

**National Evaluation
of Family Support
Programs**

**Final Report
Volume A: The Meta-
Analysis**

April, 2001

Prepared for:

Mary Bruce Webb
DHHS/ACYF
Room 2411
330 C Street, SW
Washington, DC 20201

Prepared by:

Jean I. Layzer
Barbara D. Goodson
Lawrence Bernstein
Cristofer Price

Abt Associates Inc.
55 Wheeler Street
Cambridge, MA 01238

Table of Contents

Preface and Acknowledgments	i
Introduction	ii
Chapter A1 Rationale for the Use of Meta-Analysis	A1-1
Chapter A2 Preparation for the Meta-Analysis	A2-1
Formulating the Problem	A2-1
Searching the Literature	A2-5
Selecting Research Studies and Programs for the Meta-Analysis	
Database	A2-7
Coding the Studies	A2-9
Chapter A3 Characteristics of Programs and Studies Included in the	
Meta-Analysis	A3-1
Program Characteristics	A3-1
Comparing Evaluated With Unevaluated Programs	A3-16
Chapter A4 Analytic Approach	A4-1
Constructing the Analytic Database	A4-1
Level of Analysis	A4-7
Analytic Strategy	A4-11
Chapter A5 Results of the Meta-Analysis	A5-1
What are the Short-Term Effects of Family Support Programs and	
Services on Children?	A5-1
Variation in Effects	A5-10
Which Program Characteristics are Related to Differential Effects	
of Family Support Programs and Services	A5-14
What are the Long-Term Effects of Family Support Programs and	
Services on Children and Families?	A5-34
Conclusions	A5-42
References	1

Preface and Acknowledgments

This document is the final report of the National Evaluation of Family Support Programs, undertaken for the Administration on Children, Youth and Families by Abt Associates Inc., with a subcontract to Yale University.

This evaluation benefited from the advice of many individuals. Our Senior Investigator's Team of Charles Bruner, Judy Langford Carter, Thomas Cook, Luis Laosa, Marie McCormick, Douglas Powell and Ralph Smith helped us to define family support programs, to select exemplary programs and to design an evaluation of them. Members of the Technical Work Group, including Carl Dunst, Larry Aber, Bernice Weissbourd, Sheila Kamerman, Heather Weiss, Francine Jacobs and Constance Williams, reviewed our evaluation plans and measures, and provided helpful comments on our preliminary findings.

Peter Rossi and Carol Weiss helped with the difficult task of designing a set of evaluations tailored to the specific circumstances of each of the programs included in the prospective studies. The site-specific evaluation designs benefited from the advice of Lynn McDonald, Carol Calfee, David Roth and Karen McCarthy. Thomas Cook, Larry Hedges and Robert Fisher offered guidance as we designed the meta-analysis that is the centerpiece of this report. Will Shadish provided ongoing help and advice as well as sharing with us an important software tool in the late stages of its development.

Staff of the Administration on Children, Youth and Families were responsible for providing technical input and for oversight of the evaluation. James Griffin provided support and oversight in the early design phases of the study. Mary Bruce Webb, who was responsible for the oversight of the study for the last five years, encouraged us to undertake a complex and challenging meta-analysis, and provided consistent sound advice throughout the study.

Our colleague and collaborator, Sharon Lynn Kagan, educated us about family support and was an invaluable resource in our efforts to design the prospective studies. In addition, she and other staff at Yale University were responsible for Volume C of this final report.

Finally, several staff members at Abt Associates Inc. played important roles on the project. Valuable contributions were made by Barbara Goodson, Lawrence Bernstein, Cristofer Price, Alan Werner, Cindy Creps, Helen Barnes, Linda Hailey, Gabriela Garcia, Marjorie Levin, Diane Greene, Kate Greene, Robert St. Pierre, Chris Saia, Susan St. Pierre, Mary Kay Crepinsek, Melanie Brown-Lyons, Michael Puma, Mark Rollins, Rebecca Burstein, Emily Pollakowski, Maureen Cook and Deb Thebearge.

Introduction

The Omnibus Reconciliation Act of 1993 (P. L. 103-66) amended Title IV-B of the Social Security Act and created Subpart 2—“Family Preservation and Support Services.” This legislative initiative provides funding to States, through a capped entitlement, “for the purpose of encouraging and enabling each State to develop and establish, or expand, and to operate a program of family preservation services and community-based family support services.” The legislation defines family preservation and family support services as follows:

- **Family preservation services** — services for children and families designed to help families (including adoptive, foster and extended families) at risk or in crisis, including services to help children at risk of foster care placement remain with their families, or where appropriate return to their families;
- **Family support services**—community-based services to promote the well-being of children and families, designed to increase the strength and stability of families through such services as respite care, drop-in centers, early development screening and parent training.

Funds were authorized for a five-year period, beginning with \$60 million in FY 1994 and increasing over the remaining four years. Subpart 2, Section 435 of OBRA 1993 required the Secretary of Health and Human Services to evaluate the effectiveness of family preservation and family support programs.

Although the two program types supported by the legislation share the ultimate goals of children’s and families’ well-being, they differ considerably in types of services provided, populations served, program philosophy and mission. They represent different points on the continuum of services provided for in the legislation. Family support programs offer services to all families in a given community (a neighborhood, a school, an early childhood program). Their hope is that, by supporting and building on family strengths, they can help families avoid stresses and conflict. For those families that are at risk, such services may ward off the problems that result in family break-up. There are tens of thousands of such programs and services across the country. For the most part, they are small, grass-roots efforts, developed within communities. Family preservation services, by contract, seek to assist families that are already manifesting problems or who are in crisis. The intervention is more specific, the population more defined, the programs fewer and more likely to be state-initiated.

Acknowledging these differences, and seeking to respond fully to the legislative mandate, the Department of Health and Human Services determined that the two types of programs required different evaluation approaches. In September 1994, the Administration on Children,

Youth and Families awarded a contract to Abt Associates Inc. to document, describe and assess the effectiveness of family support programs and services.

This is the first of three volumes of the Final Report for the National Evaluation of Family Support Programs. The three volumes report on distinctly different aspects of the study. This volume reports findings from a meta-analysis of existing research on programs that provide family support services. The second volume reports the findings from a small set of research studies of mature, well-implemented family support programs. The final volume is a set of case studies of the effect on the systems of family support services in three states of the additional funds for family support provided under federal legislation.

Chapter A1

Rationale for the Use of Meta-Analysis

The purpose of the National Evaluation of Family Support is to understand the effects of family support on families and children. A central task of the evaluation is to synthesize and analyze the body of research on family support, to answer the following questions,

- What are the effects of family support programs and services on families and children?
- How effective are different types of family support programs and services?
- How effective are family support programs and services for children and families with different characteristics, needs and circumstances?

There exist two strategies for summarizing the research on family support: narrative synthesis and meta-analysis. Narrative synthesis was, until recently, the strategy most commonly used in the social sciences for summarizing a set of research studies. Such reviews undertake to summarize the type and statistical significance of impacts reported in a group of studies of related programs. Conclusions are drawn about the distribution of effects, for example, “Eight of the 19 studies showed significant effects on children.” Sometimes statements are made about differential impacts related to characteristics of the programs or the types of participants. (For example, “home visiting programs that use professional home visitors more often report significant effects on maternal behavior”).

Meta-analysis, on the other hand, is a quantitative synthesis technique rarely used in the social sciences before the 1970s.¹ In a meta-analysis, outcome results from individual studies are analyzed statistically to estimate an overall effect across studies. Meta-analysis allows us to go beyond simply counting up the numbers of studies with significant effects. Once all analysis results are transformed into a single common metric, that of the “effect size” (expressed for this study as a portion of a standard deviation unit), it is possible to combine and average data across outcomes measured on varying scales. We can also test for moderators of variation in effect sizes, such as programmatic or research design characteristics that may be related to differences in effect sizes. (For example, randomized experiments may be found to have larger effect sizes.)

In the first year of the evaluation, we reported on a narrative synthesis of more than 75 family support programs (Barnes, Goodson & Layzer, 1995). While this report was widely circulated, it did not, and could not, in our opinion, constitute the “state-of-the-art knowledge

1 Cooper, H.M., & Hedges, L.V. (1994). Research synthesis as a scientific enterprise. In F. Cooper & Hedges, L. (Eds). *The handbook of research synthesis*. New York: Russell Sage Foundation: p. 6.

base” that the original Request for a Proposal had specified. In addition, we recognized that, as we continued to gather more published and unpublished studies over the remaining four years of the study, a narrative synthesis would probably be inadequate for the task of summarizing the existing research accurately. Therefore, the decision was made, with the concurrence of the ACYF project officer, to undertake a meta-analysis of research studies and evaluations of family support programs conducted within the last 25 years. In addition to providing a more accurate assessment of the collective results of this research, the strategy has the advantage of producing a publicly-available database of studies that could, if desired, be continuously updated, and that would provide an analytic tool to address a variety of research questions.

In the next chapter we describe the steps that were taken in preparation for the meta-analysis.

Chapter A2

Preparation for the Meta-Analysis

Preparation for a meta-analysis occurs in three stages, all three of which involve decisions that have important implications for the analysis and interpretation of findings. The first of these is Formulation of the Problem, in which the scope of the issue to be studied is defined. The second stage is Searching the Literature. This includes not only determining which studies should be included in the meta-analysis but also where to look for studies. The third stage is Coding the Studies which involves developing and implementing a coding system to describe the interventions, the research designs of the impact studies, and the impacts themselves. In the discussion below, each of these stages is described, together with the issues faced at each step along the way.

Formulating the Problem

Defining the issue to be studied is the starting point for a meta-analysis. The researcher is expected to start out with a “well-built” question. Such a question specifies: (a) the main interventions under investigation; (b) the participants or subjects of interest; (c) the outcomes of interest; (d) the comparison interventions of interest; and (e) the types of studies to be included (see Counsell, 1997; Meade & Richardson, 1997; Richardson, et al., 1995; and Oxman, Sackett, & Guyatt, 1993).

Defining Family Support

The first step in this process was to develop a working definition of family support, in order to determine which programs and interventions would be included in the analysis. Family support posed a number of challenges at this point. First and foremost was the lack of consensus on what constitutes a family support program or family support services. In addition, it was clear that, under any definition, a wide array of programs could be labeled family support, including programs that differ in their approaches to working with families, in what participating families actually do in the program, and in the length and intensity of family participation.

There are at least two distinct ways of formulating a definition of family support as a basis for deciding which interventions are of interest. The first, which focuses on the types of services provided and their goals, is embodied in Public Law 103-66. This legislation defines family support as “...community-based services to promote the well-being of children and families, designed to increase the strength and stability of families, ...to increase parents’ confidence and competence in their parenting abilities, to afford children a stable and supportive family environment, and otherwise enhance child development.” (GAO, 1996)

A second kind of definition, widely used by family support advocates, is more philosophical and focuses not only on the goals of services but more particularly on the ways in which programs work with families to provide these services. In this definition, family support is seen as: *normative* -- addressing issues faced by all families with young children; *preventive* -- not designed to address specific problems or replace more intensive, professional services required by families in need of mental health treatment, remedial education or job training, or substance abuse treatment; and a *mutually respectful partnership* between family and staff that does not rely on professional diagnosis and treatment of pathology to discern individual needs. Multiple sets of practice principles have been developed by practitioners and experts in the family support field to guide programs, and to distinguish family support from other ways of working with families.

The definition of family support contained in the federal legislation would include under the rubric of family support a wide array of services and interventions. It would include “traditional” family support programs whose primary mission is enhancing parents’ capacity to support children’s development, and which provide a variety of life skills workshops, parenting classes and parent support groups, parent-child groups and family activities, information and referral to other services outside the program, and advocacy for parents. In addition, the definition would include a set of more recent programs that have a primary mission other than enhancing parent capacity but which have incorporated family support into their programs as an integral part of their services. These programs may have the primary mission of serving adults through job training or other self-sufficiency skill building, or a child-focused mission such as preventing school drop-out, or a community development mission that focuses on housing or economic development. At the same time, these programs are also concerned with building families’ capacity to support their children’s development, and the services and opportunities offered to achieve the family support goals may be identical to those found in programs where enhancing parental capacities is the primary goal.

In practice, it is not feasible to apply a practice-based definitional screen to the program descriptions contained in most research studies and evaluations since the definitions frequently vary. Often, it is not possible to apply such a screen to descriptions provided by program staff. Obviously, the definitions described above and earlier in the report were never intended to be used in this way. Rather, they were intended to serve as guides for program developers and staff, as well as, possibly, tools to evaluate the implementation of family support programs.

With these considerations in mind, we elected to use a two-stage definition of family support for the purposes of the meta-analysis. First, to select programs to be included in the study, we applied a definition of family support that encompassed all services intended to improve child outcomes by strengthening the capacity of parents to support their children’s development. Under this definition, virtually all two-generation programs were included, as well as programs that provided family support services as an adjunct to adult-centered services or

child-centered services. Operationally, this meant that the meta-analysis included research studies or evaluations of any program or intervention intended to improve child outcomes that provided services to enhance parenting capacity, either as its central or as a secondary mission.

Secondly, we rated all the programs included in the study, using a measure of comprehensiveness developed in cooperation with Sharon Lynn Kagan and other family support experts. This measure was used in conjunction with written information about program design and implementation to rate each program on how closely their practices in working with families match important practice principles in the field (e.g. the extent to which the program targets specific populations rather than offering universal services, whether participation is voluntary rather than mandated). The measure is described in more detail as part of our description of the coding process.

Given the breadth of the definition we have adopted for initial inclusion of programs, this meta-analysis is vulnerable to the criticism that we are aggregating results from too diverse a sampling of studies. The fact that the review included such a diverse set of programs, varying in types of services offered and the methods of delivering these services, raises the question of whether a summary of effects across all programs and intervention will, in fact, be meaningful. One of the criticisms of meta-analysis is that it “mixes apples and oranges.” However, as one writer has noted:

To some degree any synthesis of information from multiple research studies involves the aggregation of studies that are dissimilar. The same is true at the level of repeated observations of the same object. Some degree of mixing apples and oranges must occur in the tidiest of studies. Even when studies are intended to be direct replications, exact replication probably cannot occur.

Hall et al., 1996, p.19

The same writer notes a little later:

..if the phenomenon [of interest] is conceptually broad and therefore should be demonstrated over a wide variety of contexts, then studies that vary extensively in subjects, situations and procedures may be appropriate for inclusion.

Hall et al., 1996, p.20

We believe that family support meets this test i.e., it is “conceptually broad and should be demonstrated over a wide variety of contexts”, and that our approach is defensible. In addition, and as another way of addressing concerns about the heterogeneity of programs, we developed a coding system that allowed us to distinguish among types of programs (where

“type” was defined and redefined in different ways) and to conduct separate meta-analyses within more narrow categories.

Participant Groups of Interest

In defining the participants of interest for the meta-analysis, the only limitation we imposed is that the program or intervention should be concerned primarily with the development of children between birth and age 12 years. Once this criterion was met, the review included interventions with all other participant groups. This meant that universal or non-targeted programs were reviewed for the meta-analysis, as well as programs for at-risk families and children. “At-risk” children included children at environmental risk, children at biological risk (low birth weight, premature, etc), and children who were identified as having behavioral or emotional problems.

Outcomes of Interest

Family support programs typically have a comprehensive set of goals, both in terms of the intended targets for change -- children, adults, families and, sometimes, communities as well-- and in terms of the breadth of the goals for each . This means that the research examines a wide range of outcomes for each group. It is not unusual for a single study to measure an extensive set of outcomes. A further complication is that outcomes are measured at different time points in the life of the intervention (during the intervention, at the end of the intervention, at various follow-up points) and at different points in the lives of the children involved in the intervention.

Despite the difficulty of reviewing and summarizing data for a wide variety of outcomes, the research question of interest for the meta-analysis asks broadly about the impacts of family support, and therefore we were interested in all the outcomes reported in the research.

Comparisons of Interest

Two types of comparisons are potentially of interest for the meta-analysis. The first type of comparison tests whether outcomes for families and children who received the intervention are different at the end of the intervention from outcomes for families and children who did not receive that intervention. This comparison allows us to estimate the effect or “value added” of the intervention, and is at the heart of the meta-analysis.

The second type compares the outcomes of different intervention approaches. For instance, a study might compare the effects of a home visit intervention that uses biweekly versus weekly visits, or professional versus paraprofessional staff. Studies that compare different interventions answer a different question from the one discussed above. They ask whether, for a given population, one form of intervention has larger impacts than another form of intervention. This is an interesting question, but to address it as part of a meta-analysis

requires that a sufficiently large number of studies compare the same set of program variations.

Our decision, therefore, was to include only the first type of comparison in our analyses. The main meta-analysis will exclude treatment/treatment contrasts. At the same time, as part of the coding, we described the treatment/treatment contrasts in the research literature. If there are sufficient numbers of contrasts that test the same question, e.g., the relative effectiveness of home visiting versus center-based programming, of para-professional versus professional parent education staff, of two years versus one year of intervention, it would be possible to summarize the data on specific contrasts in a secondary meta-analysis.

Types of Studies to be Included

The question we faced here was first, whether to limit the scope of the meta-analysis to randomized studies or to include other study designs and secondly, whether **any** types of studies should be excluded from the review. Since it is generally agreed that randomized studies provide the strongest evidence about the effect of an intervention, many meta-analyses in other fields confine their search to experimental studies.

While we recognized the analytic benefits of limiting the review to studies that used experimental designs, we were concerned that, in the field of family support, many of the best-known programs and certainly many of those seen as exemplifying the principles of family support have been evaluated with one or another type of quasi-experimental design. Limiting the review to randomized studies could mean that we systematically excluded studies of several different kinds of family support programs.

After discussion of the issue with a group of experts in meta-analysis, including Dr. Thomas Cook, Dr. William Shadish, Dr Larry Hedges and Dr. Robert Fisher, we elected to include in the review both randomized and non-randomized studies but to record methodological variables that would allow us to examine the relationship between methodology and effects and to determine which specific categories of design we would exclude from the final analyses.

Searching the Literature

This step involved identifying strategies for collecting studies, conducting the searches and selecting studies.

Finding Research Studies

The task of identifying all relevant research studies on family support services was a daunting one, since so many different fields of research fall under the omnibus definition of family support and because the research literature exists in so many forms and forums. It was crucial that we conduct a search that accomplished an unbiased if not complete identification of relevant studies. As one researcher notes, the data collection methods used (i.e., how relevant studies are identified) are “of primary importance to the results obtained in a systematic review or meta-analysis...[The validity of the results of statistical analyses depends on the validity of the underlying data.]” (Dickersin, Scherer, Lebevre, 1994: p. 1286)

Exhibit A2-1 lists four ways of searching for studies that have been recommended for achieving “high” recall of documents. The goal, as described by White (1994), is to consider “*all* empirical studies on a subject--not only the published but the unpublished ones--so as to capture in the syntheses the full range of reported statistical effects.” (p. 42)

All of the methods listed in Exhibit A2-1 were employed for this meta-analysis. We searched for both published research and unpublished manuscripts. Although some researchers believe that unpublished studies should be excluded because they have not been peer-reviewed, an unmeasured influence on publication status is not assessment of scientific rigor but the nature of the results themselves. If fewer studies with negative or null results are published than studies with larger, more positive results, reviews that exclude unpublished works are likely to overestimate the relationship between an intervention and its outcomes (Dickersin & Yi, 1993; Dickersin, et al., 1987; Simes, 1987; Egger & Smith, 1995). This publication bias appears to hold true for small, nonrandomized studies in particular (Dickersin & Yi, 1993; Easterbrook, et al., 1991; Newcombe, 1987). Most researchers conducting systematic reviews believe that unpublished studies should be included, and, if necessary, the results can be analyzed with and without the unpublished data (Cook, et al., 1993).

As Exhibit A2-1 suggests, our search began, as an exhaustive literature search should, with research studies in published journal literature; these studies are the “most utilized by research synthesists” and the “most heavily represented in reference databases” (Reed and Baxter, 1994). The less well-represented research literature formats include books and book chapters, research and technical reports, and conference papers. Even more “fugitive” are the results of studies which are not yet in any formal written report but exist instead as research memos or notes.

These were found to a limited extent in some printed and computer databases, including:

- ERIC, which lists published works but also some unpublished papers and papers presented at meetings and conferences;

- The National Technical Information Service (NTIS), a database of summaries of completed research sponsored by 600+ federal agencies;
- Social and Behavioral Science Documents (SBSD), published by the American Psychological Association, which contains abstracts of technical papers, reports and bibliographies;
- Social Sciences Literature Information System (SOLIS), listing monographs, reports and conference proceedings in the social sciences and humanities (since 1976); and
- Federal Research in Progress (FEDRIP), which provides access to information about current and ongoing federally funded research.

Unpublished reports were also unearthed through the methods listed in Exhibit A2-1 under “Consultation.” Another rich source of literature was dissertations, which were searched through databases such as *Dissertations Abstracts International*, *Master Abstracts*, and *Research Abstracts*. In conducting the searches, the only criteria we imposed were: (1) the research was conducted after 1965; and (2) the research was conducted in the United States, Canada or Great Britain.

Selecting Research Studies and Programs for the Meta-Analysis Database

Out of the more than 900 research reports that were collected and reviewed,¹ we selected for coding 665 studies, representing 260 programs. In selecting which of the 900 studies to code, we took the approach that every relevant research study, where relevance was defined as involving an intervention, outcomes, and participants of interest, would be coded regardless of its design. All of the research studies in the data base compare two groups of participants who receive different kinds or levels of family support services. This includes: (a) studies that compare one group of families who receive the treatment (i.e., a specific set of family support services) with another group of families who do not receive this treatment; and (b) studies that compare a group of families who receive one set of family support services with another group of families who receive a different set of family support services.

¹ In addition, we obtained over 60 research reviews that summarize multiple studies on early childhood education (25 reviews), family support and parenting education (9 reviews), parent involvement (3 reviews), two generation programs (2 reviews), home visiting programs (4 reviews), Head Start (12 reviews), and child mental health programs (6 reviews).

Recommended Strategies for Searching the Literature

Footnote Chasing^a

References in journals from nonreview papers
References from nonreview papers not published in journals
References in review papers written by others
References in books written by others
Topical bibliographies compiled by others

Consultation

Informal conversations with colleagues
Attending meetings and conferences
Communication with people who typically share information with you
Formal requests of scholars who are active in the field
Comments from readers/reviewers of past work
General requests to government agencies
Reviewing electronic networks

Searches in Subject Indexes

Manual search of abstract databases (e.g., Child Development Abstracts and Bibliography)
Computer search of abstract databases (e.g., ERIC, Psychological Abstracts)
Manual search of proceedings from relevant research conferences

Citation Searches^b

Manual search of citation index
Computer search of citation index

Browsing

Browsing through library shelves
Browsing through colleagues' shelves

a Defined by White (1994) as "the adroit use of other authors' references to the prior literature on a topic."
b Following an author's citations, forward and backward in time; examples include the *Social Sciences Citation Index*.

SOURCE: Adapted from White (1994), in turn adapted from Cooper (1985) and Wilson (1992)

In addition, we coded information on 167 family support programs for which there was descriptive literature but on which no outcome studies had been conducted (or which only provided data on participant satisfaction at the end of the program). We coded information on these programs to ensure that our descriptions of family support programs encompassed a broad spectrum of programs.

The result of these decisions about which studies to code was a comprehensive database of studies that would allow us to describe a wide variety of programs. A list of the programs in the database and the research studies associated with them is provided in Appendix A.

Coding the Studies

This project has two purposes: to describe family support programs, in order to characterize the field as a whole; and to summarize the research on the effects of family support programs using meta-analytic techniques. The coding system developed for the meta-analysis accommodates both of these needs by means of a hierarchical coding system. Exhibit A2-2 shows the five levels of the coding system and describes the information that is coded at each level. At the “lowest” level of the coding system is the individual finding (contrast), which is the datum on which the meta-analysis is conducted. Higher levels of the coding system capture the study characteristics and the characteristics of the intervention, which may be used to explain differences in effects or to characterize the field of family support programs and research.

At the base of this coding system is a *specific contrast*: numerical data describing the level of performance of two conditions (groups) on a specific measure at a specific time point. For example, a contrast might be the posttest means for a specific treatment and control group in a particular study on a child measure administered when the children were 12 months of age. The hierarchical coding system means that each contrast is linked to a large amount of information on (1) the measure itself, (2) the two groups being compared, (3) the study and its methodological characteristics, and (4) the program model being evaluated.

The actual coding scheme was developed by researchers familiar with the research domain of family support. In developing the coding categories, we also consulted with experts in the area of research synthesis methods, who advised us about which study characteristics should be coded. A copy of the coding form and instructions can be found in Appendix B.

Exhibit A2-2

Levels of the Coding System for the Meta-Analysis of Family Support

- | |
|---|
| <ul style="list-style-type: none">I. Program/Model: goals, services, target groups, length, delivery mode(s), staff qualifications<ul style="list-style-type: none">A. Study: source of report, study setting, overall design<ul style="list-style-type: none">1. Contrast Conditions: details of sampling and attrition, presence or absence of threats to internal validity, description of family support services received by treatment condition<ul style="list-style-type: none">a. Dependent Measure: name of measure, type of measure (adult vs child, self-report/observation/physiological measure), reliability of measure, level of performance of two conditions at pretest (data used to compute pretest effect size)<ul style="list-style-type: none">(1) Time Within Dependent Measure: sequence in administration of measure (1st, 2nd, 3rd, etc), age of child at administration, time elapsed since beginning of intervention, level of performance of two conditions at this administration of the measure (data used to compute posttest effect size) |
|---|

Computing Effect Sizes

In the research studies that we coded, the numerical data on which a contrast is based are presented in many different ways--as means, proportions, t-statistics, probability levels, etc. In our coding system, the data were recorded in the form that they were presented and were then converted into an effect size using specialized software developed expressly for this purpose by Dr. William Shadish. This software takes data expressed in 40 different forms and computes a standardized effect size using the appropriate statistical method. Appendix C briefly describes the forty methods handled by the software.

The Coders and the Coding Process

There were two separate coding activities. The first was the coding of the research articles in terms of the intervention itself, the study methodology, and the findings. The second activity was to rate each program in terms of its adherence to the family support practice principles.

The research articles were coded by a small group of senior staff. Each coder was trained individually or in a small group by the senior researcher who developed coding protocol. The training took approximately six hours, and, at the end of the training, coders were required to complete a practice protocol that was checked for accuracy against a master coding.

After the training, all articles were double-coded. Two coders completed a coding protocol independently, and then the two protocols were compared. Differences were discussed and, if necessary, given to the trainer for adjudication. An agreed-upon protocol was then entered into an Access database, to be converted into a SAS database for analysis. The exception is the outcome data (i.e., the actual numerical means, etc), which were entered separately into the program that computed the effect sizes.

Rating Adherence to Family Support Practice Principles

Each program or intervention studied was rated in terms of its adherence to family support principles. This rating was based on all available written information about the program. The rating system itself was developed in collaboration with Dr. Sharon Lynn Kagan of Yale University. The rating scale is discussed in Chapter A3. The actual rating of the programs in the meta-analysis was done by a small number of senior staff who are knowledgeable about family support programs and principles.

Chapter A3

Characteristics of Programs and Studies Included in the Meta-Analysis

The database constructed for the study contains information on 427 programs, all of which met our initial criteria for inclusion in the study. Of these, 260 programs had at least one evaluation study associated with the program; the remaining 167 had not been evaluated. Only the 260 programs that have evaluations are included in the meta-analysis, but we felt that it was important to code and include in the database programs that had not been evaluated. The inclusion of unevaluated programs in the larger database allowed us to represent a very wide range of program types and to compare the characteristics of programs that were and were not evaluated. The existence of systematic differences between the two groups could constrain the extent to which the findings of the meta-analysis might be expected to hold true for all types of family support program, including those not previously evaluated.

In this chapter, we first describe the 260 programs or interventions represented in the meta-analysis. Secondly, we address the question of the representativeness of these programs by comparing them with the 167 family support programs that were **not** evaluated. Finally, we describe the characteristics of the *studies* included in the analysis and compare them with studies that were excluded.

Program Characteristics

These descriptors represent characteristics of the programs or interventions themselves. In deciding on the aspects of programs about which we needed to code information, we chose those that met at least one of the following criteria: policy significance in the current debate about effective ways to work with families; relationship to program impacts demonstrated in previous research; or usefulness in describing differences across programs.

The program variables selected can be grouped into ten categories: program goals and purpose; methods of delivering services to children and families; type and qualifications of program staff ; where services are delivered; whether the program is a research or demonstration program or a community-based program; whether the program targets a specific population and who is targeted; age of children toward whom the program or intervention is directed; the intended length of the program or intervention; the services offered; and the extent to which the program is rated as “family supportive“.

Program Goals and Focus

From the written program descriptions, we extracted the central stated purpose or goals of the program. As we might expect of family support programs, it was unusual for programs to have a single goal; most often their goals reflected their two-generation focus. Almost every program or intervention asserted the twin goals of improved parenting (98%) and enhanced child development (91%). Less than 10 percent of the programs had as a goal increasing family self-sufficiency, either by increasing parents' educational or literacy skills or through job training or other employment assistance. Other kinds of assistance to parents, such as providing social support or developing self-help and leadership skills, were more likely to be articulated (22 percent and 12 percent of programs, respectively). Providing health care to children or to families was a goal for 17 percent of programs. Preventing child abuse was an explicit goal of only 14 percent of programs or interventions (Exhibit A3.1).

Exhibit A3.1

Program Goals (n=260 programs)

Goal	Programs with Goal*	
	n	%
Improved parenting	254	98
Child development	237	91
Social support for parents	57	22
Child/Family health care	43	17
Child abuse prevention	35	14
Parent self-help, empowerment	30	12
Parent literacy, employment	22	8
Parent community/school involvement	6	2
Child behavioral change	6	2

* Written materials often cited multiple goals.

Methods of Delivering Services

Programs can deliver services in a variety of ways, including visits to the home, meetings or classes at the program or at other locations, or through written information. Almost half of the programs (49%) used home visits as the primary mode of service delivery, and another 12 percent used home visits to deliver some services. More than half (59%) brought parents together in groups or for more formal classes at the program or another location and more

than a quarter (28%) provided joint activities for parents and children in a group setting. Only 18 percent provided early childhood education services in a group setting (Exhibit A3.2).

Exhibit A3.2

**Mode of Service Delivery
(n=260 programs)**

Service Delivery Mode	Primary Mode % of programs	Secondary Mode % of programs
Home visits	49	12
Parent meetings/classes/groups	45	14
Parent-child classes/groups	18	10
Group early education for children	13	5

Staff

The coding scheme for the meta-analysis allowed us to capture information on the staff who delivered services. For staff who worked with parents or families and for those who worked directly with children, we first determined their professional status (i.e. whether they were professionals, with both formal education and training, non-professionals, with a college education but no formal training, or paraprofessionals, without a degree or training before they were hired.) If professional staff were used, we determined the type, (e.g., medical, social worker, certified teacher, child development specialist).

The majority of programs (87%) used professional staff to deliver services to parents (or to families). Almost one-third used paraprofessional staff, sometimes in combination with professionals. Few programs (9%) used staff in the intermediate category, that is, staff with a college education but no specialized training (Exhibit A3.3). Less than 30 percent of the programs employed any staff to work directly and only with children. The majority of these programs, 70 out of 76, used professionals, sometimes supported by paraprofessional staff.

More than one-quarter (28%) of the programs used social workers or trained counselors. Medical personnel, who might be doctors, nurses, physician assistants or physical therapists, were the next most frequently used professional staff (22% of programs). Only 17 programs, seven percent of the sample, had trained child development specialists on staff (Exhibit A3.4).

Exhibit A3.3**Programs' Use of Trained Staff
(n=260 programs)**

Professional Status of Staff	% of Programs
Professional (degree and formal training)	87
Non-professional (degree, no formal training)	9
Paraprofessional (without training or degree)	30

Exhibit A3.4**Types of Professional Staff Used by Programs
(n=260 programs)**

Staff Type	n of programs	% of programs
Medical (doctor, nurse, physical therapist)	57	22
Social worker, counselor	72	28
Certified school teacher	50	19
Medical mix	39	15
Non-medical mix	25	10
Child development specialist	17	7
Unspecified	38	15

Location of Program Services

Each program or intervention was coded as being based (i.e., services were delivered) **primarily** in the home, in a school, university or other institution of higher education, hospital or clinic, community center, or public or private social service agency. More than half of the programs delivered services in a single location; the remainder combined delivery of services at the program location with visits to the home.

In a majority of programs (62%), some or all of the services were provided in the family's home. Hospitals or clinics and schools were the next most frequent locations for services (29% and 25% respectively). Small numbers of programs provided services in other community locations (Exhibit A3.5).

Exhibit 3.5**Primary Location of Program Services
(n=260 programs)**

Location	n of programs	% of programs
Family's home	162	62
Hospital or clinic	75	29
School	65	25
Community center	42	16
University or college	22	9
Public agency	7	3
Private agency	7	3
Other	16	6

Type of Program

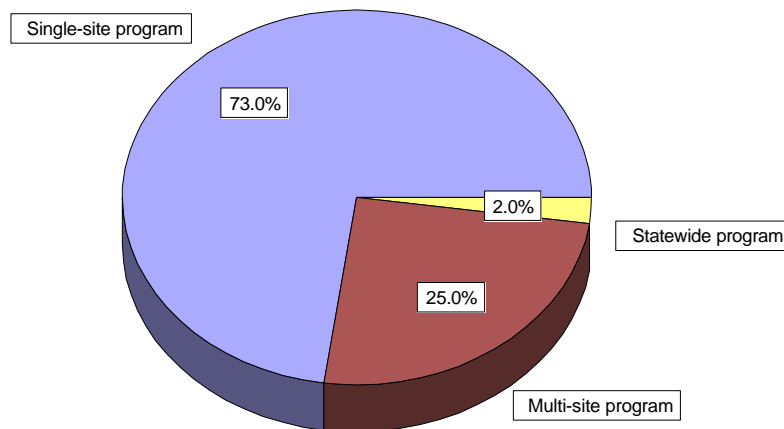
We were interested in coding whether a program was identified in the study description as a research or demonstration program. Although they were not always explicitly categorized as such, we used the information provided in the study to determine whether the program was set up with a concurrent evaluation to test a strategy or approach and operated for a defined period. Many of these were single-site programs; others, like the Child and Family Resource Program were replicated concurrently in several sites, or like the Nurse Home Visiting Program had two successive single-site replications.¹ Our interest in this distinction stemmed from evidence from earlier research, in the field of early childhood education for example, that suggests that the effects of some research-driven interventions are stronger than the effects of subsequent programs generated at the local level that attempt to replicate all or some of the features of the original intervention.

More than 70 percent of the programs included in the meta-analysis were classified as research or demonstration programs. To the best of our ability to determine, they had been funded and set up to test or demonstrate a strategy or intervention, and an evaluation was an integral part of their operation from the outset. Most ended at approximately the same time

1 We recognize that Dr. Olds' program has moved beyond the research and demonstration phase into multi-site and even state-wide replication. In the meta-analysis, however, the program is represented by studies of the original program and its two single-site replications. Since all the subsequent replications of the program must incorporate some level of research activity, in later versions of the meta-analysis, they would probably be categorized differently.

as the evaluation. While all remaining programs had been studied at least once, they were not set up for research purposes and had no defined period of operation. The majority of programs (73 %) operated in a single site. (This included most of the research and demonstration programs.) A quarter of the programs were in multiple states, and a small number of programs (2 %) were statewide multi-site programs.

Exhibit A3.6**Types of Programs in the Meta-Analysis
(n=260 programs)**

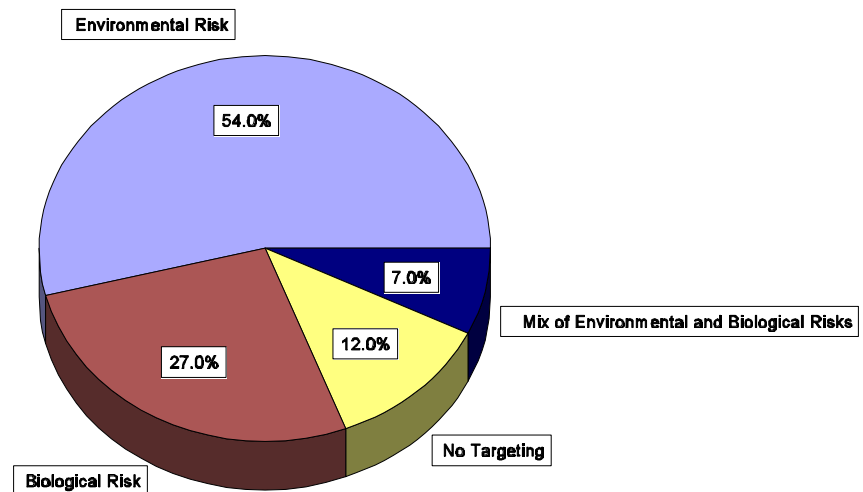
**Targeting of Specific Populations**

While many of the original family support programs were neighborhood entities that served any family in the neighborhood, their underlying principles and practices have been applied in programs that provide family support services to specific populations. We categorized programs as targeted if there was evidence that the program was designed for a specific population and if potential participants were screened for entry into the program or for receipt of services. If programs were categorized as “targeted”, the participation criteria were coded as either “environmental risk” or “biological risk”, the assumed risk being to the child’s healthy development.

Environmental risk factors include: family poverty; welfare dependency; risk of or instance of child abuse or neglect; teen parenthood; maternal depression or isolation; incarcerated parent; and recent immigrant status, among others. Biological risk factors include: low-birth weight; physical disability; developmental delay; and behavior problems, among others.

As Exhibit A3.7 shows, only 31 programs or 12 percent of the sample did not explicitly target their services to a specific population; more than half (55%) targeted families who faced one or more environmental risks. Families with a child at biological risk were targeted by 27 percent of the programs, and a few programs directed services to families with both environmental and biological risk factors (an example might be low-income families with a low-birth weight baby).

Exhibit A3.7**Program Targeting
(n=260 programs)**

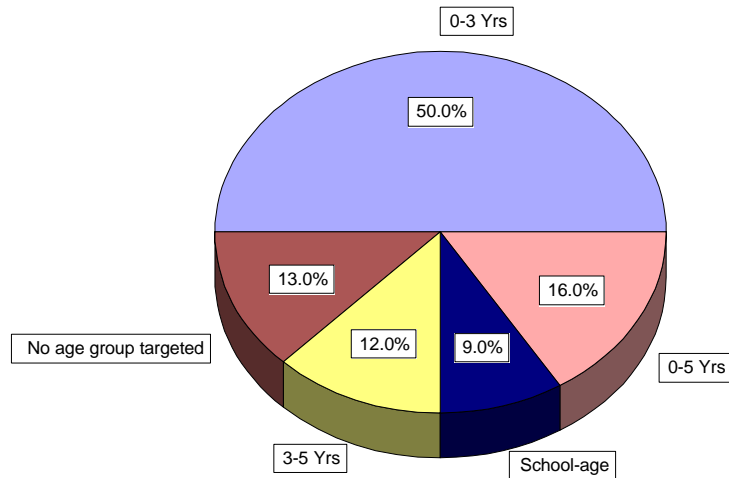


Of those programs that targeted families with environmental risks, a majority (63%) directed services to low-income or welfare populations. Families at risk for abuse and neglect were targeted by 22 percent of programs; teen parents were targeted by close to a quarter of programs. Developmental delay and low birth weight were the most frequently targeted biological risk factors (20% and 16% of programs, respectively).

Ages of Children Targeted

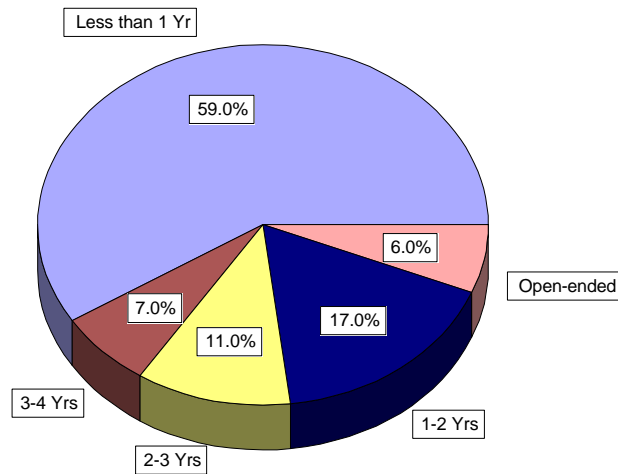
Most of the programs in the sample targeted children within a specific age-range. Half targeted children from birth (or before birth) to three years of age. Another 16 percent targeted children from birth to five years of age. Only 13 percent of the programs did not target a specific age group but provided services to children of all ages (Exhibit A3.8).

Exhibit A3.8**Ages of Children Targeted
(n=260 programs)**

**Intended Length of Program**

Programs that had specified service periods provided services for an average of fifteen months. More than half of the programs or interventions (59%) were designed to provide services for less than a year. Of these, most were designed to provide services for six months or less. A very small proportion (6%) of programs did not specify a treatment period but were prepared to provide services to families for as long as the family wanted them or until the family reached a specified goal (Exhibit A3.9).

Exhibit A3.9**Intended Length of Program or Treatment
(n=260 programs)**

**Types of Services Offered**

Almost every program (98%) offered some form of parenting education, provided in a variety of ways. Most often provided through home visits, information about parenting strategies, children's health and development and age-appropriate activities was also offered through classes for parents, classes for parents and their children, parent groups and printed materials. About half the programs organized parent groups and other activities to provide social support and reduce isolation. Almost half of the programs (44%) provided case management or counseling services or referral to social or health services. Just over one-quarter (29%) offered health care services for parents and children. Only 10 percent of programs offered adult education classes or other kinds of training aimed at improving parents' economic status (Exhibit A3.10).

Exhibit A3.10**Types of Services Offered by Programs
(n=260 programs)**

Types of Service	% of Programs Offering Service
Parenting education	98
Social support	51
Case management/counseling/referral to services	44
Support services (transportation, respite care, play groups, toy lending)	36
Health care for parents and children	28
Center-based ECE	19
Leadership/advocacy training	16
Adult basic education/job skills	10

Amount of Key Services

Programs provided, on average, a total of 60 hours of parenting education to families. However, the average amount masks wide variation among programs. As Exhibit A3.11 shows, about a third of the programs provided less than 20 hours of parenting education, another third provided between 20 and 40 hours, and the final third provided more than 40 hours. Since 40 hours is roughly equivalent to half the number of hours in a typical community college course, most programs provided a relatively weak level of this service.

Approximately 15 percent of the programs provided case management services to families. In these programs, families received, on average, less than 10 hours of case management services, with little variation across programs (Exhibit A3.12).

About one-quarter of the programs (59 programs) provided early childhood education services. In the majority of these programs (63%), children received less than 400 hours of early childhood education, or the equivalent of a day a week for a year. Another 15 percent provided between 400 and 1000 hours of early childhood education, while the remainder provided more than 1000 hours (Exhibit A3.13).

Intensity of Services

The length of the program and the amount of service provided give us some measure of the level of “treatment” or “dosage” that families received, that may be helpful in understanding

Exhibit A3.11

**Amount of Parent Education
(n=260 programs)**

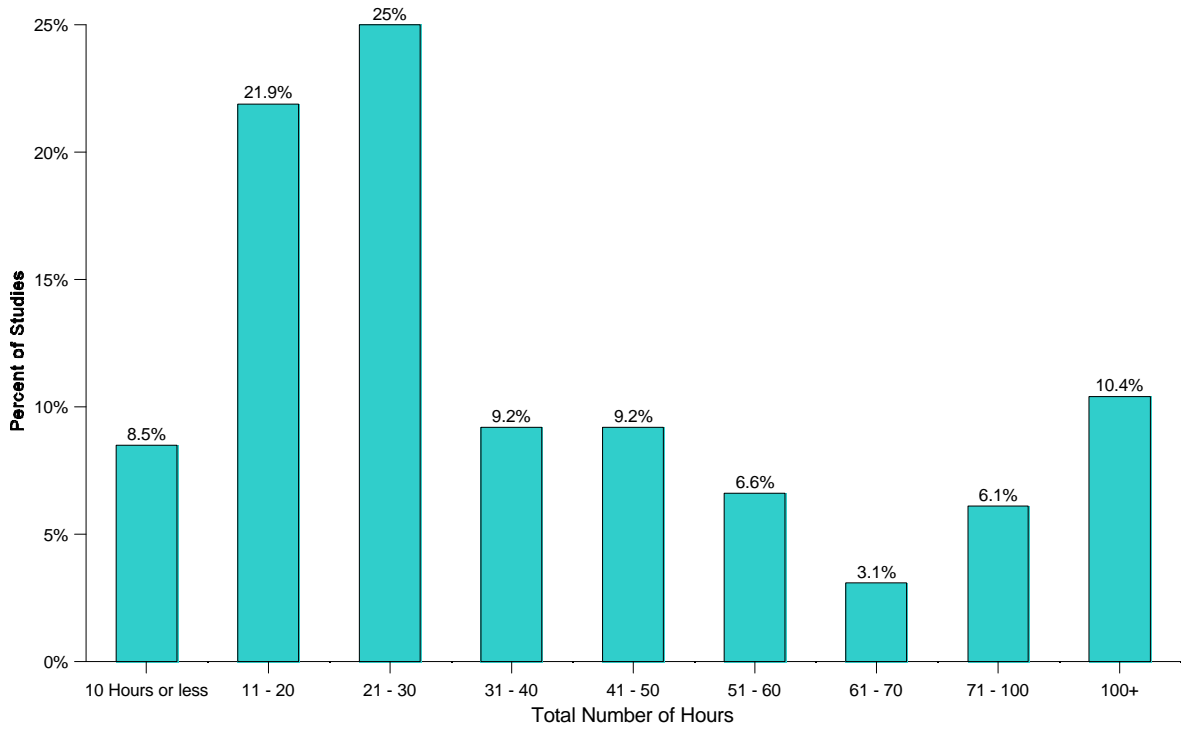


Exhibit A3.12

**Amount of Case Management
(n=37 programs providing case management)**

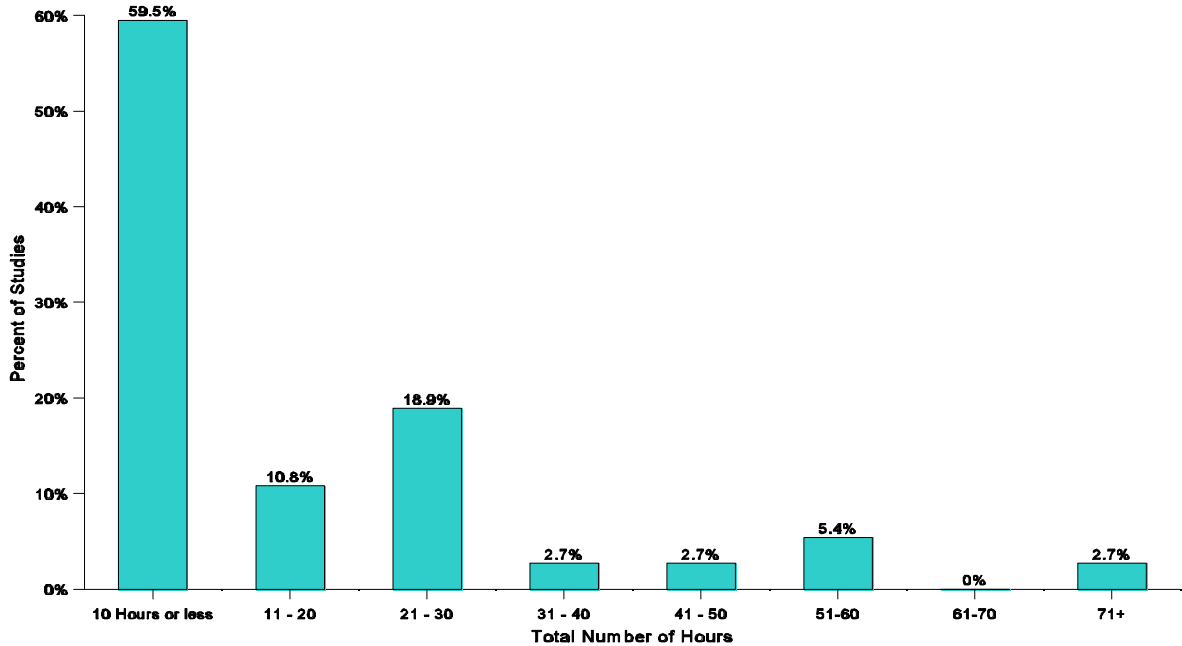
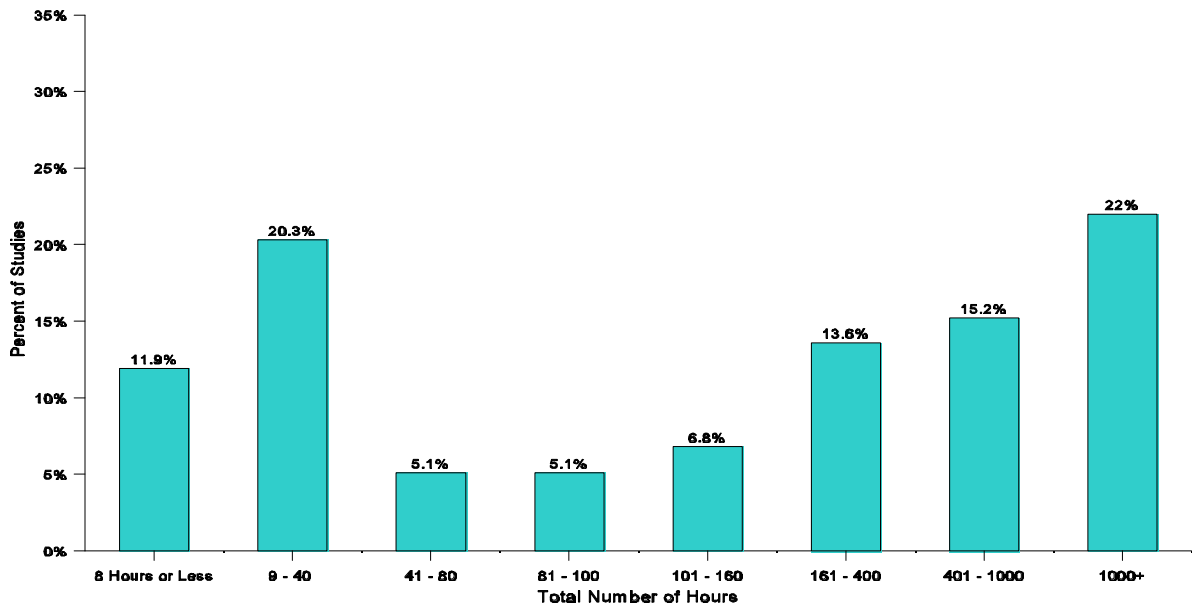


Exhibit A3.13**Amount of Early Childhood Education
(n=59 programs)**

findings about program or service outcomes. Another aspect of “dosage” is the intensity with which services are delivered, assuming, for example, that the experience of 400 hours of a service spread over three years may have a different effect from the same number of hours concentrated in a shorter time period. For this description, the intensity of services was computed as the number of hours of a service divided by the number of months over which the service was provided.

In the case of parenting education, about half the programs (51%) provided less than five hours a month of the service. Just over one-third (36%) provided between 5 and 15 hours of services and just over 10 percent provided more than 15 hours of services a month. (Exhibit A3.14).

Because the absolute amount of case management services was low, the intensity of the service was also low -- two-thirds of the programs provided one hour or less of care management services to families each month (Exhibit A3.15).

For early education services, almost 40 percent of programs provided less than half a day a month of services. At the other end of the spectrum were approximately 20 percent of programs that provided the equivalent of half-time to full-time early childhood education (Exhibit A3.16).

Exhibit 3.14

Intensity of Parent Education
(n=260 programs)

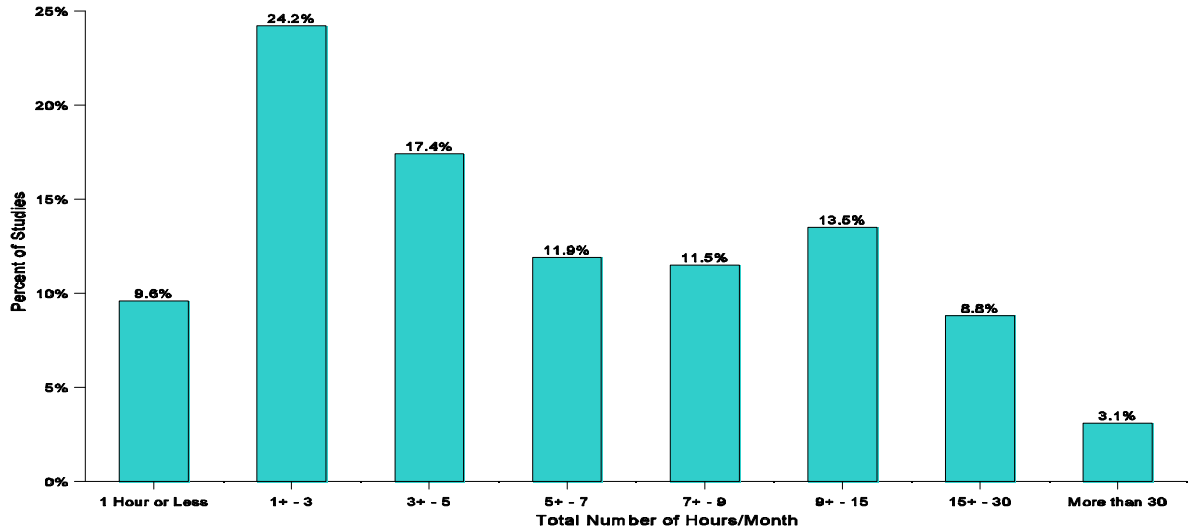


Exhibit A3.15

Intensity of Case Management
(n=37 programs)

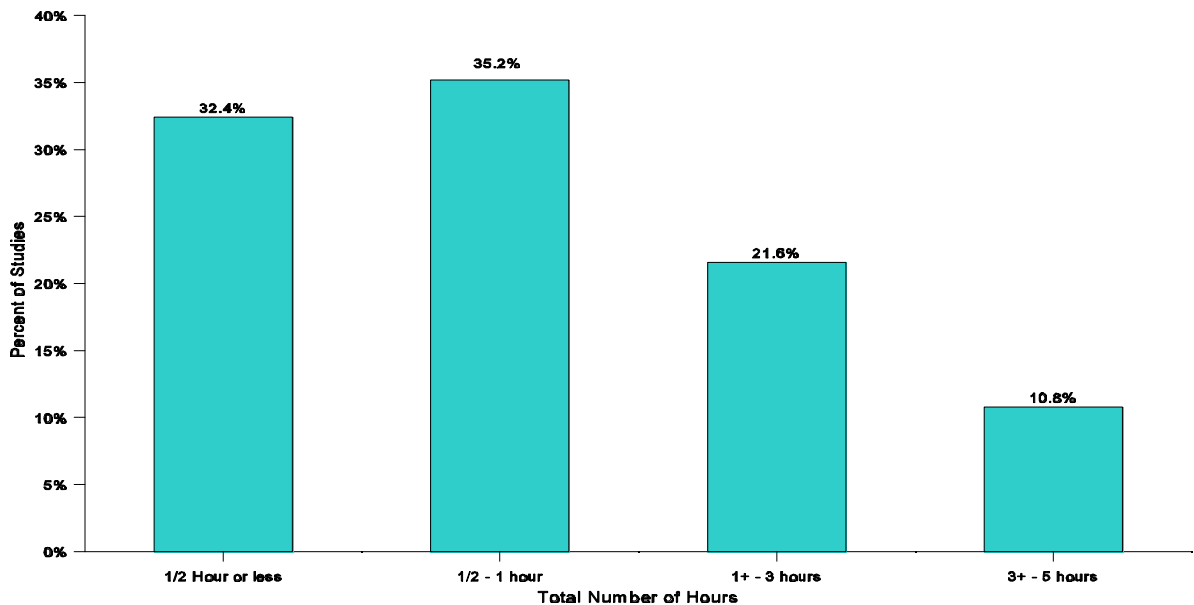
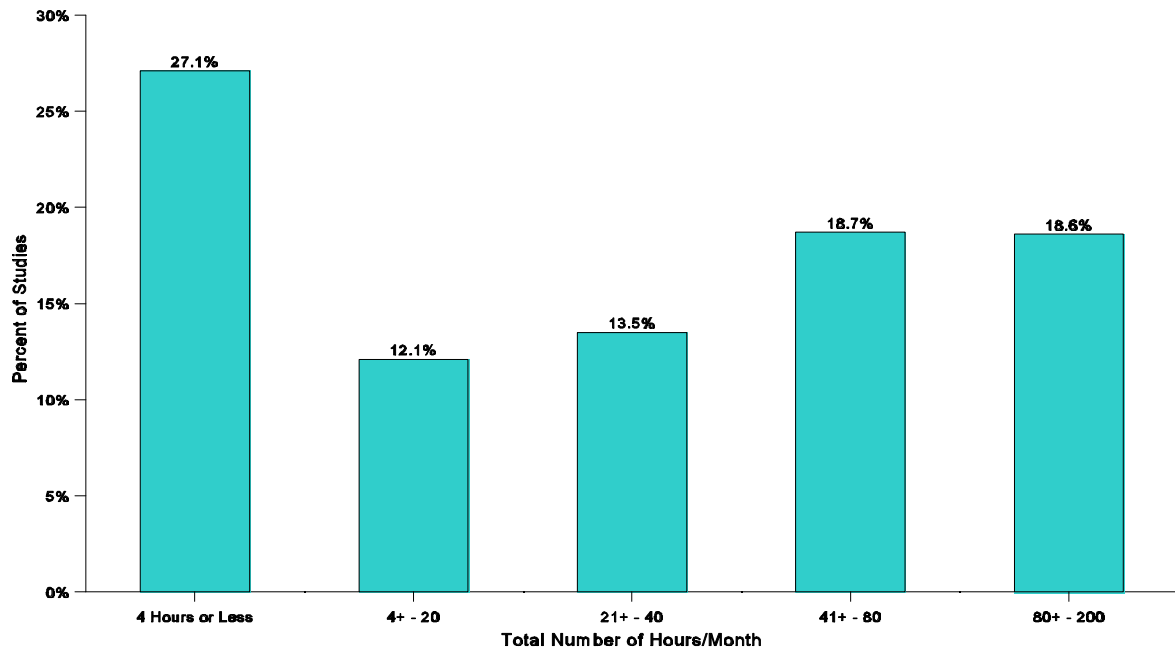


Exhibit A3.16

Intensity of Early Childhood Education (n=59 studies)



Family Supportiveness of the Program

All of the programs included in the study met the definition of family support services provided in the federal legislation. However, we recognized that they were not all equal in the extent to which they embodied the principles of family support. Each study in the sample was rated on its “family supportiveness.” On the basis of work done by the Family Resource Coalition to develop a definition of family support, we developed a rating system that scores an intervention or program on seven dimensions. These seven dimensions do not comprise all of the features that, in the FRC definition, characterize family support programs. The seven were selected because they were features that we believed could be coded from the kind of program descriptions typically provided in research reports. The dimensions are as follows:

- eligibility for the program;
- presence of services to promote child development;
- presence of services to promote adult development;
- services focused on parent/child interaction;
- services/activities that help build social support among parents;

- link to community resources (referral only or with active involvement with service providers in implementing service plan); and
- community advocacy.

Exhibit A3.17 displays the criteria and the coding rules for rating the family supportiveness of a program.

Exhibit 3.17

Definition and Scoring Rules for Features for Family Supportiveness Score

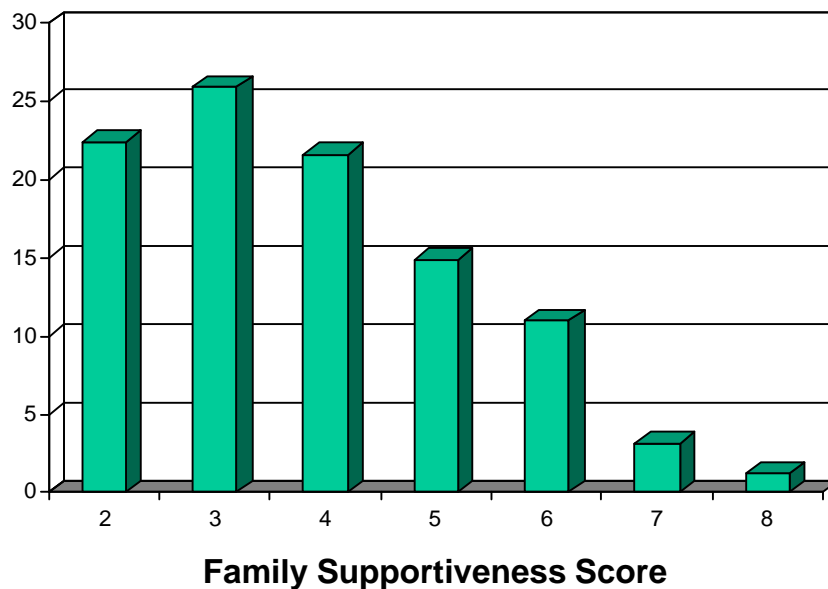
Definitions of Features	Rating
1. Eligibility for the program is not based on identified problems or deficits, and the program is not designed to remediate specific problems. <i>(For this definition, child age, ethnicity, geography are not counted as problems or risk factors).</i>	No targeted eligibility: yes = 1, no = 0
2. The program includes a component that focuses on the child's development.	Has child component: yes = 1, no = 0
3. The program includes a component that focuses on the development of the adult parent, including emotional, educational, economic development, life skills.	Has adult component: yes = 1, no = 0
4. The program includes a component that focuses on the development of the parent-child relationship.	Has parent-child component: yes = 1, no = 0
5. The program provides opportunities for peer support and builds informal networks of friends and neighbors.	Provides opportunities: yes = 1, no = 0
6. The program connects families to other resources in the community.	Connects families: 2 = referrals + interagency cooperation 1 = referrals only 0 = no connection with community resources
7. The program advocates for improved/increased/more accessible services for families in the community.	Community advocacy: yes = 1, no = 0

The total score for the measure was obtained by summing the scores of the seven individual items. Across the sample of 260 programs, the average score for family supportiveness was 3.8 out a possible score of 8.0. The standard deviation for the score was 1.48, indicating that 96 percent of the programs had scores between 2 and 6. Half of the programs had total scores of 4.0 or greater. Almost all programs included both a child development component and a component that focused on the development of the parent-child relationship. Only six

percent of programs were scored positively on the item that measured advocacy for services in the community. Exhibit A3.18 shows the distribution of scores for the 260 programs.

Exhibit A3.18

Family Supportiveness Rating of Program (n=260 programs)



Comparing Evaluated With Unevaluated Programs

As we noted earlier, information on program characteristics was coded for both the 260 programs with one or more evaluations associated with them and for 127 programs that met our criteria for inclusion, but had not been evaluated. We compared evaluated and unevaluated programs on the program characteristics described above: goals and services, targeting, service delivery methods, types of staff, length of program. Overall, the two groups of programs were significantly different on this set of characteristics (as shown in a discriminant function analysis). Specifically, although evaluated and unevaluated programs were similar with respect to their goals and types of services provided, compared with the unevaluated programs, evaluated programs:

- were more likely to target their services to specific populations;
- were more likely to use home visits as the primary mode of service delivery;
- were less likely to use center-based early childhood education as a primary mode of service delivery; and

- were less likely to use paraprofessionals or non-professionals to provide parenting education.

As Exhibit A3.19 shows, the two groups of programs had similar goals. Improved parenting and enhanced child development were goals for almost all programs. One difference is that unevaluated programs were more likely to have increased community involvement as a goal for parents (19% vs. 2% for evaluated programs). The mix of program services offered by the two groups of programs was generally similar. Somewhat larger percentages of unevaluated programs provided parent social support and leadership training services, as well as center-based early childhood education.

Only 12 percent of evaluated programs did not target services to a specific population compared with 53 percent of unevaluated programs (Exhibit A3.20). Evaluated programs were more likely to direct services to families with environmental risks or with children at biological risk, and to target specific age groups of children.

Evaluated programs made more use of home visiting as a primary mode of service delivery (49% vs. 30%), and less use of center-based early childhood education (18% vs. 32%), as Exhibit A3.21 shows. All programs had staff who provided parenting education, but Exhibit A3.22 shows that evaluated programs were more likely to use professional staff to provide this service. Evaluated programs were somewhat more likely to provide services in the home (62% vs. 50%) and less likely to base services in a school (25% vs. 43%) or community agency (16% vs. 50%) (Exhibit A3.23).

The differences between the two groups of programs have implications for how we will interpret the findings reported in Chapter 5. While we may be able to generalize findings about the effects of specific *service strategies*, *staffing patterns* or *targeting of services* to any programs that provide these services, we need to be more guarded in the statements we make about family support programs as a group, recognizing that the programs included in the meta-analysis represent a substantial part, but not the universe of family support programs.

Exhibit A3.19

Primary Goals of Evaluated Versus Unevaluated Programs

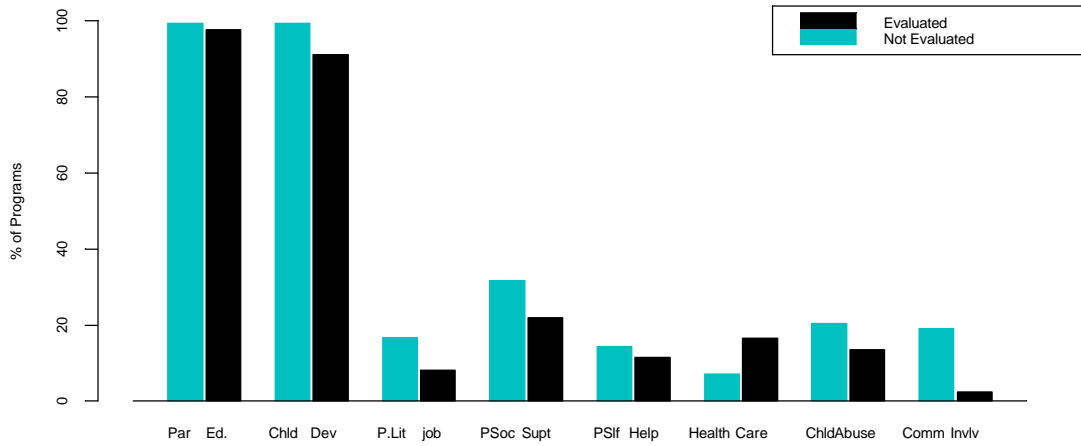


Exhibit A3.20

Targeting of Program or Intervention

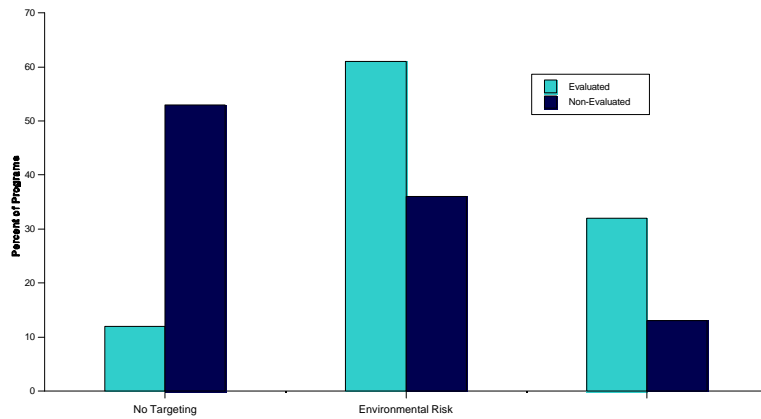


Exhibit A3.21

Service Delivery: Primary Mode

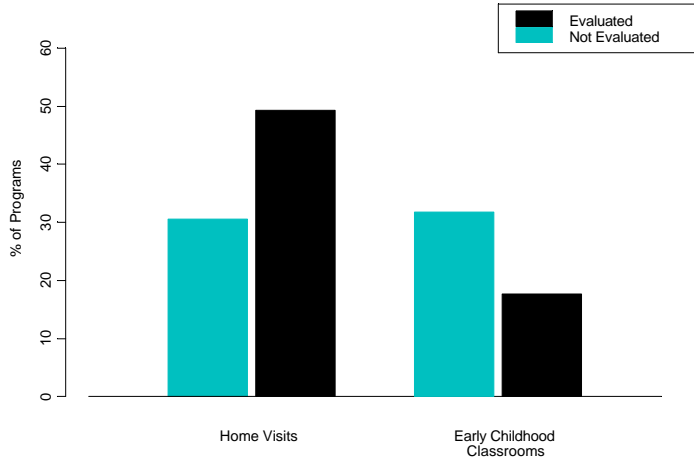


Exhibit A3.22

Staffing

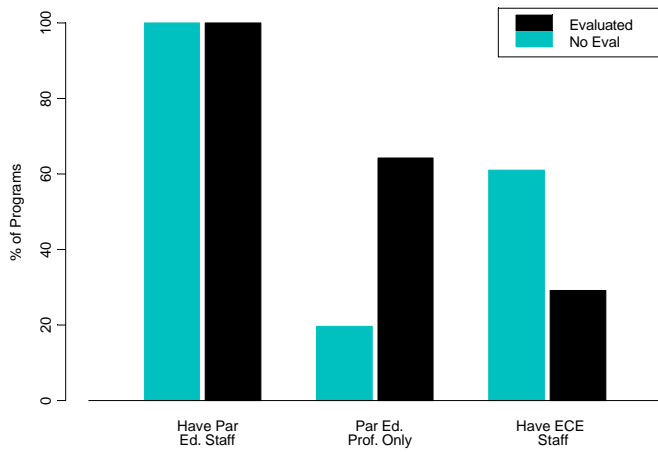
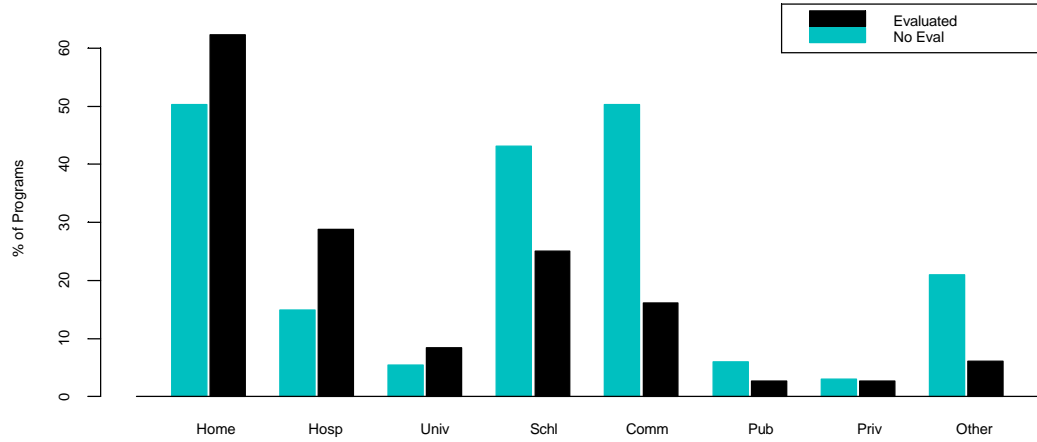


Exhibit A3.23

Location of Program Services



Chapter A-4

Analytic Approach

The meta-analysis addressed three central questions:

What is the impact (i.e., average effect size) of family support services on selected child and adult outcomes?

Are there program or treatment characteristics that are significantly related to the impacts (i.e., average effect sizes) of family support programs on child and adult outcomes?

What aspects of family support programs or services account for variation in the impacts of family support programs on child and adult outcomes?

This chapter sets forth our approach to answering these questions. The discussion begins by describing the analytic database used to conduct these analyses.

Constructing the Analytic Database

The analytic database for addressing the three research questions included only a subset of the studies on the full database. The final analytic database was constructed in three steps, moving from the full set of 665 studies and the 260 programs associated with 12,486 effect sizes, to two analytic databases: (1) a database for the **end-of-treatment** outcome data, which includes 5,681 effect sizes from 351 randomized or quasi-experimental studies of 191 programs; and (2) a database for the **follow-up** outcome data, which includes 2,224 effect sizes from 158 randomized or quasi-experimental studies of 87 programs.

In the first step, we excluded two kinds of studies from the full database. First, we eliminated studies that compared two different family support treatments. Second, we eliminated studies in which we did not have a true pretest to use in constructing an effect size. Eliminating these two types of studies reduced the database to 562 studies representing 255 programs. The data from these 562 research studies comprise a database of 11,112 effect sizes, in which each effect size measures, for a given study, the difference between two groups on a specific outcome at a particular point in time. The fact that most research studies of family support programs report data on multiple outcomes and at multiple time points accounts for the large number of effect sizes in the database. Effect sizes are grouped within studies, and, in this database, each study contributed, on average, 20 effect sizes, which represent different outcome measures, different time points, or both.

In constructing the database, we grouped the effect sizes within outcome domains as well as within studies. The effect sizes were grouped into nine broad outcome domains: four child domains--cognitive development/school performance, social-emotional development, health/physical development, and child safety; and five parent domains--parenting knowledge, parenting behavior, family functioning, adult mental health/health risks, and family economic self-sufficiency. Each of these nine outcome domains comprises a range of more fine-grained outcomes, listed in Exhibit A4.1. We then grouped effect sizes into three time points, reflecting when the outcome was measured: during the treatment period (interim); at the end of the treatment period; or at some point after the end of the treatment period (follow-up).

Exhibit A4.2 shows the distribution of effect sizes in the database, by outcome domain and timing of the outcome measure. As the exhibit shows, the data base contained more parent outcomes than child outcomes. Of the child outcomes, the domain of social-emotional behavior was most frequently assessed; of the adult outcomes, parenting attitudes and knowledge were most often measured -- nearly three times as often as actual parent behavior. Subsequent analyses to determine the effects of family support programs on children and families used the nine broad outcome domains, rather than the finer-grained sub-domains. This was both for practical purposes--to have sufficient data within a domain--and for policy purposes, since policy questions about outcomes for children and families tend to be enunciated in these same broad categories.¹ The exhibit also shows that the end-of-treatment outcomes were the most common in each of the outcome domains although, in the case of child cognitive outcomes, there were nearly as many follow-up effect sizes as there were end-of-treatment effect sizes.

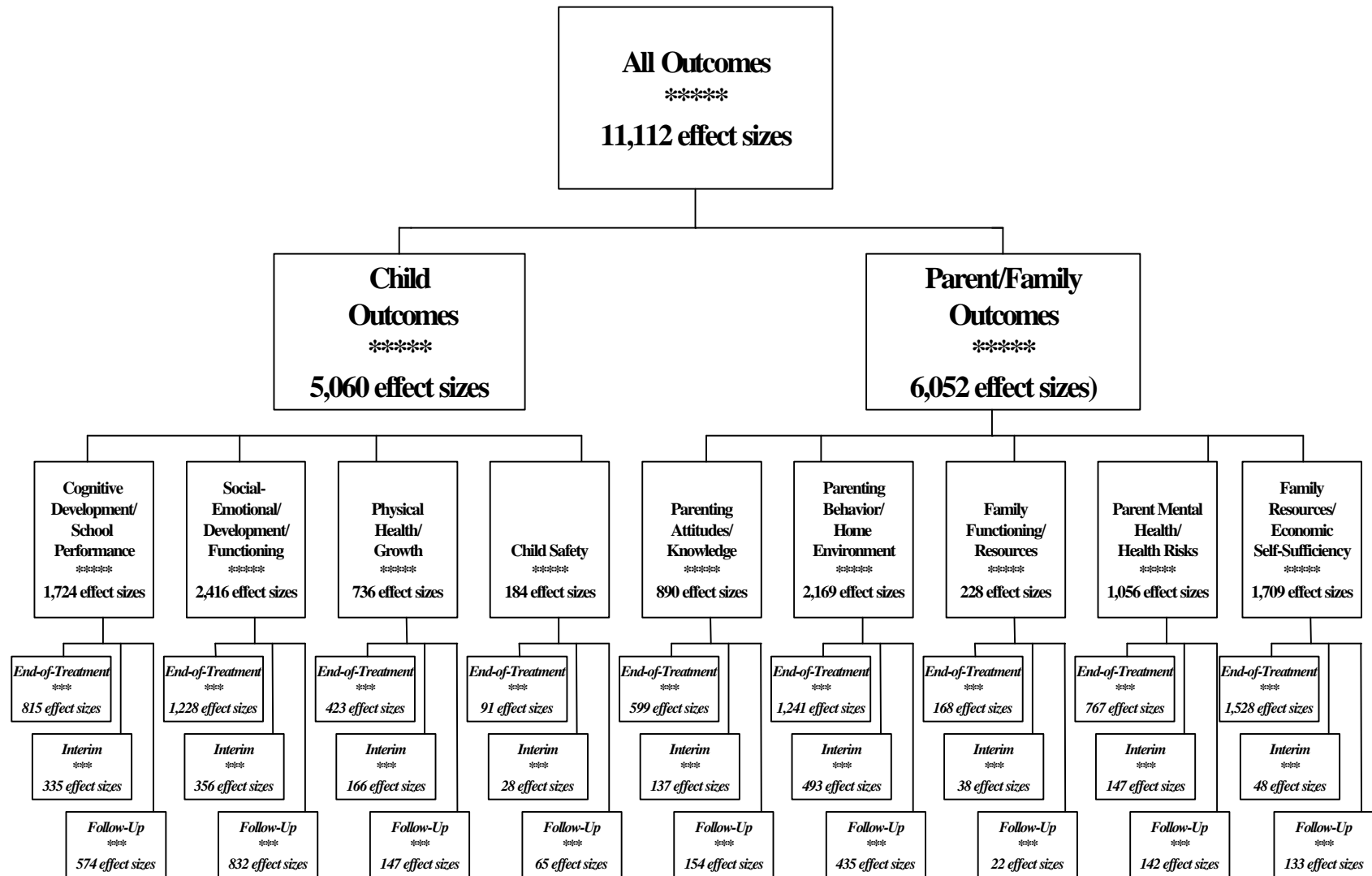
In the next step, we created two databases--one that comprised effect sizes associated with end-of-treatment outcomes (outcomes measured at the end of the intervention or service period) and a second comprising the effect sizes associated with follow-up outcomes measured after the end of the service period.² This resulted in databases of 6,860 end-of-treatment effect sizes from 488 studies and 2,502 follow-up effect sizes from 438 studies.

-
- 1 Even the finer-grained outcome domains represent a much larger number of specific measures. The six sub-domains in Exhibit A4.1 represent aggregations of 3,500 different instruments used to measure these outcomes.
 - 2 The interim outcomes were not included in these analyses, since we believed that the questions of most policy interest focused on outcomes at the end of the intervention or at an interval after the intervention ended. There are interesting research questions that could be addressed in future analyses that would examine when, during an effective intervention, the effects begin to manifest themselves.

Exhibit A4.1**Outcome Domains and Components in the Meta-Analysis Database**

Child Outcome Domains	Components
<i>Cognitive development/ school performance</i>	Tests of cognitive development, language development; measures of developmental progress; aptitude tests School performance measures: achievement measures; school performance measures (grades, ratings of level of child's performance, use of special ed services, promotion/retention/graduation)
<i>Social-emotional development</i>	Measures of social skills with peers and adults, including responsiveness, positive/negative interactions; behavior problems, emotional stability, school behavior (truancy, absence, disciplinary actions), delinquency
<i>Physical health/growth</i>	Birth circumstances, growth, illnesses, chronic health problems, receipt of preventive medical care
<i>Child safety</i>	Accidents, injuries, abuse, neglect, removal from home
Parent/Family Outcome Domains	Components
<i>Parenting knowledge/attitude</i>	Parenting knowledge (developmental milestones, developmental expectations, etc), parenting attitudes (re discipline, communication, etc)
<i>Parenting behavior</i>	Parenting behavior, including home environment
<i>Family functioning</i>	Family functioning, family resources
<i>Health</i>	Mental health: locus-of-control, self-esteem, loneliness, sense of social support, depression, coping Health risk behaviors: pregnancy outcomes, smoking and substance abuse
<i>Economic self-sufficiency</i>	Educational progress/diplomas, job training, employment

Exhibit A4.2
Distribution of Effect Sizes in the Full Meta-Analysis Database
(n=562 studies)



The final decision about the database involved which types of study designs to include in the analyses. Of the 488 studies in the end-of-treatment database, 45 percent used random assignment designs, 27 percent used quasi-experimental designs,³ and 28 percent used pre-post designs.⁴ For the primary analyses, we elected to exclude the studies with pre-post designs, for methodological reasons (see Chapter 2). We then conducted tests, using the end-of-treatment data, to determine whether excluding pre-post studies introduced a bias into the sample of studies, that is, whether the programs evaluated with less rigorous methods are systematically different than the programs evaluated more rigorously. If this were true, it would mean that excluding the pre-post studies from the meta-analysis reduces the generalizability of the findings.

To test whether the pre-post studies represented a different type of program from other studies, we compared, on a variety of characteristics, programs for which we had only pre-post studies with programs for which we had quasi-experimental and/or randomized studies. There was substantial overlap in the characteristics of programs with different types of evaluation studies, as Exhibit A4.3 shows. Nevertheless, the programs with only pre-post evaluations were, in fact, significantly different from the programs with quasi-experimental or the randomized evaluations: they were less likely to focus on children under 3 years of age or to use home visiting as the primary method of providing parent education, and they were shorter in duration, on average.⁵ Excluding pre-post studies and the programs that were studied only with pre-post designs did not mean that the remaining sample failed to include some important part of the universe of programs serving families, only that the distribution of the resulting sample was different from the distribution of the full sample of programs identified in the meta-analysis.

The final analytic databases included studies that employed either a randomized or a quasi-experimental design and that reported data from one or more of nine outcome domains. The

3 Of the 120 contrasts with quasi-experimental study designs, about 40 percent employ an internal comparison group coming from the same pool of subjects as the treatment group (e.g. all started out in the same group), while about 60 percent are based on an external comparison group obtained from a patently different pool of subjects.

4 Corresponding percentages for the follow-up database are 50 percent randomized designs, 33 percent quasi-experimental designs, and 18 percent pre-post design.

5 Programs with quasi-experimental studies differed significantly from the programs that had been studied in experiments, although the differences were less substantial than in the contrast with the pre-post studies: Programs studied with quasi-experiments tended to have higher family supportiveness ratings and to more often be targeted to at-risk families.

Exhibit A4.3

Characteristics of Programs by Type of Evaluation Design^a

Program Characteristic	Type of Evaluation Design			Significant Group Differences ^{b, c}
	Randomized (n=109 programs)	Quasi-Experiment (n=75 programs)	Pre-Post (n=60 programs)	
Average family supportiveness rating	3.55	4.18	3.69	quasi > pre-post, randomized
Focus on children < 3 years	67%	57%	44%	pre-post < randomized, quasi
Includes an early childhood education component	20%	29%	21%	no s.d.
Targeted to environmental risk	57%	72%	56%	quasi > pre-post, randomized
Targeted to child developmental/biological risk	39%	28%	33%	no s.d.
Targeted to teenage mothers	12%	22%	19%	no s.d.
Home visiting as primary method of parent education	59%	49%	27%	pre-post < randomized, quasi
Professional staff for parent education	62%	63%	68%	no s.d.
Average # months of services	13.76 months	15.36 months	6.67 months	pre-post < randomized, quasi
All characteristics combined (multivariate)				Pre-post significantly different than randomized, quasi combined (p < 01)

- a Each program was categorized in terms of the most rigorous study associated with that program; that is, if a program was studied with at least once using a randomized design, the program was placed in the Randomized group.
- b Between-group differences based on discriminant function analysis.
- c Design types separated by “ < ” are significantly different at $p < .05$ and lower; design types separated by a comma are not significantly different from each other.

end-of-treatment analytic database comprised 5,681 effect sizes from 351 studies of 190 programs (Exhibit A4.4). The follow-up analytic database comprised 2,224 effect sizes from 158 studies of 87 programs (Exhibit A4.5).

Level of Analysis

The research questions for the meta-analysis focus on the study as the critical unit of analysis. That is, we are interested in understanding the effects of family support interventions, as they are implemented and studied at a particular time and place with a particular sample of individuals, and in determining whether variation in effects across studies can be predicted. Since the research questions are at the study level, it was appropriate to analyze the data at the study level. This meant that we needed to aggregate data from the level of the individual effect size to the study level. Our strategy for aggregating individual effect sizes is to average effect sizes within a particular study *within an outcome domain*. That is, instead of constructing a *single* effect size to represent all of the effects of a particular study, e.g., by averaging all of the effect sizes within a study, we instead constructed a separate aggregated effect size for each of the five outcome domains. In any outcome domain in which a study reported outcomes, we computed the average effect size for that study within that domain. As a result, an individual study is represented in the analytic database between one and five times, depending on the number of outcome domains for which that study reported data. In the database, a study could have an average effect size associated with child cognitive outcomes, child social-emotional outcomes, parenting outcomes, parent mental and physical health outcomes, and family economic self-sufficiency outcomes.

With this strategy, theoretically each of the studies in the analytic database could contribute nine average effect sizes, one in each of the outcome domains. For the end-of-treatment database, which comprised 351 studies, this would result in a total of 3,159 average effect sizes in the database. In reality, the database included 1,004 average effect sizes across the 351 studies, which means that each study had data in 2.8 outcome domains, on average. For the follow-up database, which included 158 studies, if each study had data in each of the nine outcome domains, the database would have included 1422 average effect sizes. The actual number of effect sizes was 361.

Exhibit A4.6 shows the number of outcome domains represented in the studies in the end-of-treatment database. Over a quarter of the studies (28 percent) reported data in only one outcome domain. The majority of studies report on outcomes in three or fewer domains. Half of the studies (52 percent) reported both child and parent outcomes; just over a quarter of the studies (28 percent) reported only child outcomes, and the remaining 20 percent of the studies reported only parent outcomes. Virtually all of the remaining studies reported data on outcomes for both children and parents, with most studies reporting outcomes in only one or two domains for either group.

Exhibit A4.4

Distribution of Effect Sizes in the Analytic Database: End-of-Treatment Outcomes

(n=351 Studies; Randomized and Quasi-experimental Designs)

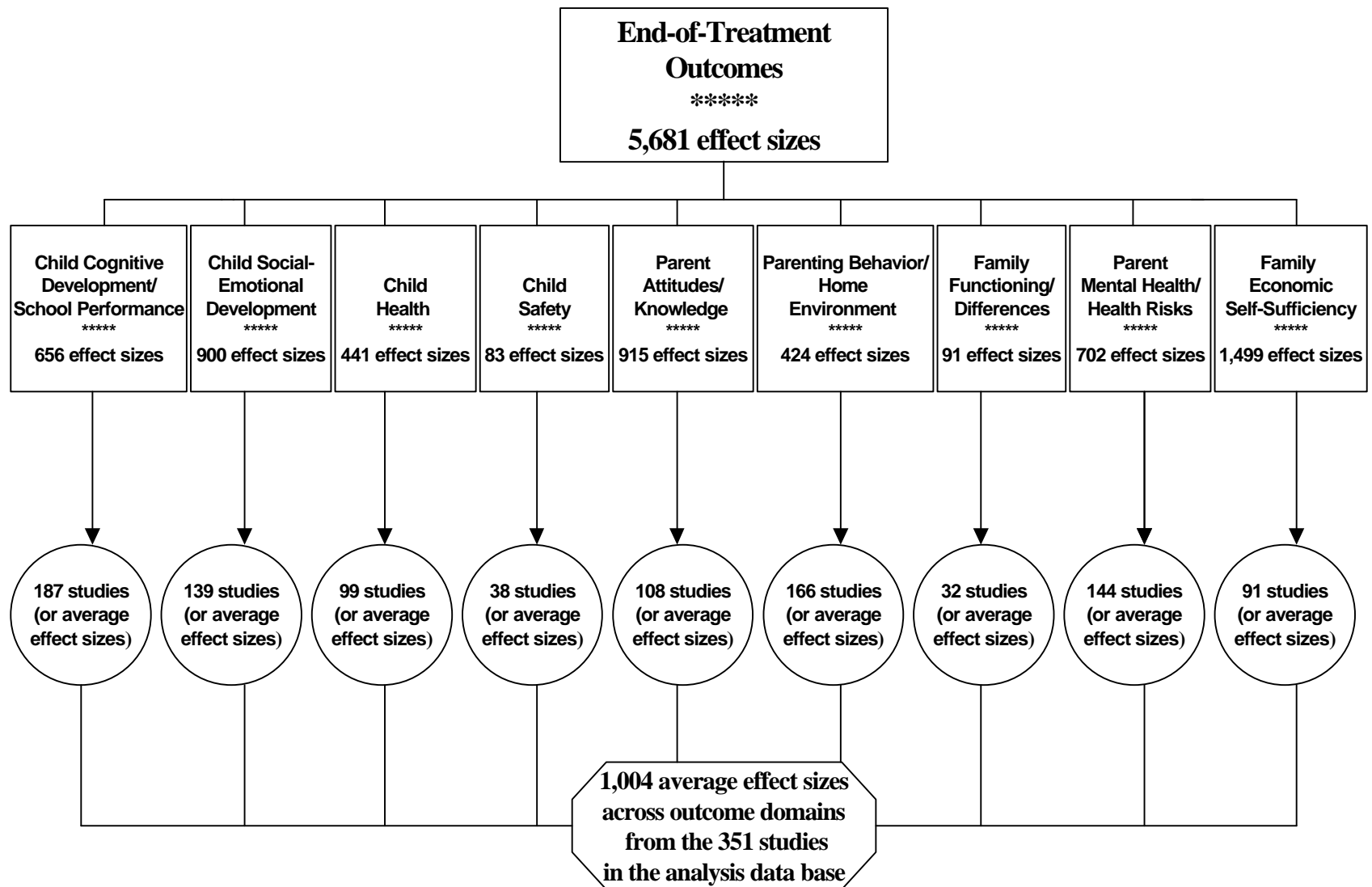
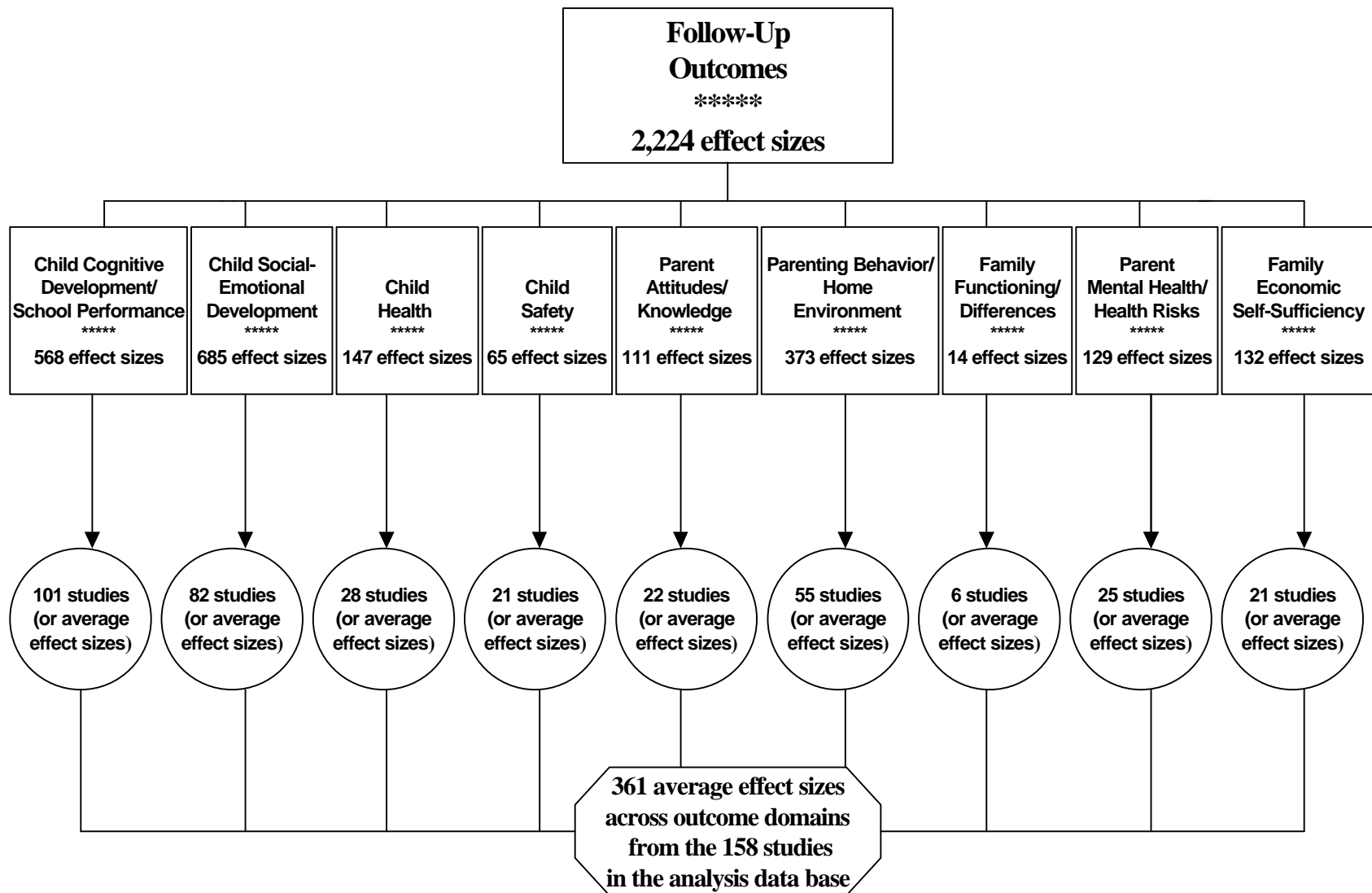


Exhibit A4.5

Distribution of Effect Sizes in the Analytic Database: Follow-Up Outcomes

(n=158 Studies; Randomized and Quasi-experimental Designs)



As a result of this approach to aggregating effect sizes, there were a different number of studies, each contributing an average effect size, in each of the outcome domains. In the end-of-treatment database, there were 187 average effect sizes in the child cognitive domain (i.e., effect sizes representing 187 of the 351 studies in the analytic database). The corresponding numbers of average effect sizes (or studies) in each of the other outcome domains were: 139 average effect sizes for child social-emotional outcomes; 99 effect sizes for child health outcomes; 38 for child safety outcomes; 108 average effect sizes for parenting attitude outcomes; 166 effect sizes for parenting behavior; 32 effect sizes for family functioning; 144 average effect sizes for parent mental health/health risks; and 91 average effect sizes for economic self-sufficiency outcomes (see Exhibit A4.4).⁶ In the follow-up database, there were 101 average effect sizes in the child cognitive domain, 82 average effect sizes in the child social-emotional domain, 28 effect sizes for child health, 21 effect sizes for child safety, 22 effect sizes for parent attitudes, 55 effect sizes for parenting behavior, 6 effect sizes for family functioning, 25 effect sizes for parent health, and 21 effect sizes for family economic self-sufficiency (Exhibit A4.5).

Exhibit A4.6

Number of Outcome Domains Reported on Studies in the Analytic Database (n=351 studies)

Number of Outcome Domains	Number of Studies (% of Sample)	
9	0	(0%)
8	3	(1%)
7	31	(9%)
6	10	(3%)
5	27	(8%)
4	30	(9%)
3	44	(13%)
2	106	(31%)
1	97	(28%)

6 It should also be noted that in a number of bases, the same programmatic model was represented by multiple studies in the database. While these studies can be viewed as replications of programmatic models with varying levels of treatment intensity and methodological approach, for the analyses reported here, studies were treated as independent tests of family support interventions. Technically, stochastic dependencies amongst both outcomes within contrasts, and contrasts within studies, need to be accounted for in the analysis. This clustering effect, although of statistical interest, is often of little practical consequence when samples sizes are very large and estimates are thus often robust to these within-study dependencies (Will Shadish, personal communication).

Single-Predictor Analyses

For each outcome domain in which we determined that there was sufficient variation, we proceeded with analyses that tested the relationship between selected program and treatment characteristics (predictors) and average effect size. In these analyses, predictors were tested one at a time for each of the nine outcome domains. The results of these analyses told us which, if any, program or treatment characteristics account for a significant amount of variance in the average effect size of family support programs in different outcome domains. As Chapter 3 showed, the coding system we developed for recording information on each study in the meta-analysis provided us with data on many programmatic and treatment characteristics. From this large pool of variables we selected the set of variables to be tested as predictors. The variables selected as predictors were chosen based on findings from previous meta-analyses, advice from technical advisors, an understanding of which predictors would be most important to policy-makers, and on hypotheses we developed prior to the analyses. The predictors are listed in Exhibit A4.7.

Treatment Predictors

Treatment variables have been tested as predictors in virtually all meta-analyses of program effects. Whether the intensity of services received has any relationship to the size of the effects of a program is a question of interest to both policy makers and program designers and planners. We elected to test separately the effect of intensity of three kinds of services. Many family support programs provide a wide array of services to parents and/or children; there was little uniformity across programs in the combination of services provided. Therefore, we felt it was important to look separately at the effects of different kinds of service. The following treatment variables were selected for the analysis:

- *length of treatment* (months elapsed from pre-test to post-test at end of treatment);
- *amount (hours) of services or treatment* received by program parents and children during the treatment period—early childhood education services, parent education services, and case management services; and
- *intensity of services or treatment received* (hours of service per month), for each type of service.

Program Predictors

The programmatic variables coded in the current meta-analysis include a large number of possible predictors. Variables in this group were selected a) so that we could address questions of specific policy interest (i.e., questions about the effectiveness of home visiting, the importance of staff qualifications, effects of different kinds of populations served), or b) because they had been found to be significant in previous meta-analyses of other types of programs, or c) because they represented some of the sources of variance in the sample of studies. The following program variables were selected for the analyses:

- program goals and focus;
- services provided to families;
- methods of delivering services;
- whether program is targeted (vs universal) and types of parents or children targeted;
- characteristics of population served by program;
- qualifications of staff who provide parent education; and
- intended amount, length and intensity of services.

Control Variables

In these regression analyses, we also included a set of control variables, in order to have a more precise estimate of the relationship between a predictor and average effect size. The analyses then produced an estimate of the relationship of the predictor to average effect size, accounting for any variation that was associated with the control variable. These are termed control variables because, although the variables individually are not policy variables, we want to take their effects into account when we estimate the effects of the treatment and programmatic variables. Previous research has shown that these types of variables often account for a significant amount of the variance in effect sizes. Five variables representing sources of within-study variation were included as control variables. They are listed below and described in more detail in Appendix E:

- accuracy of the method used to compute each effect size,
- manipulability of the measure for each effect size (how easy it would be for the respondent to tailor his/her response to be closer to the “desirable” outcome,
- number of effect sizes included in computing the average effect size for a study within each outcome domain,
- sample size associated with each individual effect size across both treatment and comparison groups, and
- where the data were reported (published or unpublished document).

In addition, in analyses in which randomized and quasi-experimental studies were combined, a sixth cohort variable was added to account for the form of the study design.

Multivariate Modeling Analyses

In the third stage of the analyses, prediction models were developed and tested in order to determine a “best fit” model for each outcome domain. That is, different combinations of program and treatment characteristics were tested to identify the smallest set of predictors

that accounted for the most variance in the outcome domain. The modeling followed three steps: the selection of program and treatment characteristics to be tested in the modeling for each outcome domain; examination of the correlations among the potential predictors for each outcome domain; and development and testing of alternate prediction models for each outcome domain.

Selecting the Variables to be Tested

The selection of variables to be tested in the modeling analyses was based on the results of the single-predictor analyses of the relationships between each individual predictor variable and average effect size. Only the predictors that individually accounted for a statistically significant amount of variance in effect size were tested in the modeling analyses. Because we did not expect that the same predictors would be significant in the nine outcome domains, the single-predictor analyses were conducted separately for each domain and the significant predictors for each domain were carried into the modeling analyses.

Examining the Correlations Among the Predictor Variables

Once we identified the subset of variables to be tested as predictors in each outcome domain, we then examined the correlations among the program and treatment variables in each domain, to determine if any of the predictor variables were highly correlated with each other. The pattern of correlations among the program and treatment variables strongly influenced what models were tested, since we avoided entering highly-correlated variables in the same model.

Developing the Prediction Models

The modeling analyses had as their goal determining the “best fit” model for each of the outcome domains. The “best fit” model is defined as the model that: (a) accounts for the greatest amount of explained variance in effect size; (b) includes only predictors that accounted for a statistically significant amount of the variance, when the significant control variables and the other program and/or treatment predictors are also entered; and © includes the smallest number of predictors that accounted for the most variable in effect size.

The process described below was repeated for data in each of the nine outcome domains. We began the modeling process by fitting models of effect sizes with two or more program variables (together with the control variables). The criterion for inclusion on the initial list was that the variable had to show a significant relationship with average effect size in the full sample. We fit a variety of models in each outcome domain, using different combinations of treatment and program variables. If a model was found to have two significant predictors, then additional models were constructed in which each of the remaining variables on the list was added to the previous model with two treatment/ program variables. Otherwise, a model with only a single significant predictor was retained as the preferred model.

The best models generally included predictors that were each statistically significant at the $p < .05$ level. The overall goodness of a model was judged by the proportion of variance explained by that model. If a significant amount of variance was unaccounted for after developing our best models, it suggests that if more, or more accurate, information were available on the measures, studies, treatments or programs, then better explanatory models could be found.

In the chapter that follows, we present the results of these analyses.

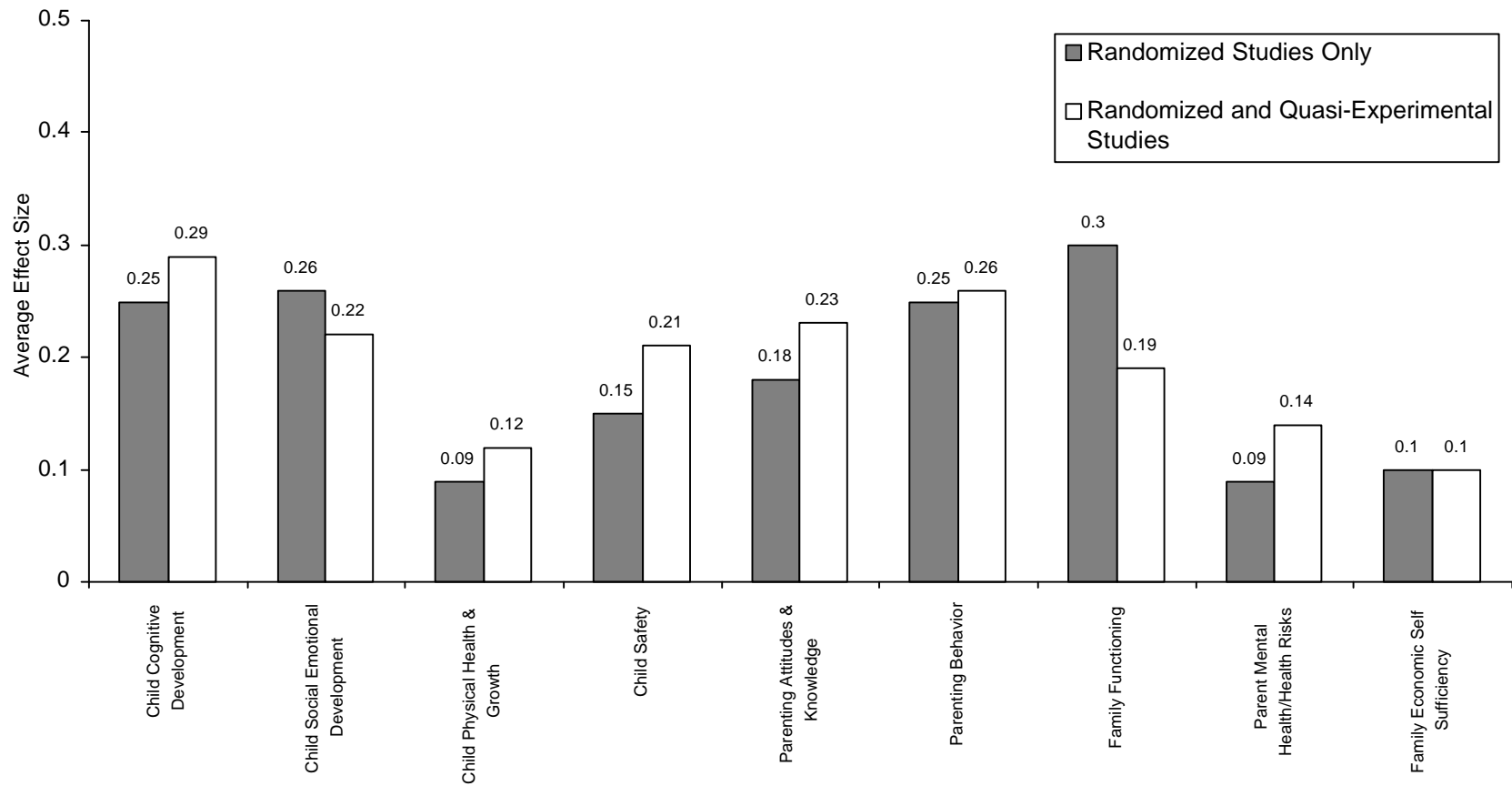
Exhibit A4.7

Program, Population and Treatment Characteristics Tested as Predictors of Effect Sizes

Program Goals/Focus	Services/Activities	Service Delivery Strategies	Staffing	Targeting	Population Served	Amount of Services
Primary program goal: – social support – adult personal development/self-help – child abuse prevention – adult economic self-sufficiency/literary – community change – child behavior change	Direct early childhood education (0/1)	Parent education delivered through home visits vs. parent groups (0/1)	Qualifications of parenting education staff: all professional vs. all paraprofessional/mix (0/1)	Universal program vs. targeted (0/1)	Percentage of families minority (0-100%)	Number of hours of: - early childhood education - parent education - case management
	Case management (0/1)					
	Health services (0/1)					
	Adult education classes (0/1)					
	Parent/child activities (0/1)			Targeted to environmental/family risk (0/1)	Percentage of families low-income (0/1)	Number of months of services
	Collaboration with community agencies (0/1)			Targeted to child special needs (0/1)		Intensity of services (hours/month) - early childhood education - parent education - case management
	Community advocacy (0/1)			Targeted to children with behavior problems (0/1)		
				Targeted to teenage parents (0/1)		
				Targeted to infants/toddlers (lt 3 years) (0/1)		

Exhibit A5.1

Average Adjusted Effect Sizes^{a, b} in Nine Outcome Domains



^a Means weighted inversely proportional to the variance of the effect size estimate (which is related to sample size).

^b Means adjusted for within-study variation associated with: number of effect-sizes in study average, sample size, type of measures on which effect sizes were based, accuracy of methods used to compute effect sizes, whether study was published or not, and for the sample combining randomized and quasi-experimental studies, the study design.

Exhibit A5.2

Weighted Mean Standardized Effect Sizes for Nine Outcome Domains: Short-Term Outcomes

Results	Child Cognitive Development	Child Social-Emotional Development	Child Physical Health & Development	Child Injury, Abuse, Neglect	Parenting Attitudes & Knowledge	Parenting Behavior	Family Functioning/ Family Resources	Parent Mental Health/ Health Risks	Family Economic Self-Sufficiency
Randomized Studies Only									
n of studies	110	96	68	22	71	118	18	98	71
Average effect size ^{a, b}	.253	.258	.091	.134^d	.182	.246	.284^d	.093	.095
Standard error	.031	.021	.024	.064	.027	.031	.092	.024	.020
95% confidence interval	(.192, .314) ^{***}	(.203, .313) ^{***}	(.046, .136) ^{***}	(.010, .260) [*]	(.125, .239) ^{***}	(.186, .306) ^{***}	(.104, .464) ^{**}	(.042, .144) ^{**}	(.046, .144) ^{***}
Homogeneity of variance test ^c	$\chi^2 = 207.2^{***}$	$\chi^2 = 101.8^{***}$	$\chi^2 = 102.0^{**}$	$\chi^2 = 131.9^{***}$	$\chi^2 = 89.2^*$	$\chi^2 = 254.3^{***}$	$\chi^2 = 81.7^{***}$	$\chi^2 = 162.2^{**}$	$\chi^2 = 64.6$
Randomized & Quasi- Experimental Studies									
n of studies	187	139	99	38	108	166	32	144	91
Average effect size ^{a, b}	.293	.223	.123	.213^d	.230	.257	.169^d	.137	.099
Standard error	.025	.025	.024	.068	.026	.026	.052	.023	.022
95% confidence interval	(.244, .342) ^{***}	(.175, .271) ^{***}	(.075, .170) ^{***}	(.080, .346) ^{**}	(.179, .281) ^{***}	(.206, .308) ^{***}	(.067, .271) ^{**}	(.092, .182) ^{***}	(.056, .142) ^{***}
Homogeneity of variance test ^c	$\chi^2 = 384.5^{***}$	$\chi^2 = 198.8^{***}$	$\chi^2 = 199.9^{***}$	$\chi^2 = 213.2^{***}$	$\chi^2 = 141.9^{***}$	$\chi^2 = 344.4^{***}$	$\chi^2 = 72.0^{***}$	$\chi^2 = 270.6^{***}$	$\chi^2 = 95.3$

* $p < .05$ ** $p < .01$ *** $p < .001$

a Means adjusted for within-study variation associated with (1) number of effect sizes in average, (2) sample size for each effect size, (3) types of measures in effect sizes, (4) accuracy of methods used to compute effect sizes, (5) whether study was published, and (6) whether the study was a randomized or quasi-experimental design.

b Means weighted inversely proportional to the variances effect size estimate (which is related to sample type).

c Homogeneity of variance test indicates whether there is significant inter-study variation among effect sizes.

d Because of small sample sizes, means adjusted for only two control variables: published/unpublished study and study design.

Exhibit A5.3

Distribution of Average Effect Sizes on Children's Outcomes

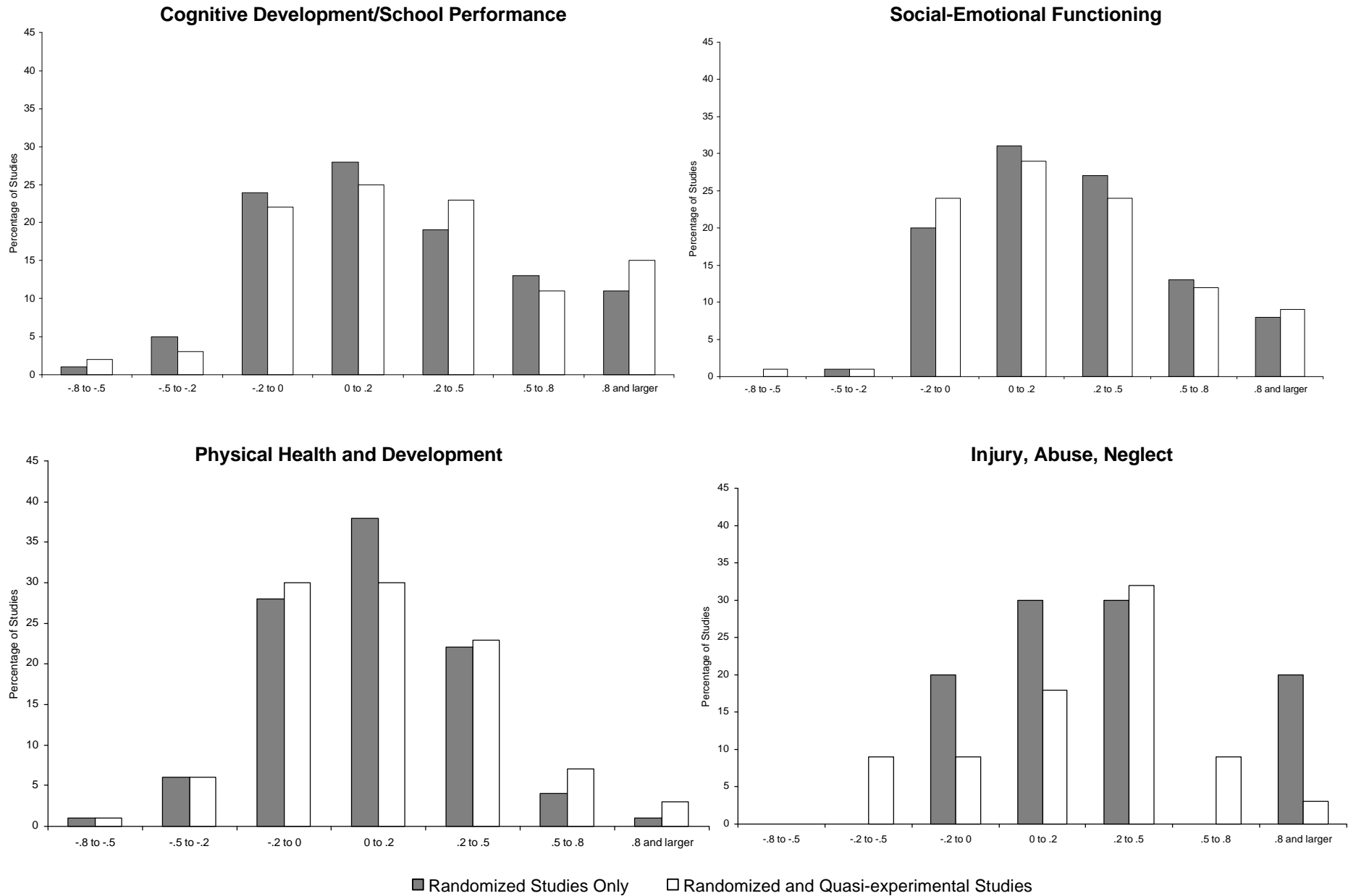


Exhibit A5.4 Distribution of Average Effect Sizes on Parenting Outcomes

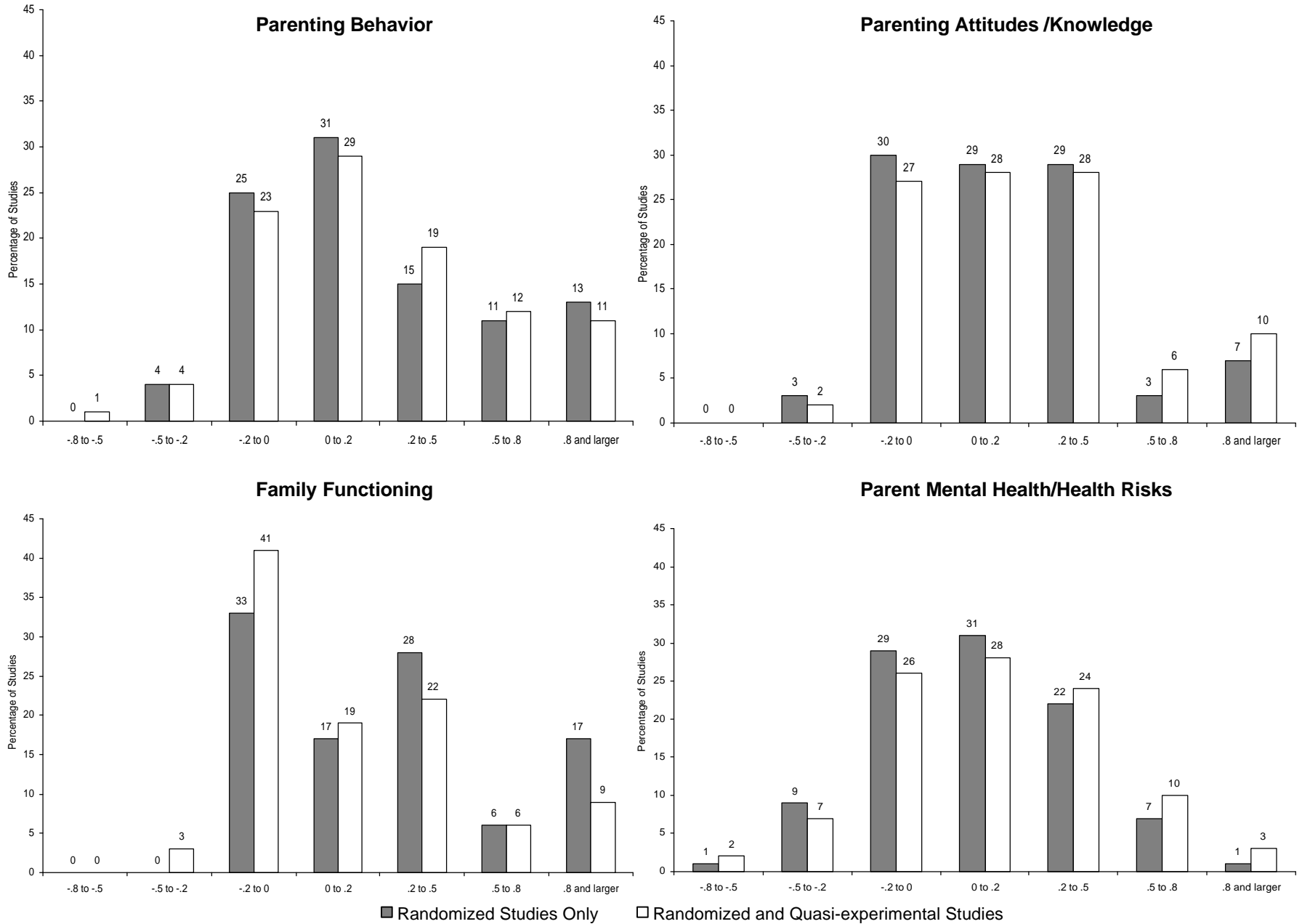


Exhibit A5.4 (continued)

Distribution of Average Effect Sizes on Parenting Outcomes

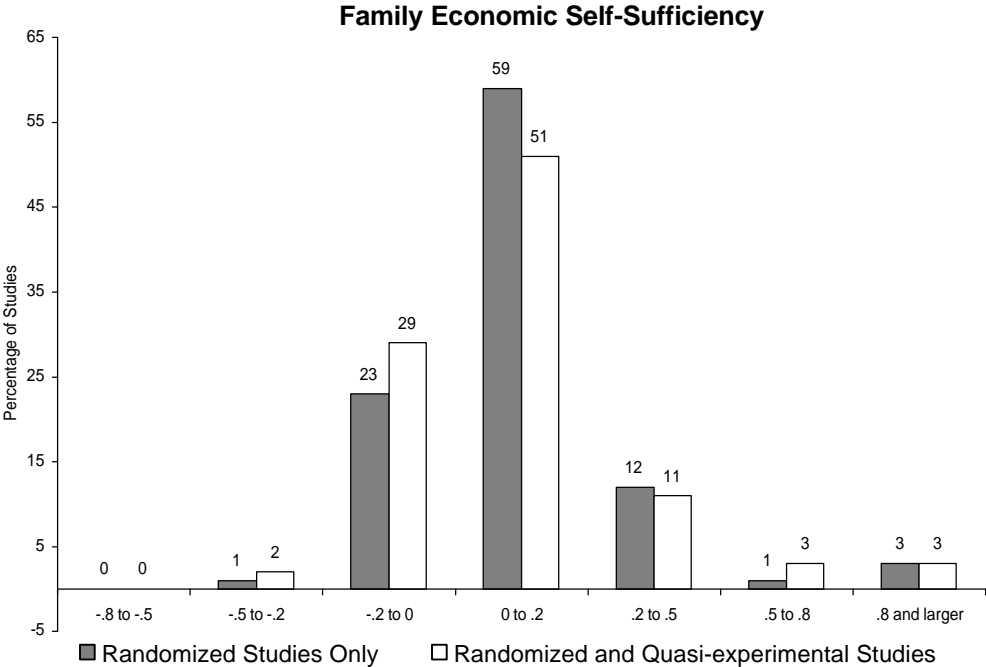


Exhibit A5.5

Significant Relationships^a between Programmatic Characteristics and Effects in Nine Outcome Domains: Randomized Studies

Program Characteristic	Outcome Domain								
	Child Cognitive	Child Social	Child Health	Child Safety	Parent Attitudes	Parent Behavior	Family Functng	Parent Health	Econ Self-Suffic
	n=110	n=96	n=68	n=22	n=71	n=118	n=18	n=98	n=71
Primary Program Goals									
Social support		insufficient variation in effect sizes to test predictors							insufficient variation in effect sizes to test predictors
Parent Self-help/self-development						+			
Prevention of child abuse/neglect							+		
Economic self-sufficiency/literacy						-		-	
Child mental health/behavior									
Community participation									
Targeting									
Universal									
Biological/developmental child risk	+							-	
Biological risks	+				-	+			
Developmental risks	+				-				
Teenage parents									
Population served									
Majority families low-income									
% minority families									
Majority teenage parents					+				
Child age: infant/toddler vs older					-				
Child age at end of services (months)					-			-	

Program Characteristic	Outcome Domain								
	Child Cognitive	Child Social	Child Health	Child Safety	Parent Attitudes	Parent Behavior	Family Functng	Parent Health	Econ Self-Suffic
Parenting education									
Staff qualifications: all professionals			- *				+		
Home visits (vs parent groups)	- **								
Services provided: Types of services									
Intended length of services									
Any early childhood education	+ **								
Any parent/child activities								-	
Peer support activities	+ *		- *		+ *				
Any adult education activities					-			- *	
Any case management				+ *		- *			
Collaboration with other agencies						-			
Any health services								- *	
Any community advocacy activities						- *			
Services provided: Amount of services									
Months services provided									
Hours of early childhood education									
Intensity of early childhood education									
Amount (hrs) of parent education									
Intensity of parent education									
Amount (hrs) of case management									
Intensity of case management									

p<.10 * p<.05 ** p<.01 *** p<.001

a Effect of each characteristic based on regression analysis (hierarchical linear modeling approach) with control variables accounted for.

Exhibit A5.6

Significant Relationships between Programmatic Characteristics and Effects in Nine Outcome Domains: Randomized and Quasi-Experimental Studies

Program Characteristic	Outcome Domain									
	Child Cognitive	Child Social	Child Health	Child Safety	Parent Attitudes	Parenting Behavior	Family Functng	Parent Health	Economic Self-Suffic	
	(n=187)	(n=139)	(n=99)	(n=38)	(n=108)	(n=166)	(n=32)	(n=144)	(n=91)	
Primary Program Goals										
Social support										insufficient variation in effect sizes to test predictors
Parent self-help/self development		+ **			+ *	+ **				
Prevention of child abuse/neglect							+ *	+		
Economic self-sufficiency/literacy		-						- **		
Child mental health/behavior						+ *				
Community participation										
Targeting										
Universal				+ *						
Biological/developmental child risk										
Biological risks				- *					-	
Developmental risks		+ *								
Population served										
Majority families low-income		- *								
% minority families										
Majority teenage parents					+ **					
Child age: infant/toddler vs older										
Child age at end of services (months)									- **	

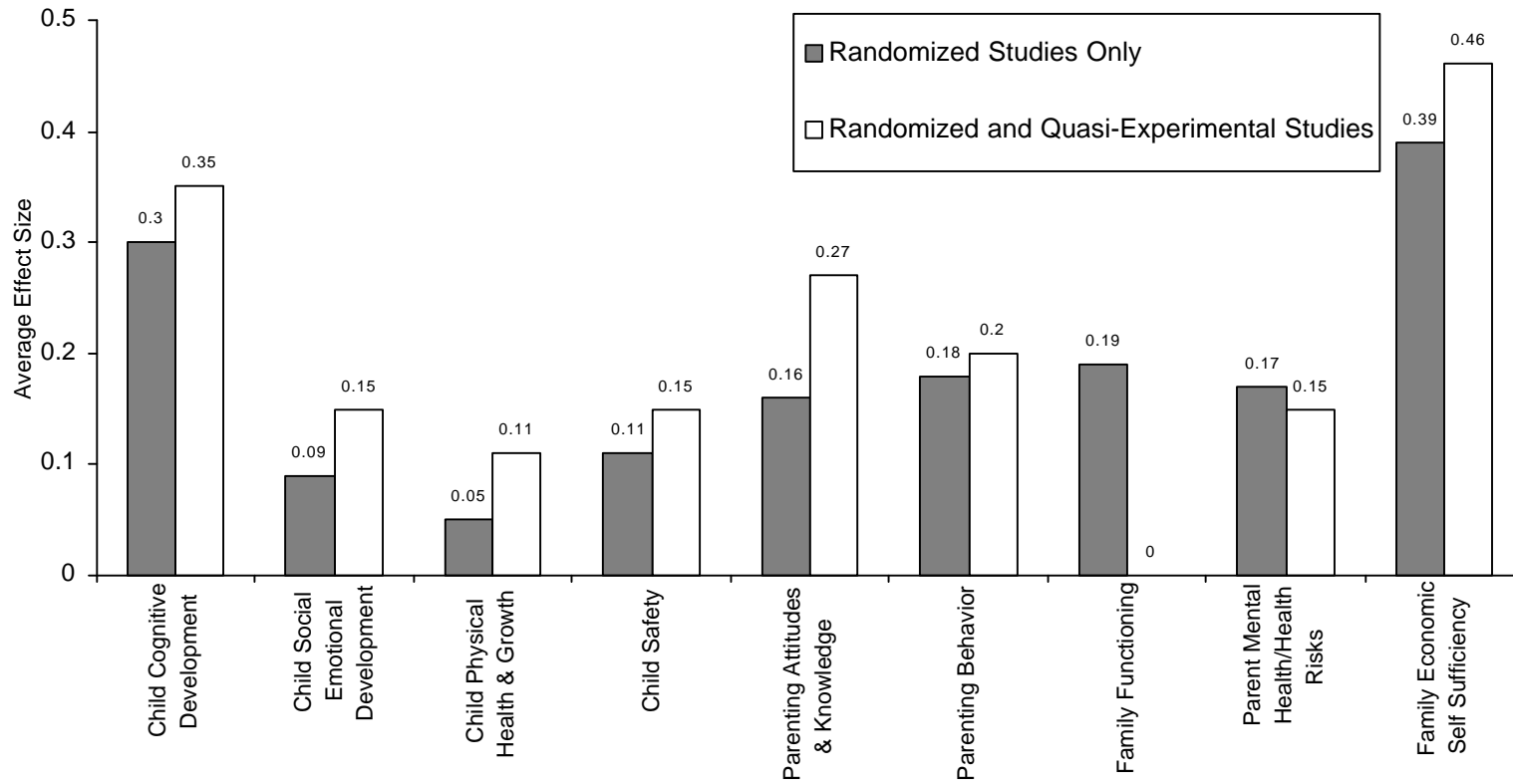
Program Characteristic	Outcome Domain									
	Child Cognitive	Child Social	Child Health	Child Safety	Parent Attitudes	Parenting Behavior	Family Functng	Parent Health	Economic Self-Suffic	
Parenting Education										
Staff qualifications: all professionals		+							+	
Home visits (vs parent groups)		-				-				
Services provided: Types of services										
Intended length of services							-			
Any early childhood education										
Any parent/child activities				+					-	
Peer support activities					+		+			
Any adult education activities					-				-	
Any case management		-		+			-		-	
Collaboration with other agencies										
Any health services										
Any community advocacy activities				+			-			
Services provided: Amount of services										
Months services provided							+			
Hours of early childhood education										
Intensity of early childhood education										
Amount (hrs) of parent education										
Intensity of parent education										
Amount (hrs) of case management									-	
Intensity of case management										

p<.10 * p<.05 ** p<.01 *** p<.001

a Effect of each characteristic based on regression analysis (hierarchical linear modeling approach) with control variables accounted for.

Exhibit A5.26

Average Adjusted Effect Sizes^{a,b} in Nine Outcome Domains: Follow-up Outcomes



^a Means weighted inversely proportional to the variance of the effect size estimate (which is related to sample size).

^b Means adjusted for within-study variation associated with: number of effect-sizes in study average, sample size, type of measures on which effect sizes were based, accuracy of methods used to compute effect sizes, whether study was published or not, and for the sample combining randomized and quasi-experimental studies, the study design.

Exhibit A5.27

Weighted Mean Standardized Effect Sizes for Nine Outcome Domains: Follow-Up Outcomes^a

Results	Child Cognitive Development	Child Social-Emotional Development	Child Physical Health & Development	Child Injury, Abuse, Neglect	Parenting Attitudes & Knowledge	Parenting Behavior	Family Functioning/Family Resources	Parent Mental Health/Health Risks	Family Economic Self-Sufficiency
Randomized Studies Only									
n of studies	48	51	20	14	12	38	4	18	12
Average effect size ^{a, b}	.304	.094^d	.049^d	.115^e	.152^d	.178	.190^f	.165^e	.386^f
Standard error	.045	.028	.031	.074	.075	.053	.198	.055	.161
95% confidence interval	(.216, .392)***	(.039, .149)***	(-.012, .104)	(-.030, .260)	(.005, .299)	(.074, .282)**	(-.198, .578)	(.038, .273)**	(.070, .702)*
Homogeneity of variance test ^c	$\chi^2 = 63.8^*$	$\chi^2 = 39.4$	$\chi^2 = 16.1$	$\chi^2 = 43.2^{***}$	$\chi^2 = 25.0^{**}$	$\chi^2 = 41.5$	$\chi^2 = 20.0^{***}$	$\chi^2 = 27.2^*$	$\chi^2 = 151.8^{***}$
Randomized & Quasi- Experimental Studies									
n of studies	101	82	28	21	22	55	6	25	21
Average effect size ^{a, b}	.345	.150^d	.112^d	.152^d	.273^d	.204	.002^f	.226^e	.464^e
Standard error	.040	.024	.036	.068	.054	.041	.045	.051	.089
95% confidence interval	(.267, .423)***	(.103, .197)***	(.041, .183)**	(.019, .285)*	(.167, .379)***	(.123, .285)***	(-.086, .010)	(.116, .236)***	(.290, .639)***
Homogeneity of variance test ^c	$\chi^2 = 184.5^{***}$	$\chi^2 = 101.4^*$	$\chi^2 = 33.3$	$\chi^2 = 68.6^{***}$	$\chi^2 = 27.1$	$\chi^2 = 65.7^*$	$\chi^2 = 9.30$	$\chi^2 = 47.8^{**}$	$\chi^2 = 58.0^{***}$

* $p < .05$ ** $p < .01$ *** $p < .001$

a Means adjusted for within-study variance associated with (1) number of effect sizes in average, (2) sample size for each effect size, (3) types of measures in effect sizes, (4) accuracy of methods used to compute effect sizes, (5) whether study was published, and (6) whether the study was a randomized or quasi-experimental design.

b Means weighted inversely proportional to the variance of the effect size estimate (which is related to sample size).

c Homogeneity of variance test indicates whether there is significant inter-study variation among effect sizes.

d Means adjusted for only two of the control variables because of collinearities.

e Means adjusted for only one of the control variables because of collinearities.

f No control variables entered as covariates because of small sample size.

Exhibit A5.28

Significant Relationships^a between Programmatic Characteristics and Effects in Nine Outcome Domains: Randomized Studies

Program Characteristic	Outcome Domain								
	Child Cognitive	Child Social	Child Health	Child Safety	Parent Attitudes	Parent Behavior	Family Functng	Parent Health	Econ Self-Suffic
	n=48	n=51	n=20	n=14	n=12	n=38	n=4	n=18	n=12
Primary Program Goals									
Social support		insufficient variation in effect sizes to test predictors	insufficient variation in effect sizes to test predictors			insufficient variation in effect sizes to test predictors	insufficient variation in effect sizes to test predictors		
Parent Self-help/self-development									
Prevention of child abuse/neglect				+					
Economic self-sufficiency/literacy									
Child mental health/behavior									
Community participation									
Targeting									
Universal									
Biological/developmental child risk									
Biological risks		+				+			
Developmental risks									
Teenage parents									
Population served									
Majority families low-income									
% minority families									
Majority teenage parents									
Child age: infant/toddler vs older									
Child age at end of services (months)								- **	
Parenting education									
Staff qualifications: all professionals									
Home visits (vs parent groups)								- *	

Exhibit A5.28

Significant Relationships^a between Programmatic Characteristics and Effects in Nine Outcome Domains: Randomized Studies

Program Characteristic	Outcome Domain								
	Child Cognitive n=48	Child Social n=51	Child Health n=20	Child Safety n=14	Parent Attitudes n=12	Parent Behavior n=38	Family Functng n=4	Parent Health n=18	Econ Self-Suffic n=12
Services provided: Types of services									
Intended length of services									
Any early childhood education									
Any parent/child activities								-	
Peer support activities					+ *				
Any adult education activities					-			- *	
Any case management	+ *								
Collaboration with other agencies									
Any health services									
Any community advocacy activities									
Services provided: Amount of services									
Months services provided									
Hours of early childhood education									
Intensity of early childhood education									
Amount (hrs) of parent education									
Intensity of parent education									
Amount (hrs) of case management									
Intensity of case management									

p<.10 * p<.05 ** p<.01 *** p<.001

a Effect of each characteristic based on regression analysis (hierarchical linear modeling approach) with control variables accounted for.

Exhibit A5.29

Significant Relationships^a between Programmatic Characteristics and Effects in Nine Outcome Domains: Randomized and Quasi-Experimental Studies

Program Characteristic	Outcome Domain								
	Child Cognitive	Child Social	Child Health	Child Safety	Parent Attitudes	Parent Behavior	Family Functng	Parent Health	Econ Self-Suffic
	n=101	n=82	n=28	n=21	n=22	n=55	n=6	n=25	n=21
Primary Program Goals									
Social support			insufficient variation in effect sizes to test predictors		insufficient variation in effect sizes to test predictors			insufficient variation in effect sizes to test predictors	
Parent Self-help/self-development						+ *			
Prevention of child abuse/neglect				+ *					
Economic self-sufficiency/literacy									
Child mental health/behavior									
Community participation									
Targeting									
Universal									
Biological/developmental child risk									
Biological risks									
Developmental risks									
Teenage parents									
Population served									
Majority families low-income		- *							
% minority families									
Majority teenage parents									
Child age: infant/toddler vs older		+ *							
Child age at end of services (months)								- **	
Parenting education									
Staff qualifications: all professionals									
Home visits (vs parent groups)								- *	

Exhibit A5.29

Significant Relationships^a between Programmatic Characteristics and Effects in Nine Outcome Domains: Randomized and Quasi-Experimental Studies

Program Characteristic	Outcome Domain								
	Child Cognitive	Child Social	Child Health	Child Safety	Parent Attitudes	Parent Behavior	Family Functng	Parent Health	Econ Self-Suffic
	n=101	n=82	n=28	n=21	n=22	n=55	n=6	n=25	n=21

Services provided: Types of services

- Intended length of services
- Any early childhood education
- Any parent/child activities
- Peer support activities
- Any adult education activities
- Any case management
- Collaboration with other agencies
- Any health services
- Any community advocacy activities

+ *

Services provided: Amount of services

- Months services provided
- Hours of early childhood education
- Intensity of early childhood education
- Amount (hrs) of parent education
- Intensity of parent education
- Amount (hrs) of case management
- Intensity of case management

p<.10 * p<.05 ** p<.01 *** p<.001

a Effect of each characteristic based on regression analysis (hierarchical linear modeling approach) with control variables accounted for.

