting through child age 2 (Mitchell-Herzfeld, Izzo, Greene, Lee, & Lowenfels, 2005; DuMont et al., 2006).

Program model. Like other HFA home-visiting programs, *Healthy Families New York* (HFNY) uses population-based screening and assessments to target pregnant women and new parents deemed to be at risk for child abuse and neglect. Home visits are conducted by specially trained paraprofessionals who live in the community being served and share the same backgrounds as program participants. Services are intended to be intensive, ideally beginning during the prenatal period and lasting until the targeted child is five years old, or enrolls in Kindergarten or Head Start. While there were no pilot studies of this program prior to conduct of the trial, all three program sites had operated the program since HFNY began offering services in New York State

© 2007 The Authors
Journal compilation © 2007 Association for Child and Adolescent Mental Health.

in 1995 and were considered to be well managed administratively.

HFNY is credentialed by Prevent Child Abuse America/Healthy Families America as a multi-site system that is recognized as providing administratively sound quality assurance, training and technical assistance, policies and evaluative support to its sites. The 28 sites within the system are also recognized as providers of high quality home visitation services.

Research design and methods. One of the strengths of this study is that it is testing a program that is operating in community context. And while one may question whether the 3 sites chosen for the trial are representative of the entire set of 28 sites that operate the program throughout New York, it is admirable that the State has allocated resources to determine whether their investment in this program is yielding returns.

The numbers reported as randomized and registered in the trial vary between the first- and second-year reports. We rely primarily on the second-year report as it includes a CONSORT-type diagram (DuMont et al., 2006). The evaluators randomized 1297 families eligible for the program based upon preliminary screening. After randomization, 43 cases (26 HFA and 17 controls) were deemed ineligible primarily because of their living outside of the catchment area and their having language barriers; of the 1254 deemed eligible for the trial, 81 were subsequently excluded (42 HFA and 39 control) because of participant refusal, loss to follow-up, fetal or pregnancy loss, or the child was removed from the mother's custody (9 of the 14 pregnancy/fetal losses were in the HFA group; all 4 of the custody losses were in the HFA group). After these secondary screens for eligibility and exclusions, 1173 of the randomized parents completed the baseline interview (579 of the 647 assigned to the HFA group and 594 of the 650 assigned to the control condition). While the exclusion of cases after randomization is a violation of the intent-to-treat principle, the authors' clear specification of exclusions and drops by treatment condition increases the value of this report. The authors did attempt to complete assessments on every parent who completed a baseline interview, and were successful in completing interviews at year 1 on 90% and at year 2 on 85% of those who completed the baseline interviews (81% and 77%, respectively, of those randomized to each treatment). Among those who completed the follow-up interviews, the groups were similar on background characteristics. The evaluators measured outcomes with maternal report of their pregnancy outcomes, parenting and life-course; and reviews of state child protective service records.

In the second-year report, the authors conducted an exploratory analysis to examine whether program effects would be greater for that small portion of their sample who were similar to the population targeted by the Nurse Family Partnership (NFP), that is, they registered during pregnancy, were 19 years of age or less, and were having a first child – a 'prevention' group (Dumont et al., 2006). While the authors sought to create a target population similar to the one addressed by the NFP, it is important to note that the NFP registers women of all ages as long as they are having a first live birth and are poor. A second exploratory analysis examined whether program effects were greater for mothers who were more psychologically vulnerable by virtue of their having higher levels of depressive symptoms and limited sense of mastery; this analysis focused on the top 10% of the sample at greatest risk and again was constructed to replicate findings from the NFP, given that program impacts on parenting and child outcomes in trials of the NFP were greater for mothers who were more psychologically vulnerable (Olds, 2002).

Major findings. At the end of the first year of program implementation, in contrast to their counterparts in the control group, the mothers in HFNY reported fewer beliefs in the value of physically punishing infants; while the effect overall was small (and significant probably because of the large sample) the program effect was greater for mothers who were under 18 at registration and who were the least depressed at registration (Mitchell-Herzfeld et al., 2005). The HFNY–control differences in mothers' beliefs about physical punishment overall were greater for one of the three counties. While there also was a trend for HFNY mothers to have greater knowledge of child development, the effect was too small to be clinically meaningful. In one of the three sites, HFNY mothers were less likely to hold inappropriate expectations for their children's development.

In the first-year report, the authors hypothesized that program effects would be greatest for the segment of the sample that was moderately depressed, but the program effects on parenting attitudes were greatest for those with the lowest depression scores. This means that these results must be treated with even more caution than usual, given that this pattern was not hypothesized or consistent with other studies.

At the end of the first year of program implementation, there were no program effects on the self-reported prevalence of severe abuse or neglect that would lead to substantiation of child abuse or neglect (Strauss, Hamby, Bonney-McCoy, & Sugarman, 1996). On a measure of the counts of incidents of severe abusive or neglectful events (what the authors refer to as 'chronicity'), those assigned to the HFNY group had significantly fewer incidents in the first year of the child's life. The effect of the program on this measure was greater for mothers less than 18 years of age at registration, and those mothers

© 2007 The Authors
Journal compilation © 2007 Association for Child and Adolescent Mental Health.

where there was no domestic violence. In analyses of subscales of the overall composite, the authors found that HFNY mothers reported less neglect of their infants. While there were significant reductions on some aspects of these measures of abuse and neglect for selected program sites and ethnic groups (Latinas in particular), these effects were neither hypothesized nor consistent with other patterns of results in the literature.

At year 2, HFNY mothers reported that they engaged in fewer seriously abusive behaviors and as a trend fewer neglectful behaviors during the year prior to the interview. Analyses of the 'prevention' and psychologically vulnerable subgroups found that program impacts were greater for these two groups (Dumont et al., 2006).

The rates of substantiated cases of abuse and neglect in the first year of the child's life were higher in the intervention than control group (9% versus 7%), but not significantly so. During the second year of the child's life, the rates of substantiated maltreatment were identical (6%). The apparently higher rate of official records of child abuse and neglect in the HFNY group during the first year is probably due to the greater detection of abuse and neglect in the homes of visited families (Olds, Henderson, Kitzman, & Cole, 1995), an interpretation that the authors supported with additional analyses of their data.

The authors report that among those women who registered for the program 2 months prior to delivery, the HFNY women delivered significantly fewer low birth-weight newborns than did their control-group counterparts (3.3% versus 8.3%, $p = .012$). The criterion employed for inclusion of participants in this analysis, however (registration 2 months prior to delivery), automatically excluded a large portion of those women who registered at the end of the second trimester and beginning of the third trimester and who delivered prematurely. Given the disappointing history of home visiting and social support programs designed to affect the rates of low birth-weight (see Hodnett & Fredericks, 2004), the authors' having no explanation for how the program may have affected this outcome, and the low rate of low birth-weight reported in the intervention group (lower in fact than the 5.0% targeted rate for the nation as a whole (U.S. Department of Health and Human Services, 2000), this finding also needs to be treated with great caution.

Families in the HFNY group, within a year of registration in the study, were more likely to have secured health insurance and there was a trend for them to report having taken their child to an emergency department for an injury or ingestion more frequently than their control group counterparts. There were subgroup effects for certain outcomes (e.g., in one county women were more likely to have used the Women Infants and Children (WIC) nutritional supplementation program; mothers who had more than one child at intake reported higher rates of breast feeding), but these subgroup effects have little consistency with the literature and were not predicted at the start of the program. Given that randomization was conducted within program sites, however, these site-based findings do deserve careful examination to elucidate why the program would be successful in some locations for some outcomes but not others, an examination the authors have begun.

Overall, women in the control group were more likely to be employed one year after registration in the program. The authors reason that this is not necessarily an adverse program effect, given that returning to work too soon after delivery is sometimes associated with negative effects for children (Mitchell-Herzfeld et al., 2005). There were no overall effects on maternal use of welfare, educational achievement, maternal depression, personal mastery, use of substances, contraceptive use, or rate of subsequent pregnancy or childbirth, although there was a trend for the intervention group to report less alcohol abuse. In one of the counties, mothers in the intervention group had lower rates of depression than those in the control group and in another county the mothers in the intervention group reported having fewer drinks than their counterparts in the control group; mothers less than 18 years of age at registration reported smoking less frequently. Again, these subgroup effects need to be interpreted cautiously. There is no report to date on program effects on maternal life-course at the 24-month follow-up.

In general, the HFNY evaluation shows small effects for some self-reports of parenting outcomes that are associated with child abuse and neglect, moderate effects on self-reported neglect, and moderate effects for some outcomes concentrated in certain subgroups, such as mothers <18 year of age at registration and living in households without domestic violence. Program effects also were greater for the 'prevention' and psychologically vulnerable subgroups. While these effects are consistent with findings in trials of the NFP (Olds, 2002), the HFNY subgroups are small; these findings should be replicated with hypotheses formulated a priori. The program impact on low birth-weight is questionable.

San Diego trial of HFA

Landsverk and colleagues conducted a trial of a version of the HFA model in San Diego (HFSD) that attempted to optimize the HFA program and its implementation.

Program model. This augmented version of the HFA model included the standard HFA home visiting services, but also included parent-group meetings and elevated educational requirements for home visitors. All visitors had at least associates degrees. This team employed outcome measures consistent with those used by Duggan and colleagues in the Oahu trial of HSP (Landsverk et al., 2002). The HFSD

© 2007 The Authors
Journal compilation © 2007 Association for Child and Adolescent Mental Health.

visitors were able to deliver a significantly higher number of home visits than were their counterparts in Hawaii, Alaska, and New York. Over the first three years of program implementation, families received an average of 43 visits. At the end of year 1, over 75% HFSD were still receiving home visits compared to 49% of those in Hawaii; at the end of year 3, 50% of the HFSD families were still receiving services, compared to only 23% of the families enrolled in the program on Oahu. Thus, the HFSD trial was closer to an efficacy trial of this model in that the model was optimized and its implementation extraordinary.

Research design and methods. While Landsverk et al. used a procedure for recruiting and randomizing participants that was similar to that used by Duggan and her team, the San Diego team provided a CONSORT diagram that accounted for participants after randomization, including reasons for loss of participants within treatment assignment. Of those deemed eligible and offered participation ($n = 735$), 515 were randomized. Baseline interviews were conducted on about 95% of those assigned to treatment conditions. The authors provided a detailed analysis of the equivalence of the treatment conditions at baseline and found no significant differences. The investigators were remarkably successful in achieving high rates of participant retention through the end of the trial (83% for each treatment group).

Major findings. At years 2 and 3 (but not at year 1), intervention mothers reported that their children had significantly more well-child visits and at year 2 reported significantly greater compliance with well-child care than did control families. There were no program effects throughout the first 3 years of the target child's life, however, on number of sick visits and immunizations.

There was a trend for mothers in the intervention group to have fewer subsequent pregnancies at the 24-month follow-up evaluation, a difference that was statistically significant for Anglo women but not for women in other racial/ethnic groups. Women in HFSD also reported a significantly greater decline in depressive symptoms over the first two years compared to those in the control group and at the third-year follow-up they reported greater school attendance during the preceding year than did mothers in the control group.

At year 2, a significantly lower percent of HFSD mothers reported use of any psychological aggression in dealing with their children (e.g., demeaning, angry verbalization), but this difference was not sustained at child age 3. At year 3, there were trends for HFSD families to report less corporal punishment and less physical abuse of their children. Finally, the target children in HFSD tested significantly higher on mental functioning than did control children at years 1 and 2, but not at child age 3.

In general, the San Diego trial found a larger number of statistically and clinically significant program effects on mothers and children than did the trial of the program as implemented in Hawaii. These effects nevertheless were relatively small and rarely enduring, which is particularly disappointing given that the HFSD program model studied in this trial was augmented and delivered so well.

Enhanced Healthy Families

In 2002, Daphne Bugental and her colleagues published the results of a small trial of home visiting based upon an augmented version of the HFA model.

Program model. The augmented version of the HFA program was based upon findings that parents who abuse or neglect their children often attribute negative intent to their babies' crying (e.g., 'She's only trying to get back at me.' 'He's mean just like his father.'). In this program, the visitors were trained to focus explicitly on helping parents accurately interpret their infants' communicative signals and devise strategies for managing care-giving problems (Bugental et al., 2002). The home visiting program was modeled after the Hawaii Healthy Start program, but used the Parents as Teachers parenting curriculum to guide their work with parents, with enhancements that included specific strategies for helping parents accurately diagnose problems in parent–infant interaction, develop a specific plan to address the problem, and to pay particular attention to helping parents reframe their negative attributions of their children's behavior and intent.

Research design and methods. The authors recruited 107 participants during pregnancy or soon after the birth of the child and 96 of those agreed to participate in the study. Families were assigned to 1 of 3 treatment conditions in rotating order: a no-treatment control group, HFA as usual, and the augmented HFA program. This report does not include a CONSORT diagram. According to the author (Daphne Bugental, personal communication, July 14, 2006), the randomization was conducted prior to families' accepting participation and completing informed consent. Of the 96 participating families who started the study by completing baseline interviews, 73 (76%) completed outcome assessments at child age 12 months, with roughly equivalent completion rates – 80% of the families in the enhanced program, 77% of the families in the un-enhanced home visit condition, and 70% of the families in the control condition. As many of the study participants were migrant workers, almost all of the drops from the study were due to participants' moving from the study catchment area (Bugental, personal communication, June, 2006).

Major findings. Compared to mothers in the control condition and the un-enhanced home visiting pro-

© 2007 The Authors
Journal compilation © 2007 Association for Child and Adolescent Mental Health.

gram, mothers in the enhanced home visiting program reported less harsh parenting, superior child health, and lower rates of physical abuse in the first year of the child's life (4% in the enhanced home visiting program, 26% in the control condition, and 23% in the un-enhanced home visiting program). The benefits of the program were greatest among families in which the children had low Apgar scores or preterm delivery, a finding which suggests that families with a greater sense of vulnerability are more likely to engage and make use of these kinds of programs.

While this trial is small, it illustrates the kind of formative work that the field (and Healthy Families in particular) needs to move forward. By focusing on a single risk factor (negative attributions) with a clearly delineated set of interventions, the enhanced home visiting program has produced promising effects that should be replicated.

Comment on Healthy Families America

Using the results of the Duggan trial of HFA in Hawaii, the Landsverk trial of HFA in San Diego, a trial of HFA in Hampton Beach, Virginia (which the authors were unable to produce for us), and Daro's report on quasi-experimental evaluations of HFA nationally, the Washington State Institute for Public Policy (Aos et al., 2004) has estimated that an investment in this program produces a loss of $4,500 for each family served. (The Rand report (Karoly et al., 2005) does not monetize HFA costs and savings.) It is unlikely that the results of the HFNY and Alaskan trials produced to date will shift this estimated loss to a gain.

There is some consistency in program effects on children's mental development in the San Diego and Alaskan trials. While the effects in San Diego did not endure, this suggests that this program model, if properly designed, might be able to produce effects on clinically important outcomes. The long-term impact on cognitive outcomes deserves careful examination. However, in light of the substantial site-to-site variation resulting in part from the leadership of HFA actively eschewing a 'monolithic' approach (Daro & Harding, 1999), the results of any single trial (either positive or negative) will be hard to apply to other HFA sites. Given the disappointing results from most trials of HFA home visiting programs on clinically important outcomes and the variation in how HFA is operationalized programmatically, it will be hard for the HFA program model to find a level of scientific evidence that can be confidently applied to all HFA program models.

It is noteworthy that HFA over time has shifted the proportion of home visitors who have bachelors' degrees. In a recent report, 37% of the visitors were found to have bachelors' degrees or higher (Diaz, Oshana, & Harding, 2004). Given that the San Diego trial had modest effects even though all visitors had associates' degrees and the program was implemented well, it is unlikely that upgrading staff qualifications alone will lead to the kind of public health impact promised by this program. Nevertheless, it will be important to see whether doing so improves program influence on participant engagement, retention, and functioning.

In our judgment, the Healthy Families America program will benefit from supporting the kind of research conducted by Bugental and her colleagues (2002), using the results of these kinds of studies to delineate a clearly defined program, testing that program in rigorous RCTs, and then, assuming that the results show sufficient impact, using the clearly defined evidence-based model as the foundation for a national dissemination effort. With an evidence-based model identified, the national replication effort could promote dissemination, requiring adherence to the model tested.

Parents as Teachers

Parents as Teachers (PAT) is a universal parent-education program delivered by home-visitors who begin working with parents either during pregnancy or soon after birth and continue through the child's third birthday. Several quasi-experimental evaluations of PAT, which qualify as pilot studies of the intervention, conducted early in its development suggested that the program increased the school readiness of children whose parents participated in the program (Pfannenstiel & Seltzer, 1985; Winter & McDonald, 1997). These studies typically evaluated the program when it served working-class or middle-class white families who lived in non-urban areas. In 1992, investigators from SRI International began two RCTs of PAT in California (Wagner & Clayton, 1999). Subsequently, a multi-site trial of PAT was conducted by the same principal investigator in 3 communities in the United States (Wagner & Spiker, 2001). Beginning in 1999, the PAT national office began delivering a new curriculum known as Born to Learn (BTL). Drotar, Hurwitz and Kirchner, (2006) subsequently conducted an RCT of PAT in which visitors employed the BTL curriculum.

Program model. PAT grew out of the work of Burton White, who emphasized the importance of the first 3 years of life as a foundation for later learning and success in school and society and the importance of parenting in those earliest years (White, 1985). The first program was started in 1981 in Missouri and by early 2001 had grown to serve families in over 3000 sites, in all 50 states, and 7 foreign countries (Parents as Teachers National Center, 2005). The program consists of 4 elements: 1) home visits; 2) parent group meetings; 3) developmental screenings for the children; and 4) referral to other services. In most PAT programs, home-visits are scheduled once a month through the child's first 3 years of life. The

© 2007 The Authors
Journal compilation © 2007 Association for Child and Adolescent Mental Health.

BTL curriculum evaluated in the 4th trial of PAT is delivered through handouts and videos.

Given its early foundations in the pilot work conducted in Missouri and its reliance on educational activities for parents following White's work, the original educational curriculum was well articulated by the time the trials were conducted. Unfortunately, the original pilot work was not conducted as a small-scale RCT and participant engagement appears not to have been examined. Whether this program resonated to deeply felt vulnerabilities that would lead to participant engagement is not clear. Prior to Drotar's trial of PAT with BTL, a non-randomized trial of the Born to Learn curriculum focused on parents' knowledge, attitudes, and behaviors, and found a high drop-out rate among low-income single-parent families (reported in Drotar et al., 2006). The primary question addressed in the 4th RCT was whether the BTL curriculum would enhance PAT and produce an impact on children's development.

Design and methods of Northern California PAT. The first trial evaluated an augmented version of the Northern California PAT program, which served primarily Latino parents in the Salinas Valley of Monterey County. The Northern California PAT program delivered the standard PAT home-visit curriculum provided by parent educators and provided parents with group meetings to discuss parenting. The augmentation consisted of their providing a 'drop-in' center where parents might go to play with their children.

In this trial, 497 participants were randomly assigned to PAT (199) or a control condition (298). Sixty-eight percent of the participants in PAT and 74% of the controls completed outcome assessments at child age 3. The sequence of randomization is not clear nor is there a CONSORT diagram that accounts for those randomized.

Program participants received an average of 20 home visits during the first 3 years of the child's life. Less than 15% of the families attended any of the group meetings. Forty-three percent of the families dropped out of the program prior to the child's third birthday.

Design and methods of Teen PAT Demonstration. The second trial served adolescent parents in 4 counties in Southern California (the Teen Parents as Teachers Demonstration). In this program the home-visits only extended through the child's second birthday, but the parent educators' efforts were augmented by case managers who focused on helping the teen mothers complete their educations and postpone subsequent pregnancies. In order to sort out the impact of each of these components of the program, separate intervention groups consisted of PAT parent education alone and case management alone. Participating families were enrolled through demonstration sites in Los Angeles, San Bernadino, San Diego, and Santa Barbara. Across these sites, families were randomly assigned to one of 4 groups: 1) PAT alone ($n = 177$); 2) case management alone ($n = 174$); 3) PAT plus case management ($n = 175$); 4) control ($n = 178$). Assessments relied on parent report of parenting and child development, observations of the home environment, tests of children's development, and reviews of children's medical records and child protective service records. Outcome assessments at child age 2 were conducted on 54% of the controls, and 51% of each of the 3 intervention groups. There is no CONSORT diagram and the sequence of randomization in relationship to obtaining consent and completing baseline assessments is not given.

On average, those assigned to the PAT service conditions received 10 visits over the 2-year period. Case management services were modeled after California's Adolescent Family Life Program and were made available as often as the mothers requested them, but face-to-face meetings were supposed to be held at least 4 times a year. Those assigned to receive case management services received an average of 10 case-management contacts in the 2-year intervention period. An average of 57% of the families in the 3 intervention conditions dropped out of the program before the child's second birthday.

Design and methods of multi-site PAT evaluation. The third study consists of a 3-site randomized trial of PAT in 3 communities in the United States (Wagner, Iida, & Spiker, 2001). As with the California trials, this study examined PAT impact with a diverse sample of parents, many of whom were impoverished. In this trial, 665 participants were randomly assigned in the first 8 months of the child's life to PAT ($n = 275$) or a control group ($n = 390$) in which families were sent age-appropriate toys at regular intervals. Aside from the provision of the toys, the control group participants were a 'treatment-as-usual' group. While the groups were essentially equivalent at baseline and at the 1- and 2-year follow-ups, the rates of completed assessments at follow-up were poor: 52% at year 1 and 40% at year 2. This increases the chance that the groups differed in unobserved ways at the follow-up assessment periods.

There are no CONSORT diagrams for any of these trials to account for the loss of participants. Nor is there a discussion of the randomization process or its location in the sequence of recruitment, obtaining informed consent, conducting baseline interviews. We cannot ascertain whether the sample on which baseline assessments were conducted included all cases randomized. The authors indicate that the allocation to treatment and control conditions was conducted by SRI staff, and show tables that indicate the essential equivalence of the treatment groups at baseline on a variety of socio-demographic characteristics.

© 2007 The Authors
Journal compilation © 2007 Association for Child and Adolescent Mental Health.

The authors of each of these trials reported the significance of effects using 1-tailed tests, a non-conservative approach that increases the likelihood that program–control differences are due to chance and that precludes examining possible adverse program effects. In the multi-site evaluation, the effects also were reported in standard deviation units.

Northern California and Teen PAT findings. California trials. After adjusting for differences in background characteristics of the groups that remained in the trials at the follow-up assessments, there were no intervention effects in the 2 California trials on parents' understanding of child development and parenting, or observed or reported quality of the home environment. Those mothers who had been assigned to the combined intervention (PAT plus case management) had no 'open' child abuse or neglect reports over the period covered by the intervention compared to 2.4% in the control group, a difference that was significant using a 1-tailed test. Mothers in the Salinas Valley PAT program and those in the case management and PAT + case management conditions in the Teen Demonstration trial reported that their children had more advanced cognitive development at the end of the program than did their control group counterparts. In the Salinas trial, however, there were no program effects on children's Bayley test scores, which challenges the apparent program impact on children's development based upon parent report. In the Salinas trial, mothers in the PAT conditions reported that their children had better social development and self-help skills than did mothers in the control group.

The authors report extensive analyses of program effects broken down by women's ethnicity and primary language and find several indications that the program failed to help non-Latina families but helped Latinos and especially monolingual Spanish speakers. Data presented suggest that the program actually harmed the parenting practices and child development of the non-Latina mothers and children, but because the authors chose to use 1-tailed tests we are unable to make inferences about harmful effects. Given that these subgroup effects were not hypothesized and that the authors have employed 1-tailed tests, these findings are highly suspect.

Multi-site trial findings. In the multi-site trial, there were few statistically significant effects on parenting or child outcomes. PAT mothers reported being 'very happy' while caring for their child at the 2-year assessment. There were no overall differences in mothers' knowledge of child development or parents' reports of their observations of their children. PAT mothers had higher scores on the HOME language-and-literacy-promoting behaviors scale, using the 1-tailed test, at the year-2 assessment. There were no overall program effects on any of the child outcomes.

There were indications that program effects on some parenting outcomes were greater for very low-income parents, but none of the effects achieved statistical significance. For 2 outcomes (parents' happiness in caring for their children and parents' acceptance of their children's behavior), effects were greater for those parents in the higher-income stratum. There also were some indications that program effects were greater in one particular site.

Design and methods of PAT with the Born to Learn curriculum. The 4th trial of PAT, using the BTL curriculum, randomized 459 families (PAT $n = 227$ and control $n = 232$, Drotar et al., 2006). Randomization was conducted after the investigators obtained the participants' consent. The PAT and control groups were remarkably similar on a range of background characteristics at baseline; sample retention was reasonably high, with final participant attrition at the end of the study similar for PAT (27.3%) and control families (23.7%). The analysis was conducted separately among low SES and middle/high SES families, with a much smaller portion of families falling into the low-SES subgroup (low SES PAT $n = 64$, low SES control $n = 66$). Participant attrition was much higher in the low SES families; final assessments were conducted on 32 children in each of the low-SES groups, or about 50% of those randomized in the low-SES group. One of the strengths of this study is its use of a variety of direct measures of child development as the primary outcomes at 12, 24, and 36 months after study enrollment. Data gatherers were masked to treatment assignment.

Born to Learn. Overall, children in PAT, compared to those in the control group, demonstrated higher mastery motivation on selected measures and their teachers reported that they had higher levels of assertion. There were no overall program effects, however, on children's cognitive development, attachment security, conceptual skills, early reading readiness, expressive language development, and parents' ratings of their social skills. There were no significant program effects on parents' knowledge of child development or sense of competence. Among low-SES families, on the other hand, PAT children exceeded their control-group counterparts in cognitive development and adaptive behavior at 24 months, but not 36 months. Among the small number of low-SES children rated by preschool teachers at 36 months (PAT $n = 11$, control $n = 16$), PAT children were rated as having substantially better social skills than their control-group counterparts. The attenuation of program impact on cognitive development among the low-SES children is not unlike what was observed in the San Diego trial of HFA.

© 2007 The Authors
Journal compilation © 2007 Association for Child and Adolescent Mental Health.

Comment on PAT

The pattern observed here, that is, disappointing findings in RCTs following initially promising findings from quasi-experimental evaluations of interventions, is not unlike what we saw with the evaluations of Hawaii Healthy Start and Healthy Families America. The effort to improve the PAT model through curriculum development and testing those improvements in RCTs is exemplary.

The promising effects found for the low-SES group in the Born-to-Learn trial through 24 months suggests that an enhanced attractive curriculum delivered with pamphlets and videos may be sufficiently engaging at first to produce at least temporary effects on the development of children from low-SES household. These findings are sufficiently promising that a larger trial of PAT with the BTL curriculum focused on low-SES participants seems warranted. Prior to the conduct of such a trial, it would seem prudent to conduct formative work on the intervention to find ways of improving the engagement of low-SES participants. The Washington State Institute for Public Policy, based upon its review of the first two of these trials and the quasi-experimental studies of PAT, has estimated that the investment in the program produces an $800-per-family return on investment. It will be important to update the cost analysis of PAT with the more recent trials reviewed here.

Comprehensive Child Development Program

In 1988, Congress passed the Comprehensive Child Development Act (Public Law 100–297), which authorized the Administration on Children, Youth, and Families (ACYF) in the Department of Health and Human Services to conduct a 5-year initiative designed to improve the life-chances of children born into low-income families.

CCDP program model. CCDP was established to employ case management and parent education services delivered by paraprofessional home visitors as a means of improving parents' economic self-sufficiency, care of their child, and child health and development. This $240 million effort was mandated by the United States Congress and illustrates dramatically the kinds of problems that can be created when funds are made available to test policies before programs have been developed carefully with formative evaluations and tested in pilot studies to determine whether such investments are warranted. The randomization of participants into the trial began in 1990, within 2 years after Congress passed the authorizing legislation.

CCDP's services included case management, delivered by CCDP staff in home visits; parenting education for mothers, also delivered by CCDP staff through home visits; center-based early childhood education accessed primarily through local providers; and developmental screening for children. Given the high cost of center-based child care, CCDP sites typically paid for early childhood education through existing providers. By child age 4–5, about half of the children had been enrolled in center-based early childhood education. CCDP was conducted in 24 sites administered through universities, hospitals, public and private non-profit organizations, and school districts.

CCDP programs were mandated to serve families continuously and comprehensively from the child's first year until the child entered kindergarten or first grade. The legislation passed these broad mandates but there were no clear models upon which ACYF or local grantees might base their programmatic efforts. While ACYF established a contract with a technical assistance contractor to monitor CCDP implementation, local grantees were expected to devise their 5-year comprehensive interventions within a short time after being awarded their grants. Not surprisingly, CCDP programs were characterized by substantial heterogeneity.

CCDP research design and methods. CCDP was established as a 24-site demonstration project, tested in the form of a 21-site randomized controlled trial, with a sample of 4410 low-income families (CCDP $n = 2197$; Control $n = 2213$) registered during the first year of the target child's life (St. Pierre & Layzer, 1999; Goodson, Layzer, St. Pierre, Bernstein, & Lopez, 2000).

The randomization was conducted in a non-uniform way, with some of the randomization conducted by local staff and some conducted by an independent contractor. Most sites stratified the sample by family demographic characteristics, such as race/ethnicity, maternal age, and service site. There is no CONSORT diagram that accounts for participant retention in the trial. Among those who remained in the trial at the 5-year follow-up, the groups were essentially equivalent on background characteristics measured prior to randomization.

Retention of families for research was relatively good for a study of this size and complexity: At the 5-year follow-up assessment, 74% of those assigned to the program and 78% of those in the control group participated in the follow-up evaluation. The outcome evaluation consisted of direct assessments of the children's cognitive and language development and achievement in reading and math, parents' ratings of their children's health and social and emotional development, parents' self-reports of their parenting, observations of parenting behavior and the home environment, and reports of family economic self-sufficiency and parents' educational achievements.

Major findings. There were no treatment–control differences in the children's cognitive or language

© 2007 The Authors
Journal compilation © 2007 Association for Child and Adolescent Mental Health.

development, achievement, behavior problems, or adaptive behavior. There also were no program effects on measures of the home environment or parenting behavior, with one exception: CCDP mothers were less likely to endorse the value of physical punishment ($p = .05$), although the effect was very small clinically. Finally, there were no program effects on measures of parents' employment, family income, use of welfare, participation in job training programs, or educational advancement. Additional analyses found that there were no program effects present for subgroups of interest.

This program of research has been challenged on many grounds, including its difficulties with the randomization and testing the initiative prematurely (Gilliam, Ripple, Zigler, & Leiter, 2000). We agree that CCDP was tested prematurely, but not in the sense that Gilliam et al. meant. We believe that the entire initiative was premature and that it should not have been undertaken until promising models of early intervention had been tested in pre-trial research to provide some assurance that the investment had a reasonable chance of succeeding. Gilliam et al. (2000) have noted that the authors of CCDP and some policy makers have been too quick to interpret these results as indicating that early interventions focused on parents cannot work. We agree.

Early Head Start

Early Head Start (EHS) is the federal initiative that built on some of the lessons of CCDP. Those lessons, however, were incompletely distilled at the time EHS was initiated. Begun in 1995, EHS is a federally funded program for low-income families with young children in the birth-to-three-year age range in the United States (Love et al., 2005).

Program models. In 2004, EHS served 62,000 families through a variety of modalities: center-based infant child care, home visiting, case management, and parenting education. Local programs select a program model: either home-based, center-based or mixed (both home- and center-based) from which to deliver services. The home visits are typically delivered by paraprofessionals. All programs must meet federally mandated performance standards, but there is no single operational or theoretical model to guide the home-visiting or infant day care components of these programs. There is thus a focus on good implementation, but no clarity on what exactly is to be implemented.

Research design and methods. EHS has been evaluated in the form of a 17-site randomized controlled trial that randomized 3001 families (EHS $n = 1513$, control $n = 1488$). Four of the sites consisted of center-based models, 7 were home-based models, and 6 consisted of 'mixed' (i.e., home and center-based) approaches. Programs differed in the degree to which they were fully implemented and were classified as early implementers, late implementers, or incomplete implementers for purposes of secondary analyses. Data were collected on approximately 70% of the participants for the 28-month parent services interview and 55% for the child evaluations conducted at 36 months of age. Data on the rates of completed assessments by treatment assignment can be found in the EHS final report (Administration on Children, Youth, and Families [ACYF], 2002a, 2002b), but are not included in the published article.

The primary analysis reported in the published article was conducted on those who completed a minimum number of home visits (two visits) or attended the center-based program for at least two weeks; an analysis on the entire sample for which data are available, which produced similar results, can be found in an Appendix of the final report (ACYF, 2002b). Analyses of the degree to which the treatment and control groups were equivalent for those who were assessed at follow-up found few significant differences between the groups that might bias the estimates of treatment differences. While both the published article and technical report give the impression that the randomization was conducted before parents agreed to participate, baseline data were gathered and informed consent obtained prior to random assignment by the evaluation contractors (personal communication with John Love, May 19, 2006), minimizing the likelihood of treatment-related post-assignment participant selection.

While the authors chose to present results from analyses in which they controlled statistically for 48 covariates on which there were some differences between treatment and control conditions at baseline, a procedure that has been questioned by some statisticians (Assmann, Pocock, Enos, & Kasten, 2000), they presented the results in their technical report that did not include all of these statistical adjustments. These alternative methods of analysis produced similar results, indicating that the findings were not the result of a single analytic strategy, thus strengthening the findings.

Major findings. Children enrolled in EHS overall, compared to children assigned to the control group, had better cognitive and language development, fewer cases of delayed cognitive and language development, better emotional engagement with their parents and more sustained attention with objects during play observations, and less aggressive behavior reported by their parents. EHS parents, compared to controls, were more emotionally supportive, provided more language and learning stimulation in the home, read to their children more and spanked them less. Overall, these effect sizes were relatively small (in the .10–.20 standard deviation range).

© 2007 The Authors
Journal compilation © 2007 Association for Child and Adolescent Mental Health.

There were no program effects on parents' physical or mental health, parenting stress, or family conflict at child age 3. There were, however, small, but consistent program effects on parents' employment and enrollment in education/job-training in the first 26 months after enrollment in the trial. This did not translate into greater numbers of hours they were working or enrolled in an educational program, high school graduation rates or completion of college degrees, or lower use of welfare (Love et al., 2002). There was, however, a program effect on the percentage of parents with a subsequent birth after study enrollment, a highly important outcome that may affect other aspects of maternal and child health as the target children grow older. Rates of subsequent pregnancies and abortions are not given so it is difficult to understand the mechanisms through which this outcome was achieved.

Parents in the home-based programs were observed to be more supportive of their children and their children were observed to engage them more effectively during observations of parent–child interaction. There were no other statistically significant program effects reported for selected child and family outcomes measured at child age 3 for the home-based programs alone (Love et al., 2005).

Program effects were greatest for the EHS sites that combined home visiting and center-based programs and especially those that implemented the performance standards early. It is possible that the mixed-model approaches were most effective because they contained program components that were able to be adapted to the participants' needs, thus increasing the degree to which parents valued the program and made effective use of its services.

One of the major strengths of this evaluation is that it examines a large-scale program in what may be close to an effectiveness trial, in that programs were conducted in community settings, where the full set of EHS participants were enrolled, and with no involvement of the researchers in program implementation.

Comment on EHS. The Washington State Institute for Public Policy has estimated that for every family served by EHS, there is a $5,665 loss on investment. The Rand report excluded an estimate of program cost impacts for EHS because the most recent follow-up period reported to date is at child age 3 and the Rand report only included studies with follow-ups at child age 5 and beyond. While some have challenged the WSIPP report on the grounds that there is not enough evidence to support long-range estimates of program cost savings, the Institute has responded by noting that their long-range estimates of economic impact are based upon accepted econometric methods of projecting long-term economic outcomes from early effects reported to date. They ask why government should invest in programs when there is insufficient evidence of its functional or economic impacts (Aos et al., 2004).

The Administration on Children, Youth, and Families has made the most of this evaluation by funding a technical assistance network that is established to support sites' efforts to improve EHS program operations. But by moving into wide-scale dissemination before program models were developed and articulated in earlier phases of research, the EHS initiative has missed an important opportunity to refine those interventions that show the great promise, to cull out those that are not working, and to disseminate only those known to work.

New Zealand Early Start program

David Fergusson and his colleagues (Fergusson, Grant, Horwood, & Ridder, 2005, 2006) have conducted a trial in New Zealand of a well-crafted program of postnatal home visiting services delivered by 2 nurses and 4 social workers known as the Early Start program. The impetus for this program was the Hawaii Healthy Start program, but the design was unique to New Zealand.

Program design and methods. This program was designed to be integrated in the New Zealand system of community health nurses (Plunkett nurses) who serve newborns and their parents. Following the Hawaii approach, families were screened to assess their risk for having difficulties in caring for their children using the Kempe Family Stress Index (Kempe, 1976). The visitors in this program each managed caseloads of 10–20 families (Fergusson et al., 2005). There are several distinctive features of the Early Start program. The first is that the program starts with a careful assessment of family needs and strengths. Second, the home visitors develop partnerships with families that focus on addressing families' challenges in a collaborative way. The home visitors provide mentoring, support, and advice to help families mobilize their strengths and resources. The visitors are trained professionals who focus on the following goals: 1) improving child health; 2) reducing child abuse and neglect; 3) improving parenting skills; 4) promoting parental physical and mental health; 5) promoting family economic and material well-being; and 6) promoting stable, positive partner relationships. The visitors are scheduled to begin visiting as soon after birth as is possible and to continue through the child's third birthday. A crucial feature of this program is that it was developed and tested in a pilot study before the more expensive full-scale trial was conducted. The pilot study was employed to determine the extent to which the program procedures and infrastructure were working properly and whether preliminary evidence supported program impact on those aspects of maternal and child outcomes targeted by the program. Armed with promising preliminary results, the investigators proceeded to conduct the field trial of the program. This approach is exemplary.

© 2007 The Authors
Journal compilation © 2007 Association for Child and Adolescent Mental Health.

Research design and methods. The community nurses identified 588 families who qualified for the trial; 443 consented to participate and were randomized; 427 completed the baseline interview. Families were randomized to the intervention and control conditions at the point of their being referred to the trial, prior to their completing the baseline interview, a procedure that increases the likelihood of participant selection factors influencing the composition of the intervention and control groups. Moreover, those assigned to the intervention group were given a 1-month introductory period during which participants could learn about Early Start without having to make a long-term commitment and staff could determine the degree to which they actually needed the service, a process that included administration of the Kempe Family Stress Index (Fergusson et al., 2005). Ninety-nine percent of those assigned to the control condition (221/223) registered in the trial, while 96% of those assigned to the Early Start condition (206/220) registered.

Those in the control condition were paid $50 for every completed assessment (at 6, 12, and 24 months of the child's life), while those assigned to Early Start were not paid. By the end of the trial, the rates of completed assessments were higher in the control group (207/223 – 93%) than in the Early Start group (184/220 – 84%). The authors conducted extensive analyses of background characteristics and found no statistically significant differences in the characteristics of those in the intervention and control groups who remained in the study at child age 36 months. Nevertheless, this study is vulnerable from the standpoint of selective registration, and higher risk families being retained more in the control group, given that only the control group was reimbursed for completing assessments.

Major findings. By the child's third birthday, children in the ES program had paid a larger number of visits to the general practitioner; were more likely to be up to date with their well-child checks; were less likely to have gone to the hospital for injury or ingestions; and were more likely to be enrolled for dental care. ES children had longer enrollment in early childhood education and ES families had larger numbers of community service contacts. Parents in the ES program reported better parenting attitudes and practices. Parents in the ES program also reported fewer behaviors on the Conflict Tactics scale that represent severe physical assault, but there were no ES–control differences in the rates with which the groups had contact with the agency responsible for following up on cases of child maltreatment, possibly because of surveillance bias.

At the end of the program at 36 months, there were no differences in maternal depression, contraception use, pregnancy, and substance use and no differences in exposure to stressful life events, family functioning, and family economic circumstances (Fergusson et al., 2006).

Comment on Early Start. While the differential rates of participant enrollment and attrition raise some concerns about the validity of program effects reported at child age 3, this program shows promise as a means of improving child health and development. The generally favorable results of this trial are likely due to the careful crafting of the intervention itself, the professional implementation of the program by nurses and social workers, and the integration of the program into the child health care system, which may help account for the high level of acceptance of and engagement in the intervention. No estimates are given for the cost of the program and its possible cost-savings.

In the United States, a promising program known as Healthy Steps integrated developmental specialists into pediatric care (Minkovitz et al., 2003). The specialists provided advice to parents and held parent-group meetings. This intervention was evaluated in a 15-site study in which 6 of the sites used randomization and 9 used quasi-experimental procedures. While there were no effects on emergency department visits, hospitalizations, child behavior problems, or home safety practices, parents in the intervention sites were more satisfied with care, and their babies were more likely to receive well-child care and immunizations in the office. In general, programs for pregnant women and infants that are integrated into primary care deserve to be examined in greater depth because they carry with them the legitimacy of the health-care system and the relative absence of stigmatization that, at least in theory, can increase participant engagement and behavior change.

UCLA Family Development Project

Christoph Heinicke and his colleagues have conducted a trial of a relationship-based home-visiting program and parent–infant support group for families bearing first children who were at risk for poor parenting; the visiting began during the third trimester of pregnancy and extended through the child's second year of life (Heinicke et al., 1999; Heinicke, Fineman, Ponce, and Guthrie, 2001).

Program model. This program grew out of Professor Heinicke's research laboratory. It includes well-crafted family assessment procedures and clinical interventions, and was carefully developed prior to its being subjected to a randomized controlled trial. Families enrolled in the Family Development Program were selected on the basis of their exhibiting characteristics that would increase their response to the intervention, such as their being poor and lacking social support. Home visits were conducted by mental health professionals with backgrounds in

© 2007 The Authors
Journal compilation © 2007 Association for Child and Adolescent Mental Health.

child development and family systems who focused on enhancing the mothers' and fathers' relationship-building skills between members of the spousal dyad, parents and child, and parents and other family members. The overriding focus of the intervention was on developing positive, trusting relationships. In 42% of the cases, the fathers were often involved in the home visits, a remarkable accomplishment for an intervention with high-risk families. A weekly mother–infant group was made available from infant age 3–15 months. The visitors completed an average of 3 visits per month in the first year and 1.5 visits in the second year of the child's life.

Research design and methods. To evaluate this thoroughly developed intervention, 70 families were randomly assigned to one of two groups: 1) this intensive home visiting and parent-group intervention ($n = 35$); or 2) a pediatric follow-up group that involved developmental evaluations and referral for other services as needed ($n = 35$). The randomization process is described completely and although the baseline interviews were conducted after the randomization was completed there was virtually no immediate loss of study participants. Complete data were available at both the 12- and 24-month data gathering time points on 64 (91%) of those randomized. All data were gathered by staff masked to treatment assignment. The sample is accounted for in CONSORT-like tables.

Major findings. At the 12th month of the child's life, there were no significant treatment–control differences in the background characteristics of those who remained in the study ($n = 31$ in home visiting; $n = 33$ in the control group). As predicted, those in the intervention group experienced greater partner and family support at the 12-month assessment than did their control-group counterparts. Mothers in the intervention group were more responsive to their infants, who simultaneously were rated as exhibiting greater attachment security during testing procedures. Home-visited infants at 12 months also were rated as having more than three times as many secure as opposed to insecure attachment classifications in the Ainsworth Strange Situation (Ainsworth, Blehar, Waters, & Wall, 1978). Synchrony of mother–infant play was greater, as was maternal encouragement of child task involvement and child endurance in the test situation. Over time, home-visited mothers were observed to encourage their infant's autonomy and to be less intrusive and their infants exhibited a greater sense of separate self and less non-compliance than their pediatric follow-up counterparts. There were no program effects on the infants' developmental indices at 12 months of age.

At 24 months of age, those in the intervention group, compared to those in the pediatric follow-up condition, used more persuasive as opposed to coercive and intrusive methods of controlling their children's behavior. Children in the intervention group responded more positively to their parents' control efforts.

It is important to note that Heinicke and his team (2006) have been studying factors that predict engagement in the intervention and have found that parents' attachment representations derived from the Adult Attachment Interview (Main, Goldwyn, & Hesse, 2003) administered at the end of pregnancy predict variations in involvement in the intervention, and together with that involvement, predict family social-emotional outcomes at 2 years (Heinicke et al., 2006).

Comment on Family Development Project. This program was refined in a mental health laboratory over many years and illustrates the potential of this kind of service, if time is invested in carefully developing the intervention. It will be important to find out whether the salient intervention elements can be reproduced and if the effects observed in this efficacy trial can be replicated in larger effectiveness trials. Replication is needed to gain some understanding of the transportability of the intervention, its range of clinical impacts, and cost.

Nurse home visiting

There is a long and rich history of home visiting by nurses for over a hundred and fifty years in Europe and the United States (Wald, 1915; Buhler-Wilkerson, 1993). One of the earliest randomized trials to show the promise of home visiting by nurses was published by Gutelius and colleagues in the early 1970s (Gutelius et al., 1972) in which she found that a program of prenatal and infancy home visiting for teen mothers bearing first children augmented by mobile pediatric care produced significant impacts on children's injuries and behavior, and maternal life-course (Gutelius et al., 1972; Gutelius & Kirsch, 1975). In the past decade, a number of articles on nurse home visiting have been reported that reveal the potential of this method of service delivery when the right program methods and content are incorporated in the clinical model. Part of the salience of programs that employ nurses for this work is likely due to their being acknowledged as having legitimate agendas and skills to address the concerns of pregnant women and parents of young children, a fact that increases their acceptance and persuasive power with vulnerable families at this phase in the life cycle (Olds, 2002). More generally, nurses are consistently rated by the US public as professionals with the highest levels of honesty and ethics (Gallup Organization, 2005), followed by other medical professions. Whether nurses are viewed as favorably in other societies is not clear. Earlier reviews of this literature indicted that simply employing nurses as home visitors, however, was insufficient (Olds & Kitzman, 1990; 1993).

© 2007 The Authors
Journal compilation © 2007 Association for Child and Adolescent Mental Health.

Nurse–Family Partnership

This program builds upon the history of public health nursing in the United States and Europe with a program of nurse home visiting that has been tested in three separate RCTs since 1977: first in Elmira, New York, with a sample of 400 low-income primarily white families; then in Memphis, Tennessee, with a sample of 1138 low-income primarily African American families; and most recently in Denver with a sample that includes a large portion of Hispanics. The Denver trial ($n = 735$) systematically examined the impact of the program when delivered by paraprofessionals (individuals who shared many of the social characteristics of the families they served) and by nurses. The paraprofessionals in the Denver trial all had high school educations, but no college education in the helping professions. The paraprofessionals had extensive training and supervision to support their work with families. All were parents themselves and most had spent some part of their lives receiving welfare benefits. Olds and colleagues (1986) conducted a pilot study to refine the NFP program plan and research design prior to beginning the original Elmira study. The pilot study (a small randomized trial) preceded corresponding phases of the main study and was used to refine the program and research design before the full-scale trial was conducted. This approach was employed in each of the RCTs of the NFP.

Program model. This program model consists of a clear delineation of the target population, program content, methods of engaging and bringing about adaptive behavior change, and the importance of employing nurses in serving families during pregnancy and the early years of the child's life.

Developmental and epidemiologic research guided decisions about the families to be served and the content of the program. All of the studies examined program impact with women who had no previous live births, and each focused recruitment on women who were either low income, unmarried, or adolescents, as the problems the program was designed to address (e.g., poor birth outcomes, child abuse and neglect, and diminished economic self-sufficiency of parents) are concentrated in those populations (Elster & McAnarney, 1980; Overpeck, Brenner, Trumble, Trifiletti, & Berendes, 1998; Furstenberg, Brooks-Gunn, & Morgan, 1987). In the Elmira trial, any woman bearing a first child was allowed to register, although those who were poor, unmarried and teens were actively recruited. Given that many of the program effects in Elmira, as we shall see, were greater for families in which the mothers were poor and unmarried at registration, the subsequent trials in Memphis and Denver focused recruitment more exclusively on those with overlapping risks (i.e., being both unmarried and from a low-income family). This team focused on women who had no previous live births because they hypothesized that such women would be more receptive to home-visitation services concerning pregnancy and child rearing than would women who had already given birth (Olds, 2002). Moreover, as parents learn parenting and other skills through the program, they should be better able to care for themselves and subsequent children, thereby increasing the public health impact of the intervention.

The nurses in this program are charged with improving the outcomes of pregnancy by helping women improve their prenatal health-related behaviors (e.g., reducing use of cigarettes, alcohol, illegal drugs, identifying and obtaining treatment for emerging obstetric complications); with improving children's postnatal health by helping parents provide more responsible and competent care of the child early in life; and with improving parents' economic self-sufficiency by helping parents develop a vision for their futures, plan subsequent pregnancies, complete their educations, and find work. In attempting to accomplish these goals, the nurses systematically involve fathers, grandmothers, and other concerned family members or friends in the program and link families with needed health and human services.

Research designs and methods. In each of the trials, women were randomized either to home visiting during pregnancy and the first two years of their children's lives or comparison services that consisted of some combination of free transportation for prenatal and well-child care and sensory and developmental screening and referral of the child for potential problems. All three studies employed a variety of data sources, including maternal interviews, developmental tests of the child, observations of parent–child interaction and the home environment, and medical and administrative data.

Each of the trials included CONSORT-like tables to account for study participants following randomization. Prospective participants agreed to participate in the research with the understanding that they might be assigned to receive home visiting or comparison services, and completed informed consent and the baseline interviews before random assignment. Randomization was conducted after stratification by maternal race, marital status and geographic region of residence. Except where indicated, all analyses were conducted on an intention-to-treat basis. The investigators in the Elmira trial registered 400 women and assigned them to 4 conditions: 1) sensory and developmental screening and referral ($n = 90$); 2) screening and referral + free transportation for regular prenatal and well-child care ($n = 94$); 3) screening, referral, transportation + prenatal home visiting ($n = 100$); 4) screening, referral, transportation, prenatal and infancy home visiting ($n = 116$). The investigators in the Memphis trial registered 1139 for the pregnancy phase and 743 for the

© 2007 The Authors
Journal compilation © 2007 Association for Child and Adolescent Mental Health.

infancy and toddler follow-up phase ; they randomly assigned participants to 2 conditions for the prenatal phase (intervention $n = 458$ and control $n = 681$) and to 2 conditions for the postnatal phase (intervention $n = 228$; control $n = 515$). Finally, investigators in the Denver trial registered 735 participants and randomly assigned them to 3 treatment conditions: 1) a control condition (screening and referral $n = 255$); 2) a condition in which parents were visited by paraprofessional visitors ($n = 245$); or 3) a nurse-visited condition ($n = 235$).

High rates of sample retention have been achieved at each of the follow-up periods (e.g., 81% of those mothers randomized for the 15-year follow-up of the Elmira trial; 86% of those randomized in the 6-year follow-up of the Memphis trial; and 86% of those mothers randomized for the 4-year follow-up of the Denver trial). The groups remained essentially equivalent in each of the follow-ups.

Given that program effects on child maltreatment and injuries in the Elmira trial were concentrated in children born to mothers with limited sense of control over their lives (see Olds, 2002), this team hypothesized that program effects on care-giving and child outcomes in the Denver and Memphis trials would be greater in families characterized by the mothers having limited psychological resources (limited sense of control, intellectual functioning, and more symptoms of depression and anxiety).

Major findings from Elmira trial

Early program effects. During pregnancy and the first four years after delivery, the program was found to improve women's prenatal health-related behaviors, pregnancy outcomes, the rates of state-verified reports of child abuse and neglect in the first 2 years post partum ($p = .07$), the quality and safety of the home environment, injuries detected in the medical record, rates of subsequent pregnancies, intervals between births, and maternal employment. Many of these effects were greater for families in which the mother was at greater risk by virtue of her being from a low-income household, unmarried, and teenaged at registration. Moreover, program effects on child maltreatment and injuries were concentrated in those children born to mothers who at registration had limited sense of mastery. The earliest analyses (through child age 4) focused on the women who were white (89% of the sample) because the sample was too small to allow complete cross-classification of race with other factors included in the analysis (maternal age, marital status, and socioeconomic status of family), an acceptable procedure given that the randomization was stratified by maternal race. The analysis of pregnancy outcomes (birth-weight and length of gestation; Olds et al., 1986), on the other hand, can be challenged on the grounds that women with certain high-risk obstetrical conditions were removed from the analysis and these conditions had not been used as exclusion criteria at the stage of sample recruitment. The authors reported that the effects of the program on pregnancy outcomes were the same irrespective of these exclusions, but the published report is based on the sample with obstetrical complications removed.

Adolescent follow-up. Results from a 15-year follow-up of the Elmira sample (Olds et al., 1997) indicate that the nurse-visited comparison differences in rates of state-verified reports of child abuse and neglect grew between the children's fourth and fifteenth birthdays. Overall, during the 15-year period after delivery of their first child, in contrast to women in the comparison group, those visited by nurses during pregnancy and infancy were 48% less likely to be identified as perpetrators of child abuse and neglect, an effect that was greater for families in which the mother was unmarried and from a low-SES household at registration (Olds et al., 1997).[1]

While this reduction in child abuse and neglect was promising, the program did not eliminate maltreatment. The investigators therefore analyzed why the program was not successful in preventing child abuse and neglect for certain families, and hypothesized that the presence of domestic violence in the home would attenuate the preventive effects of the program. The program had no impact on the incidence of domestic violence, but domestic violence did moderate the impact of the program on child abuse and neglect. The program effect on child abuse and neglect was reduced in those households in which domestic violence was higher during the 15-year period following the birth of the first child. The moderating influence of domestic violence was specific to child abuse and neglect and did not attenuate program effects on any other outcome (Eckenrode et al., 2000). As a result of this analysis, this team intensified the nurses' efforts to help women cope with domestic violence and to promote effective partner communication.

At the 15-year follow-up, nurse-visited women overall, compared to their control-group counterparts, had significantly fewer arrests, convictions, and numbers of days incarcerated, effects that were greater for women who were at greater risk by virtue of their being unmarried and from low-SES families at registration.[1] No differences were found for the full sample on fertility, employment, or receipt of welfare, but among women who were unmarried and from poor families at registration, the program produced statistically and clinically significant reductions in

[1] Olds and his team have recently completed an updated analysis of the Elmira 15-year follow-up data. The findings differ some from the results published originally; apparent impacts on the number of times children ran away from home and on alcohol use are not significant under the reanalysis. The findings reported here, which are being submitted for publication, reflect these updated analyses.

© 2007 The Authors
Journal compilation © 2007 Association for Child and Adolescent Mental Health.

the rates of subsequent pregnancies and births, increases in inter-birth intervals and employment, and reductions in receipt of welfare (Olds et al., 1997).

The follow-up study also assessed children of the participants, when the children turned 15 years of age. In contrast to 15-year-old adolescents born to women in the comparison group, those visited by nurses reported 59% fewer arrests and were found to have 90% fewer adjudications as persons in need of supervision (PINS) for incorrigible behavior. There was a trend for nurse-visited youth to have fewer convictions and violations of probation and fewer sexual partners by their 15th birthday. The effects on crime and antisocial behavior were greater for children born to mothers at greater risk by virtue of their being poor and unmarried at registration. There were no program effects on other behavioral problems, such as teachers' reports of adolescents' acting out in school; suspensions; initiation of sexual intercourse; and parents' or children's reports of major or minor acts of delinquency (Olds et al., 1998).[1]

Major findings from Memphis trial

The Memphis trial was closer to an effectiveness trial in that the program was administered through the Memphis/Shelby County Health Department and the study investigators were less involved with the implementation of the program given their distance from the study site (Kitzman et al., 1997). Given that program effects on child maltreatment and health-care encounters for injuries were concentrated on children born to mothers with limited sense of mastery, the investigators in this trial hypothesized that the impact of the program on care-giving and child health and development would be concentrated in mothers who had limited psychological resources – limited sense of mastery, intellectual functioning, and higher levels mental health symptoms (primarily depression and anxiety).

There were no program effects on women's use of standard prenatal care or obstetrical emergency services after registration in the study. By the 36th week of pregnancy, nurse-visited women were more likely to use other community services than were women in the control group. There were no program effects on women's cigarette smoking, probably because the rate of cigarette use was only 7% in this sample, compared to 55% of the sample in Elmira.

In contrast to women in the comparison group, nurse-visited women had fewer yeast infections after randomization and fewer instances of pregnancy-induced hypertension. Among women with pregnancy-induced hypertension, those who received a nurse home visitor had lower mean arterial blood pressures during labor than those in the comparison group, an indication of less severe cases (Kitzman et al., 1997). Despite these differences in PIH, there were no program effects on average birth weight, percent low birth weight, length of gestation, spontaneous preterm delivery, indicated preterm delivery, or Apgar scores.

Nurse-visited mothers reported that they attempted breast-feeding more frequently than did women in the comparison group, although there were no differences in duration of breast-feeding. By the 24th month of the child's life, in contrast to their comparison-group counterparts, nurse-visited women held fewer beliefs about child-rearing associated with child abuse and neglect. Moreover, the homes of nurse-visited women were rated on the HOME scale (Bradley & Caldwell, 1979) as more conducive to children's development (Kitzman et al., 1997). While there was no program effect on observed maternal teaching behavior, children born to nurse-visited mothers with low levels of psychological resources were observed to be more communicative and responsive toward their mothers than were their comparison-group counterparts.

The rate of substantiated child abuse and neglect in the population of two-year old, low-income children in Memphis was too low (3–4%) to serve as a valid indicator of child maltreatment in this study. The authors therefore hypothesized that they would see a pattern of program effects on childhood injuries that would be similar to the pattern observed in Elmira, reflecting a reduction in dysfunctional care of children. During their first two years, nurse-visited children overall had 23% fewer health-care encounters in which injuries and ingestions were detected than did children in the comparison group. Nurse-visited children also were hospitalized for fewer days with injuries and/or ingestions than were children in the comparison group.

The effect of the program on both total health-care encounters and number of days children were hospitalized with injuries and ingestions was greater for children born to women with few psychological resources, an effect similar to the one observed in Elmira for child abuse and neglect and emergency department encounters. An examination of the children's hospital records indicated that nurse-visited children were hospitalized for fewer days than were children in the comparison group and that when they were hospitalized they tended to be older and to have less severe conditions (Kitzman et al., 1997).

At the 24th month of the first child's life, nurse-visited women reported fewer second pregnancies and fewer subsequent live births than did women in the comparison group. Nurse-visited women and their first-born children relied upon welfare for slightly fewer months during the 2nd year of the child's life than did comparison-group women and their children, although there were no differences during the children's first year of life. There were no program effects on mothers' reported educational achievement or length of employment for either the whole sample or for those with few psychological resources (Kitzman et al., 1997).

© 2007 The Authors
Journal compilation © 2007 Association for Child and Adolescent Mental Health.

By child age 6, in contrast to counterparts assigned to the comparison group, women visited by nurses had fewer subsequent pregnancies and births; longer intervals between births of first and second children; longer relationships with current partners; and fewer months of using welfare and food stamps (Olds, Kitzman et al., 2004). Nurse-visited mothers were more likely to be cohabiting with the biological father of the child two and a half years after the program had ended, and there were trends for them to be married more frequently and to have fewer therapeutic abortions (Kitzman et al., 2000). There were no statistically significant program effects on maternal educational achievement or employment after the program ended.

By child age 6, children visited by nurses had higher intellectual functioning and receptive vocabulary scores and fewer behavior problems in the borderline or clinical range. Nurse-visited children born to mothers with low psychological resources had higher arithmetic achievement test scores and expressed less aggression and incoherence in response to story stems (Olds, Kitzman et al., 2004).

Denver Major findings from trial

In the Denver trial, the investigators were unable to use the women's or children's medical records during pregnancy and the first two years of the child's life to assess outcomes because the health-care delivery system was too complex to enable them to abstract reliably all of the health-care encounters as they had done in Elmira and Memphis. This limited the number of health outcomes the investigators could examine in this trial. Moreover, as in Memphis, the rate of state-verified reports of child abuse and neglect was too low in this population (again, 3–4% for low-income children birth to two years of age) to allow them to use Child Protective Service records to assess the impact of the program on child maltreatment. Consequently, the investigators introduced new measures to detect program impacts on parents' care-giving, reflected in objective observations of infants' early emotional expressions in laboratory settings (Olds et al., 2002).

Denver results for paraprofessionals. There were no paraprofessional effects on women's prenatal health behavior, maternal life-course, or child development, although at 24 months, paraprofessional-visited mother–child pairs in which the mother had low psychological resources interacted more responsively than did control-group counterparts. Moreover, while paraprofessional-visited women did not have statistically significant reductions in the rates of subsequent pregnancy, the reductions observed were clinically significant (Olds et al., 2002).

Two years after the program ended, women visited by paraprofessionals, compared to controls, were less likely to be married and to live with the biological father of the child, but worked more and reported greater sense of mastery and better mental health. Mothers and children visited by paraprofessionals, compared to controls, displayed greater sensitivity and responsiveness toward one another and, in those cases in which the mothers had low psychological resources at registration, had home environments that were more supportive of children's early learning (Olds, Robinson et al., 2004).

Denver results for nurses. The nurses produced effects consistent with those achieved in earlier trials of the program (Olds et al., 2002; Olds, Robinson et al., 2004). In contrast to their control-group counterparts, nurse-visited smokers had greater reductions in urine cotinine (the major nicotine metabolite) from intake to the end of pregnancy. Use of other substances (e.g., cocaine, marijuana) was too low (by urinalysis) to serve as valid outcomes.

During the first 24 months of the child's life, nurse visited mother–infant dyads interacted more responsively than did control pairs, an effect concentrated in the low-resource group. At the four-year follow-up, nurse-visited mothers with low psychological resources at registration, compared to control-group counterparts, provided home environments that were more supportive of children's learning.

At 6 months of age, when tested in a laboratory situation designed to elicit emotions, nurse-visited infants, in contrast to control-group counterparts, were less likely to exhibit emotional vulnerability in response to fear stimuli and those born to women with low psychological resources were less likely to display low emotional vitality in response to joy and anger stimuli. At 21 months, nurse-visited children were less likely to exhibit language delays than were children in the control group, an effect concentrated among children born to mothers with low psychological resources. Nurse-visited children born to women with low psychological resources also had superior language and mental development in contrast to control-group counterparts.

By 24 months after delivery, nurse-visited women, compared to controls, were less likely to have had a subsequent pregnancy and birth and had longer intervals until the next conception. Women visited by nurses were employed longer during the second year following the birth of their first child than were controls.

At the follow-up conducted at child age four, nurse-visited women reported greater intervals between the birth of their first and second children, and less domestic violence. The impact of the program on domestic violence is important, but needs to be replicated before great confidence can be placed on this finding. Nurse-visited 4-year-old children whose mothers had low psychological resources at registration, compared to control-group counter-

© 2007 The Authors
Journal compilation © 2007 Association for Child and Adolescent Mental Health.

parts, had more advanced language, superior executive functioning and better behavioral adaptation during testing.

Estimates of nurse vs. paraprofessional effects. While the program was in operation, for most outcomes on which there was an effect for either program, paraprofessionals produced effects that were approximately half the size of those produced by nurses. Paraprofessional-visited mothers began to experience benefits from the program 2 years after the program ended at child age 2, but their children were not statistically distinguishable from their control-group counterparts on most outcomes.

Nurse-visited families were less likely to drop out of the program and they were more likely to show up for scheduled appointments (Korfmacher, O'Brien, Hiatt, & Olds, 1999), but these differences in quantity of program delivered did not account for fact that the nurses produced a large number of maternal and child effects while the paraprofessionals did not. The investigators concluded that the differences in outcomes observed were a reflection of the mothers' finding greater value in being visited by nurses and thus following the nurses' suggestions and coaching more completely than did mothers visited by paraprofessionals. Visitor background thus likely affects parents' engagement in such programs and likely affects the visitors' ability to motivate parents' adaptive behavioral change.

Policy implications

Using data from the three trials of the NFP, the Washington State Institute for Public Policy estimated that for every family served by nurses, society experiences a $17,000 return on the investment (Aos et al., 2004).

National replication of NFP. In 1996, the Department of Justice invited representatives of the Nurse Family Partnership to replicate the program outside of research contexts. With the results of the Memphis trial showing essential replication of many of the major findings from the Elmira trial, this team accepted the invitation, but built into the replication process a structure for helping to ensure that the program would be conducted with fidelity to the model tested in the trials. These replication efforts focused on ensuring that the organization and community are good contexts for program replication, that the nurses experience thorough education and technical assistance in the details of the program model, and that nurses have detailed visit-by-visit program guidelines to structure their work with families (Olds, Hill, O'Brien, Racine, & Moritz, 2003). A significant part of the national replication process consists of a requirement that every NFP program site use the national office's web-based clinical information system (CIS) to monitor ongoing program implementation and performance. Even with these efforts, this team believes that there inevitably will be slippage in implementation when community replication is compared to implementation in the earlier trials.

Comment on NFP. Researchers involved with the NFP have learned that there are certain features of the program that require attention, including participant attrition rates that are higher than found in the original trials, nurses' challenges in implementing some aspects of the parenting curriculum, and addressing maternal mental illness and intimate partner violence in the home. As a result of the higher rates of participant attrition, nurses in the replication sites are completing about 75% of the number of home visits completed by nurses in the trials. Understanding these program vulnerabilities has led to both continuous quality improvement activities with sites, but also the development of promising augmentations of the basic program model that are now being tested in site-based RCTs.

An intervention to increase participant retention and the number of completed home visits is currently being tested in the form of a 30-site cluster randomized trial. A separate trial of an intervention to increase nurses' skills in managing intimate partner violence is under review. Other trials that are in various stages of development and submission for funding will focus on improving nurses' management of parents' mental illness, their skills in promoting competent care-giving, and involving fathers in their children's and partners' lives. Given that all of the NFP national program sites operate under a common program model and employ a common web-based information system that monitors program implementation and outcomes, the local sites are linked in a virtual research network that is being used to improve both implementation and the program model itself.

Trial of nurse home visiting for adolescent mothers

Deborah Koniak-Griffin and colleagues (Koniak-Griffin, Anderson, Verzemnieks, & Brecht, 2000; Koniak-Griffin et al., 2002; Koniak-Griffin, Mathenge, Anderson, & Verzemnieks, 1999; Koniak-Griffin et al., 2003; Lesser & Koniak-Griffin, 2000) designed and evaluated a program of nurse home visiting known as the Early Intervention Program (EIP).

Program model. The nurses served adolescent mothers (62% Latina, 13% African American) bearing their first children and focused on improving pregnancy outcomes, child health and development by improving parents' care of their children, and maternal life-course, including planning future pregnancies. Nurses began visiting in mid pregnancy (20 weeks) and continued through the infants' first

© 2007 The Authors
Journal compilation © 2007 Association for Child and Adolescent Mental Health.

12 months of life. This program is thus similar in conceptual background and implementation to the Nurse Family Partnership (NFP), but with several distinct modifications: the nurses had other caseload responsibilities in addition to the EIP; the program used a group education model in pregnancy; the program was shorter in duration (12 vs. 24 months postpartum); and the program had a less intensive schedule of home visiting.

Research design and methods. The EIP was carefully developed and monitored for intervention fidelity and was tested in a randomized trial where intervention families were compared to a group of comparable adolescent mothers receiving traditional public health nursing (TPHN) care consisting of three home visits (intake at mid pregnancy, one prenatal visit, and one postnatal visit). A sample of 144 adolescent first-time mothers were recruited and randomly assigned to groups, and by the 24-month evaluation point there were 56 mothers remaining in the intervention group and 45 mothers in the control group. There is no CONSORT diagram for this trial. Among those on whom subsequent assessments were conducted, the treatment and control groups were essentially equivalent on baseline variables.

Major findings. At six weeks postpartum ($n = 126$), compared to infants assigned to the TPHN group, EIP infants had fewer days in the hospital and fewer total episodes of hospitalizations involving injuries than did the TPHN infants (Koniak-Griffin et al., 2000; Koniak-Griffin et al., 1999). This program effect on childhood injuries continued through the 12- and 24-month follow-ups (Koniak-Griffin et al., 2002; Koniak-Griffin et al., 2003). At 12 months the EIP infants demonstrated significantly higher rates of immunizations, a finding that did not endure at 24 months. At 12 months there were no significant program effects on mother–child interaction, subsequent pregnancies, maternal depression, or substance use. At the 24-month follow-up, EIP mothers used marijuana less than those assigned to the TPHN (Koniak-Griffin et al., 2003). While program effects on rates of subsequent pregnancy did not achieve statistical significance, the rates found in the intervention versus control groups were remarkably similar to those found in the trials of the NFP. The absence of effect is thus probably due to low statistical power.

Comment on EIP. Given the shorter duration (and corresponding reduced cost) of this program in comparison to the NFP, it would be useful to see whether it produces effects on other clinically important aspects of maternal and child functioning, such as mental health, maternal employment, father involvement, and qualities of parent–child interaction, and whether the cost of the program can be recovered with lower government expenditures and societal costs (see Karoly et al., 2005).[1]

Baltimore trial of nurse home visiting for drug exposed infants

Butz et al. (2001) conducted a trial of a program of home visiting by nurses for parents whose infants had been exposed to drugs during the mothers' pregnancy.

Program model. The program consisted of 16 home visits from birth through the child's 18th month of life, with more frequent visits provided during the first 6 months. Visits were conducted by 2 pediatric nurse specialists who had considerable experience in conducting home visits with inner-city populations. They were trained and supervised in pediatric assessments for infants exposed to cocaine or opiates in-utero. The nurses established a caring relationship with mothers, educated them about their infants' needs using the Hawaii Early Learning Profile and the Carolina preschool curriculum, promoted mother–infant interaction, and monitored their children's health and development. Mothers received an average of 12.8 visits (SD, 3.2, range 1–20).

Research design and methods. Mothers between the ages of 19 and 40 were recruited for the study if they self-reported or had positive urine screens for cocaine or opiates. Preterm babies or those hospitalized for more than 24 hours were excluded, as were infants discharged into non-kinship care. Mothers with major psychiatric disorders also were excluded. The authors provide a diagram to account for participants in the trial: 248 mother–infant dyads were deemed eligible; 15 were discharged before consent was obtained; 233 were invited to participate; and 204 consented, were enrolled, and randomized. The diagram does not, however, show the allocation to treatment conditions and the corresponding rates and reasons for study attrition by treatment assignment. While the number of cases on whom follow-up assessments were conducted is quite similar by treatment condition, we have no way of discerning the extent to which reasons for loss to follow-up may be related to treatment assignment.

Between the time of enrollment and the completion of the follow-up assessments, 87 participants were lost to follow-up, primarily because mothers did not attend the follow-up assessment session ($n = 66$). One hundred parent–child dyads provided complete outcome data when the children were between 2 and 3 years of age ($n = 49$ for the nurse-visited group and $n = 51$ for the controls). This means that 49% of those originally randomized completed follow-up assessments; while this rate is much lower than desirable, some allowances need to be made for the particularly high-risk status of the sample. Analyses were conducted on study participants irrespective of the number of home visits completed.

© 2007 The Authors
Journal compilation © 2007 Association for Child and Adolescent Mental Health.

Major findings. Nurse-visited children were reported by their parents to have fewer total behavioral/emotional problems on the Child Behavior Checklist than were children in the control group, reflected primarily in fewer problems with anxiety, depression, withdrawal, and aggression. There was a trend ($p = .06$) for nurse-visited mothers to report lower levels of parenting stress than did mothers in the control group.

Comment on nurse visiting for drug-exposed infants. This study would have been strengthened if the investigators had included objective measures of maternal and child functioning rather than parental report as their primary source of information on outcomes. The findings are sufficiently promising, however, and the problems associated with children born to substance-using mothers so great, that this trial deserves to be replicated.

Conclusions

We now turn our attention to synthesizing the evidence and drawing conclusions regarding factors that are likely to increase our success in promoting parents' capacities to care for their children. In addition, we examine aspects of research design that can be improved to increase the validity of intervention trials conducted in the future.

Summary of trials

Use of nurses. Of all of the parenting interventions studied, those that send nurses into the homes of high-risk families, focusing on the improvement of prenatal health, the child's health and development, and parents' own economic self-sufficiency, have the strongest evidentiary foundation. Simply using nurses as home visitors was insufficient to affect important maternal and child outcomes (Olds & Kitzman, 1990; 1993). This conclusion is driven home by Koniak-Griffin's trial of the EIP where the control group received a minimal dose of traditional public health nursing and the EIP produced clinically important outcomes on maternal substance use and childhood injuries. One of the reasons nurses work so well at this phase in the life cycle (beginning during pregnancy or the perinatal period) is that families find nurses valuable. This conclusion is driven home in the Denver trial of the NFP, in which families visited by paraprofessionals dropped out of the program more frequently if they were visited by paraprofessionals and they opened their doors less to paraprofessionals (Korfmacher et al., 1999). Given that the differences in program impact between nurses and paraprofessionals were not explained simply by the amount of program received, we believe that nurses carry with them greater persuasive power by virtue of their well-established roles as caring and competent service providers for pregnant women and parents of young children. Simply having nurses deliver the service, however, is insufficient. They need guidance as to what to focus on, how to bring about adaptive behavior change that is based upon solid epidemiology and theory, and sufficient contact with families for change to occur.

This is not to say that other service providers cannot do a reasonable job of serving parents during pregnancy and infancy. If programs are carefully crafted and piloted following the procedures described by the IOM, they will have greater likelihood of success. This observation is supported by the results of trials of thoroughly developed programs such as the NIDCAP intervention for very low birth-weight newborns (Als et al., 2003), the Infant Health and Development Program (McCormick et al., 2006), Heinicke's Family Development Program (Heinicke et al., 2001), and the New Zealand Early Start program (Fergusson et al., 2005). These programs show promise, but still need to be examined thoroughly from the standpoint of their replicability, costs and cost-savings to determine their practical value. It is our judgment, nevertheless, that those interventions that employ non-nurses at this phase in the life cycle will have greater difficulty in achieving success because families will not have the same level of trust that non-nurses will competently address issues of concern to them as will nurses.

Grounding in epidemiology and theory. The nurse home visiting programs with the strongest evidentiary foundations are those that focus on at-risk pregnant women having no previous live births and parents in at least the first year after delivery, and do so with program guidelines that have been crafted to reliably alter specific risk and protective factors predictive of prenatal health, infant care-giving, and maternal life-course. The Nurse–Family Partnership and Koniak-Griffin's EIP nurse home visiting program for teenage mothers both contain these essential elements, as did the Gutelius trial of nurse home visiting plus mobile pediatric care.

Targeting on populations that need, want, and can benefit from the service. Together, Gutelius's work, the NFP, Koniak-Griffin's Early Intervention Program, and Butz's work provide consistent evidence that nurse home visiting during pregnancy and infancy has substantial potential to affect clinically important outcomes in the lives of vulnerable children and families. Each of these interventions targeted vulnerable populations during pregnancy or the postpartum period using nurses. Now the question is whether those programs can be replicated reliably and can produce corresponding effects when delivered in typical community contexts.

© 2007 The Authors
Journal compilation © 2007 Association for Child and Adolescent Mental Health.

Improving programs and research for policy and practice

In the past decade, we have witnessed significant growth in well-designed research on programs designed to improve child health and development by helping parents improve their care of their children early in life. Our ability to translate the results of basic research into effective public programs will depend upon our crafting programs that can engage parents and motivate them to provide the best possible care for themselves during pregnancy and their children during the early years of life. Our success in realizing this vision over the next decade will depend upon our continued use of formative evaluations of promising interventions, our pretesting them in small-scale RCTs, and then testing them in rigorously designed RCTs to help ensure that we are on the right path. Simply knowing that a program can work in a research setting does not mean that it can be effectively translated into community practice. Substantial work remains to be conducted to better understand how to make this translation effective. The process we are recommending takes time and runs counter to the heightened sense of urgency often voiced by advocates who believe that we need to act now. The history of policy making in parenting and early childhood interventions suggests that acting with insufficient data can be wasteful of resources and hope. In the long run, building solid programmatic and scientific foundations will get the desired results more efficiently. Let us review the essential elements of this strategy.

Develop interventions before testing them

Clear population targets. Some researchers and program developers have been reluctant to consider targeting their interventions on those segments of the population with higher rates of putative adverse influences on the outcomes targeted for change and thus likely to benefit the most from the service (see, for example, Zigler & Styfco, 2006). One result of this is that many programs with diffuse foci have been found to have limited impact on the samples examined in RCTs. In the United States, we have witnessed large investments in major national initiatives, such as Early Head Start and its predecessor the Comprehensive Child Development Program, which did not contain clearly defined populations and programs developed and tested using these approaches. In the United Kingdom, we find a similar approach in its Sure Start initiative, a program with broad mandates and population targets. There is no randomized trial of Sure Start to provide an estimate of program impact, but a quasi-experimental evaluation shows mixed results that give little indication that the investment is producing its intended effects (Belsky, Melhuish, Barnes, Leyland, & Romaniuk, 2006).

Epidemiology, theory, and pilot testing. To date, there is no federal parenting or early childhood intervention initiative for low-income parents and children that has followed the IOM strategy for developing effective preventive interventions. The practice followed so far has been to develop interventions around a set of broad principles rather than as targeted interventions with theoretically and empirically grounded strategies for altering specific risk and protective factors linked to the outcomes the intervention is designed to affect. Some intervention trials have been preceded by pilot studies and pretest work, but it is rare. In some cases, the pilot work was not conducted as a small-scale RCT, with the result that unbiased estimates of program effects were not possible; the result is that initiatives were launched that probably should have been put on hold until the ships were more seaworthy. As pilot studies become the norm, more attention needs to be given to ensuring participant engagement, retention, and adaptive behavioral change as prerequisites to the conduct of large-scale RCTs.

Engagement and retention. Almost every review of parenting interventions and home visiting programs has lamented the challenges of engaging and retaining parents. As we noted above, the Health Belief Model holds considerable promise as a means of addressing this issue. Urie Bronfenbrenner (1979) noted that major life transitions create a sense of vulnerability that increase individuals' receptivity to offers of help. These transitions create opportune times for preventive intervention. The challenge is to offer interventions that are viewed by intervention participants as helpful in reducing that vulnerability.

A recent evaluation of a home visiting program for socially disadvantaged families with newborns delivered by volunteers in the United Kingdom found significant challenges with family recruitment and engagement (Barnes, MacPherson, & Senior, 2006). Two-thirds of the women offered support from neighborhood volunteer Home Start visitors declined the service, a finding that resonates with the Denver trial of the Nurse Family Partnership, where families assigned a paraprofessional visitor dropped out of the program more frequently and missed more scheduled visits than did their nurse-visited counterparts (Korfmacher et al., 1999). Unless interventions are designed in ways that reliably ensure parental investment in the service and corresponding adaptive behavior change, they will fail.

Before large investments are made in RCTs or in programs delivered at the community level, these fundamental aspects of interventions need to be pretested, piloted, and demonstrated to have potential. The importance of following these steps in intervention development is further emphasized by a suggestion in the Home Start cluster-based trial in the United Kingdom that children in the intervention

© 2007 The Authors
Journal compilation © 2007 Association for Child and Adolescent Mental Health.

group performed worse on measures of child development than their control-group counterparts (Barnes et al., 2006). In spite of our best intentions, it is possible to harm those we attempt to help.

The potential to improve children's health and development by improving parental care has spurred great interest on the part of policy advisory groups such as the U.S. Advisory Board on Child Abuse and Neglect (1991) and the Council on Child and Adolescent Health of the American Academy of Pediatrics (1998). Conducting the kind of formative work on promising interventions prior to moving to larger-scale trials will allow investigators and program developers to determine whether programs show sufficient promise to warrant larger-scale trials. By following this kind of strategy, we can abort lackluster programs before they are tested in full scale trials and thus avoid squandering time, expense, and hope on programs that show little promise.

An example. Minding the Baby (MTB) is an intervention being developed in New Haven, Connecticut, that is following this model. It integrates advanced practice nursing and mental health care in home visits for at-risk first-time parents to prevent an array of negative maternal and infant outcomes. This pilot program is being developed using the kind of formative research methods that necessarily precede larger and more costly intervention trials. The MTB program has been built on the conceptual base and lessons learned from the NFP, as well as the work of (Heinicke and colleagues 1999, 2001), with some interesting additions (Olds, Kitzman, Cole, & Robinson, 1997; Slade, Sadler, & Mayes, 2005). The MTB intervention aims to enhance the parental capacities of young mothers (specifically maternal *Reflective Functioning (RF)* or the mother's capacity to understand her infant's mental and emotional needs), the infant attachment relationship, the health and development of infants and their young mothers, and the mental health of young mothers and infants who live in communities affected by the many stressors associated with poverty (Fonagy & Target, 1997; Slade, 2002; Slade, Grienenberger, Bernbach, Levy, & Locker, 2005). Mothers are coached to keep both the *physical and the emotional needs of their infants in mind* as they provide daily care and nurturing. It is an approach that acknowledges the importance of the mother's cognitive processes relative to the evolving parent–child relationship. Explicit attention is also given to the mothers' own mental health issues, and families are visited by a team consisting of a master's prepared nurse practitioner and a social worker.

The manualized MTB program begins in mid-pregnancy with weekly visits, and like the NFP extends through the child's second birthday. The visitors focus on the promotion of maternal reflective functioning (maternal attachment capacity), maternal self-efficacy and life-course outcomes (subsequent pregnancy timing and educational success), infant attachment, and child health. Clinical experience and initial neuro-biological evidence (Mayes et al., 2005) support the advantages of working with young mothers during a first pregnancy as an ideal time to engage them and to change negative patterns of behavior and thinking and to build on existing strengths. The program has been developed and evaluated for feasibility and is currently being evaluated in the form of a phase II pilot study, exploring preliminary effect sizes for key outcomes and intervention fidelity.

Preliminary findings suggest that this intensive relationship-based home visiting program may be of help to high risk families for whom significant attachment and relationship disruption may preclude response to less intensive interventions. Careful follow-up, a more rigorous randomization scheme for the next phase of research, and larger samples will help to determine whether these preliminary findings hold up and justify public investment in this intensive and relatively costly intervention.

Improve conduct of randomized trials

The increased reliance on RCTs to examine parenting interventions over the past decade is a substantial step forward, but there are several fundamental features of these designs that have been overlooked frequently in research design and implementation.

Intention-to-treat analyses. The most frequently violated design feature is employing intention-to-treat analyses. Violation of this principle is more frequently occurring in those studies that randomize participants at the stage of eligibility rather than after informed consent is obtained and baseline interviews conducted. This has led to possible selective refusals to participate in intervention or control groups in some studies, which undermines a basic feature of the randomized trial. Such designs could yield useful data if the measurement designs relied upon administrative data for which the investigators had complete access and the research question was whether offering such a service would make a difference in outcomes measured with these administrative data, but this scenario rarely occurs. While randomizing only those who agree to receive the intervention or control condition may limit the generalizability of the inferences derived from such trials, in studies of interventions that rely upon voluntary receipt of the service, this is not likely to be a significant problem as long as the rate of non-participation due to concerns about participating in research is minimized (Olds, 1988). It will be important, nevertheless, to examine the rate of participation in the trial when randomization is conducted after enrollment into the trial and to use this as a basis for understanding the generalizability of findings.

© 2007 The Authors
Journal compilation © 2007 Association for Child and Adolescent Mental Health.

Examine outcomes with clear public health significance. Some studies have examined outcomes that have questionable public health significance (such as parents' reports of their attitudes and behavior) and have used measurement designs that do not include multiple sources of information on the outcome domains of interest. The use of parent report without corroboration from independent data sources, such as developmental tests of the child, administrative and medical records, undermines the validity of trials that rely upon parent report alone.

Avoid over-interpretation of subgroup analyses. The reliance on results derived from subgroup analyses is likely to lead to incorrect answers and to failed policies and practices. A number of evaluations have reported subgroup effects as if they were on equal par with those findings that are derived from the full sample. Unless subgroup effects were predicted, were based upon stratified randomization procedures, or were based upon replicated findings across trials, they have an unacceptable likelihood of being wrong.

Replication. One of the basic principles advocated by the Society for Prevention Research (Flay et al., 2004) and the Blue Prints for Violence Prevention (Elliott, 2002) is that evidence needs to be replicated before it can serve as a foundation for policy and practice. To the extent that findings are replicated with different populations that are likely to be the focus of intervention in public policy, researchers and policy makers can have greater confidence in the program chosen for investment. This principle has not guided federal policy and practice in the early intervention arena to date, although in the United States some states, such as Pennsylvania, Colorado, and Louisiana are moving in this direction. Eventually, it will be important to conduct randomized trials of programs with strong evidentiary foundations as they are replicated in community settings (i.e., to conduct effectiveness trials). First, however, it will be important to ensure that those evidence-based programs replicated in community settings are functioning well, as it is possible to test them before they are implemented as well as they could be.

Olds and his team have taken the position that critical features of program implementation need to be addressed and resolved to the extent possible before an effectiveness trial of the NFP is conducted. These features include increasing participant retention, making the parenting program easier to implement, and addressing maternal mental illness and intimate partner violence more effectively. This approach to working out the challenges of implementing evidence-based interventions and improving the model itself before it is tested in effectiveness trials will increase the likelihood that interventions that do get tested will demonstrate their potential.

Estimate costs and cost-savings. As research in this field progresses, it will be important to gain additional insight into the costs of the interventions examined and their economic benefits. As a society we need to know not only whether the interventions we develop and test produce functional benefits, but whether the interventions make economic sense. James Heckman, an economist from the University of Chicago, is leading a consortium of early intervention researchers that is examining this issue with several classic early intervention models that have longitudinal outcome data. The economic returns to society resulting from investments in early parenting interventions are not well known, but could be substantial (Cunha, Heckman, Lochner, & Masterov, in press).

Correspondence to

David Olds, 1825 Marion Street, Denver, CO, USA; Email: Olds.David@tchden.org

References

Achenbach, T.M., Howell, C.T., Aoki, M.F., & Rauh, V.A. (1993). Nine-year outcome of the Vermont intervention program for low birth weight infants. *Pediatrics*, *91*, 45–55.

Administration on Children, Youth, and Families. (2002a). *Making a difference in the lives of infants and toddlers and their families: The impacts of Early Head Start. Volume I: Final technical report*. Washington, DC: Department of Health and Human Services.

Administration on Children, Youth, and Families. (2002b). *Making a difference in the lives of infants and toddlers and their families: The impacts of Early Head Start. Volume II: Final technical report appendixes*. Washington, DC: Department of Health and Human Services.

Ainsworth, M., Blehar, E., Waters, E., & Wall, S. (1978). *Patterns of attachment: A psychological study of the Strange Situation*. Hillsdale, NJ: Erlbaum.

Als, H. (1982). Toward a synactive theory of development: Promise for the assessment and support of infant individuality. *Infant Mental Health Journal*, *3*, 229–243.

Als, H., Duffy, F.H., McAnulty, G.B., Rivkin, M.J., Vajapeyam, S., Mulkern, R.V. et al. (2004). Early experience alters brain function and structure. *Pediatrics*, *113*, 846–857.

Als, H., Gilkerson, L., Duffy, F.H., McAnulty, G.B., Buehler, D.M., Vandenberg, K. et al. (2003). A three-center, randomized, controlled trial of individualized developmental care for very low birth weight preterm infants: Medical, neurodevelopmental, parenting, and caregiving effects. *Developmental and Behavioral Pediatrics*, *24*, 399–408.

Aos, S., Lieb, R., Mayfield, J., Miller, M., & Pennucci, A. (2004). *Benefits and costs of prevention and early intervention programs for youth*. Olympia: Washington State Institute for Public Policy.

© 2007 The Authors
Journal compilation © 2007 Association for Child and Adolescent Mental Health.

Assmann, S.F., Pocock, S.J., Enos, L.E., & Kasten, L.E. (2000). Subgroup analysis and other (mis)uses of baseline data in clinical trials. *Lancet*, *355*, 1064–1069.

Barnes, J., MacPherson, K., & Senior, R. (2006). Factors influencing the acceptance of volunteer home-visiting support offered to families with new babies. *Child and Family Social Work*, *11*, 107–111.

Begg, C.B., Cho, M.K., Eastwood, S., Horton, R., Moher, D., Olkin, I. et al. (1996). Improving the quality of reporting of randomized controlled trials: The CONSORT statement. *Journal of the American Medical Association*, *276*, 637–639.

Belsky, J., Melhuish, M., Barnes, J., Leyland, A., & Romaniuk, H. (2006). Effects of Sure Start local programmes on children and families: Early findings from a quasi-experimental, cross-sectional study. *British Medical Journal*, *332*, 1476–1481.

Bohman, M. (1996). Predisposition to criminality: Swedish adoption studies in retrospect. In G.R. Bock & J.A. Goode (Eds.), *Genetics of criminal and antisocial behavior* (pp. 99–114). Chichester, UK: Wiley.

Bornstein, M. (1995). Parenting infants. In Marc H. Bornstein (Ed.), *Handbook of parenting* (pp. 3–39). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Bowlby, J. (1969). *Attachment and loss.* Vol 1. *Attachment*. New York: Basic Books.

Bradley, R.H., & Caldwell, B.M. (1979). Home observation for measurement of the environment: A revision of the preschool scale. *American Journal of Mental Deficiency*, *84*, 235–244.

Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Cambridge, MA: Harvard University Press.

Brooks-Gunn, J., McCarton, C.M., Casey, P.H., McCormick, M.C., Bauer, C.R., Bernbaum, J.C. et al. (1994). Early intervention in low-birth-weight premature infants: Results through age 5 years from the Infant Health and Development Program. *Journal of the American Medical Association*, *272*, 1257–1262.

Bugental, D.B., Ellerson, P.C., Lin, E.K., Rainey, B., Kokotovic, A., & O'Hara, N. (2002). A cognitive approach to child abuse prevention. *Journal of Family Psychology*, *16*, 243–258.

Buhler-Wilkerson, K. (1993). Bringing care to the people: Lillian Wald's legacy to public health nursing. *American Journal of Public Health*, *83*, 1778–1786.

Butz, A.M., Pulsifer, M., Marano, N., Belcher, H., Lears, M.K., & Royall, R. (2001). Effectiveness of a home intervention for perceived child behavioral problems and parenting stress in children with in utero drug exposure. *Archives of Pediatric Adolescent Medicine*, *155*, 1029–1037.

Campbell, F.A., & Ramey, C.T. (1995). Cognitive and school outcomes for high-risk African-American students at middle adolescence: Positive effects of early intervention. *American Educational Research Journal*, *32*, 743–772.

Caspi, A., Moffitt, T.E., Morgan, J., Rutter, M., Taylor, A., Arseneault, L. et al. (2004). Maternal expressed emotion predicts children's antisocial behavior problems: Using monozygotic-twin differences to identify environmental effects on behavioral development. *Developmental Psychology*, *40*, 149–161.

Caspi, A., Sugden, K., Moffitt, T.E., Taylor, A., Craig, I.W., Harrington, H. et al. (2003). Influence of life stress on depression: Moderation by a polymorphism in the 5-HTT gene. *Science*, *301*, 386–389.

Chaffin, M. (2004). Is it time to rethink Healthy Start/Healthy Families? *Child Abuse and Neglect, 28,* 589–95.

Clutton-Brock, T.H. (1991). *The evolution of parental care*. Princeton, New Jersey: Princeton University Press.

Council on Child and Adolescent Health – The American Academy of Pediatrics. (1998). The role of home-visitation programs in improving health outcomes for children and families. *Pediatrics*, *101*, 486–489.

Cunha, F., Heckman, J., Lochner, L., & Masterov, D. (in press). Interpreting the evidence on life cycle skill formation. In E. Hanushek & F. Welch (Eds.), *Handbook of economics and education*. Amsterdam: North-Holland.

Daro, D.A., & Harding, K.A. (1999). Healthy Families America: Using research to enhance practice. *Future of Children*, *9*, 152–176.

Diaz, J., Oshana, D., & Harding, K. (2004). *Healthy Families America: 2003 annual profile of program sites*. Chicago, IL: Prevent Child Abuse America.

Drotar, D., Hurwitz, H.M., & Kirchner, H.L. (2006). *The Cleveland eastern suburban Born to Learn program: Final report*. Unpublished report.

Duggan, A., Caldera, D. Rodriguez, K., Burrell, L., Shea, S., & Rohde, C. (2005). *Evaluation of the Healthy Families Alaska Program*. Retrieved August 15, 2006 from http://www.hss.state.ak.us/ocs/Publications/JohnsHopkins_HealthyFamilies.pdf.

Duggan, A., McFarlane, E., Fuddy, L., Burrell, L., Higman, S.M., Windham, A., & Sia, C. (2004). Randomized trial of a statewide home visiting program: Impact in preventing child abuse and neglect. *Child Abuse and Neglect*, *28*, 597–622.

Duggan, A.K., McFarlane, E.C., Windham, A.M., Rohde, C.A., Salkever, D.S., Fuddy, L. et al. (1999). Evaluation of Hawaii's Healthy Start Program. *Future of Children*, *9*, 66–90.

DuMont, K., Mitchell-Herzfeld, S., Greene, R., Lee, E., Lowenfels, A., & Rodriguez, M. (2006). *Healthy Families New York (HFNY) randomized trial: Impacts on parenting after the first two years*. New York: New York State Office of Children & Family Services.

Eckenrode, J., Ganzel, B., Henderson C.R. Jr, Smith, E., Olds, D.L., Powers, J. et al. (2000). Preventing child abuse and neglect with a program of nurse home visitation: The limiting effects of domestic violence. *Journal of the American Medical Association*, *284*, 1385–1391.

Elliott, D.S. (1998). *Blueprints for Violence Prevention series*. Boulder, CO: Center for the Study and Prevention of Violence.

Elliott, D.S. (2002). *Blueprints for Violence Prevention replications: Factors for implementation success*. Boulder, CO: Center for the Study and Prevention of Violence.

Elster, A., & McAnarney, E. (1980). Medical and psychosocial risks of pregnancy and childbearing during adolescence. *Pediatric Annals*, *9*, 13.

© 2007 The Authors
Journal compilation © 2007 Association for Child and Adolescent Mental Health.

Fergusson, D.M., Grant, H., Horwood, L.J., & Ridder, E.M. (2005). Randomized trial of the Early Start program of home visitation. *Pediatrics*, *116*, 803–809.

Fergusson, D.M., Grant, H., Horwood, L.J., & Ridder, E.M. (2006). Randomized trial of the Early Start program of home visitation: Parent and family outcomes. *Pediatrics*, *117*, 781–786.

Fewell, R.R., & Scott, K.G. (1997). The cost of implementing the intervention. In R.T. Gross, D. Spiker, & C.W. Haynes (eds.), *Helping low birth weight, premature babies: The Infant Health and Development Program*. (pp. 479–502; 624–625). Stanford, CA: Stanford University Press.

Fixsen, D.L., Naoom, S.F., Blase, K.A., Friedman, R.M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature*. Tampa: University of South Florida.

Flay, B.R., Biglan, A., Boruch, R.F., Castro, F.G., Gottfredson, D., Kellam, S.G. et al. (2004). *Standards of evidence: Criteria for efficacy, effectiveness and dissemination*. Falls Church, VA: Society for Prevention Research. Available at http://www.preventionresearch.org/StandardsofEvidencebook.pdf.

Flay, B.R., Biglan, A., Boruch, R.F., Castro, F.G., Gottfredson, D., Kellam, S. et al. (2005). Standards of evidence: Criteria for efficacy, effectiveness, and dissemination. *Prevention Science*, Online First.

Fonagy, P., & Target, M. (1997). Attachment and reflective function: Their role in self-organization. *Development and Psychopathology*, *9*, 679–700.

Francis, D.D., Diorio, J., Liu, D. & Meaney, M.J. (1999). Variations in maternal care form the basis for a nongenomic mechanism of inter-generational transmission of individual differences in behavioral and endocrine responses to stress. *Science*, *286*, 1155–1158.

Furstenberg, F.F., Brooks-Gunn, J., & Morgan, S.P. (1987). *Adolescent mothers in later life*. Cambridge: Cambridge University Press.

Gallup Organization. (2005). *Nurses remain atop honesty and ethics list*. Retrieved August 15, 2006 at http://poll.gallup.com/content/default.aspx?ci=20254&pg=1.

Gardner, J.M., Walker, S.P., Powell, C.A., & Grantham-McGregor, S. (2003). A randomized controlled trial of a home-visiting intervention on cognition and behavior in term low birth weight infants. *Journal of Pediatrics*, *143*, 634–639.

Gilliam, W.S., Ripple, C.H., Zigler, E.F., & Leiter, V. (2000). Evaluating child and family demonstration initiatives: Lessons from the Comprehensive Child Development Program. *Early Childhood Research Quarterly*, *15*, 41–60.

Gomby, D. (2005). *Home visitation in 2005: Outcomes for children and parents*. Washington, DC: Committee for Economic Development: Invest in Kids Working Group.

Gomby, D.S., Culross, P.L., & Behrman, R.E. (1999). Home-visiting: Recent program evaluations – analysis and recommendations. *The Future of Children*, *9*, 4–26.

Goodson, B.D., Layzer, J.I., St. Pierre, R.G., Bernstein, L.S., & Lopez, M. (2000). Effectiveness of a comprehensive five-year family support program on low-income children and their families: Findings from the Comprehensive Child Development Program. *Early Childhood Research Quarterly*, *15*, 5–39.

Greenwald, P., & Cullen, J.W. (1984). The scientific approach to cancer control. *CA: A Cancer Journal for Clinicians*, *34*, 328–333.

Gutelius, M.F., & Kirsch, A.D. (1975). Factors promoting success in infant education. *American Journal of Public Health*, *65*, 384–387.

Gutelius, M.F., Kirsch, A.D., MacDonald, S., Brooks, M.R., McErlean, T., & Newcomb, C. (1972). Promising results from a cognitive stimulation program in infancy. A preliminary report. *Clinical Pediatrics*, *11*, 585–593.

Hahn, R.A., Bilukha, O.O., Crosby, A., Fullilove, M.T., Lieberman, A., Moscicki, E.K. et al. (2003). First reports evaluating the effectiveness of strategies for preventing violence: Early child home visitation. Findings from the Task Force on Community Preventive Services. *Morbidity and Mortality Weekly Report*, *52*(RR14), 1–9.

Heinicke, C.M., Fineman, N.R., Ponce, V.A., & Guthrie, D. (2001). Relationship-based intervention with at-risk mothers: Outcome in the second year of life. *Infant Mental Health Journal*, *22*, 431–462.

Heinicke, C.M., Fineman, N.R., Ruth, G., Recchia, S.L., Guthrie, D., & Rodning, C. (1999). Relationship-based intervention with at-risk mothers: Outcome in the first year of life. *Infant Mental Health Journal*, *20*, 349–374.

Heinicke, C.M., Goorsky, M., Levine, M., Ponce, V., Ruth, G., Silverman, M., & Sotelo, C. (2006). Pre- and postnatal antecedents of a home-visiting intervention and family developmental outcome. *Infant Mental Health Journal*, *27*, 91–119.

Hodnett, E., & Fredericks, S. (2004). Support during pregnancy for women at increased risk of low birthweight babies. Cochrane Pregnancy and Childbirth Group. *Cochrane Database of Systematic Reviews*, *3*.

Infant Health and Development Program. (1990). Enhancing the outcomes of low-birth-weight, premature infants: A multisite, randomized trial. *Journal of the American Medical Association*, *263*, 3035–3042.

Jacobs, S., Sokol, J., & Ohlsson, A. (2002). The newborn individualized developmental care and assessment program is not supported by meta-analyses of the data. *Journal of Pediatrics*, *140*, 699–706.

Karoly, L.A., Kilburn, M.R., & Cannon, J.S. (2005). *Early childhood interventions: Proven results, future promise*. Santa Monica, CA: RAND.

Kempe, H. (1976). *Child abuse and neglect: The family and the community*. Cambridge, MA: Ballinger Publishing Company.

Kitzman, H., Olds, D.L., Henderson, C.R. Jr, Hanks, C., Cole, R., Tatelbaum, R. et al. (1997). 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*, 644–652.

Kitzman, H., Olds, D.L., Sidora, K., Henderson, C.R. Jr, Hanks, C., Cole, R. et al. (2000). Enduring effects of nurse home visitation on maternal life course: A 3-year follow-up of a randomized trial. *Journal of the American Medical Association*, *283*, 1983–1989.

Koniak-Griffin, D., Anderson, N.L., Brecht, M., Verzemnieks, I., Lesser, J., & Kim, S. (2002). Public

© 2007 The Authors
Journal compilation © 2007 Association for Child and Adolescent Mental Health.

health nursing care for adolescent mothers: Impact on infant health and selected maternal outcomes at one year post birth. *Journal of Adolescent Health*, *30*, 44–54.

Koniak-Griffin, D., Anderson, N., Verzemnieks, I., & Brecht, M. (2000). A public health nursing early intervention program for adolescent mothers: Outcomes from pregnancy through 6 weeks postpartum. *Nursing Research*, *49*, 130–138.

Koniak-Griffin, D., Mathenge, C., Anderson, N., & Verzemnieks, I. (1999). An early intervention program fro adolescent mothers: A nursing demonstration project. *Journal of Obstetric, Gynecologic and Neonatal Nursing*, *28*, 51–59.

Koniak-Griffin, D., Verzemnieks, I., Anderson, N., Brecht, M., Lesser, J., Kim, S. et al. (2003). Nurse visitation for adolescent mothers. *Nursing Research*, *52*, 12–135.

Korfmacher, J., O'Brien, R., Hiatt, S., & Olds, D. (1999). Differences in program implementation between nurses and paraprofessionals providing home visits during pregnancy and infancy: A randomized trial. *American Journal of Public Health*, *89*, 1847–1851.

Landsverk, J., Carrilio, T., Connelly, C.D., Ganger, W.C., Slymen, D.J., Newton, R.R. et al. (2002). *Healthy Families San Diego Clinical Trial Technical Report*. San Diego, CA: Child and Adolescent Services Research Center, San Diego Children's Hospital and Health Center.

Lesser, J., & Koniak-Griffin, D. (2000). The impact of physical or sexual abuse on chronic depression in adolescent mothers. *Journal of Pediatric Nursing*, *15*, 378–387.

Love, J.M., Kisker, E.E., Ross, C., Raikes, H., Constantine, J., Boller, K. et al. (2005). The effectiveness of Early Head Start for 3-year-old children and their parents: Lessons for policy and programs. *Developmental Psychology*, *41*, 885–901.

Love, J.M., Kisker, E.E., Ross, C.M., Schochet, P.Z., Brooks-Gunn, J., Paulsell, D. et al. (2002). *Making a difference in the lives of infants and toddlers and their families: The impacts of early head start*. Princeton, NJ: Mathematica Policy Research, Inc.

MacMillan, H.L., Thomas, B.H., Jamieson, E., Walsh, C.A., Boyle, M.H., Shannon, H.S., & Gafni A. (2005). Effectiveness of home visitation by public-health nurses in prevention of the recurrence of child physical abuse and neglect: A randomised controlled trial. *Lancet*, *365*, 1786–1793.

Main, M., Goldwyn, R., & Hesse, E. (2003). Adult attachment scoring and classification systems. *Manual in preparation*.

Mayes, L.C., Swain, J.E., & Leckman, J. (2005). Parental attachment systems: Neural circuits, genes and experiential contributions to parental engagement. *Clinical Neuroscience Research*, *4*, 301–313.

McCarton, C.M., Brooks-Gunn, J., Wallace, I.F., Bauer, C.R., Bennett, F.C., Bernbaum, J.C. et al. (1997). Results at age 8 years of early intervention for low-birth-weight premature infants. The Infant Health and Development Program. *Journal of the American Medical Association*, *277*, 126–132.

McCormick, M.C., Brooks-Gunn, J., Buka, S.L., Goldman, J., Yu, J., Salganik, M. et al. (2006). Early intervention in low birth weight premature infants: Results at 18 years of age for the Infant Health and Development Program. *Pediatrics*, *117*, 771–780.

McGauhey, P.J., Staffield, B.H., Alexander, C., & Ensminger, M.E. (1991). Social environment and vulnerability of low birth weight children: A social-epidemiological perspective. *Pediatrics*, *88*, 943–953.

Miller, W.R., & Rollnick, S. (2002). *Motivational interviewing: Preparing people for change* (2nd edn). New York: Guilford Press.

Minkovitz, C.S., Hughart, N., Strobino, D., Scharfstein, D., Grason, H., Hou, W., Miller, T., Bishai, D., Augustyn, M., McLearn, K.T., & Guyer, B. (2003). A practice-based intervention to enhance quality of care in the first 3 years of life: The Healthy Steps for Young Children Program. *Journal of the American Medical Association*, *290*, 3081–3091.

Mitchell-Herzfeld, S., Izzo, C., Greene, R., Lee, E., & Lowenfels, A. (2005). *Evaluation of Healthy Families New York (HFNY): First year program impacts*. New York: Office of Children and Family Services Bureau of Evaluation and Research.

Moher, D., Schulz, K.F., & Altman, D., for the CONSULT Group. (2001). The CONSORT statement: Revised recommendations for improving the quality of reports of parallel-group randomized trials. *Journal of the American Medical Association*, *285*, 1987–1991.

Mrazek, P.J., & Haggerty, R.J. (1994). *Reducing risks for mental disorders: Frontiers for preventive intervention research*. Washington, DC: Committee on Prevention of Mental Disorders, Institute of Medicine.

National Commission to Prevent Infant Mortality. (1989). *Home visiting: Opening doors for America's pregnant women and children*. Washington, DC: National Commission to Prevent Infant Mortality.

Olds, D., Henderson C.R. Jr, Kitzman, H., & Cole, R. (1995). Effects of prenatal and infancy nurse home visitation on surveillance of child maltreatment. *Pediatrics*, *95*, 365–372.

Olds, D., Kitzman, H., Cole, R., & Robinson, J. (1997). Theoretical foundations of a program of home visitation for pregnant women and parents of young children. *Journal of Community Psychology*, *25*, 9–25.

Olds, D., Kitzman, H., Cole, R., Robinson, J., Sidora, K., Luckey, D. et al. (2004). Effects of nurse home visiting on maternal life-course and child development: Age-six follow-up of a randomized trial. *Pediatrics*, *114*, 1550–1559.

Olds, D.L. (1988). Common design and methodological problems encountered in evaluating family support services: Illustrations from the Prenatal/Early Infancy Project. In H.B. Weiss & F.H. Jacobs (Eds.), *Evaluating family programs* (pp. 239–265). New York: Aldine De Gruyter.

Olds, D.L. (2002). Prenatal and infancy home visiting by nurses: From randomized trials to community replication. *Prevention Science*, *3*, 153–172.

Olds, D.L. (2006). The Nurse–Family Partnership: An evidence-based preventive intervention. *Infant Mental Health Journal*, *27*, 5–25.

Olds, D.L., & Kitzman, H. (1990). Can home visitation improve the health of women and children at environmental risk? *Pediatrics*, *86*, 108–116.

Olds, D.L., & Kitzman, H. (1993). Review of research on home visiting for pregnant women and parents of young children. *The Future of Children*, *3*, 53–92.

© 2007 The Authors
Journal compilation © 2007 Association for Child and Adolescent Mental Health.

Olds, D.L., & Korfmacher, J. (1997). The evolution of a program of research on prenatal and early childhood home visitation: Special issue introduction. *Journal of Community Psychology*, *25*, 1–7.

Olds, D.L., Eckenrode, J., Henderson, C.R. Jr, Kitzman, H., Powers, J., Cole, R. et al. (1997). Long-term effects of home visitation on maternal life course and child abuse and neglect: 15-year follow-up of a randomized trial. *Journal of the American Medical Association*, *278*, 637–643.

Olds, D.L., Henderson, C.R. Jr, Tatelbaum, R., & Chamberlin, R. (1986). Improving the delivery of prenatal care and outcomes of pregnancy: A randomized trial of nurse home visitation. *Pediatrics*, *77*, 16–28.

Olds, D.L., Henderson, C.R. Jr, Cole, R., Eckenrode, J., Kitzman, H., Luckey, D. et al. (1998). Long-term effects of nurse home visitation on children's criminal and antisocial behavior: 15-year follow-up of a randomized trial. *Journal of the American Medical Association*, *280*, 1238–1244.

Olds, D.L., Hill, P., Robinson, J., Song, N., & Little, C. (2000). Update on home visiting for pregnant women and parents of young children. *Current Problems in Pediatrics*, *30*, 105–148.

Olds, D.L., Hill, P.L., O'Brien, R., Racine, D., & Moritz, P. (2003). Taking preventive intervention to scale: The Nurse–Family Partnership. *Cognitive and Behavioral Practice*, *10*, 278–290.

Olds, D.L., Robinson, J., O'Brien, R., Luckey, D.W., Pettitt, L.M., Henderson, C.R. Jr et al. (2002). Home visiting by paraprofessionals and by nurses: A randomized, controlled trial. *Pediatrics*, *110*, 486–496.

Olds, D.L., Robinson, J., Pettitt, L., Luckey, D.W., Holmberg, J., Ng, R.K. et al. (2004). Effects of home visits by paraprofessionals and by nurses: Age-four follow-up of a randomized trial. *Pediatrics*, *114*, 1560–1568.

Overpeck, M.D., Brenner, M.D., Trumble, A.C., Trifiletti, L.B., & Berendes, H.W. (1998). Risk factors for infant homicide in the United States. *New England Journal of Medicine*, *339*, 1121–1216.

Parents as Teachers National Center. (2005). *The shape of things to come: 2005 annual report*. St. Louis, MO: Parents as Teachers National Center, Inc.

Pfannenstiel, J., & Seltzer, D. (1985). *Evaluation report: New Parents as Teachers project*. Overland Park, KS: Research & Training Associates.

Ramey, C.T., Bryant, D.M., Wasik, B.H., Sparling, J.J., Fendt, K.H. & LaVange, L.M. (1992). Infant Health and Development Program for low birth weight, premature infants: Program elements, family participation, and child intelligence. *Pediatrics*, *89*, 454–465.

Raynor, P., Rudolf, M.C.J., Cottrell, D., Marchant, P., & Cooper, K. (1999). A randomised control trial of specialist health visitor intervention for failure to thrive. *Archives of Disease in Childhood*, *80*, 500–506.

Rosenstock, I.M. (1990). The Health Belief Model: Explaining health behavior through expectancies. In K. Glanz, F.M. Lewis, & B.K. Rimer (Eds.), *Health behavior and health education: Theory, research, and practice* (pp. 39–62). San Francisco: Jossey-Bass.

Sanders, M.R., Markie-Dadds, C., & Turner, K.M.T. (2003). *Theoretical, scientific, and clinical foundations of the Triple P-Positive Parenting Program: A population approach to the promotion of parenting competence. Parenting Research and Practice Monograph No. 1*. Queensland, Australia: The Parenting and Family Support Centre.

Slade, A. (2002). Keeping the baby in mind: A critical factor in perinatal mental health. *Zero to Three*, *22*, 10–16.

Slade, A., Grienenberger, J., Bernbach, E., Levy, D., & Locker, A. (2005). Maternal reflective functioning and attachment: Considering the transmission gap. *Attachment and Human Development*, *7*, 283–292.

Slade, A., Sadler, L.S., & Mayes, L. (2005). Minding the baby: Enhancing parental reflective functioning in a nursing/mental health home visiting program. In L. Berlin, Y. Ziv, L. Amaya-Jackson, & M. Greenberg (Eds.), *Enhancing early attachments* (pp. 152–177). New York: Guilford Publications.

Sparling, J., & Lewis, I. (1979). *Learning games for the first three years*. New York: Walker & Co.

Spoth, R., Redmond, C., & Shin, C. (2000). Modeling factors influencing enrollment in family-focused preventive intervention research. *Prevention Science*, *1*, 213–225.

St. Pierre, R.G., & Layzer, J.I. (1999). Using home visits for multiple purposes: The Comprehensive Child Development Program. *Future of Children*, *9*, 134–151.

Strauss, M.A., Hamby, S.L., Bonney-McCoy, S., & Sugarman, D.B. (1996). The revised Conflict Tactics Scale (CTS2). *Journal of Family Issues*, *17*, 283–316.

Sweet, M.A., & Appelbaum, M.I. (2004). Is home visiting an effective strategy? A meta-analytic review of home visiting programs for families with young children. *Child Development*, *75*, 1435–1456.

Tienari, P., Wynne, L.C., Moring, J., Lahti, I., Naarala, M., Sorri, A. et al. (1994). The Finnish Adoptive Family Study of Schizophrenia: Implications for family research. *British Journal of Psychiatry*, *23* (Suppl.), 20–26.

U.S. Advisory Board on Child Abuse and Neglect (1991). *Creating caring communities: Blueprint for an effective federal policy on child abuse and neglect*. Washington, DC: U.S. Government Printing Office.

U.S. Department of Health and Human Services. (2000). *Healthy people 2010* (2nd edn). Washington, DC: U.S. Government Printing Office.

Wagner, M., & Spiker, D. (2001). *Multisite Parents as Teachers evaluation: Experiences and outcomes for children and families (SRI Project P07283)*. Menlo Park, CA: SRI International.

Wagner, M., Iida, E., & Spiker, D. (2001). *The multisite evaluation of the Parents as Teachers home visiting program: Three year findings from one community*. Menlo Park, CA: SRI International.

Wagner, M.M., & Clayton, S.L. (1999). The Parents as Teachers program: Results from two demonstrations. *The Future of Children*, *9*, 91–115.

Wald, L. (1915). *The house on Henry Street*. New York: Henry Hold and Company, Inc.

Webster-Stratton, C., Reid, M.J., & Hammond, M. (2001). Preventing conduct problems, promoting social competence: A parent and teacher training partnership in Head Start. *Journal of Clinical Child Psychology*, *30*, 238–302.

© 2007 The Authors
Journal compilation © 2007 Association for Child and Adolescent Mental Health.

Webster-Stratton, C., Reid, M.J., & Hammond, M. (2004). Treating children with early-onset conduct problems: Intervention outcomes for parent, child, and teacher training. *Journal of Clinical Child and Adolescent Psychology*, *33*, 105–124.

White, B. (1985). *The first three years of life*. Indianapolis: Prentice Hall Professional Technical Reference.

Winter, M., & McDonald, D. (1997). Parents as Teachers: Investing in good beginnings for children. In G. Albee & T. Gullotta (Eds.), *Primary prevention works* (pp. 119–145). Thousand Oaks, CA: Sage Publications.

Zigler, E., & Styfco, S.J. (2006). Epilogue. In N.F. Watt, C. Ayoub, R.H. Bradley, J.E. Puma, & W.A. LeBoeuf (Eds.), *The crisis in youth mental health: Early intervention programs and policies* (pp. 347–371). Westport, CT: Praeger.

Manuscript accepted 24 August 2006

© 2007 The Authors
Journal compilation © 2007 Association for Child and Adolescent Mental Health.