

**Effectiveness of Early Intervention for Infants and their Families:
Relating the Working Alliance to Program Outcomes**

by

Robert G. Santos

A Thesis Submitted to
the Faculty of Graduate Studies
in Partial Fulfillment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Department of Psychology
University of Manitoba
Winnipeg, Manitoba

© Robert G. Santos, May 2005



Library and
Archives Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

0-494-08796-X

Your file Votre référence

ISBN:

Our file Notre référence

ISBN:

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

**
Canada

THE UNIVERSITY OF MANITOBA
FACULTY OF GRADUATE STUDIES

COPYRIGHT PERMISSION

**Effectiveness of Early Intervention for Infants and their Families:
Relating the Working Alliance to Program Outcomes**

BY

Robert G. Santos

**A Thesis/Practicum submitted to the Faculty of Graduate Studies of The University of
Manitoba in partial fulfillment of the requirement of the degree
Of
Doctor of Philosophy**

Robert G. Santos © 2005

**Permission has been granted to the Library of the University of Manitoba to lend or sell copies of
this thesis/practicum, to the National Library of Canada to microfilm this thesis and to lend or sell
copies of the film, and to University Microfilms Inc. to publish an abstract of this thesis/practicum.**

**This reproduction or copy of this thesis has been made available by authority of the copyright
owner solely for the purpose of private study and research, and may only be reproduced and copied
as permitted by copyright laws or with express written authorization from the copyright owner.**

Abstract

Meta-analyses support the *efficacy* of early childhood intervention studied under optimal research conditions (Bakermans-Kranenburg, van IJzendoorn, & Juffer, 2003; MacLeod & Nelson, 2000; Nelson, Westhues, & MacLeod, 2003; Sweet & Appelbaum, 2004). However, less is known about (a) the *effectiveness* of early childhood programs under real-world service delivery conditions and (b) the *effective ingredients* that make these programs work. Programs theorize that the *working alliance* (the relationship between program participant and provider) is the key ingredient. Although supported by meta-analyses on psychotherapy (Horvath & Symonds, 1991; Martin, Garske, & Davis, 2000), alliance theory has rarely been tested empirically in early childhood intervention. The present study conducted secondary analyses of data from a 3-year, province-wide, home visiting program for families with newborns, targeted through population screening and assessment by public health nurses. Paraprofessional home visits facilitated problem-solving, parent-child interaction, and community referrals. The effectiveness of home visiting after 1 year was tested using regression-discontinuity design analyses ($N = 187$ program families and 63 comparison families) for (a) family (social support), (b) parent (depression, psychological well-being [PWB]), and (c) parent-child (positive parenting, hostile/ineffective parenting, reading with children) outcomes. Theorized relationships between parents' interpersonal histories, parents' early alliances with home visitors, alliance improvements over 1 year, and outcomes were tested using correlation and multiple regression analyses. Home visiting improved parental PWB, particularly autonomy, environmental mastery, and self-acceptance; and positive parenting; but also unintentionally increased hostile/ineffective parenting. Stronger early alliances and improvements in the alliance predicted improvements in parental PWB, positive parenting, and hostile/ineffective parenting. Results suggest that parents' interpersonal histories influence the development of their relationships with home visitors, which influence parents' relationships with their children, their environments, and themselves (e.g., self-acceptance). The working alliance appears central to the effectiveness of home visiting.

Acknowledgements

Nothing is more life-changing than our relationships with each other. The central lesson of this thesis is that working together in a caring and collaborative way can promote human growth and development. I have learned this lesson personally from the following individuals, to whom my debt is deep and my gratitude deeply felt:

To my PhD thesis advisor, Dr. Bruce Tefft, for inspiring my turning point from clinical to community psychology in 1997 and for showing me psychology's potential for preventing human suffering and promoting well-being and social justice for all. To my PhD thesis committee, Dr. Marvin Brodsky, Dr. John Whiteley, and Dr. John O'Neil (Department of Community Health Sciences), for their faith in me and their sage advice, especially to Marvin for teaching me since my third undergraduate year in 1990. To my external examiner, Dr. Geoff Nelson (Wilfred Laurier University), for our conversation about community psychology in 2004, and his thorough and thoughtful review of this thesis. My special thanks to Dr. Marianne Johnson and Dr. David Martin, for teaching me how essential the alliance is in helping others be who they were meant to be.

To Leanne Boyd, my mentor at the Manitoba Children and Youth Secretariat (CYS) (1998-2000) and colleague at the Healthy Child Manitoba Office (HCMO) (2000-present) for teaching me that a clinician could change public policy and population health for the better. To Doris Mae Oulton, CEO of CYS (1998-2000), for taking a chance on an unproven young student who knew next to nothing about government and showing me what it could become, and to Bonnie Mitchelson, Minister responsible for CYS (1998-2000) for her support. To Kary Alexander, Dale Brownlee, Mariette Chartier, Darlene Girard, Andrea Friesen, Kathleen McMillan, Patricia Phelps, and Jeanne Whalen, for

their invaluable work in the implementation of the provincial BabyFirst evaluation and the data that it provided for this thesis. To Jan Sanderson, Executive Director of HCMO and Secretary to the Healthy Child Committee of Cabinet (HCCC) (2001-present), for helping me grow into a civil servant and strengthening me through the most difficult years of balancing work, school, and family. To the HCCC Chairs, Minister Tim Sale (2000-2003), Minister Jim Rondeau (2003-2004), and Minister Theresa Oswald (2004-present), and the Healthy Child Deputy Ministers' Committee Chairs, Tannis Mindell (2000-2003), Debra Woodgate (2003-2004), and Milton Sussman (2004-present) for their enduring commitment to this research (and to my finally finishing school!). You have all shown me how a government can accomplish great things for the children that it serves.

To all the BabyFirst public health nurses and home visitors, for their profound dedication to the program and its evaluation. To the Manitoba parents (and their babies, born 2000-2003) who graciously participated in the provincial evaluation of the BabyFirst program, for sharing their experiences so that other families might benefit.

To my wife, Kim Santos, for taking a chance on a dance with an unemployed graduate student back in 1998, and staying, almost always patient and always loving, with me through thousands of pages of books and articles, over 64,000 hours to today.

To my parents, Dr. Helen Guzman and Dr. Conrado Santos, especially my Mom, who showed me firsthand during my first five years, with immeasurable help from their families and friends, how a good early childhood can last a lifetime. Thank you for being my lifelong role models in learning and service, and for your love and forbearance. This thesis, the culmination of exactly half my life (17 years) spent as a university student, is dedicated to you. Above all, thanks to God, who has blessed me with all of you.

Table of Contents

Abstract	ii
Acknowledgements	iii
List of Tables	xii
List of Figures	xiii
Introduction	1
<i>Efficacy and Effectiveness of Intervention</i>	1
<i>What We Know: Efficacy of Early Childhood Intervention</i>	2
<i>What We Don't Know: Effectiveness of Early Childhood Intervention</i>	9
<i>What We Don't Know: Effective Ingredients of Early Childhood Intervention</i>	16
A Brief Overview of Home Visiting Programs	19
<i>Efficacy Studies of Home Visiting Programs</i>	21
<i>Effectiveness Studies of Home Visiting Programs</i>	26

Manitoba's BabyFirst Program	30
<i>BabyFirst Program Inclusion Criteria</i>	32
<i>BabyFirst Program Evaluation</i>	36
Evaluation Logic of Prevention and Early Intervention Programs	38
Program Theoretical Model	38
<i>Family Outcome</i>	41
<i>Social support</i>	41
<i>Parent Outcomes</i>	42
<i>Parental depression</i>	42
<i>Parental psychological well-being</i>	42
<i>Parent-Child Outcomes</i>	43
<i>Positive parenting</i>	43
<i>Hostile/ineffective parenting</i>	43
<i>Parental reading with children</i>	44
Working Alliance: Lessons from the Psychotherapy Literature	46

Toward a Theory of the Working Alliance in Early Childhood Intervention	47
<i>A Framework for Early Relationship-Based Intervention</i>	49
<i>Attachment relationships and the alliance</i>	50
<i>Relationship development and alliance development</i>	52
<i>The Trans-theoretical Model of the Alliance</i>	53
<i>Goals and tasks</i>	53
<i>Bonding</i>	54
<i>Empirical Validation of the Trans-theoretical Model of the Alliance</i>	54
<i>Summary of the Theoretical Framework for the Present Study</i>	55
<i>Hypotheses</i>	56
<i>Hypothesis 1: The effectiveness hypothesis</i>	56
<i>Hypothesis 2: The internal working model-early alliance hypothesis</i>	56
<i>Hypothesis 3: The early alliance-outcome hypothesis</i>	56
<i>Hypothesis 4: The alliance improvement-outcome hypothesis</i>	56
Method	57
<i>Participants</i>	57
<i>Design</i>	57
<i>Selection of Study Variables</i>	61

<i>Baseline Variables</i>	62
<i>Child characteristics</i>	62
<i>Parent characteristics</i>	63
<i>Family characteristics</i>	67
<i>Neighbourhood characteristics</i>	68
<i>Program Variables</i>	69
<i>Working alliance: Parent and home visitor ratings</i>	69
<i>Reliability and validity of the Working Alliance Inventory (WAI)</i>	70
<i>Age 1 Outcome Variables</i>	71
<i>Family outcome</i>	71
<i>Parent outcomes</i>	71
<i>Parent-child outcomes</i>	72
<i>Summary and Research Framework</i>	74
<i>Statistical Considerations: Significance Tests, Effect Sizes, and Confidence Intervals</i>	74
<i>Statistical Considerations: Type I and Type II Error Rates</i>	77
<i>Statistical Power Analysis</i>	79
<i>Using pretests as covariates</i>	80
<i>Using unequal sample sizes</i>	80
<i>Statistical power for effectiveness analyses</i>	81
<i>Statistical power for alliance-outcome analyses</i>	82

<i>Data Analyses</i>	83
Results	85
<i>Group Comparisons Regarding Sociodemographic Characteristics at Baseline</i>	85
<i>Reliability of Participant Measures at Baseline and 1 Year</i>	88
<i>Group Comparisons Regarding Participant Measures at Baseline</i>	91
<i>Testing Hypothesis 1: Does Participation in the BabyFirst Home Visiting Program Improve Family, Parent, and Parent-Child Outcomes at 1 Year?</i>	94
<i>Effect of home visiting on parental psychological well-being (PWB) at 1 year</i>	94
<i>Effect of home visiting on positive parenting at 1 year</i>	96
<i>Effect of home visiting on parental depression at 1 year</i>	98
<i>Effect of home visiting on parental reading with children at 1 year</i>	100
<i>Effect of home visiting on family social support at 1 year</i>	102
<i>Effect of home visiting on hostile/ineffective parenting at 1 year</i>	104
<i>Parent and Home Visitor Ratings of the Working Alliance</i>	106
<i>Analyses of available and missing working alliance data</i>	106
<i>Reliability of Early and 1 Year Working Alliance Ratings</i>	109
<i>Testing Hypothesis 2: Do More Positive Internal Working Models of Parents Predict More Positive Parent Ratings of the Early Alliance in Home Visiting?</i>	110
<i>What other baseline family and parent variables predict the early alliance?</i>	112

<i>Testing Hypotheses 3 and 4: The Early Alliance-Outcome Hypothesis and the Alliance Improvement-Outcome Hypothesis</i>	113
<i>Hypothesis 3: Does a stronger early alliance between parents and home visitors predict better family, parent, and parent-child outcomes at 1 year?</i>	113
<i>Hypothesis 4: Does a home visiting alliance that improves over the first year predict better family, parent, and parent-child outcomes at 1 year?</i>	113
<i>Alliance and parental psychological well-being (PWB) at 1 year</i>	114
<i>Alliance and parental autonomy at 1 year</i>	114
<i>Alliance and parental environmental mastery at 1 year</i>	115
<i>Alliance and parental self-acceptance at 1 year</i>	115
<i>Alliance and positive parenting at 1 year</i>	115
<i>Alliance and hostile/ineffective parenting at 1 year</i>	116
<i>Summary of alliance-outcome correlations</i>	116
<i>Comparison of results using available data with results using imputed data</i>	117
Discussion	120
<i>What Influences the Development of the Working Alliance in Home Visiting?</i>	121
<i>How Does Home Visiting Improve Parental Psychological Well-Being and Positive Parenting?</i>	123
<i>Why Did Home Visiting Increase Hostile/Ineffective Parenting?</i>	123
<i>What Are the Limitations of Home Visiting Programs?</i>	127
<i>Limitations on the effectiveness of early childhood intervention</i>	130

<i>How Can Home Visiting Programs Be Strengthened?</i>	132
<i>Improving the effectiveness of home visiting on hostile/ineffective parenting</i>	135
<i>Improving the effectiveness of home visiting on parental depression</i>	135
<i>Improving the effectiveness of home visiting on family social support</i>	135
<i>Strengths and Limitations of the Present Study</i>	136
<i>Implications and Future Directions for Research</i>	137
<i>Measuring positive outcomes</i>	137
<i>Pursuing the program, its participants, and their interaction</i>	138
<i>Learning what works in the real world</i>	139
<i>Understanding neighbourhood and community influences</i>	140
<i>Studying and valuing culture, race, and ethnicity</i>	140
<i>Moving toward meta-analysis</i>	140
<i>Charting the dimensions of the alliance</i>	141
<i>Looking at the alliance across time</i>	141
<i>Looking across generations</i>	141
<i>Lessons from the litany in psychotherapy research</i>	142
References	145

List of Tables

<p>Table 1 <i>A Framework for Understanding How Early Interventions Work: The Program, its Participants, and their Interaction</i></p> <p>Table 2 <i>Definitions of Theory-Guided Dimensions of Well-Being</i></p> <p>Table 3 <i>Sociodemographic Characteristics of BabyFirst and Comparison Families at Baseline (Continuous Variables)</i></p> <p>Table 4 <i>Internal Consistency Reliability Coefficients for Participant Measures at Baseline and 1 Year</i></p> <p>Table 5 <i>Baseline Participant Measures for BabyFirst Families and Comparison Families</i></p> <p>Table 6 <i>Intercorrelations for Baseline Child, Parent, Parent-Child, Family, and Community Variables</i></p> <p>Table 7 <i>Predictors of Parental Psychological Well-Being at 1 Year</i></p> <p>Table 8 <i>Predictors of Positive Parenting at 1 Year</i></p> <p>Table 9 <i>Predictors of Parental Depression at 1 Year</i></p> <p>Table 10 <i>Predictors of Parental Reading with Children at 1 Year</i></p> <p>Table 11 <i>Predictors of Family Social Support at 1 Year</i></p> <p>Table 12 <i>Predictors of Hostile/ineffective Parenting at 1 Year</i></p> <p>Table 13 <i>Correlations of Family and Parent Variables at Baseline with Early Alliance Ratings</i></p> <p>Table 14 <i>Early Alliance and Alliance Improvement as Predictors of Parent and Parent-Child Outcomes at 1 Year: Comparison of Results Using Available Data and Imputed Data</i></p>	<p style="margin-top: 100px;">18</p> <p>66</p> <p>86</p> <p>89</p> <p>92</p> <p>93</p> <p>95</p> <p>97</p> <p>99</p> <p>101</p> <p>103</p> <p>105</p> <p>111</p> <p>119</p>
---	---

List of Figures

Figure 1. BabyFirst flow chart: From universal screening to program evaluation.	33
Figure 2. Assignment to BabyFirst and non-BabyFirst groups using cut-off score.	37
Figure 3. Conceptual framework of BabyFirst program (after Duggan et al., 1999).	39
Figure 4. Illustration of program effectiveness in the regression-discontinuity design.	59
Figure 5. Illustration of program non-effectiveness in the regression-discontinuity design.	59

Effectiveness of Early Intervention for Infants and their Families: Relating the Working Alliance to Program Outcomes

Many earnest workers in charity feel that social conditions could be wonderfully improved if, to every family in distress, could be sent a volunteer visitor, who would seek out and, with patience and sympathy, strive to remove the causes of need (Richmond, 1899, pp. 12-13, quoted in Gomby, Larson, Lewit, & Behrman, 1993, p. 6.)

At the turn of the twentieth century, the idea that families with very young children (i.e., under age 6 years) could benefit from early intervention¹ began to take hold in North America (Gomby et al., 1993). The development of home visiting in the United States was greatly influenced by the work of Florence Nightengale, who was gaining renown for her efforts to train both nurses and paraprofessionals ("rural health missionaries") to provide home visits to families in need (Wasik, 1993). Also renowned was Nightengale's conviction that the impact of home visiting needed to be evaluated. In her own writings she exhorted that "results shown are the only test" (Nightingale, 1894, quoted in Gomby, 1999).

Efficacy and Effectiveness of Intervention

Many researchers have distinguished the efficacy of an intervention from its effectiveness (Brooks-Gunn, 2000; Flay, 1986; Greenberg & Weissberg, 2001; Hoagwood, Hibbs, Brent, & Jensen, 1995; Hohmann & Shear, 2002; Olds, Robinson, Song, Little, & Hill, 1999; Ramey & Ramey, 2002; Streiner, 2002; Weisz, Donenberg, Han, & Weiss, 1995; Weisz & Jensen, 1999). Other researchers disagree. For example: "a lot has been said about effectiveness versus efficacy outcomes, and this becomes a very difficult issue in terms of the integrity with which programs are implemented as they are broadly disseminated. A two-level distinction between effectiveness and efficacy is unrealistic, and researchers need instead to simply start routinely reporting information on settings, implementation integrity, and staffing characteristics" (Dadds, 2001).

¹ For this study, the term "intervention" is defined as actions taken to improve the health, well-being, and/or development of others.

The present study used the following definitions of efficacy and effectiveness: *Efficacy* refers to the effects of intervention delivered under ideal research conditions, wherein “the investigator exerts considerable control over sample selection, delivery of the intervention, and the settings in which the intervention takes place” (Weisz & Jensen, 1999, p. 125), usually in randomized experiments of model programs. In this phase of research, “conditions are highly conducive to the intervention being delivered with great fidelity to the model, and the population served is at risk for the problem to be prevented and quite receptive to participating in the program. This phase of research asks whether, under optimal conditions, the program can produce the desired outcome” (Olds, Robinson, et al., 1999, p. 49). *Effectiveness* refers to the effects of intervention delivered under real-world conditions of practice, usually to heterogeneous samples in everyday service delivery settings (Weisz & Jensen, 1999). This phase of research asks “whether the intervention can produce the desired results when the program is conducted under less than optimal conditions, akin to those that are likely to prevail if the program were conducted on a larger scale” (Olds, Robinson, et al., 1999, p. 49). The efficacy-effectiveness distinction can be regarded as the endpoints of a continuum, with most intervention outcome studies in the literature falling on the efficacy end of the continuum (Weisz & Jensen, 1999).

What We Know: Efficacy of Early Childhood Intervention

Today, at the turn of the twenty-first century, it is clear that Florence Nightengale's empirical test has been well met. Scores of high-quality studies support the efficacy of early childhood intervention (for reviews, see Anderson et al., 2003; Bakermans-Kranenburg, van IJzendoorn, & Juffer, 2003; Brooks-Gunn, Berlin, & Fuligni, 2000; Cowan, Powell, & Cowan, 1998; Durlak & Wells, 1997; Farran, 2000; Greenberg, Domitrovich, & Bumbarger, 2001; Hahn et al., 2003; Karoly et al., 1998; Lamb, 1998; MacLeod & Nelson, 2000; National Research Council [NRC] & Institute of Medicine [IOM], 2000; Nelson, Laurendeau, & Chamberland, 2001; Nelson, Westhues, & MacLeod, 2003; Sweet & Appelbaum, 2004; Weissberg & Greenberg, 1998; see also American Academy of Pediatrics Committee on Children with

Disabilities, 2001). In a landmark meta-analysis of 177 outcome studies, Durlak and Wells (1997) found that the efficacy of primary prevention programs for children and adolescents was comparable, in terms of the magnitude of effects (standardized mean differences or *ds* ranged from .24 to .93), to the efficacy of psychological, educational, behavioral, and medical treatments reported in an earlier meta-analysis (overall *d* = .47, *SD* = .28) (Lipsey & Wilson, 1993). Brooks-Gunn (2003) has suggested that "We have had more success in early childhood than at later points in the life course" (p. 9).

A recent meta-analysis of 34 long-term studies found that early childhood intervention effects persisted through school and adolescence: (a) effects on children's cognitive development endured during Kindergarten through grade 8 (*d* = .30); (b) effects on children's social-emotional development endured during Kindergarten through grade 8 (*d* = .27) and high school and beyond (*d* = .33); and (c) effects on parent-family wellness endured during Kindergarten through grade 8 (*d* = .30) (Nelson et al., 2003). This meta-analysis identified several characteristics of early childhood programs associated with enduring effects: (a) programs with direct teaching components and follow-through educational components through elementary school predicted larger enduring effects on children's cognitive development; (b) longer program duration (more than 1 year) predicted larger enduring effects on children's cognitive and social-emotional development; and (c) greater program intensity (more than 300 sessions) predicted larger enduring effects on children's cognitive development and parent-family wellness (Nelson et al., 2003). Finally, long-term effects of early childhood intervention were larger for programs that served predominantly African-American children and families (average *d* = .45), compared to those of other ethnoracial backgrounds (average *d* = .14). Nelson et al. (2003) suggested that this may indicate that early childhood interventions benefit those who are more disadvantaged. However, because ethnoracial status was associated with program intensity, it was unclear which was more important in predicting the long-term success of early childhood programs (Nelson et al., 2003).

There are several exemplary efficacy studies in the research literature, including the Carolina Abecedarian Project (Campbell & Ramey, 1994; Ramey et al., 2000), the High/Scope Perry Preschool Program (Schweinhart & Weikart, 1980; Schweinhart et al., in press), and the Infant Health and Development Program (IHDP, 1990; McCarton, Brooks-Gunn, Wallace, & Bauer, 1997). A fourth set of exemplary efficacy studies focus on the Nurse-Family Partnership (Olds, 2002), which will be described in a later section (see *A Brief Overview of Home Visiting Programs* below).

The Carolina Abecedarian Project was a prospective randomized experiment ($N = 111$, 57 children in the intervention group, 54 in the non-treated control group) that evaluated the potential benefits of early childhood education for children from low-income families (98% African-American). Children in the intervention group received full-time, high-quality educational intervention in a childcare setting from infancy through age 5 years. Each child had an individualized prescription of educational activities, which consisted of “games” incorporated into the child’s day. These activities focused on social, emotional, and cognitive areas of development, with a special emphasis on language. Children’s progress was monitored over time, with follow-up studies conducted at ages 12, 15, and 21 years. Longitudinal findings demonstrated that statistically significant ($p \leq .05$), practically important, and long-lasting benefits were associated with the program from the early years through to age 21 (Campbell, Pungello, Miller-Johnson, Burchinal, & Ramey, 2001; Campbell & Ramey, 1994, 1995; Campbell, Ramey, Pungello, Sparling, & Miller-Johnson, 2002; Ramey et al., 2000). These benefits included higher cognitive test scores from age 3 to 21 (age-appropriate Stanford-Binet and Wechsler scales), higher academic achievement in reading and math from age 8 to age 21 (age-appropriate Woodcock-Johnson tests), decreased likelihood of grade retention (47% less likely) and special education placement (75% less likely) at age 15, more completed years of education at age 21, and a three-fold increased likelihood to attend a four-year college in intervention group children (36%) compared to control group children (12%) at age 21. The

program also reduced the rate of teenage births: As young adults, the intervention group was, on average, one year older (19.1 years) when their first child was born, compared with those in the control group (17.7 years). Mothers of children in the intervention group achieved higher educational and employment status than mothers of children in the control group, and results were especially pronounced for teen mothers. Over the lifetime of the program participants, a benefit-cost analysis (in U.S. dollars) showed that the Carolina Abecedarian Project returned \$3.78 in child, family, and public benefits for every dollar invested in the program (Masse & Barnett, 2002). The Carolina Abecedarian Project is notable for its scientifically rigorous design (prospective randomized trial); timing, intensity, and duration of intervention (beginning in infancy, lasting all-day, year-round, over 5 years; to date, no other early intervention program reporting post-high school results has provided treatment intensity and duration of this magnitude, Campbell et al., 2001); the range and magnitude of intervention effects it achieved (medium to large effect sizes in cognitive, academic, educational, and parental life course outcomes); its multiple longitudinal assessments (ages 8, 12, 15, and 21); and its high follow-up rate to age 21 years (104 of the original 111 participants, or 94%). A major limitation was the small study sample size, offset in part by extensive longitudinal data. The authors state that findings generalize to African-Americans from low-income families (Campbell et al., 2001).

Perhaps the most well-known efficacy study of early childhood intervention is the *High/Scope Perry Preschool Program*. This study assessed whether high-quality, active learning preschool programs could provide both short-term and long-term benefits to children living in poverty and at risk for school failure. From 1962 through 1967, a group of 123 African-American children, aged 3 to 4 years, living in Ypsilanti, Michigan, were randomly assigned to either an intervention group ($n = 58$) based on the High/Scope active learning approach or a control group ($n = 65$) that received no preschool program. Intervention included daily classes (2.5 hours each) comprising child-planned (i.e., self-initiated) learning activities with low staff-child ratios (1:6) and weekly home visits (1.5 hours each) to engage parents in their children's learning.

Intervention lasted over two 30-week school years. Over three decades, researchers followed the children's outcomes, annually from ages 3 to 11 years (e.g., Weikart, Bond, & McNeil, 1978), at ages 14-15 years (Schweinhart & Weikart, 1980), at age 19 years (Berrueta-Clement, Schweinhart, Barnett, Epstein, & Weikart, 1984), at age 27 years (Schweinhart, Barnes, Weikart, Barnett, & Epstein, 1993) and most recently at age 40 years (Schweinhart, 2004; Schweinhart et al., *in press*), and demonstrated statistically significant ($p \leq .05$), practically important, and long-lasting program effects. Over the years, compared with the control group, the intervention group had higher scores in intellectual performance (IQ) from the end of the first year of the preschool program (67% vs. 28% with IQ of 90 or higher) to the end of first grade at age 7 (Weikart et al., 1978), higher school achievement at age 14 (49% vs. 15%) (Schweinhart & Weikart, 1980), and higher general literacy at age 19 (61% vs. 38%) (Berrueta-Clement et al., 1984). Moreover, the intervention group spent fewer years in special education programs (15% vs. 34%) and had higher rates of homework completion at age 15 (68% vs. 40%) (Schweinhart & Weikart, 1980). Compared with the control group at age 27, the intervention group had higher monthly earnings (29% vs. 7% earning \$2,000 or more per month), higher percentages of home ownership (36% vs. 13%) and second-car ownership (30% vs. 13%), higher levels of completed schooling (71% vs. 54% completing grade 12 or higher), lower rates of welfare or social service use over the previous 10 years (59% vs. 80%), and fewer arrests (7% vs. 35% having 5 or more arrests), including fewer arrests for crimes of drug-making or drug-dealing (7% vs. 25%) (Schweinhart et al., 1993). Over the lifetime of the program participants through age 27, a benefit-cost analysis (in U.S. dollars) showed that the Perry Preschool Program returned a remarkable \$7.16 in public benefits for every dollar invested in the program (Schweinhart et al., 1993). This efficacy study is notable for its high degree of internal validity, its multiple longitudinal assessments over three decades, and its high follow-up rate (117 of the original 123 participants or 95%) at age 27. A forthcoming follow-up at age 40 will show that these outcomes endured through the subsequent

decade, and returned a staggering \$17.07 in benefits for every dollar invested in the program (Schweinhart, 2004; Schweinhart et al., in press).

The *Infant Health and Development Program (IHDP)* began in 1984 and was “the largest multisite randomized trial testing the efficacy of early childhood intervention upon children’s well-being” during the twentieth century (Brooks-Gunn, 2003, p. 4). Across eight U.S. sites serving diverse populations, a random sample of 985 low birthweight (LBW), premature (≤ 37 weeks gestation) infants, stratified by two LBW groups (lighter or ≤ 2000 g and heavier or 2001 – 2500 g) were given pediatric care, and parents were referred to services available in their communities to help them and their newborns. Of the sample, 377 infants and their parents were randomly assigned to the intervention group, which participated in a three-year early education program, consisting of home visits from trained workers who facilitated parent-child activities and provided family support (from neonatal discharge to age 3). Children in the intervention group also participated in day care at special centres where they further engaged in educational activities (from age 1 to age 3) and parents also enrolled in support groups. The IHDP measured the children’s outcomes at ages 1, 3, 5, and 8 years. At age 3, compared to the comparison group ($n = 608$), children in the intervention group scored higher on IQ tests (14 points higher, $p \leq .05$), demonstrated better language and spatial skills, and had fewer behaviour problems (IHDP, 1990). By age 5 years, the intervention group had full-scale IQ scores similar to children in the comparison group² and there were no statistically significant differences between groups in terms of behaviour or health status (Brooks-Gunn et al., 1994). By age 8 years, this pattern of results continued, with no statistically significant differences³ between groups in terms of cognitive and academic skills (McCarton et al., 1997). However, similar to the Carolina Abecedarian and Perry

² However, children in the heavier LBW intervention subgroup (2001 to 2500 g) had higher full-scale IQ scores (3.7 points) and verbal IQ scores (4.2 points) (Brooks-Gunn et al., 1994).

³ The heavier LBW intervention subgroup had higher full-scale IQ scores (4.4 points), verbal IQ scores (4.2 points), performance IQ scores (3.9 points) mathematics achievement scores (4.8 points) and receptive vocabulary scores (6.7 points) (McCarton et al., 1997).

Preschool programs, a long-term follow-up of the IHDP⁴ has shown remarkable longitudinal outcomes that were not anticipated at the beginning of the efficacy studies, including higher rates of school completion, and reduced crime and delinquency in late adolescence and early adulthood. A recent effectiveness study of the Chicago Child-Parent Centres has also found similar sustained effects of early intervention into young adulthood. This study is described in the next section on effective studies.

In a recent paper entitled, "Early Childhood Education: The Journey from Efficacy Research to Effective, Everyday Practice," Ramey and Ramey (2002) concluded that "the efficacy question of whether early childhood education can have a meaningful positive impact on at-risk children has been answered with a clear yes" (p. 20, italics in original). A similar conclusion was reached by the National Research Council's Committee to Integrate the Science of Early Childhood Development (NRC & IOM, 2000). In a commentary on a recent systematic review of early childhood interventions (Anderson et al., 2003), the co-editor of the landmark *From Neurons to Neighborhoods* report on early childhood development (NRC & IOM, 2000), impatiently expressed:

As one steps back and examines the empiric knowledge base from a broader perspective, however, it is difficult to avoid the nagging conclusion that the wheels of progress have been spinning more than rolling . . . Why, after almost 4 decades of carefully evaluated demonstration projects and multiple literature reviews, do we continue to feel a need to repeatedly address the generic question of whether early childhood interventions can produce positive developmental outcomes? The consistency of the research findings . . . make the persistence of this question even more curious (Shonkoff, 2003, p. 4).

In summary, there is considerable evidence that early childhood interventions studied under carefully controlled research conditions (efficacy) can be beneficial to children's development.

⁴ In a recent interview, Dr. Marie McCormick, the current co-principal investigator of the IHDP, indicated that a follow-up of the IHDP children in adolescence (not yet published) found less delinquency and higher rates of school completion in the intervention group ("Bringing Up Baby – Better," 2002). The IHDP cohort turned age 17 in 2001 ("Grant Given to Study," 2001).

What We Don't Know: Effectiveness of Early Childhood Intervention

In contrast to efficacy, very little is known about the real-world effectiveness of early childhood intervention (Black, 2000; Masten, 1999; NRC & IOM, 2000). Filling this knowledge gap represents a major task for future research. “Despite repeated calls for a more differentiated approach to the question of program effectiveness, we have made relatively modest progress in our understanding of how impacts vary by program model and for different types of children and families” (Shonkoff, 2003, p. 4). This is due in part to widely disseminated interventions that lack rigorous evaluation or demonstrated efficacy. Where evaluation is included, adequate comparison groups are often absent (Duggan et al., 2000; Gomby, 1999). Our limited knowledge about effectiveness is also due in part to the current state of early childhood intervention research. Questions of effectiveness are usually not addressed until after efficacy has been established, and clear evidence of efficacy, especially from long-term follow-up studies of program participant outcomes, has appeared only in recent years (e.g., Albee & Gullotta, 1997; Beckwith, 2000; Cowen, 2000b; Guralnick, 1997; Nelson et al., 2003; Yoshikawa, 2000). Early childhood intervention has “come of age” (Cowen, 1997, 2000a). As the field entered a new millennium, high-quality studies of the effects of early childhood intervention delivered on a large scale under real-world conditions of service delivery (effectiveness) began to emerge. These effectiveness studies include the Chicago Child-Parent Centres (Reynolds, 2000), the Early Head Start National Research and Evaluation Project (Administration on Children, Youth and Families [ACYF], 2002), and the National Institute of Child Health and Human Development (NICHD) Study of Early Child Care (NICHD Early Child Care Research Network [ECCRN], 1993). A fourth effectiveness study—a randomized evaluation of the Hawaii Healthy Start home visiting program—is described in a later section (see *A Brief Overview of Home Visiting Programs*).

The long-term effectiveness of a U.S. centre-based preschool and school-based intervention program for urban low-income children was recently assessed in a 15-year follow-up of a nonrandomized, matched-group cohort of 1539 low-income, mostly African-American

children born in 1980, who were enrolled in a variety of early childhood programs in 25 sites in Chicago (Reynolds, Temple, Robertson, & Mann, 2001). The *Chicago Child-Parent Centers (CPCs)* ($n = 989$ children) provided comprehensive education, family, and health services, and included half-day preschool at ages 3 to 4 years, half- or full-day kindergarten, and school-age services in elementary schools at ages 6 to 9 years. The comparison group ($n = 550$ children) consisted of children who participated in alternative early childhood programs (full-day kindergarten): 374 children in the preschool comparison group from five randomly selected schools plus two other schools that provided full-day kindergarten and additional instructional resources and 176 children who attended full-day kindergartens in six CPCs without preschool participation. Age 21 follow-up comparisons of these two groups (Reynolds et al., 2001) showed several statistically significant ($p < .05$), practically important, and enduring program effects: children in the intervention group had a higher rate of high school completion (50% vs. 39%), more years of completed education (10.6 years vs. 10.2 years), lower rates of juvenile arrests (17% vs. 25%) and violent arrests (9% vs. 15%), and lower rates of school dropout (47% vs. 55%). The CPC interventions were associated with lower rates of grade retention and special education placement. Further analyses showed that the CPCs prevented later child abuse and neglect (based on substantiated referrals to child protective services and the juvenile court from ages 4-17 years) by 33% to 52% (Reynolds & Robertson, 2003). Results suggested that increased parental involvement in school and reduced school mobility mediated these intervention effects (Reynolds & Robertson, 2003). Over the lifetime of the program participants through age 21, a benefit-cost analysis (in U.S. dollars) showed that the Chicago Child-Parent Centres returned up to \$7.14 in public benefits for every dollar invested in the program (Reynolds, Temple, Robertson, & Mann, 2002). Delivered through the 1980s, the CPCs are notable because most research on the long-term effects of early childhood interventions are efficacy studies of demonstration programs that began in the 1960s (e.g., see Carolina Abecedarian Project and High/Scope Perry Preschool Program above) rather than effectiveness studies of contemporary,

large-scale public programs. The CPC study is, therefore, the first effectiveness study of a contemporary, early childhood intervention to report long-term outcomes in early adulthood (Zigler & Styfco, 2001). However, it is important to note that even the most effective early childhood intervention is “not an inoculation against all past and future developmental risks imposed by living in poverty. Just as 1 year of good nutrition is not expected to make a child healthy for life, it is foolish to assume that any brief intervention will lead to academic success and a good-paying job” (Zigler & Styfco, 2001, p. 2379).

The *Early Head Start National Research and Evaluation Project* (ACYF, 2002) is the largest effectiveness study of early childhood intervention conducted to date (Brooks-Gunn, 2000). In 1994, Early Head Start was designed as a “two-generation program” provided during the first three years of life to enhance children’s development and health, strengthen family and community partnerships, and support the staff delivering new high-quality services to low-income families with pregnant women, infants, or toddlers (ACYF, 2002). Early Head Start grantees are charged with tailoring their program services to meet the needs of families in their local communities and may select among program options specified in the Early Head Start program standards (home-based, centre-based, combination, and locally designed options). The first 143 Early Head Start programs were implemented in 1995 and 1996. From 1996 through 1998, 17 programs from across the United States were selected to participate in a rigorous, large-scale randomized evaluation. The 17 sites include all the major program approaches, are located in all regions of the U.S. in both urban and rural settings, and serve highly diverse populations. A research sample of 3,001 families from across the 17 sites was randomly assigned to the Early Head Start intervention group ($n = 1,513$) or the control group ($n = 1,488$), which could receive other services available to families in the community. Comprehensive, longitudinal, multimethod data were collected on baseline family characteristics; service use (6, 15, and 26 months after random assignment); program implementation; family partnerships (e.g., parent involvement in program policy, operations, or governance); program quality (of both home-based and centre-

based interventions); and parent, parent-child, and child outcomes (at ages 14, 24, and 36 months). Outcome measures included direct child assessments, observations of children's behaviour by in-person interviewers, standardized ratings of videotaped parent-child interactions, parent ratings of children's behaviours, and parent self-reports of their own behaviours, attitudes, and circumstances (ACYF, 2002).

Early Head Start evaluation results showed a consistent pattern of statistically significant ($p \leq .10$) and positive impacts across a range of outcomes when children were ages 2 and 3 years, with larger impacts in key subgroups. Effect sizes were modest, ranging from 10% to 20% improvement, but higher in some subgroups, ranging from 20% to 50% improvement (ACYF, 2002) as follows. Positive child outcomes at ages 2 and 3 years included improved cognitive development (Bayley Scales of Infant Development), language development (Peabody Picture Vocabulary Test), and social-emotional development (including observed increases in children's engagement with their parents, increased attentiveness during play, and lower parent ratings of children's aggressive behaviour). Positive parent-child outcomes included improvements in observed emotional support and Home Observation for Measurement of the Environment (HOME) scores regarding language and learning support, higher rates of daily reading with children, and lower rates of negative parenting behaviours (including detachment, spanking, and other punitive strategies). Positive parent outcomes included improvements in self-sufficiency (increased participation in education, job training, and employment during the first 26 months after random assignment, but no significant improvements in parents' income), maternal life course (decreased likelihood of subsequent births during the first 2 years after enrollment), and father-child interactions (decreased likelihood of spanking, intrusiveness, and higher levels of children's engagement with their fathers and attentiveness during play). The subgroups in which impacts were relatively large (effect sizes ranging from 20% to 50% improvement) included mixed-approach programs (combination of home-based and centre-based services), African-American families, Hispanic families, families who enrolled during pregnancy, and families with

a moderately high (vs. either a low or very high) number of demographic risk factors. In addition to income (all Early Head Start families were low-income families), five demographic risk factors were considered: (a) being a single parent, (b) receiving public assistance, (c) being neither employed nor in school or job training, (d) being a teenage parent, and (e) lacking a high school diploma or equivalent. Low risk was defined as less than 3 risk factors, moderate risk as 3 risk factors, and high risk as greater than 3 risk factors. Two subgroups that other programs have found difficult to serve also benefited from participation in Early Head Start: parents at risk for depression (decreases in parental depressive symptoms, improvements in parenting outcomes and children's social-emotional outcomes) and teenage parents (increased school attendance of teenage parents, improved child outcomes). Early Head Start did not result in improvements in parental safety practices (e.g., correct use of child car seats) and did not increase parents' access to needed mental health services or improve overall parent mental health. The longitudinal follow-up study of Early Head Start, to be completed in 2004, will assess outcomes as children enter Kindergarten (ACYF, 2002).

For the highest-risk families, participation in Early Head Start had a wide range of unfavourable effects on children's cognitive and language development, social and emotional development, quality of the home environment, parenting behaviour, parent knowledge of safety practices and discipline strategies, parent health, and paternal participation. In reviewing these harmful effects, the Early Head Start authors posited some possible explanations, including (a) problematic working relationships between parents and program providers and (b) problems in program implementation:

The services provided by Early Head Start programs may not be sufficient to meet the needs of the families at greatest risk and may not be as effective as other community programs that target these families. The difficulties program staff reported in working with these families may be reflected in the less-favorable outcomes. In addition, the families with the most risks were more likely to be in home-based or mixed-approach programs that were not fully implemented early, and it is possible that the staff turnover and disruptions in staff-family relationships experienced in some of these programs had an adverse effect on the most vulnerable families (ACYF, 2002, p. 344).

Finally, the Early Head Start authors (ACYF, 2002) also posited a “major life change hypothesis” as an explanation for the unfavourable program impacts:

This hypothesis suggests that low-income families who have experienced high levels of instability, change, and risk may be overwhelmed by the changes that a new program introduces into their lives, even though the program is designed to help. As a result, the program requirements may create unintended negative consequences for these families (p. 355).

It may be noted that these different explanations are not mutually exclusive. The wide range of negative impacts on child, parent, parent-child, and family outcomes could indicate that several factors may have contributed to the negative and unintended consequences of participation for the highest-risk families in the Early Head Start program. In summary, the Early Head Start study provides the best available data on the effectiveness of early intervention during the first three years of life, and serves as an important reminder of some significant challenges in serving the highest-risk families under real-world conditions, including the possibility of harmful effects.

The NICHD Study of Early Child Care is a prospective, longitudinal, multisite study of approximately 1,300 children, with data collected from birth through grade 1 (NICHD ECCRN, 1993). It is the largest and most rigorous effectiveness study conducted to date of the relationships between early nonparental child care and child development. The majority of infants in the U.S. now spend time in nonparental child care, typically beginning in the early months of life (NICHD ECCRN, 1996). The NICHD sample was socioeconomically and geographically diverse (10 U.S. sites), and was drawn from area hospitals at birth rather than child care settings, permitting the observation of children in the entire range of child care settings used by parents. It should be noted that, unlike other programs, there was no consistent or planned intervention across sites. Child care characteristics were assessed on multiple occasions using direct and intensive observation. Information about the quantity of early child care, defined as regular care each week by anyone other than the mother, was collected through telephone interviews with mothers at 3-month intervals through age 36 months. Cumulative amount of time in nonparental care was calculated as the mean number of hours per week (NICHD ECCRN, 1993). The quality

of early child care was measured through observational assessments of the primary nonparental care arrangements at 6, 15, 24, and 36 months. Composite measures of quality were calculated for each age level. At 6, 15, and 24 months, these composite scores were the mean of five 4-point qualitative ratings of caregiver behaviour (sensitivity to child's non-distress signals, stimulation of cognitive development, positive regard for child, emotional detachment, flatness of affect). At 36 months, these five scales plus two additional scales (fosters child's exploration and fosters child's intrusiveness) were included in the composite measure of quality (NICHD ECCRN, 1993, 1996).

NICHD study findings have confirmed that the quantity and quality of early nonparental child care predict maternal sensitivity and child engagement at ages 15, 24, and 36 months (NICHD ECCRN, 1999); attachment insecurity under conditions of low maternal sensitivity and responsiveness at age 15 months (NICHD ECCRN, 1997) and age 36 months (NICHD ECCRN, 2001b); social competence and behavior problems at ages 24 and 36 months (NICHD ECCRN, 1998a, 1998b); cognitive and language development at ages 15, 24, and 36 months (NICHD ECCRN, 2000); and children's peer interaction at ages 24 and 36 months (NICHD ECCRN, 2001a). These findings are consistent with the larger extant research literature on the positive effects of high quality child care (Lamb, 1998; NRC & IOM, 2000).

Notwithstanding the wide-ranging and positive effects of early nonparental child care found in the NICHD study, other findings from the NICHD study show that the quality of maternal caregiving remains the strongest predictor of a child's cognitive competence and a moderate predictor of social competence (NICHD ECCRN, 2002). The NICHD study found no main effects of child care on the security of infant-mother attachment, i.e., nonparental child care did not appear to detrimentally affect the relational bond between parent and child, a longstanding concern in the literature (Lamb, 1998; NRC & IOM, 2000).

In summary, the new Chicago CPCs, Early Head Start, and NICHD studies provide the first rigorous data on the effectiveness of early childhood intervention. Findings suggest that interventions which begin early in childhood, are well implemented with high standards of

quality, and are sustained over several years can result in a wide range of positive child and family outcomes. At the same time, the available data also suggest that the effect sizes of programs delivered under real-world conditions (in effectiveness studies) are more modest in magnitude than those delivered under optimal research conditions (in efficacy studies). Even so, the Chicago CPCs study suggests that programs delivered under real-world conditions can have sustained effects over decades into early adulthood. The Early Head Start study provides a reminder that effective early childhood intervention may also lead to unintended negative effects for the highest-risk families. As noted above at the beginning of this section, determining the real-world effectiveness of early childhood interventions is a major task for future research. As the Chicago CPCs, Early Head Start, and NICHD effectiveness studies continue, a second major task for future research is to understand how early childhood intervention programs lead to positive outcomes by identifying their effective ingredients.

What We Don't Know: Effective Ingredients of Early Childhood Intervention

Currently, little is known about the *effective ingredients* of early childhood intervention: those program elements, features, characteristics, or components that are most important in determining outcomes (Benard, 1999; Berlin, O'Neal, & Brooks-Gunn, 1998; Durlak, 2003; Gillham, Shatté, & Reivich, 2001; Kitzman, 2004; Masten & Coatsworth, 1998; Tolan, Hanish, McKay, & Dickey, 2002; Werner & Johnson, 1999; Westhues, Nelson, & MacLeod, 2003). Korfmacher (2001) lamented: "You would think that after 30 years of studying early childhood intervention programs there would be some pretty impressive things to say about these issues. . . . Sadly, no such scenario exists in the real world" (p. 1). Campbell et al. (2001) argued that future research "has to move beyond outcome measure follow-up designs and examine more closely the intervening process variables that mediate the effects of early intervention and enable effects to be maintained over time" (p. 240). For successful replication and dissemination of efficacious early childhood interventions, effectiveness must be established and effective ingredients must be identified (Olds, O'Brien, Racine, Glazer, & Kitzman, 1998).

What makes early childhood interventions work? Berlin et al. (1998) suggested a useful tripartite framework for understanding how early childhood interventions work that includes characteristics of the program, its participants, and their interaction. This framework is presented in Table 1. *Program characteristics* include the program approach (theory of change, targets of service, curricula or specific activities, method of service delivery, quantity and timing of services), the program staff (personal characteristics and training, program-staff relationship), and the cultural relevance of services. *Participant characteristics* include child characteristics, family characteristics, community characteristics, and cumulative risk, i.e., “the accumulation of biological and environmental conditions that have been shown to decrease children’s opportunities for achieving their full potential and increase the likelihood of developmental delays and problems” (Berlin et al., 1998, p. 11). Finally, *characteristics of interactions between program and participants* include participant engagement and participant-staff relationships, as well as the match between participants’ needs and program services.

From the foregoing reviews resound two recommendations for the next generation of early childhood intervention research: the need to (a) evaluate the effectiveness of early childhood intervention under real-world conditions and (b) identify its effective ingredients. Studying effectiveness (“Does the program work in the real world?”) and effective ingredients (“What makes the program work?”) are distinct yet interrelated research agendas. Ideally, effective ingredients are a focus in efficacy studies, where they can be identified and replicated under optimal research conditions. Effectiveness studies then follow, wherein these effective ingredients are disseminated under real-world conditions of practice, to test for replication of outcomes found in efficacy studies. More common, however, is widespread dissemination of interventions with scarce evidence of efficacy, much less of effectiveness or of effective ingredients (NRC & IOM, 2000). This has been especially true of home visiting programs, for reasons outlined in the next section, including contradictions and lack of clarity among studies, and tension between advocacy and science (Olds, O’Brien, et al., 1998; see also Chaffin, 2004).

Table 1

A Framework for Understanding How Early Interventions Work: The Program, its Participants, and their Interaction

Characteristics of the Program	Characteristics of the Participants	Interactions between Program and Participants
Program approach <ul style="list-style-type: none"> • Theory of change • Targets of service • Curricula or specific activities • Method of service delivery • Quantity and timing of services 	Child characteristics Family characteristics Community characteristics Cumulative risk	Participant engagement and participant-staff relationships Match between participants' needs and program services
Program staff <ul style="list-style-type: none"> • Personal characteristics and training • Program-staff relationship 		
Cultural relevance of services		

Source: Berlin et al. (1998).

A Brief Overview of Home Visiting Programs

"Home visiting is not a single, uniform intervention, but rather a strategy for service delivery" (Gomby, Culross, & Behrman, 1999, p. 6). As in most early childhood interventions, home visiting programs generally seek to reduce risk factors and promote protective factors for healthy child development. A *risk factor* generally refers to a variable that increases the odds of a negative outcome (e.g., child maltreatment) and/or decreases the odds of a positive outcome (e.g., improved cognitive development). In contrast, a *protective factor* generally refers to a variable that decreases the odds of a negative outcome and/or increases the odds of a positive outcome. Kraemer et al. (1997) have provided a useful framework for conceptualizing risk (and protective) factors. Most important for studies of early childhood intervention is their concept of a causal risk factor—a risk factor that, when changed, is shown to change the outcome (Kraemer et al., 1997). Effective early childhood interventions are successful in changing causal risk and protective factors for healthy child development (Kraemer, Stice, Kazdin, Offord, & Kupfer, 2001).

Home visiting programs share some common elements: the goal of improving the lives of children, the use of service providers who enter the homes of families with young children, and the encouragement of changes in the attitudes, knowledge, and/or behaviour of parents (Gomby et al., 1999; Hahn et al., 2003; NRC & IOM, 1999; Sweet & Appelbaum, 2004). To facilitate such changes, most home visiting programs use interventions such as the provision of social support, practical assistance (e.g., referral to community resources), and education regarding parenting and/or child development. Most programs theorize that the creation of a trusting relationship between parent and home visitor "can be a first step in developing the parent's ability to form and sustain secure relationships with others, including their own children" (Gomby et al., 1999, p. 7).⁵ Beyond these common characteristics, home visiting programs differ in their specific goals, levels

⁵ This theoretically crucial proposition is elaborated in a later section of the present study.

of service, duration of service, staffing, and clientele (Gomby et al., 1999; Hahn et al., 2003; NRC & IOM, 1999; Sweet & Appelbaum, 2004).

Early literature reviews of home visiting in the 1970s suggested a lack of promise, followed by reviews in the 1980s that overstated the evidence for positive effects (Olds, O'Brien, et al., 1998). Subsequent reviews (e.g., Olds & Kitzman, 1993) did find accumulating evidence of promising programs. In the 1990s, several major policy organizations in the U.S. began national efforts to prevent child maltreatment, including the endorsement and promotion of home visiting programs that lacked a rigorous evidence base. "In their understandable zeal to help children and families, advocates for particular health and social welfare programs often overstate[d] the potential benefits of these services" (Olds, O'Brien, et al., 1998, p. 90; see also Chaffin, 2004; Duggan et al., 2000). For their part, policymakers must balance the availability of definitive evidence against the necessity of immediate action, whereas researchers must recognize the pressing information needs of policymakers, as well as the service needs of children and families (Olds, O'Brien, et al., 1998). A recent commentary on home visiting programs summarizes these fundamental and ongoing tensions:

While it is hard to find fault with the good intention that every home visiting program has at its core—to use home visitors to help improve the lives of at risk children and their parents—it is equally difficult to deny the fact that the majority of these programs do not work as advertised (Alper, 2002).

A review of recent program evaluations of home visiting models (Gomby et al., 1999) included six representative programs. All programs focus on improving parenting skills to facilitate healthy child development and most seek to prevent child maltreatment. Four of the programs are model or demonstration programs and comprise efficacy research: the Home Instruction Program for Preschool Youngsters, the Nurse-Family Partnership, the Parents as Teachers program, and the Comprehensive Child Development Program. The other two programs are larger in scale and widely delivered in the field and, thus, comprise effectiveness research: the Hawaii Healthy Start program and Healthy Families America.

As the Nurse-Family Partnership efficacy studies are exemplary, they will be described in greater detail in the next section (see *Efficacy Studies of Home Visiting Programs* below). The Hawaii Healthy Start program model has been replicated in Manitoba as the BabyFirst program. Because the BabyFirst program was the source of data for the present study, the Hawaii Healthy Start theoretical model is described in greater detail in a subsequent section (see *Program Theoretical Model* below).

Efficacy Studies of Home Visiting Programs

The *Home Instruction Program for Preschool Youngsters (HIPPY)* is a two-year, home-based, early education program designed to help parents with limited formal education to prepare their 3- to 5-year-old children for kindergarten and first grade (Baker, Piotrkowski, & Brooks-Gunn, 1999). Based on the work of Lombard (1981), the founder of HIPPY, the program attempts to enhance the home literacy environment, the quality of parent-child verbal interaction, and parents' ability to help their children learn, through home visits by paraprofessionals. HIPPY programs are usually funded and administered by local agencies (e.g., public schools, community organizations) (Baker et al., 1999).

The first of two efficacy studies of HIPPY found that, compared to children in the control group, children who had been enrolled in HIPPY scored higher on measures of cognitive skills, classroom adaptation, and standardized reading. Further, HIPPY children were more likely to be promoted to first grade (Baker et al., 1999). However, the second efficacy study failed to replicate these results, and found that children in the control group outperformed HIPPY children on school readiness and standardized assessment (Baker et al., 1999). The authors conducted analyses to determine whether the samples in the first and second studies differed in some way that might explain their different results. The authors searched for differential attrition rates, variations in program implementation, and differences in child and family background characteristics. However, none of these possible explanations was supported by their analyses and

the authors concluded that further research is needed to clarify the mixed results of the two studies (Baker et al., 1999).

The *Parents as Teachers (PAT)* program promotes the development of children from birth to age three. PAT is a parent-education program that includes home visiting. Through home visits, professional "parent educators" help parents enhance their parenting skills and knowledge of child development to help prepare their children for school. Two efficacy studies were conducted, one with a largely Latino group of families, the other with a group of teen parents and their children (Wagner & Clayton, 1999). These studies revealed small and inconsistent positive effects on parent knowledge, attitudes, and behaviour, with no gains in children's health or development (Wagner & Clayton, 1999). Some sample characteristics may have limited the efficacy of the PAT program, including the poverty, household instability, and poor education of both groups, and the age and developmental stage of the teen parents (Wagner & Clayton, 1999).

The *Comprehensive Child Development Program (CCDP)* was a five-year, U.S. federal demonstration program that provided case management and early childhood education through home visiting with 4,410 poor families in 24 sites to promote children's development, parents' ability to parent, and family self-sufficiency. Paraprofessional CCDP case managers connected families to a wide range of educational, health, and social services. They also conducted home visits to facilitate early childhood education (ECE), either through home-based parent education regarding ECE (e.g., parent-child activities) or referral to centre-based ECE. Comparisons of CCDP families ($n = 2,213$) and control families ($n = 2,197$) in the evaluation of the CCDP found no statistically significant impacts in any of the domains assessed: early childhood education, child and family health, parenting education, family economic self-sufficiency, or maternal life course (St. Pierre & Layzer, 1999). The authors concluded that, in light of these results, "it appears that social programs will be more effective if they focus on direct service provision rather than on the organization of existing services" (St. Pierre & Layzer, 1999, p. 149).

The *Nurse-Family Partnership (NFP)* (previously called the *Prenatal and Early Intervention Program* and the *Nurse Home Visitation Program*) is a university-based demonstration program developed in Elmira, New York, and studied again in Memphis, Tennessee and Denver, Colorado. It is now being replicated nationally across the United States (Olds, Henderson, et al., 1999; Olds, 2004). The NFP is perhaps the best-known home visiting program, due in no small part to its rigorous, long-term program of efficacy research. This includes (a) an extensive 15-year follow-up of study participants in the first randomized trial of the NFP in Elmira, New York (Olds et al. 1997; Olds, Henderson, et al., 1998) and a study of moderating variables (Eckenrode et al., 2000); (b) a second randomized trial replicating the program with a different population of families in Memphis, Tennessee (Kitzman et al., 1997) and its 3-year follow-up (Kitzman et al., 2000); (c) a third randomized trial in Denver, Colorado that directly compared the efficacy of paraprofessional and professional home visitors (Olds et al., 2002); and (d) an ongoing study of program dissemination in 25 sites (described in Olds, O'Brien, et al., 1998, and Olds, 2002). To illustrate the efficacy of the NFP, the Elmira and Memphis trials are described below.

The NFP has its theoretical basis in ecological theory (Bronfenbrenner, 1979), self-efficacy theory (Bandura, 1977), and attachment theory (Bowlby, 1969). First, children's development is viewed as embedded within interdependent systems at the level of families, neighbourhoods, communities, and society. Second, parents are viewed as actively participating in the selection and shaping of their own life experiences and contexts. Third, children's capacity for empathy and responsiveness to their own children when they become parents is viewed as strongly influenced by the extent to which they develop secure attachments with caring, responsive parents and adults through infancy, childhood, and adolescence. These theoretical frameworks were the basis for interventions that (a) linked parents to resources in their families and communities; (b) facilitated parents' establishment of realistic goals and small, achievable

objectives; and (c) promoted sensitive, responsive, and engaged caregiving in the early years of children's lives (Olds, Henderson, et al., 1999).

During NFP home visits (75 to 90 minutes each), the nurses carried out three major activities: (a) promoting improvements in women's behaviour thought to affect pregnancy outcomes (e.g., diet, smoking, and substance use during pregnancy), their children's health and development (e.g., parent-child interaction, household safety), and parents' life course (e.g., education, work, planning future pregnancies); (b) helping women in building supportive relationships with family members and friends; and (c) linking women and their family members with other needed health and human services (Olds, Henderson, et al., 1999).

In the Elmira trial, families in the intervention group ($n = 216$) received nurse home visits during pregnancy through to the child's second birthday. Families in the control group ($n = 184$) received standard prenatal and well-child care in a clinic. Families were from a primarily white, semirural population. The 15-year follow-up of the Elmira trial reported data on 324 of the original 400 participants (81%). Results showed that, compared to the control group, women in the intervention group (a) had fewer state-verified reports of child abuse and neglect (0.29 vs. 0.54), (b) had fewer subsequent births (1.3 vs. 1.6), (c) had longer intervals between the birth of the first and second child (65 months vs. 37 months), (d) shorter durations of welfare use (60 months vs. 90 months), (e) fewer behavioural impairments due to the use of alcohol and other drugs (0.41 vs. 0.73), and (f) fewer arrests by self-report (0.18 vs. 0.58) and verified arrest records (0.16 vs. 0.90) (Olds et al., 1997). A follow-up of the children in the intervention group at age 15 years (Olds, Henderson, et al., 1998) found (a) fewer instances of running away (0.24 vs. 0.60), (b) fewer arrests (0.20 vs. 0.45), (c) fewer convictions and violations of probation (0.09 vs. 0.47), (d) fewer lifetime sex partners (0.92 vs. 2.48), (e) fewer cigarettes smoked per day (1.50 vs. 2.50), and (f) fewer days having consumed alcohol in the past 6 months (1.09 vs. 2.49). Parents of nurse-visited children reported that their children had fewer behavioural problems related to the use of alcohol and other drugs (0.15 vs. 0.34) (Olds, Henderson, et al., 1998). A benefit-cost

analysis of the Elmira trial found that by age 15, the program had returned \$5.06 in benefits for every dollar invested in the program (Karoly et al., 1998). Further analysis of the Elmira data found that the presence of domestic violence may limit the effectiveness of interventions to reduce child abuse and neglect. While home visiting through pregnancy and infancy reduced child maltreatment over a 15-year period in families with 28 or fewer incidents of domestic violence (79% of the sample), it had no such effect among families reporting more than 28 incidents of domestic violence (21% of the sample) (Eckenrode et al., 2000).

In the Memphis trial, the efficacy of the NFP was replicated in a sample of 1139 primarily African-American, low-income women. It was found to reduce pregnancy-induced hypertension, childhood injuries, and subsequent pregnancies. There were no program effects on preterm delivery or low birthweight; children's immunization rates, mental development, or behavioural problems; or mothers' education and employment (Kitzman et al., 1997). A 3-year follow-up found enduring effects in the intervention group, compared to the control group, including (a) fewer subsequent pregnancies (1.15 vs. 1.34), (b) fewer closely spaced pregnancies (0.22 vs. 0.32), (c) longer intervals between the birth of the first and second child (30.25 months vs. 26.60 months), and (d) shorter durations of welfare use (32.55 months vs. 36.19 months) and food stamp use (41.57 months vs. 45.04 months) (Kitzman et al., 2000). While these results were smaller in magnitude than those achieved in the Elmira trial, the direction of effects was consistent across the two studies (Kitzman et al., 2000).

Reflecting on the findings of the Elmira and Memphis efficacy studies, Olds and colleagues have concluded that

the use of nurses as home visitors is key; that services should be targeted to the neediest populations, rather than being offered on a universal basis; that clinically tested methods of changing health and behavioral risks should be incorporated into program protocols; and that services must be implemented with fidelity to the model tested if program benefits found in scientifically controlled studies are to be reproduced as the program is replicated in new communities (Olds, Henderson, et al., 1999, p. 44).

Some have commented that, for families living in dire conditions of socioeconomic disadvantage and risk, it is remarkable that home visits for a few hours per month over a few years have any positive or lasting effects at all (Bryant, 2003; Gomby, 2000; NRC & IOM, 1999). Home visiting may represent a case where statistically small effects are nonetheless impressive, because they are achieved “even under the most inauspicious circumstances” (Prentice & Miller, 1992, p. 163).

Effectiveness Studies of Home Visiting Programs

The *Hawaii Healthy Start (HHS)* program operates statewide in Hawaii and serves families identified, through population-based screening and assessment at birth, as highly stressed and/or at risk for child abuse (Duggan et al., 1999). HHS is a model of paraprofessional home visiting to improve family functioning, promote child health and development, and prevent child abuse and neglect (Duggan et al., 2000). Home visitors are trained paraprofessionals working under professional supervision. They are recruited from the community, for qualities considered essential for working with vulnerable families: warmth, self-assurance, cultural sensitivity, and good parenting skills (Duggan et al., 1999). The minimum education required for a home visitor is a high school diploma or equivalent (Duggan, McFarlane, et al., 2004). The minimum education for a supervisor is a master's degree in public health or a health or human service field plus 3 years of client service and administrative experience in human services programs, or a bachelor's degree plus 5 years of relevant experience (Duggan, McFarlane, et al., 2004).

Home visits are provided during the child's first 3 years of life and include direct support services, parenting education, and case management to connect families to pediatric primary care and other community resources. “Direct service includes providing emotional support to parents, encouraging them to seek needed professional help, teaching about child development, and role-modeling parenting skills and problem-solving techniques” (Duggan et al., 2000, p. 500).

The expected number of visits (1-2 hours in duration) varies by level, beginning at enrollment with weekly visits (Level 1), and sequential promotions to service levels with lower expected frequency as family functioning improves, from biweekly (Level 2), to monthly (Level

3), to quarterly (Level 4). "There are explicit criteria for promotion. For example, criteria for promotion to Level 2 include home stability with no crisis for 30 days and ability to identify a positive support system or person other than the home visitor" (Duggan, McFarlane, et al., 2004, p. 600). Level X is an option for families with whom the home visitor has great difficulty meeting. It focuses on creative outreach to establish contact with the family and does not have an expected number of visits. "A family was classified as receiving a high dose of service for a given year if the family met three criteria: (1) active in the program at the end of the year; (2) had \geq 75% of expected visits; and (3) on Level X for \leq 3 months. A family was considered to have a high dose of service for the full 3 years if it met similar criteria: (1) active in or graduated from the program at the end of the third year; (2) had \geq 75% of expected visits over the full period of enrollment; and (3) on Level X for \leq 3 months total" (Duggan, McFarlane, et al., 2004, p. 603). The HHS authors note that this measure of high dosage does not incorporate home visit content or service quality (Duggan, Fuddy, et al., 2004).

Started in 1975 as a single-site test on the island of Oahu, HHS is the oldest continuing home visitation program in the United States (Alper, 2002). Although early evaluations of HHS were problematic (e.g., lacked control groups), child abuse and neglect rates were lower after participation in the program, compared with historical maltreatment data. This limited evidence base was sufficient for the Hawaii legislature to expand the program statewide in 1989 (Alper, 2002; Duggan et al., 1999), a classic illustration of the tension between policy and science, as discussed in an earlier section of the present study (*see A Brief Overview of Home Visiting Programs above*).

A 3-year randomized, controlled effectiveness trial of HHS found that, after two years of intervention, compared with a control group ($n = 270$), parents in the HHS intervention group ($n = 373$) reported better linkage with pediatric medical care, improved parenting efficacy, decreased parenting stress, more use of nonviolent discipline, and decreased injury due to partner violence in the home (Duggan et al., 1999). No overall benefits were found in other major

outcome areas (i.e., child development, children's home learning environment, parent-child interaction, well-child health care, pediatric health care use for illness or injury, child maltreatment, maternal life skills, maternal mental health, maternal social support, or maternal substance use). However, results varied across sites according to the degree of program implementation, and some families benefited in terms of parent-child interaction and child development (Duggan et al., 1999).

In order to explain the lack of HHS impact on child maltreatment, a recent follow-up report examined the effects of HHS on eight risk factors for child maltreatment (i.e., maternal depression, severe parenting stress, poor general maternal mental health, maternal illicit drug use, maternal problem alcohol use, psychological abuse by partner, physical abuse by partner, and injury resulting from partner violence). Overall, there were no statistically significant or practically meaningful HHS impacts on any of these risk factors (Duggan, Fuddy, et al., 2004).

Of the 373 HHS group families, 84 had a high dose of service in their first year of enrollment, 55 in their second year, and 42 in their third year. There were 53 families with a high dose over all 3 years combined (Duggan, Fuddy, et al., 2004). Another recent follow-up report indicated that HHS families who received a high dose of service showed both favourable and unfavourable outcomes: increases in maternal responsiveness and acceptance on the Home Observation Measure of the Environment (HOME) but also increases in minor physical assaults, common corporal/verbal punishment, and severe physical abuse on the Conflict Tactics Scale (Duggan, McFarlane, et al., 2004). In high-dose HHS families, relative to the control group, HHS appeared to reduce the odds of two risk factors: (a) maternal problem alcohol use (adjusted odds ratio = .41, $p < .05$) and (b) physical abuse by partner (adjusted odds ratio = .51, $p < .05$) (Duggan, Fuddy, et al., 2004). The authors suggested that a mix of high attrition, home visit content, and service quality may account for the differential outcome observed across the full HHS group, the high-dose HHS subgroup, and the control group (Duggan, Fuddy, et al., 2004).

Healthy Families America (HFA) is a child abuse prevention program that evolved from Hawaii Healthy Start and is currently the subject of a pioneering, multisite research network. A synthesis of available data from 17 of the 35 ongoing effectiveness studies (including 8 randomized studies, 11 comparison group studies, and 16 single-group pretest-posttest studies) in the HFA research network suggested that HFA programs may have the most success at improving parent-child interaction (Daro & Harding, 1999). HFA appears to be less promising in other areas, with limited or mixed results in improving health care status and utilization, preventing child abuse and neglect, and improving maternal life course outcomes. To date, HFA programs have not improved children's developmental outcomes or maternal social support (Daro & Harding, 1999; Landsverk et al., 2002).

In general terms, both Hawaii Healthy Start (HHS) and Healthy Families America programs seek to prevent child abuse and neglect by reducing causal risk factors and promoting causal protective factors for child maltreatment. The programs are conceptually embedded in an ecological perspective that views human development across multiple contexts (e.g., child, parent, family, neighbourhood) (Bronfenbrenner, 1979; Masten, 2001). Because the HHS program theoretical model is important for the present research, it is described in greater detail in a later section (see *Program Theoretical Model* below).

From a province-wide replication⁶ in 1999 of the Hawaii Healthy Start program model in Manitoba—the BabyFirst program, a 3-year early childhood intervention for newborns and their families—emerged a unique opportunity to study both effectiveness and effective ingredients in early childhood intervention. The present study comprises secondary analyses of data (2000–2004) from the ongoing provincial evaluation of Manitoba's BabyFirst program.

⁶ The replication of HHS in Manitoba originally began with the creation of a small number of three-year BabyFirst research sites in 1998 (Government of Manitoba, 1998), with plans for intensive study prior to larger-scale dissemination. However, with the establishment of the Government of Canada's National Child Benefit (NCB) in 1999, a significant proportion of NCB transfers to Manitoba was reinvested in a province-wide implementation of the BabyFirst program that year.

Manitoba's BabyFirst Program

A flagship program in Manitoba's early childhood development continuum of supports, *BabyFirst* is a community-based intervention designed to support families with children up to age three years, living in conditions of risk for poor child health and development (Government of Manitoba, 2001). The primary goal of BabyFirst is to build strong families and healthy children, thereby reducing the risk for child abuse and neglect. There are four main program objectives: (a) to facilitate families' abilities to ensure the physical health and safety of their children; (b) to enable parents to develop and enhance their parenting skills and build on their strengths; (c) to support parents in nurturing and providing appropriate social, physical, and cognitive stimulation for their children; and (d) to facilitate families' connections with community resources and sense of belonging to their communities (Healthy Child Manitoba, 2000).

The program began province-wide implementation in 1999 through the public health program of Manitoba's Regional Health Authorities. It includes universal postpartum screening (about 13,000 births annually) and targeted in-depth assessment (about 2,600 families annually) by public health nurses (Government of Manitoba, 2004). Paraprofessional home visitors establish trusting, nurturing relationships with families; promote problem-solving skills; assist in strengthening the family support system; and facilitate referral to community-based services.

Following the Hawaii Healthy Start approach (previously described above), BabyFirst home visitors (a) have a minimum education of a high school diploma or equivalent, (b) are recruited from local Manitoba communities, (c) are hired for personal and interpersonal qualities deemed essential for working with vulnerable families (e.g., warmth, cultural sensitivity, good parenting skills, good problem solving skills, knowledge of community resources); and (d) have satisfactory criminal record and child abuse registry checks. Home visitors are extensively trained in the Hawaii Healthy Start curriculum (briefly described above; for details see Duggan et al., 2000) and, more recently, the Growing Great Kids curriculum (Great Kids, Inc., 2000), which

focuses on facilitating parents' empathy with their children. Home visitors are also trained in risk management (e.g., related to parental mental health, partner abuse, and child abuse and neglect).

As in Hawaii Healthy Start, home visits are offered to families identified via the in-depth assessment as having high levels of need (inclusion criteria are described in the following section), beginning on a weekly basis from enrollment through the first year of service. Promotions to higher levels (of biweekly, monthly, and quarterly visits) are based on family functioning. Most families receive weekly visits for the first 1-2 years of service delivery. Home visits are 1-2 hours in duration and are guided by the goals outlined in a family plan jointly developed with parents during the first two months of visits. Specific home visiting interventions include the establishment of a working alliance with parents, role-modelling of problem-solving skills (including crisis management related to housing, relationship problems, and domestic violence problems), curriculum-based education regarding effective parent-child interaction, and linking families to community resources (Healthy Child Manitoba, 2000). As in Hawaii Healthy Start, creative outreach is used to (a) build trust, (b) engage hesitant families, (c) re-engage families who previously received services but have refused or avoided service, and/or (d) maintain connections and service to families who have informed the home visitor that they will be temporarily out of the area for over one month. Creative outreach strategies include the use of mail, telephone, and drop-by visits.

Participation in the home visiting component of the BabyFirst program increased from 450 families in 1999/2000 to 1,150 families in 2003/2004 (Government of Manitoba, 2004). Each home visitor maintains a caseload of 15 families. Home visitors receive regular, intensive clinical supervision from public health nurses (minimum of 2 hours of supervision per week), and between 33% to 50% of each caseload is reviewed weekly. On a quarterly basis, public health nurse supervisors accompany their home visitor supervisees on at least one home visit, to evaluate the home visitor's use of the curriculum, complete developmental reviews and

performance appraisals of the home visitor, and assess additional training needs of the home visitor.

BabyFirst Program Inclusion Criteria

Families become eligible for the BabyFirst program through a two-stage screening process, illustrated in Figure 1. In the first stage, all postpartum referrals in Manitoba (about 13,000 births annually) are screened throughout Regional Health Authorities by public health nurses using the BabyFirst Screening (BFS) form, a brief measure of biological (e.g., congenital or acquired disability, low birth weight, prematurity, alcohol use or smoking during pregnancy), social (e.g., isolation, relationship distress, partner violence), and demographic (e.g., teenage parent, single parent, less than grade 12 education, low income) risk factors for child maltreatment. Approximately 17% of families score above the BFS cut-off score for the second stage of assessment. This second stage comprises an in-depth assessment using the Family Stress Checklist (Kempe, 1976; Murphy, Orkow, & Nicola, 1985; Orkow, 1985) (see Figure 1).

The Family Stress Checklist (FSC) is a 10-item rating scale used to assess parent risk for caregiving difficulties based upon responses to a semi-structured interview. The interview covers several risk domains, including psychiatric and criminal history, childhood history of care, emotional functioning, attitudes toward and perceptions of children, discipline of children, and level of stress in the parent's life. The FSC requires careful review of the parent's history and current functioning, and requires the use of clinical judgement in applying ratings of 0 (*no risk*), 5 (*moderate risk*), or 10 (*high risk*) per item, for a range of total scores from 0 to 100 (Murphy et al., 1985; Orkow, 1985).

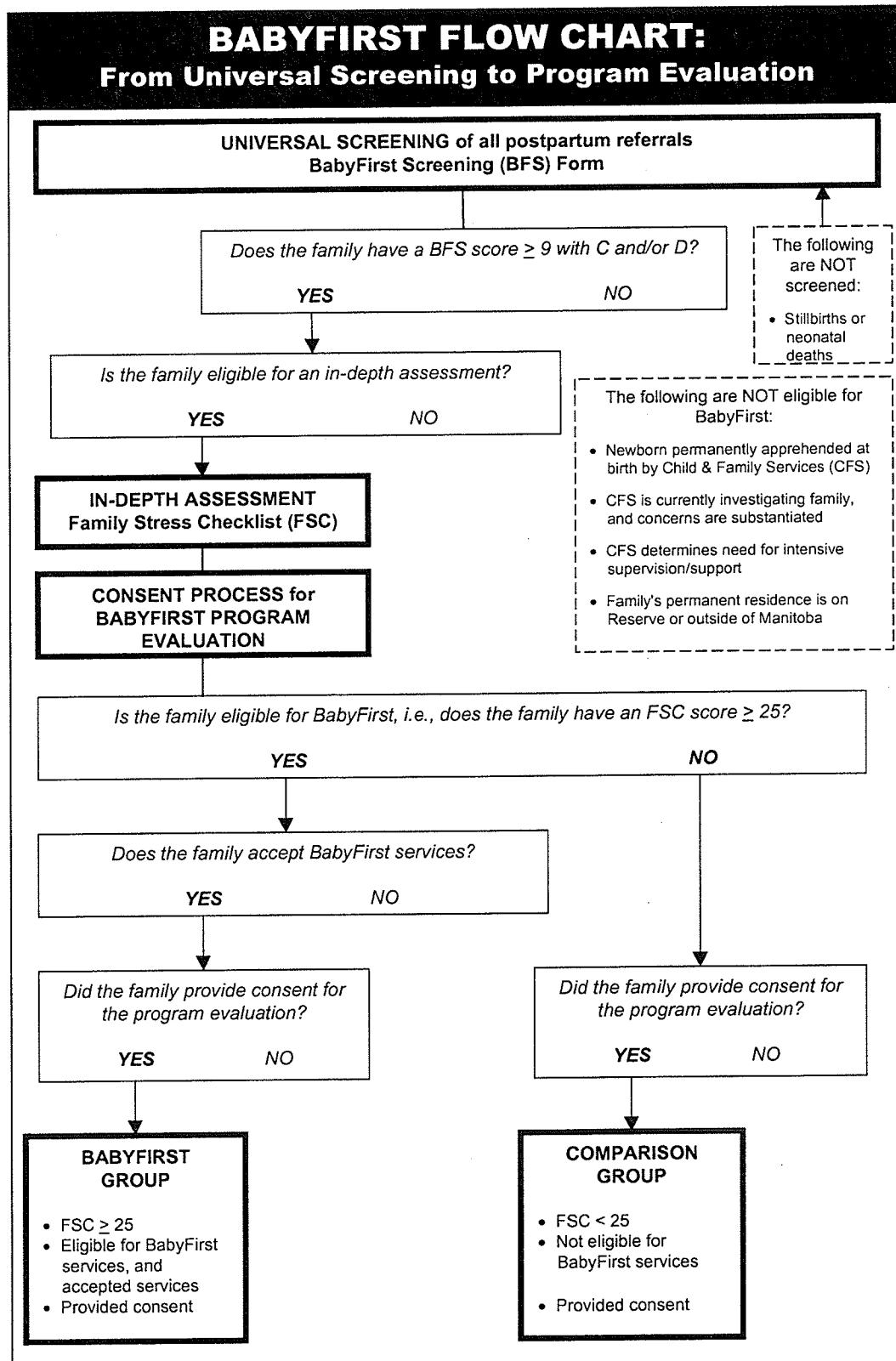


Figure 1. BabyFirst flow chart: From universal screening to program evaluation.

Developed in the early 1970s to assist in treatment planning (Kempe, 1976), the FSC has been used extensively as a tool for predicting family functioning and child maltreatment (Hiatt, 2000). It has been an integral part of the screening and assessment process for Hawaii Healthy Start (Duggan et al., 1999) and Healthy Families America (Daro & Harding, 1999). Despite its widespread use, a recent review found almost no studies of the reliability and validity of the FSC (Korfmacher, 2000). One study found a high level of correspondence ($r = .97$) between FSC interview classifications (low, mild, or severe risk) and independent file reviews of these interview data (Katzev, Henderson, & Pratt, 1997, cited in Korfmacher, 2000). Nonetheless, "for an instrument that relies heavily upon clinical judgement, not establishing inter-rater reliability is a major oversight" (Korfmacher, 2000, p. 134).

The predictive validity of an instrument can be measured by its relative performance on four measures: sensitivity, specificity, positive predictive value, and negative predictive value. Regarding the FSC, *sensitivity* refers to the percentage of maltreating parents who were scored as high-risk on the FSC (true positive rate); (b) *specificity* refers to the percentage of non-maltreating parents who were scored as low-risk on the FSC (true negative rate); (c) the *positive predictive value* refers to the percentage of parents scoring high on the FSC who were later found to maltreat their children; and (d) the *negative predictive value* refers to the percentage of parents scoring low on the FSC with no future evidence of maltreatment (Korfmacher, 2000).

Korfmacher (2000) found three validity studies of the FSC. The original validation study (Murphy et al. 1985) found that, using child abuse and/or neglect (based on hospital chart review) as the criterion, the FSC had a sensitivity of 80%, a specificity of 89.4%, a positive predictive value of 52.5%, and a negative predictive value of 96.8%. Korfmacher (2000) concluded that these are generally acceptable numbers, but noted that the positive predictive value (52.5%) indicates that a high percentage of parents scoring high on the FSC (47.5%) had no evidence of child maltreatment 1 to 2 years later, "suggesting caution in making decisions regarding the abuse potential" for parents (p. 135).

The second study (Center on Child Abuse Prevention Research [CCAPR], 1996, cited in Korfmacher, 2000) did not report sensitivity, specificity, positive predictive, or negative predictive values. Based on calculations using data obtained from the CCAPR study authors, Korfmacher (2000) found sensitivity rates of 84% – 89%, specificity rates of 28% – 35%, positive predictive values of 25% – 37%, and negative predictive values of 85% – 89%, using 6- and 12-month outcome scores on the Child Abuse Potential Inventory (Millner, 1994) as the criterion. The third study (Katzev et al., 1997, cited in Korfmacher, 2000) found a similar pattern of results: high sensitivity (97%) and low specificity (21%). A subsequent validity study of the FSC again found similar results: (a) high-risk 1-year-olds were 8.4 times more likely to have been maltreated, with a sensitivity of 86%, a specificity of 58%, a positive predictive value of 11%, and a negative predictive value of 97%; and (b) high-risk 2-year-olds were 5.2 times more likely to have been maltreated, with a sensitivity of 79%, a specificity of 41%, a positive predictive value of 22%, and a negative predictive value of 95% (Stevens-Simon, Nelligan, & Kelly, 2001).

Korfmacher (2000) urged further research on the reliability and validity of the FSC and concluded that the available research “suggests that on a group level it can identify higher risk families, many of whom may receive benefit from services or support even if they are not in danger of maltreating their children. The best recommendation regarding this measure, then, is that anyone who uses it be aware of the issues surrounding it, include it as part of a larger assessment of family functioning using multiple measures, and be encouraged to study its use to whatever extent possible” (Korfmacher, 2000, p. 138).

In the BabyFirst program, approximately 75% of the families scoring above the cut-off in the first stage screening (about 12.5% of the total sample of annual births) score at or above the cut-off on the FSC (≥ 25) (based on Murphy et al., 1985) in the second stage (about 9% of the total sample of annual births). These families are classified as being at the highest risk and are offered BabyFirst home visiting services. These families are eligible for the program from the third trimester until 3 months postpartum and comprise the BabyFirst group. Families that score

less than 25 on the FSC are not eligible for the program and comprise the comparison group for the BabyFirst program evaluation (described in the following section). This cut-off score-based assignment process, illustrated in Figure 2, lends itself to a powerful evaluation design (the regression-discontinuity design, described in the *Method* section below). Participation in the BabyFirst program is voluntary. As shown in Figure 1, program staff obtain informed consent to participate in the program evaluation from both BabyFirst and comparison groups of families.

BabyFirst Program Evaluation

The present researcher was contracted by the Government of Manitoba from 1999-2001 to coordinate the initial development and implementation of the provincial BabyFirst evaluation. Based on program goals and objectives co-developed by the Government of Manitoba and the province's Regional Health Authorities, the researcher developed an outcome evaluation framework to assess the program's success in achieving its goals and objectives. This framework development included proposals to the Government of Manitoba and Regional Health Authorities regarding the selection of evaluation design, methods, and measures. The final evaluation design and methodology reflected a consensus of these different stakeholders (government funder, service delivery agents, program evaluator). For example, some proposed evaluation measures were removed from the final framework because of concerns regarding data collection requirements (e.g., time burden), perceptions of low face validity (notwithstanding established construct validity), or inadequate levels of staff training for data collection. Moreover, stakeholders proposed and included additional measures (e.g., the Denver Developmental Screening Test II) in the final framework. These measures were already in use in the field and service providers were already widely trained and proficient in their use.

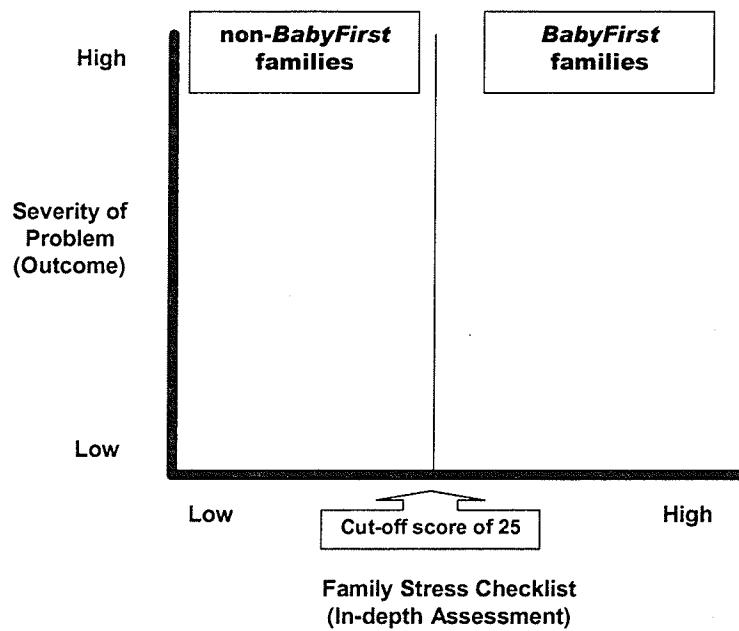


Figure 2. Assignment to *BabyFirst* and non-*BabyFirst* groups using cut-off score.

In the BabyFirst program, a number of evaluation measures are collected when infants are approximately 3 months of age. Some of these measures serve as baselines for program outcomes at age 1 year. Several measures are scales from the National Longitudinal Survey of Children and Youth (NLSCY) (Human Resources Development Canada [HRDC] & Statistics Canada, 1996). Measures included in the present study were selected from the larger set of measures collected in the BabyFirst program evaluation and are described in the *Method* section below. BabyFirst public health nurses collected these measures through structured interviews with families participating in the evaluation. Consistent with the evaluation logic of prevention and early intervention programs, the BabyFirst evaluation measures focus on variables theorized and empirically demonstrated to influence program outcomes.

Evaluation Logic of Prevention and Early Intervention Programs

The BabyFirst evaluation uses a basic strategy common in prevention and early intervention research that involves measurement of short-term, proximal outcomes which have been established as predictors of longer-term, distal outcomes of interest. Achievement of these proximal outcomes is interpreted as initial and promising evidence of efficacy or effectiveness (Tolan, 1999; Weissberg & Greenberg, 1998). As Tolan (2002) elaborates, “because prevention is focused on status at some point in the future, the type of prevention, inclusion criteria, and the methods to be used must be based on a theoretical (and, it is hoped, empirically demonstrated) relationship between the risk factors identified, the targets of intervention, the intervention activities, and the proximal and distal outcomes” (p. 204). Moreover, “one needs to articulate a theory about the program’s immediate impact and how it will lead to long-term effects” (Tolan, 2002, p. 204). In the next section, the theory underlying Manitoba’s BabyFirst program, an enhanced replication of the Hawaii Healthy Start program, is described.

Program Theoretical Model

The program theoretical model for the BabyFirst program is illustrated in Figure 3. It is generally based on the Hawaii Healthy Start program theoretical model (Duggan et al., 1999). In this model, there are four key program components:

1. Home visitor establishes a trusting relationship with parents.
2. Home visitor role models problem-solving skills.
3. Home visitor provides parenting education and models effective parent-child interaction.
4. Home visitor links families with needed community resources.

The theoretical sine qua non of the model is the working relationship or *alliance* between parents and home visitors, which serves as a *secure base* (Bowlby, 1988; Waters & Cummings, 2000) from which to explore new experiences and to serve, “when necessary, as a haven of safety and a source of comfort” (Waters & Cummings, 2000, p. 165).

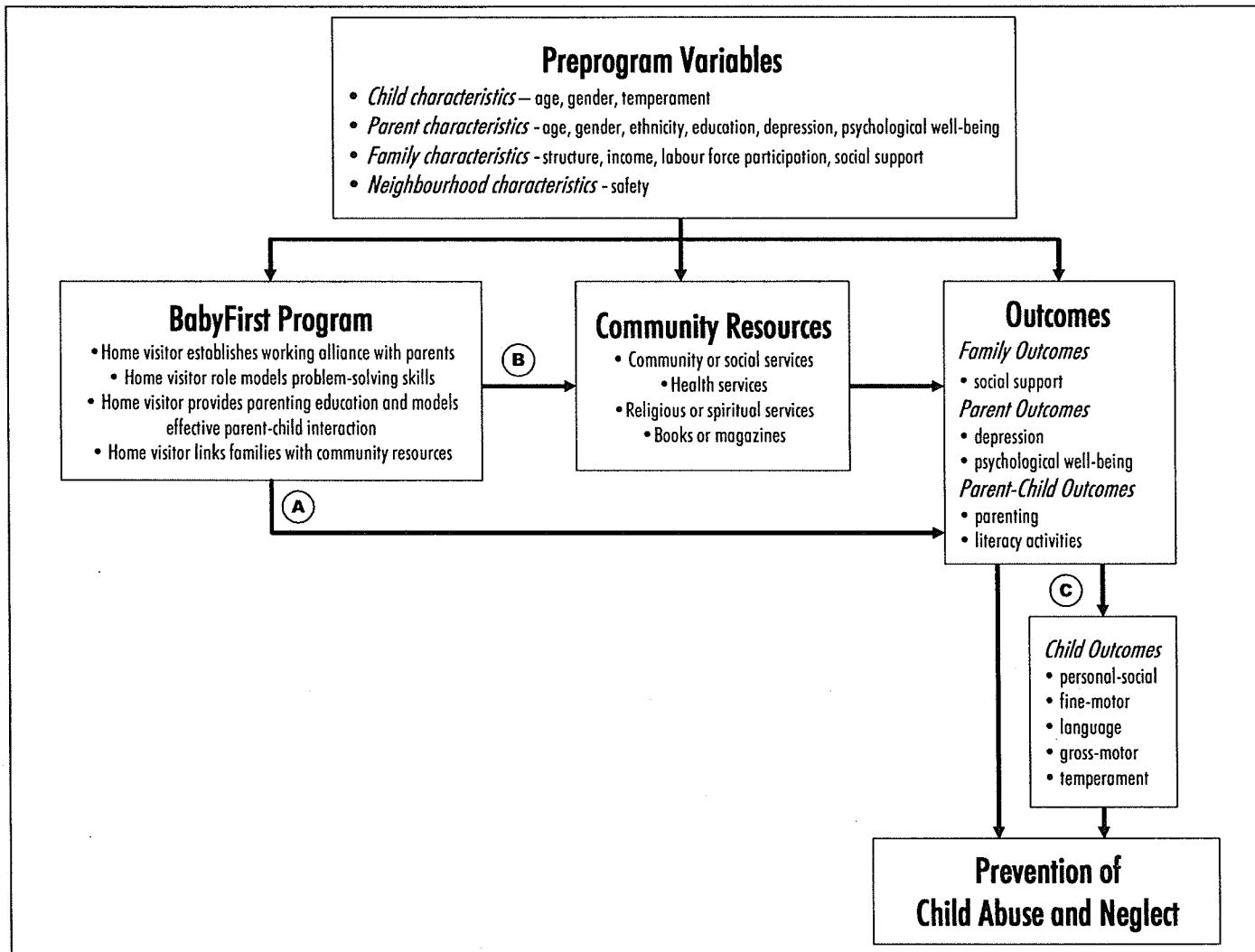


Figure 3. Conceptual framework of BabyFirst program (after Duggan et al., 1999).

The secure base of the alliance is essential for home visiting interventions (e.g., role modelling problem-solving, providing parenting education, modelling effective parent-child interaction, linking families with needed community resources). These interventions are intended to help families identify and build on their strengths to improve family functioning (Duggan et al., 1999). They are theorized to have direct effects on (a) family, parent, and parent-child outcomes (see Figure 3, A) and (b) community resource outcomes (see Figure 3, B), as well as (c) indirect effects on child outcomes (see Figure 3, C). In other words, program interventions are theorized to have proximal effects on family outcomes (e.g., increased social support), parent outcomes (e.g., improved psychological well-being, reduced depression), parent-child outcomes (e.g., improved parenting, increased parent-child literacy activities), and community resource outcomes (e.g., improved access to primary health care providers), which then lead to distal outcomes of enhanced child development and decreased child abuse and neglect.

In this program model, there is a theoretical chain from the development of a positive home visiting working alliance, to proximal outcomes including improved parental self-efficacy and secure parent-child attachment, to distal outcomes of improved child developmental outcomes, to ultimate outcomes centering on the prevention of child abuse and neglect (see Figure 3). Each link in this program theoretical model constitutes a testable hypothesis for empirical investigation. The present study investigated the first link in this theoretical chain from home visiting working alliance to proximal family outcomes, parent outcomes, and parent-child outcomes⁷ after one year. It is important to note that few prevention studies have reported effects as early as 12 months (Culp et al., 2004), so the present study may be regarded as a fairly rigorous test of program effectiveness. The importance of each of these proximal family, parent, and parent-child outcomes for healthy child development is described in the following sections.

⁷ Testing other theoretical links with proximal community resource outcomes, and between proximal, distal, and ultimate outcomes of preventing child maltreatment is outside the scope of the present study, but represents next steps in a longer-term research program.

Family Outcome

Social support. Social support is widely regarded as a major protective factor for both healthy adult development (Barrera, 2000; Berkman & Glass, 2000; Cohen, 2004; Glass, 2000; House, Landis, & Umberson, 1988; Uchino, Cacioppo, & Kiecolt-Glaser, 1996) and healthy child development (Arnold & Doctoroff, 2003; Earls & Carlson, 2001; Leventhal & Brooks-Gunn, 2000; Parke, 2004; Runyan et al., 1998; Sampson, Morenoff, & Gannon-Rowley, 2002; Serbin & Karp, 2004). It has been suggested that “parents in supportive social networks may be more likely to have positive relationships with their children, which in turn may positively affect the child’s social adjustment both within and outside the family” (Parke, 2004, p. 378). In the first cycle of the National Longitudinal Survey of Children and Youth (NLSCY) in 1994, low social support was reported by 3.6% of families with children under age 2 years, and as one of the risk factors (alongside family dysfunction and parental depression) with the greatest impact on children’s development (Statistics Canada & HRDC, 1996). Extant evidence suggests that social support has buffering effects (Barrera, 2000; Cohen, 2004; Glass, 2000), counteracting the negative effects of other variables. For example, Crockenberg and Leerkes (2003) found that the negative impact of difficult infant temperament on the parent-child relationship decreased as the social support available to the parent increased. Social support appears to be a particularly important protective factor for families living in social and economic disadvantage (Lachman, 2004; Ryff, Singer, & Palmersheim, 2004) and one of the potential mechanisms through which neighbourhoods influence children’s development (Leventhal & Brooks-Gunn, 2000).

Parent Outcomes

Parental depression. Parental depression is a major risk factor for poor child development and likely operates through a number of genetic, biological, psychological, and interpersonal mechanisms, including parent-child interaction (Atkinson et al., 2000; Beardslee, Versage, & Gladstone, 1998; Beck, 1999; Connell & Goodman, 2002; Goodman & Gotlib, 1999, 2002; Lovejoy, Graczyk, O'Hare, & Neuman, 2000). Findings from extant research on parent-focused home-based programs "as often as not indicated no significant treatment-control group differences" on maternal mental health (Brooks-Gunn et al., 2000, p. 553). Few early childhood intervention programs have reported effects on maternal depressive symptoms (Brooks-Gunn et al., 2000; see also Barnet, Duggan, Devoe, & Burrell, 2002). Some home visiting programs have led to initial but only temporary improvements in parental depression. For example, in one randomized efficacy trial, positive effects on parental depression were found at 6 weeks (Armstrong, Fraser, Dadds, & Morris, 1999), but these were not maintained at 4-month (Armstrong, Fraser, Dadds, & Morris, 2000) or 12-month follow-up (Fraser, Armstrong, Morris, & Dadds, 2000).

Parental psychological well-being. Two decades ago, the available evidence suggested that being a parent can be costly in terms of parental psychological well-being (PWB). In a classic review, McLanahan and Adams (1987) asked "How can we reconcile the fact that parenthood, a highly valued social position, is associated with lower levels of psychological well-being?" (p. 237). They suggested that future research would need to chart the effects of psychological well-being on parent behaviour. It is now clear empirically that how parents think and feel about themselves and their lives influences their parenting and, in turn, their children's development (Bugental & Johnston, 2000). There is extensive evidence that socioeconomic stress and disadvantage can indirectly affect parenting behaviour through adverse impacts on parental PWB (Mistry, Vandewater, Huston, & McCloyd, 2002; Taylor, Roberts, & Jacobson, 1997).

Parent-Child Outcomes

Positive parenting. Positive parenting refers to warm, accepting, and engaging parent behaviours toward their children. Considerable evidence confirms the importance of improving positive parenting for healthy early childhood development (Gopnik, Meltzoff, & Kuhl, 1999; NRC & IOM, 2000). The benefits of positive parenting on children's psychological adjustment appears to be universal, a conclusion based on nearly 400 studies conducted over four and a half decades across "cultures, genders, ages, geographic boundaries, ethnicities, and other defining conditions of the world" (Rohner, 2004, p. 835). Perhaps the strongest evidence comes from intervention studies showing that experimental manipulations that improve parenting lead to predicted improvements in children's behaviour (see reviews by Collins, Maccoby, Steinberg, Hetherington, & Bornstein, 2000; Maccoby, 2000; O'Connor, 2002; Rutter, 2005).

Hostile/ineffective parenting. Hostile/ineffective parenting refers to harsh and rejecting parent behaviour toward their children. There is also considerable evidence on the importance of reducing hostile/ineffective parenting for healthy early childhood development (NRC & IOM, 2000; Repetti, Taylor, & Seeman, 2002; Rohner, 2004), particularly in helping children to overcome the odds of living in social, economic, and cultural conditions of risk and disadvantage (Chen, Matthews, & Boyce, 2002; Duncan & Brooks-Gunn, 2000; Masten & Coatsworth, 1998, Rutter, 2000). A recent meta-analysis found a strong relationship between socioeconomic disadvantage and hostile/ineffective parenting in both cross-sectional ($d = .48$) and longitudinal ($d = .55$) studies (Grant et al., 2003). A follow-up meta-analytic path analysis by the authors showed that hostile/ineffective parenting mediates the relationship between poverty and children's psychological symptoms (Grant et al., 2003; see also Bradley & Corwyn, 2002; O'Connor, 2002; Rutter, 2000). Previous studies using NLSCY data have highlighted the negative influence of hostile/ineffective parenting on child development (Landy & Tam, 1998; Lipman, Boyle, Dooley, & Offord, 2002; Thomas, 2004). In children aged 2-3 years, hostile/ineffective parenting was associated with a three-fold risk of emotional disorder, a four-fold risk of hyperactivity, and a

seven-fold risk of aggressive behaviour (Landy & Tam, 1998). In a recent NLSCY study, compared to children living in families with low levels of hostile parenting, children aged 2-3 years who were living in families with high levels of hostile parenting in 1994 scored 39% higher on aggressive behaviour, and scored 85% higher on aggressive behaviour at six-year follow-up in 2000 (Statistics Canada, 2004; Thomas, 2004). In contrast, a decrease in hostile parenting from 1994 to 2000 was associated with a decrease in children's aggressive behaviour over this time period (Statistics Canada, 2004; Thomas, 2004).

Parental reading with children. Meta-analytic results indicate that parental reading with children promotes their language, literacy, and socioemotional development (Bus, van IJzendoorn, & Pellegrini, 1995). Maternal speech patterns predict children's vocabulary growth during the first three years, as well as pre-kindergarten measures of emergent literacy and print-related skills (NRC & IOM, 2000; Whitehurst & Lonigan, 1998). In the NLSCY, reading with children was shown to be more important than a composite measure of parent-child activities (laughing, praising, talking, playing, doing special activities, reading) in predicting children's learning and behaviour outcomes (Cook & Willms, 2002). Given that early parent-child literacy activities decrease in frequency with decreasing socioeconomic status (Arnold & Doctoroff, 2003; Bradley & Corwyn, 2002; NRC & IOM, 2000), early childhood interventions that promote parental reading with children in disadvantaged families may be essential to reduce socioeconomic inequalities in child development.

Findings from extant research on parent-focused, home-based programs find "at least some positive effects on parent-child interaction" (Brooks-Gunn et al., 2000, p. 553). Such programs appear to be "more effective in altering parental insensitivity than in changing the quality of the relationship attachment per se" (Brooks-Gunn et al., 2000, p. 553). This conclusion is supported by a more recent meta-analytic review of sensitivity and attachment interventions in early childhood (Bakermans-Kranenburg et al., 2003). This review of 70 studies comprising 88 intervention effects on sensitivity ($n = 7,636$) and/or attachment ($n = 1,503$) found a larger effect

size for changing insensitive parenting ($r = 0.17$) compared to infant attachment insecurity ($r = 0.10$). "In summary, quality of the mother-child interaction is a frequent family outcome. The processes underlying program effects on the mother-child relationship, however, have not been specified in great detail" (Brooks-Gunn et al., 2000, p. 554).

In the present study, the working alliance between parent and home visitor is posited as the key process underlying the effects of home visiting on the family, parent, and parent-child outcomes described above. The working alliance is a central concept in the program theoretical model described above. Indeed, it is a fundamental tenet of most home visiting programs and of most early childhood interventions with parents:

Providers' relationships with parents as the scaffolds for building or rebuilding parenting strengths [are] the best means for transforming the caregiving environment for children at risk. . . . When paraprofessionals step in to fulfill surrogate mother roles . . . they help undo past pain, enabling young mothers to break cycles of poor parenting by finding new solutions and "doing it better" than their own mothers did (Musick & Stott, 2000, p. 443).

Brooks-Gunn et al. (2000) highlighted "the importance of the quality of program participants' engagement in program services as well as the interaction between participant characteristics and their potential to benefit from intervention" (p. 554). However, they also cautioned that the concept of the alliance may have been "reified to the point that empirical verification is not considered necessary . . . parental engagement is seen as axiomatic to early intervention, which led to an implicit acceptance of its importance" (Brooks-Gunn et al., 2000, p. 562).

This axiomatic nature of the alliance may explain why alliance theory and research are almost absent from the prevention and early childhood intervention literature. However, a great deal of theoretical and empirical work regarding the working alliance has been accomplished in the counselling and psychotherapy literature (Gelso, 2000; Horvath, 2000, 2001; Horvath & Bedi, 2002; Horvath & Greenberg, 1994; Martin, Garske, & Davis, 2000; Safran & Muran, 2000). The next section bridges these two literatures, toward a theory of the working alliance in early childhood intervention, and generates testable hypotheses for the present study.

Working Alliance: Lessons from the Psychotherapy Literature

In a recent article, Horvath (2000) traced the conceptual and theoretical history of the therapeutic relationship or alliance. The alliance construct can be followed from its origins in psychoanalytic and psychodynamic theories, through behavioral theories, into the experiential theories that guided the first generation of psychotherapy research, prompted largely by Carl Rogers' (1951, 1957) insistence that his theoretical hypotheses regarding the therapeutic relationship be subjected to empirical testing.

In 1975, Edward Bordin delivered a milestone presidential address to the Annual Conference of the Society for Psychotherapy Research entitled "The Working Alliance: Basis for a General Theory of Psychotherapy" that would form the basis for a new trans-theoretical concept of the alliance (Horvath, 2000). In contrast to Rogers' "therapist-offered conditions," Bordin (1975) argued that the alliance was a true bidirectional relationship, comprising three interdependent components: *bonds* (the emotional, interpersonal connections between clients and therapists), *tasks* (agreements between clients and therapists on the central activities to be undertaken in therapy), and *goals* (agreements between clients and therapists as to the short- and long-term purposes and expectations of therapy). Bordin's concept of the alliance was twofold: the alliance was both a facilitative context for the tasks of therapy as well as a therapeutic agent in its own right (Horvath, 2000). In other words, Bordin viewed the alliance as having both indirect and direct effects on therapeutic outcome.⁸

Today, the therapeutic alliance is both one of the most widely researched clinical constructs (see Horvath & Bedi, 2002, and Horvath & Greenberg, 1994 for reviews) and one of the best predictors of outcome across several brands of psychotherapy (Henry, Strupp, Schacht, & Gaston, 1994; Norcross, 2002; Orlinsky, Grawe, & Parks, 1994). Bordin's (1975) trans-theoretical model launched the development of reliable measures of the alliance and extensive research into

⁸ An early debate on whether the alliance was a direct or indirect factor in psychotherapy outcome (for an overview, see Butler & Strupp, 1986) has given way to considerable evidence that the alliance is both (Henry et al., 1994; Orlinsky et al., 1994).

its construct validity and predictive validity (Horvath, 2000). A landmark meta-analysis of the first 15 years of investigation (Horvath & Symonds, 1991) concluded that the alliance was a robust predictor of psychotherapy outcome, as early as the third session of therapy, across various treatment approaches, clinical diagnoses, and client populations, using different measures and different measurement perspectives. Recent research has replicated the predictive power of the alliance not only in individual therapies, but in couple, family, and group therapies, as well as in supervision and training contexts (Horvath, 2000). The alliance represents "one of the relatively few therapy topics for which there *is* clinical consensus and research verification" (Goldfried, 2000, p. 157, emphasis in original). Prevention and early childhood intervention research could benefit greatly from the lessons learned from the psychotherapy literature. Psychotherapy research on the alliance has important implications for understanding the efficacy and effectiveness of early childhood intervention.

Toward a Theory of the Working Alliance in Early Childhood Intervention

To illustrate the absence of progress in this theoretically crucial area in the field of early childhood intervention, the work of Robert Emde (1988a) offers a longitudinal perspective. Seventeen years ago, Emde introduced a special section in the journal *Psychiatry* on preventive intervention and early development, containing three efficacy studies of early intervention for infants that, at the time, represented the state of the art (Barnard et al., 1988; Beckwith, 1988; Osofsky, Culp, & Ware, 1988). Reflecting on these studies, Emde observed that, "Although this area of investigation is as yet without sufficient theoretical guidance, it promises to be central in future mental health research. An important conclusion of [these three studies] is that *interventions occur by virtue of human relationships and that it behooves us to understand better how intervention relationships affect other sets of relationships*" (1988a, p. 235, emphasis added). A closer look at the above three studies reveals their very early stage of development. Initial concepts of the alliance were drawn, as they were in the psychotherapy literature, from clinical experience. For example, one investigator noted her research team's belief that "the effectiveness

of intervention depends more on the relationship between the parent and the intervenor than on the curriculum" (Beckwith, 1988, p. 242), but later concluded that "we cannot be sure that we achieved this [relationship]" (p. 246). All of the investigators from the three studies remarked on the need for both theoretical and methodological development in investigating the helping relationship in early childhood intervention.

In his conclusion to the special section, Emde (1988b) offered strong directions for future inquiry that are worth quoting at length:

We will see that our attention is drawn to the need for theory. One aspect of intervention, in particular, calls for theory, namely, the effects of relationship experiences on other relationships (p. 254).

There seems to be a clear sense that the subject/intervener relationship is a key moderating variable for intervention outcome. We might say that this is the most salient lesson coming from the research. . . . In assessing intervention effects, therefore it became necessary to assess the degree to which the interventionist/mother relationship was working (p. 256).

Still, there is a paucity of systematic theory focusing on the ways dyadic relationships influence other dyadic relationships both inside and outside the family (p. 257).

Interventions are seen as operating through the effects of relationships on other relationships. The interventionist/subject relationship is deserving of more general theoretical attention (p. 258).

. . . these three studies represent an early phase of the state-of-the-art goals for infancy prevention-intervention research. Such research is extraordinarily complex . . . It is also clear that more theory-based intervention designs are needed (p. 258).

Over a decade and a half later, almost no progress has been made in this area. In a recent book chapter, "Toward a Theory of Early Relationship-Based Intervention," Emde, Korfmacher, and Kubicek (2000) concluded that "because of a lack of empirical knowledge, this chapter can only point to a framework rather than a theory in this area. It is our hope that future research on the topic of a theory of infant mental health intervention will build on forthcoming inquiry and investigation to develop a more specific model" (p. 24; see also Korfmacher, 2001, 2002).

The framework of Emde and colleagues (Emde, Everheart, & Wise, 2004; Emde et al., 2000; Emde & Robinson, 2000) provided the conceptual foundation for the present study, which was further shaped by lessons learned from three decades of alliance theory and research on psychotherapy. In particular, the present research draws upon the theoretical work of the late

Edward Bordin (1975, 1979, 1994) and the empirical work of Adam Horvath (1981, 1994, 2000, 2001), who has led the investigation of Bordin's trans-theoretical model of the alliance and the development of a reliable and well-validated alliance measure, the Working Alliance Inventory.

In the next section, Emde's framework, Bordin's theory, and Horvath's research are linked, toward a theory for future research on early childhood intervention. Based on this synthesis, the role of the working alliance in early childhood intervention was tested in the present study through a secondary analysis of data from an evaluation of the BabyFirst program.

A Framework for Early Relationship-Based Intervention

According to Emde's framework (Emde et al., 2000, 2004; Emde & Robinson, 2000), successful mental health interventions for children and families are characterized by four features:

1. A focus on the experiences of both children and their caregivers
2. A focus on developmental processes over time
3. A focus on the relationship-based nature of early intervention
4. A focus on the influence of relationships on other relationships

First, the focus on the *experiences* of both children and their caregivers refers to a consideration of the meaning of interventions for the individual at both psychological and interpersonal levels, both to minimize suffering and to facilitate positive growth and adaptation (Emde et al., 2000).

Second, the focus on *developmental processes* over time refers to the importance of going beyond immediate symptom relief and enabling individual adaptation over the life course, with a special emphasis on "preventing future problems and strengthening future adaptations" (Emde et al., 2000, p. 5). Third, the focus on the *relationship-based* nature of early intervention refers to the dependence of interventions on "a sense of trust and availability as experienced over time with the intervener" (Emde et al., 2000, p. 4). The present study of the alliance in early childhood intervention can, therefore, benefit from a brief review of the growing body of attachment research (Main, 1999; R. A. Thompson, 2000; Waters & Cummings, 2000) and relationship science (Berscheid, 1994, 1999; Berscheid & Reiss, 1998; Reiss, Collins, & Berscheid, 2000).

(These literatures are briefly discussed below.) Fourth, the focus on the *influence of relationships on other relationships* refers to the influence of the intervention relationship on both the parent's *internalized* (psychologically represented) relationships (e.g., family of origin relationships) and *interpersonal* relationships outside the intervention relationship (e.g., relationships with children or other family members) (Emde et al., 2000).

This focus on the influence of relationships on other relationships can also refer to the influence of pre-intervention relationships on the intervention relationship. Parents with histories of positive relationships with significant others and adequate social support may be better prepared for establishing a new relationship in the context of intervention. In contrast, parents with histories of conflictual or hostile relationships with significant others and inadequate social support may find it challenging to establish an effective relationship in the context of intervention. In other words, the interpersonal history of parents (i.e., as represented in their internalized relationship experiences or "internal working models") may provide a measure of parents' openness to change and their capacity to form a working alliance with providers of early childhood intervention (Emde et al., 2004; see also Andersen & Chen, 2002; Baldwin, 1992).

Attachment relationships and the alliance. The concept of internal working models has been central to modern attachment theory, which has been central to theories of early childhood intervention (Gomby et al., 1999). According to attachment theory, internal working models are continuously revised to reflect ongoing interpersonal experience (Thompson, 2000) and are driven by the fundamental human motivation for interpersonal connectedness (Baumeister & Leary, 1995). Yet "despite enthusiasm for the concept of internal working models, there is surprisingly little consensus about the precise nature and operation of attachment schemata" (Berscheid & Reiss, 1998, p. 219). Leading attachment researchers have concluded: "We have yet to understand the formation of new attachments in adulthood" (Main, 1999, p. 861; see also Ainsworth, 1989; and Waters & Cummings, 2000). A growing body of evidence from psychotherapy research links clients' attachment styles to the therapeutic alliance and to

therapeutic outcome (Horvath, 2001; Horvath & Bedi, 2002; Mallinckrodt, 1991, 1996, 2000; Meyer & Pilkonis, 2001, 2002). From this theoretical perspective, the development of the working alliance in home visiting can be regarded as the development of an attachment relationship in adulthood. That is, the home visitor may emerge as an attachment figure for parents (Berlin & Cassidy, 1999).

According to (a) the relationship-based framework of Emde (1988b; Emde & Robinson, 2000; Emde et al., 2000, 2004), (b) Bordin's (1975, 1979, 1994) trans-theoretical model of the alliance, and (c) the secure base concept from attachment theory (Bowlby, 1988; Waters & Cummings, 2000), the relationship between home visitor and parent provides leverage (Emde et al., 2004) for intrapersonal changes in the parent, which beget interpersonal changes in caregiving behaviour between parent and child. The alliance facilitates a shared narrative understanding between parent and home visitor, including a shared view of intervention goals and the necessary tasks to achieve those goals (Bordin, 1994; Emde et al., 2004).

In the context of the relationship between parent and child, their dyadic experiences are seen as influencing the child's developmental tasks (i.e., activity, self-regulation, social fittedness, affective monitoring, and cognitive assimilation) (Emde et al., 2000; Emde et al., 2004). Through repeated experiences wherein the parent is sensitive, responsive, consistent, and emotionally available, a secure attachment is developed and the child's developmental tasks are furthered. In parallel fashion, a helping relationship (e.g., between home visitor and parent) that too is sensitive, responsive, consistent, and emotionally available is seen as facilitating the parent-child relationship and, therefore, the child's healthy development (Emde et al., 2000, 2004).

Two sets of longitudinal studies underscore the importance of the quality of the early parent-child relationship. The first set of studies have demonstrated the stability of attachments from early childhood to adulthood (Hamilton, 2000; Waters, Hamilton, & Weinfield, 2000; Waters, Merrick, Treboux, Crowell, & Albersheim, 2000; Waters, Weinfield, & Hamilton, 2000; Weinfield, Sroufe, & Egeland, 2000; but see Lewis, Feiring, & Rosenthal, 2000). The second set

of studies has shown that hostile/ineffective parenting is transmitted intergenerationally; children who grow up with hostile/ineffective parenting are more likely to parent their own children in a similarly hostile manner (Capaldi, Conger, Hops, & Thornberry, 2003; Capaldi, Pears, Patterson, & Owen, 2003; Hops, Davis, Leve, & Sheeber, 2003; Newcomb & Locke, 2001; Serbin & Karp, 2004; Thornberry, Freeman-Gallant, Lizotte, Krohn, & Smith, 2003). Given the enduring influence of the early child-caregiver relationship, it is crucial that early childhood interventions facilitate the development of secure parent-child attachment and positive parenting.

Relationship development and alliance development. Also important for the present study of the working alliance are empirical efforts to address the question of, "What causes some relationships to endure and others to fail?" (Berscheid & Reiss, 1998, p. 197). Relationship development has been defined as "the process by which two people may progress from their first encounter to develop a close relationship" (Berscheid & Reiss, 1998, p. 197). In early childhood intervention, as in other interpersonal contexts, it is expected that "building a good relationship takes time" (Halpern, 2000, p. 378). It is now well-established that initial interpersonal expectations influence future interpersonal interactions (Berscheid, 1994, 1999; Berscheid & Reiss, 1998; Reiss et al., 2000). Recent relationship research finds that individuals "are predisposed to appraise their support experiences in ways that are consistent with their chronic working models of attachment, especially when the support message is ambiguous" (Collins & Feeney, 2004, p. 363). Applied to the support offered through early childhood intervention, this finding suggests the importance of minimizing the ambiguity of interpersonal process between parent and service provider. Research on interpersonal process in psychotherapy has found that well-intended but ambiguous interventions, such as interpersonally complex communications that convey "mixed messages" (e.g., discrepancies between verbal and nonverbal communicative behaviours) predict negative outcome, partly because of client predispositions to negative interpretation of such interventions (Henry & Strupp, 1994).

From the foregoing relationship-based theoretical framework follows an evaluation framework for early childhood intervention programs that focuses on measuring the experiences of families, addresses development, examines the intervention relationship, and explores the influence of prior relationships on the intervention relationship as well as the influence of the intervention relationship on other relationships, especially the parent-child relationship (Emde et al., 2000; Emde et al., 2004). The authors note that "although a premium is placed on the helping relationship in early childhood programs . . . this relationship is very difficult to measure" (Emde et al., 2000, p. 21). They go on to conclude:

In summary, although there is converging evidence to support the importance of the helping relationship in infant mental health programs, we do not yet understand what are the critical dimensions of the helping relationship to study or how to examine them best. (Emde et al., 2000, p. 21; see also Korfsmacher, 2002, p. 283)

The Trans-theoretical Model of the Alliance

Edward Bordin's (1975, 1979, 1994) trans-theoretical model of the alliance provides promising directions for conceptualizing the critical dimensions of the helping relationship in early childhood intervention. In this model, one person (e.g., the parent) seeks change and another person (e.g., the intervener) offers to facilitate this change. Toward this change, these partners develop a mutual understanding and agreement about the goals for change and the necessary tasks to accomplish these goals. The development of a bond between these two persons provides the foundation to maintain the partners' work together.

Goals and tasks. In psychotherapy, patients may have a number of goals for therapeutic change. Bordin emphasized that "a careful search with the patient for the change goal that most fully captures the person's struggles with pains and frustration relative to the story of his or her life is a key part of the building of a strong therapeutic alliance" (Bordin, 1994, p. 15). Further, in psychotherapy, the therapist is "the major source in the selection of therapeutic tasks, but the patient must understand the relevance of these activities to change in order to maintain the role of

an active partner" (Bordin, 1994, p. 16). Bordin's concepts of goals and tasks are also applicable to early childhood intervention.

Bonding. According to Bordin, "the bonding of the persons in a therapeutic alliance grows out of their experience of association in a shared activity. Partner compatibility (bonding) is likely to be expressed and felt in terms of liking, trusting, respect for each other, and a sense of common commitment and shared understanding in the activity. Thus, the specific nature of the bonds will vary as a function of the shared activity" (Bordin, 1994, p. 16).

Although home visitors are not therapists, the concepts of goals, tasks, and bond are consistent with the home visiting approach outlined in Figure 3. In keeping with Bordin's (1994) formulation, it can be hypothesized that, in home visiting, "reaching an understood and mutually agreed-on change goal is the key process in building an initial, viable alliance" (p. 21). In many respects this also defines the tasks of home visiting. The goals and tasks of home visiting include (a) enhancing parental psychological well-being by role modelling and building problem-solving skills, (b) strengthening the parent-child relationship by providing parenting education and modelling effective parent-child interaction, and (c) improving family social support by linking families to community resources (see Figure 3).

Empirical Validation of the Trans-theoretical Model of the Alliance

Horvath's (1981, 1994, 2000, 2001; Horvath & Bedi, 2002) research has provided reliable and well-validated measures of these three critical dimensions of the working alliance—goals, tasks, and bond—and shows great promise for use in studying the alliance in early childhood intervention.⁹ Horvath's Working Alliance Inventory (WAI) is described in greater detail in the *Program Variables* subsection of the *Method* section below. Empirical validation of the trans-theoretical model of the alliance is extensive. In a meta-analysis of 80 effect sizes from 22 studies from 1977-1997, the average weighted alliance-outcome correlation of the WAI was .24 (Martin

⁹The WAI was included in the measurement protocol for the Early Head Start national evaluation (Brooks-Gunn et al. 2000) and results are forthcoming.

et al., 2000). This was similar to the average r for all alliance measures (.22, $SD = .12$) across the 79 studies included in the Martin et al. (2000) meta-analysis, the average r (.21) reported in an extension to the Martin et al. (2000) meta-analysis that added 10 studies from 1997-2000 (Horvath, 2001; Horvath & Bedi, 2002), and the average r (.26) reported in an earlier meta-analysis (Horvath & Symonds, 1991). Considerable evidence indicates that the alliance-outcome correlation is not moderated by type of outcome measure, type of alliance rater, type of outcome rater, time of alliance rating, methodological quality, or type of psychotherapy (Horvath, 2001; Horvath & Bedi, 2002; Martin et al., 2000).

Summary of the Theoretical Framework for the Present Study

The home visiting working alliance is posited as the theoretical foundation of all interventions in the BabyFirst home visiting program, including the role modelling of problem-solving skills, the provision of parenting education, the modelling of effective parent-child interaction, and the linking of families with needed community resources. In the present study, the home visiting working alliance was related to the proximal outcomes theorized to arise directly from program interventions, namely improvements in family, parent, and parent-child outcomes (see Figure 3, A). In summary, Emde's framework (Emde et al., 2000, 2004) and Bordin's (1994) trans-theoretical model of the alliance, as measured by Horvath's (1994) empirically-validated scale, the Working Alliance Inventory (WAI), guided the present study. Further, the present study followed recent recommendations for the next generation of prevention and early childhood intervention research. It investigated (a) the effectiveness of early childhood intervention under real-world conditions and (b) the effective ingredients of early childhood intervention, in particular the role of the working alliance in predicting outcome.

Hypotheses

Below are the research hypotheses derived from the foregoing theoretical framework:

Hypothesis 1: The effectiveness hypothesis. The *effectiveness hypothesis* asserts that participation in the BabyFirst home visiting program for one year after the birth of a child will improve family, parent, and parent-child outcomes at 1 year.

Hypothesis 2: The internal working model-early alliance hypothesis. The *internal working model-early alliance hypothesis* asserts that, in the BabyFirst program group, more positive internal working models of parents, represented by their pre-intervention ratings of relationships with others, will predict more positive parent ratings of the early alliance in home visiting.

Hypothesis 3: The early alliance-outcome hypothesis. The *early alliance-outcome hypothesis* asserts that, in the BabyFirst program group, a stronger early alliance between parents and home visitors, will predict better family, parent, and parent-child outcomes at 1 year.

Hypothesis 4: The alliance improvement-outcome hypothesis. The *alliance improvement-outcome hypothesis* asserts that, in the BabyFirst program group, a home visiting alliance that improves over the first year of home visiting will predict better family, parent, and parent-child outcomes at 1 year.

Method

Participants

Non-identifying data for secondary analysis in the present study were available regarding two groups of families who have (a) participated in the in-depth assessment using the Family Stress Checklist (FSC), (b) provided informed consent to participate in the BabyFirst evaluation, and (c) have completed baseline and age 1 year outcome measures. *BabyFirst families* scored at or above the cut-off score (i.e., ≥ 25 on the FSC), meeting the eligibility criterion for home visiting, and voluntarily accepted these services (see Figure 1). *Comparison families* scored below the cut-off score (i.e., < 25 on the FSC) and so were ineligible for home visiting (see Figure 1). Baseline and age 1 year outcome data were available for 250 families (187 BabyFirst families and 63 Comparison families). In the BabyFirst program, on average, about 75% of families score above the cut-off, creating unequal sample sizes between groups. As described in a later section, sample sizes for the present study were determined using the *unequal n* approach (Lipsey, 1990) for increasing statistical power (see *Statistical Power Analysis* below).

Design

The use of a cut-off score on the BabyFirst in-depth assessment to determine program eligibility makes possible the use of a powerful, yet infrequently used, quasi-experimental design, the *regression-discontinuity design (RDD)* (Campbell & Stanley, 1966; Cook & Campbell, 1979; Cook & Shadish, 1994; Marcantonio & Cook, 1994; L. B. Mohr, 1995; Trochim, 1984, 1990, 2000). The RDD¹⁰ is a pretest-posttest control group design that is characterized by its unique method of assignment to intervention: Participants are assigned to either the intervention group or control group solely on the basis of a cut-off score on a pretest measure. The RDD is one of two “convincing quasi-experiments” (Marcantonio & Cook, 1994, pp. 133-134) that stand out because

¹⁰ I am indebted to Jeannette Filion-Rosset for introducing me to the RDD.

of their high quality of causal inference.¹¹ The RDD is “a special case that stands between experimental and quasi-experimental designs” (L. B. Mohr, 1995, p. 133; see also Judd & Kenny, 1981, p. 81). Inferences drawn from a well-implemented RDD are comparable in internal validity to conclusions from randomized experiments (Cook & Shadish, 1994; L. B. Mohr, 1995; Trochim, 2000), because in both designs the selection process (i.e., the assignment of participants to intervention and control groups) is completely known. In true experiments, random assignment is used, and in the RDD, a pretest cutoff score is used. Cook and Shadish (1994, p. 563) cite earlier studies (Goldberger, 1972a, 1972b; Rubin, 1977) that provide formal statistical proofs that the RDD provides an unbiased estimate of treatment effects, just like the randomized experiment (see also Cappelleri & Trochim, 2003).

The RDD is so named because a regression line is plotted to relate the assignment and outcome variables. If the treatment is effective, a discontinuity in the regression line should occur at the cutoff point (Cook & Shadish, 1994, p. 563), as illustrated in Figure 4. This discontinuity is an estimate of the effect of the treatment for individuals near the cutoff point (Reichardt & Bormann, 1994). The direction of the effect is interpreted based on the sign of the regression coefficient for the group membership variable (intervention or control) and the direction of scale of the pretest (Trochim, 2000). By comparison, the absence of a discontinuity is interpreted as a null effect, as illustrated in Figure 5.

¹¹ The other is the interrupted time series design; Campbell (1969) was the first to designate the two designs with this special status.

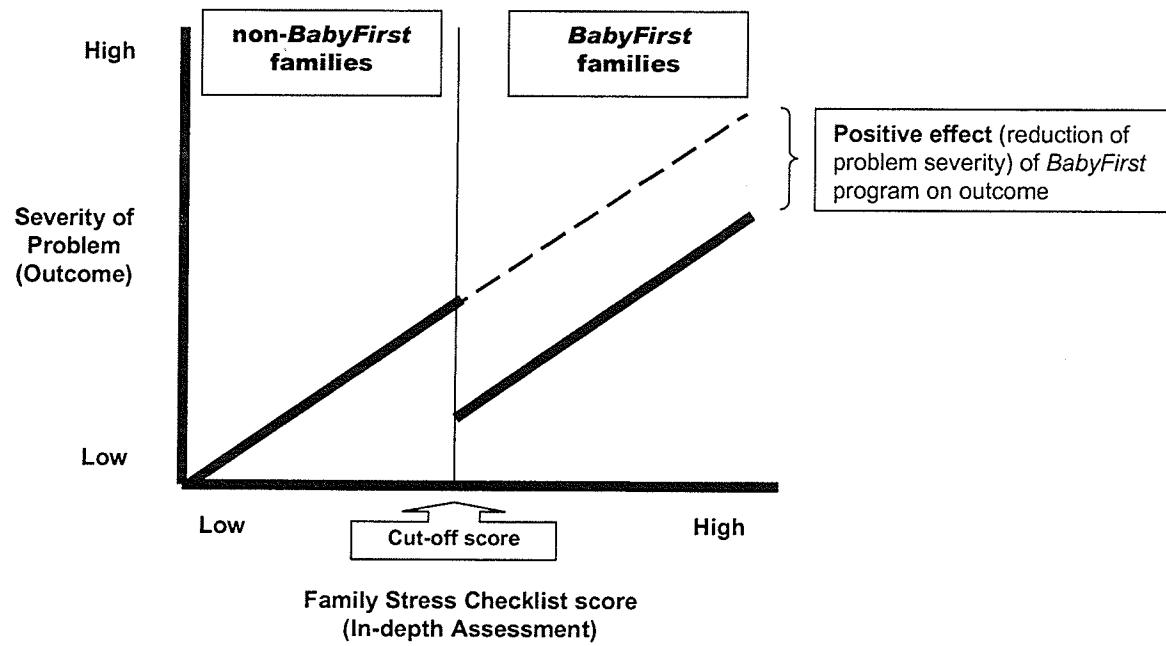


Figure 4. Illustration of program effectiveness in the regression-discontinuity design.

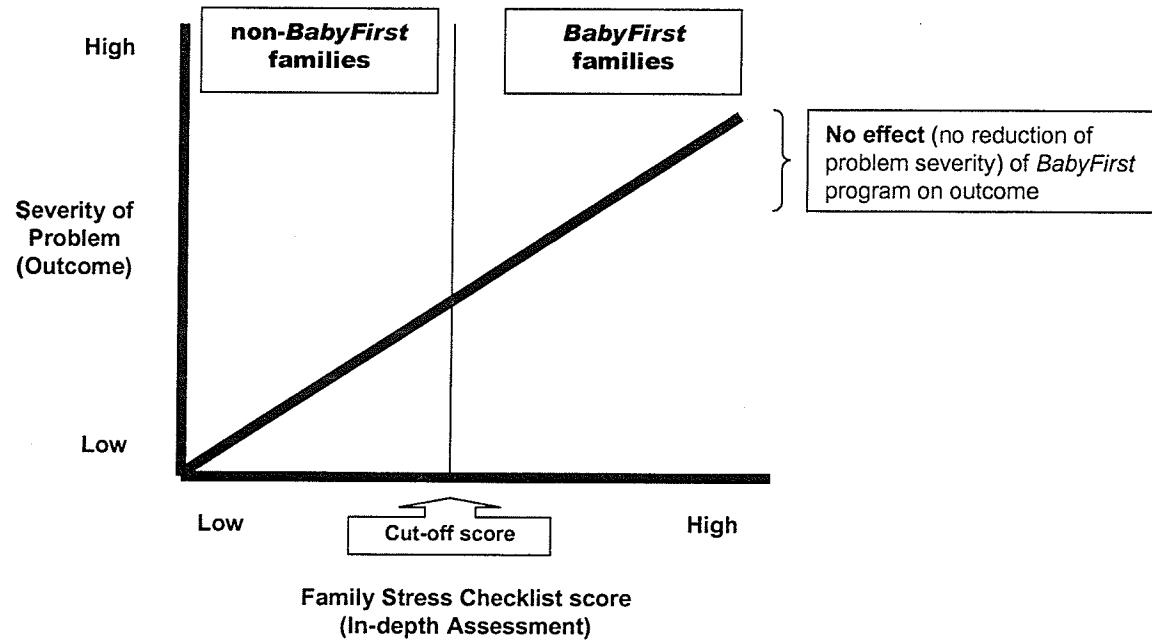


Figure 5. Illustration of program non-effectiveness in the regression-discontinuity design.

Although the RDD dates back to the 1960s (see Campbell & Stanley, 1966; Thistlethwaite & Campbell, 1960), its first major field tests did not occur until the mid-1970s (Trochim, 2000). A major reason for its infrequent use is that “at first glance, the design doesn’t seem to make sense” (Trochim, 2000). “Despite the design’s flexibility and impeccable logic, many practicing researchers are still skeptical. It seems implausible to them that a cutoff-based assignment process that necessarily creates a selection difference can rule out selection! Yet this is the case” (Cook & Shadish, 1994, p. 565). Another disincentive for its use relates to sample size: The RDD requires 2.75 times more participants than the randomized experiment to attain similar levels of statistical power (Goldberger 1972; cited in Cappelleri, Darlington, & Trochim, 1994, and Cook & Shadish, 1994). In the RDD, with alpha = .05 and power = .80, (a) to detect a large effect size (partial r of .51 or greater), a sample size of 68 per group is required; (b) to detect a medium effect size (partial r of .36 or greater), a sample size of 150 per group is required; and (c) to detect a small effect size (partial r of .14 or greater), an enormous sample size of 1078 per group is required (Cappelleri et al., 1994).

As in randomized experiments, using covariates in the RDD can dramatically increase statistical power (Judd & Kenny, 1981, p. 97). Covariates can increase the effect size, in some cases by as much as twofold or threefold (Lipsey, 1990) and, in turn, reduce sample size requirements. Covariates accomplish this by adjusting the outcome variable for predictable but unwanted sources of variability. According to Judd and Kenny (1981), a covariate in the RDD must not be the assignment variable and must not be caused by the outcome variable. Moreover, “its partial correlation with the treatment, controlling for the assignment variable, should on the average be zero” (Judd & Kenny, 1981, p. 87). The use of covariates in the RDD is discussed in greater detail and with examples in the *Statistical Power Analysis* subsection below.

The principal advantage of the RDD is ethical (Trochim, 2000): “[I]t can be used whenever policy dictates that special need or merit should be a prerequisite for access to the particular services whose effectiveness is to be evaluated” (Cook & Shadish, 1994, p. 564). The

RDD is, therefore, an optimal evaluation design for large-scale programs, such as BabyFirst, that are targeted to families with the greatest needs.

Selection of Study Variables

BabyFirst evaluation measures were selected to reflect theoretically important and empirically established determinants and outcomes of children's well-being at multiple ecological levels of the child, parent, family, and community (Bronfenbrenner, 1979), including both risk and protective factors, as outlined in the BabyFirst program theoretical model (see Figure 3). Several variables were measured by scales from the landmark National Longitudinal Survey of Children and Youth (NLSCY) (HRDC & Statistics Canada, 1996). The NLSCY is an initiative of the Government of Canada to develop a national database on the characteristics and life experiences of children and youth in Canada as they grow from infancy to adulthood. The NLSCY includes a national, stratified, multistage probability sample of approximately 23,000 Canadian children who were ages 0 to 11 years in 1994/1995. Comprehensive data on children's individual, family, preschool, school, and community characteristics, as well as children's physical, social, emotional, cognitive, language, academic, and behavioural outcomes, are collected every two years from multiple informants (parents, teachers, principals, and children ages 10 years and over). The NLSCY will continue until the youngest children in the longitudinal cohort (currently numbering about 15,000 children) reach age 25 in the year 2018 (Brink & McKellar, 2000).

Selection of measures for the NLSCY was based on several criteria (Statistics Canada & HRDC, 1995). Measures were to be (a) concise, (b) successfully used or considered suitable for use in a household survey within the general population (i.e., not just with clinical or institutional samples), (c) suitable for use by lay interviewers with Canadians of various ethnocultural and socioeconomic backgrounds, (d) comparable to studies conducted both within Canada and abroad, (e) comparable across different age groups at one point in time to permit cross-sectional analyses, (f) applicable throughout each child's development to permit longitudinal analyses, (g)

psychometrically acceptable with complete references available (with modified or adapted measures tested for reliability and validity), and (h) available in both official languages (Statistics Canada & HRDC, 1995). Use of NLSCY parent-rated measures in the BabyFirst evaluation ensured that they had been extensively field-tested in Canadian populations and allowed for statistical comparison of BabyFirst evaluation data with statistically representative provincial and national NLSCY data. Specific NLSCY measures and additional measures are described next.

Baseline Variables

Six baseline variables were included in the present analyses. These variables assessed child, parent, family, and neighbourhood characteristics at age 3 months (see Figure 3, top box).

Child characteristic. The key child characteristic was difficult temperament, which refers to frequent and intense expression of negative affect (e.g., fussing and crying, soothability, irritability). It was measured by the *NLSCY Child Temperament scale*, based on the 10-item Difficult subscale of the Infant Characteristics Questionnaire (ICQ) (Bates, 1984). The child's parent rates each item using its respective 7-point scale. Anchor points are usually item-specific. For example, the item regarding the child's degree of temperamental difficulty (e.g., "How easy or difficult is it for you to calm or soothe your child when he/she is upset?") is rated on a 7-point scale ranging from *very easy* (1) to *difficult* (7); whereas the item regarding the frequency of difficult child behaviours (e.g., "How many times per day, on average, does your child get fussy and irritable—for either short or long periods of time?") is rated on a 7-point scale ranging from *never* (1) to *15 times per day or more* (7). The Difficult subscale score is computed by averaging the 10 item scores. The Difficult subscale of the ICQ was confirmed by factor analyses of data regarding 4- to 6-month-olds (Bates, Freeland, & Lounsbury, 1979) and 13- and 24-month-olds (Bates & Bayles, 1984). The Difficult subscale has an internal consistency of .79 and a test-retest reliability of .70 over an average of 30 days (Bates et al., 1979). A major review of the reliability (internal consistency, test-retest, stability) and validity (convergent, concurrent, predictive) of 26 temperament measures concluded that "in studies where the major goal is correlates of outcomes

of difficult temperament, the use of the difficult factor score from the Infant Characteristics Questionnaire would be recommended" (Hubert, Wachs, Peters-Martin, & Gandour, 1982, p. 581). Considerable evidence suggests that "infant scales with different names are actually measuring similar constructs" (Rothbart & Bates, 1998, p. 112), with intercorrelations across scales averaging .63 for mother reports and .64 for day care teacher reports (Goldsmith & Rieser-Danner, 1986, cited in Rothbart & Bates, 1998). Amidst vigorous and ongoing debate regarding the best method for measuring temperament, the evidence to date supports the validity and use of parent-report measures of temperament (Rothbart & Bates, 1998).

In the NLSCY, the internal consistency coefficients of the Difficult subscale were .77, .79, .79, and .79 for children aged 0, 1, 2, and 3 years, respectively (Japel, Normand, Tremblay, & Willms, 2002), and construct validity was supported by correlations with parent and family variables (Normand et al., 1996). Several recent longitudinal studies have confirmed that difficult temperament during the first years of life predicts psychosocial maladjustment in childhood, adolescence, and adulthood (Caspi, 1998; Rothbart & Bates, 1998).

Parent characteristics. Parent characteristics included parental depression (measured in the *NLSCY Adult Health scale*), parental psychological well-being, and parental Aboriginal status. As in the NLSCY, parental depression was measured with a shortened 12-item version of the Centre for Epidemiological Studies–Depression Scale (CES-D) (Radloff, 1977). The original CES-D is a 20-item self-report instrument developed at the U.S. National Institute of Mental Health. It is a widely used screening measure for designating depressed and non-depressed individuals in nonclinical populations. Respondents rate each item on a 4-point scale ranging from *rarely or none of the time* (1) to *most or all of the time* (4) according to the frequency with which they experienced that symptom during the preceding week. Four CES-D items are reverse coded. A total score is derived by averaging the item scores, with higher scores indicating higher levels of depressive symptoms. It has high internal consistency (.85 in a general population sample and .90 in a clinical sample) and good test-retest reliability over 2 to 8 week intervals (.61

to .67) in its validation study (Radloff, 1977). In the NLSCY, the shortened 12-item CES-D (self-rated by parents) demonstrated good internal consistency (.82) in a large Canadian probability sample ($n = 13,140$) (Statistics Canada & HRDC, 1995) and was strongly and positively correlated with children's social impairment and emotional and behavioural problems, in both single-mother and two-parent families across Canada (Lipman et al., 2002).

It has been widely acknowledged that there are inherent limitations to using self-report instruments like the CES-D as indicators of major depression (Coyne, Thompson, Klinkman, & Nease, 2002). In a representative study, the CES-D had a sensitivity of 79.5% and a specificity of 71.1%, but only a modest positive predictive value of 28.5% using a *Diagnostic and Statistical Manual of Mental Disorders* (DSM-III-R) (American Psychiatric Association, 1987) diagnosis of major depression as the comparison (Fechner-Bates, Coyne, & Schwenk, 1994). Diagnosis was determined using the Structured Clinical Interview for the DSM-III-R (Spitzer, Williams, Gibbon, & First, 1989), which assesses current and lifetime psychiatric status for major Axis I disorders using DSM-III-R criteria, including major depression. In practical terms, the CES-D identified about 1 in 3 patients as needing further in-depth diagnostic evaluation. However, less than one-third of these patients were actually depressed, and 1 in 5 depressed patients would have been missed if a clinician had relied exclusively on the CES-D to detect depression (Fechner-Bates et al., 1994). Even by raising cutoff scores, modifying scoring, and using advanced psychometric analyses designed to identify and optimally combine only the most effective CES-D items, there is inevitably some trade-off between sensitivity and specificity, so limitations persist (Santor & Coyne, 1997). Most adults who have been designated as depressed on the basis of self-report measures such as the CES-D are not clinically depressed. It may be more appropriate to describe measures like the CES-D as measures of self-reported psychological distress rather than as measures of clinical depression (Coyne et al., 2002; Fechner-Bates et al., 1994).

Parental psychological well-being was measured by the brief 18-item version of the *Scales of Psychological Well-Being (PWB)* (Ryff, 1995; Ryff & Keyes, 1995). The original scales

(Ryff, 1989) included 120 items that comprised 20 items for each of six dimensions of psychological well-being: Autonomy, Environmental Mastery, Personal Growth, Positive Relations, Purpose in Life, and Self-Acceptance. Definitions of the six PWB dimensions (from Ryff & Keyes, 1995, p. 727) are presented in Table 2. To accommodate time and cost restrictions of a national U.S. survey, 3 of the original 20 items were selected to measure each of the six PWB dimensions (Ryff & Keyes, 1995). Items are self-rated on a 6-point scale ranging from *strongly disagree* (1) to *strongly agree* (6). Negative items are reverse-scored so that higher scores on each scale reflect higher levels of PWB. The six PWB scales are averaged to yield an overall PWB score. The brief 18-item version continued to meet psychometric criteria, with each item correlating strongly and positively with only its own scale. Modest scale intercorrelations ranged from .13 to .46 and the shortened scales correlated from .70 to .89 with their 20-item parent subscales (Ryff & Keyes, 1995). Confirmatory factor analyses of data from a national probability sample in the United States ($n = 1,108$) supported the proposed 6-factor model, with a single second-order super factor (Ryff & Keyes, 1995).

Estimates of internal consistency for the brief 18-item version were low to modest, ranging from .33 to .56 in the aforementioned study (Ryff & Keyes, 1995) and from .37 to .59 in a recent U.S. national probability sample ($n = 3,032$) (Keyes, Shmotkin, & Ryff, 2002). "The modest alpha coefficients likely reflect the small number of indicators per scale and the fact that items were chosen to represent the conceptual breadth within each construct rather than to maximize internal consistency" (Ryff & Keyes, 1995, p. 721). That is, an a priori decision was made to select items that represented the theoretical scope of each well-being construct. "The diversity in item content reduces the internal consistency coefficients for the 3-item scales but keeps them meaningfully connected to the theoretical starting points for the parent scales" (Keyes et al., 2002, p. 1011). Such low subscale reliabilities could limit the degree to which PWB scales correlate with other variables, but this has not occurred (Keyes et al., 2002).

Table 2***Definitions of Theory-Guided Dimensions of Well-Being******Autonomy***

High scorer: is self-determining and independent, able to resist social pressures to think and act in certain ways, regulates behavior from within, evaluates self by personal standards.

Low scorer: is concerned about the expectations and evaluations of others, relies on judgements of others to make important decisions, conforms to social pressures to think and act in certain ways.

Environmental Mastery

High scorer: has a sense of mastery and competence in managing the environment, controls complex array of external activities, makes effective use of surrounding opportunities, able to choose or create contexts suitable to personal needs and values.

Low scorer: has difficulty managing everyday affairs, feels unable to change or improve surrounding context, is unaware of surrounding opportunities, lacks sense of control over external world.

Personal Growth

High scorer: has a feeling of continued development, sees self as growing and expanding, is open to new experiences, has sense of realizing his or her potential, sees improvement in self and behavior over time, is changing in ways that reflect more self-knowledge and effectiveness.

Low scorer: has a sense of personal stagnation, lacks sense of improvement or expansion over time, feels bored and uninterested with life, feels unable to develop new attitudes or behaviors.

Positive Relations with Others

High scorer: has warm, satisfying, trusting relationships with others; is concerned about the welfare of others; capable of strong empathy, affection, and intimacy; understands give and take of human relationships.

Low scorer: has few close, trusting relationships with others; finds it difficult to be warm, open, and concerned about others; is isolated and frustrated in interpersonal relationships; not willing to make compromises to sustain important ties with others.

Purpose in Life

High scorer: has goals in life and a sense of directedness, feels there is meaning to present and past life, holds beliefs that give life purpose, has aims and objectives for living.

Low scorer: lacks a sense of meaning in life; has few goals or aims, lacks sense of direction; does not see purpose in past life; has no outlooks or beliefs that give life meaning.

Self-Acceptance

High scorer: possesses a positive attitude toward the self; acknowledges and accepts multiple aspects of self, including good and bad qualities; feels positive about past life.

Low scorer: feels dissatisfied with self, is disappointed with what has occurred in past life, is troubled about certain personal qualities, wishes to be different than what he or she is.

Construct validity studies of the scales of PWB show convergent and discriminant correlations with measures of personality (Schmutte & Ryff, 1997); happiness, life satisfaction, and depression (Ryff & Keyes, 1995); and subjective well-being (Keyes et al., 2002). Moreover, other frequently used indicators of psychological well-being neglect key theoretical aspects, such as Positive Relations with Others, Purpose in Life, Personal Growth, and Autonomy (Ryff & Keyes, 1995).

The Canadian Study of Health and Aging (CSHA) (Clarke, Marshall, Ryff, & Rosenthal, 2000) replicated the PWB reliability and validity findings from the two U.S. samples (Keyes et al., 2002; Ryff & Keyes, 1995) in the large national probability sample of the CSHA ($n = 4,960$) using the brief 18-item version. Scale intercorrelations ranged from .04 to .39, with a median correlation of .19, indicating that "the six dimensions are not plagued by excessive construct overlap" (Clarke et al., 2000, p. 145). Similar to the U.S. findings (Keyes et al., 2002; Ryff & Keyes, 1995), the internal consistency of the brief scales of PWB in the CSHA was modest, ranging from .26 to .52 (Clarke et al., 2000). Nonetheless, Canadian data replicated U.S. data in terms of the sociodemographic correlates, health correlates, and cross-sectional patterns of PWB in the general population (Clarke et al., 2000).

Parental Aboriginal status was measured with a self-report item used by Statistics Canada in the NLSCY and the Census. Aboriginal peoples include First Nations, Métis, and Inuit peoples. Compared to the general population in Manitoba (Hallett, Nemeth, Stevens, & Stewart, 2000; Martens et al., 2002) and Canada (MacMillan, MacMillan, Offord, & Dingle, 1996; Tookenay, 1996; Young, 2003), Aboriginal people are at significantly greater risk of mortality and morbidity across the life course, especially during childhood (Canadian Institute of Child Health [CICH], 2000; Statistics Canada, 2004a; Turcotte & Zhao, 2004).

Family characteristics. Family characteristics included perceived social support. This was measured by the *NLSCY Social Support scale*, a brief version of the Social Provisions Scale (SPS) (Cutrona, 1986; Cutrona & Russell, 1987). The SPS is a 24-item self-report measure of

perceived social support. It was developed to assess the six relational provisions identified by Weiss (1974, cited in Cutrona, 1986): attachment, social integration, reassurance of worth, reliable alliance, guidance, and opportunity for nurturance. Each provision is assessed by four items, two that describe the presence and two that describe the absence of the provision, rated on 4-point scales (ranging from *strongly disagree* to *strongly agree*). Each social provision subscale is scored by averaging reversed negative items and positive items. An overall social support score is formed by averaging the six social provision subscale scores. Internal consistency for the overall score ranged from .85 to .92 across several population samples, with alpha coefficients for individual subscales ranging from .64 to .76 (Cutrona, 1986). Factor analysis has confirmed a six-factor structure that corresponds to the six social provisions (Cutrona, 1986). There is extensive validity evidence for the scale in both adult and adolescent populations (Cutrona, 1986; Cutrona & Russell, 1987). In the NLSCY sample ($n = 13,253$), the internal consistency of the 6-item parent-rated NLSCY Social Support scale was .82 (Statistics Canada & HRDC, 1995).

Neighbourhood characteristics. Neighbourhood characteristics included neighbourhood safety, measured by a section of the *NLSCY Neighbourhood Scale*, a brief version of the Neighborhood Characteristics Questionnaire (NCQ) (Barnes McGuire, 1997). This instrument is a modification of the Neighborhood Questionnaire (Simcha-Fagan & Schwartz, 1986)¹², shortened and modified to be relevant to parents with young children, to be used in community needs assessments, and to be applicable to families from a wide range of sociocultural backgrounds (Barnes McGuire, 1997). Confirmatory factor analysis of the NCQ revealed four factors: Street Crime and Neighborhood Quality (e.g., concerns regarding personal safety), Local Social Networks (e.g., informal neighbouring), Neighborhood Attachment (e.g., likelihood of moving), and Disorder (e.g., litter, trash, graffiti) (Barnes McGuire, 1997). In the latter study,

¹² The original Neighborhood Questionnaire had 54 items and nine scales derived from factor analysis: informal neighboring, neighborhood attachment, network size and breadth, neighborhood-level organizational involvement, neighborhood anomie, local personal ties, social disorder, conflict substructure, and illegal economy. It was developed for a study of 553 residents of 12 neighbourhoods in New York City during the mid-1980s (Simcha-Fagan & Schwartz, 1986). The study found a significant negative relationship between the rate of delinquency and rates of organizational participation by local residents.

internal consistency coefficients for the four factors were .85, .82, .81, and .77, respectively (Barnes McGuire, 1997). NCQ items are rated on 3-point scales (*yes, maybe, no*). Construct validity was supported by multitrait-multimethod correlations with self-report measures of parental mental health and social support, knowledge of community services and providers, and independent neighborhood observations (Barnes McGuire, 1997). The NLSCY Neighbourhood Scale was modified to be a 4-point scale (*strongly disagree, disagree, agree, strongly agree*) completed by parents (Statistics Canada & HRDC, 1995).

Neighbourhood safety appears to be an important influence on parenting and child development (Leventhal & Brooks-Gunn, 2000). For example, in the NLSCY, neighbourhood safety was associated with children's developmental outcomes, even after controlling for demographic characteristics (e.g., child gender, parental age, parental education, family income), family social support, parental health, and parental depression (Kohen, Brooks-Gunn, Leventhal, & Hertzman, 2002). Based on this evidence, the NLSCY Neighbourhood subscale related to neighbourhood safety (3 items) was included in the BabyFirst evaluation, and selected for use in the present study. Neighbourhood variables are an important focus in emerging research on child development (Leventhal & Brooks-Gunn, 2000) and early childhood intervention (McGuigan, Katzev, & Pratt, 2003a, 2003b; Olds, 2003).

Program Variables

Working alliance: Parent and home visitor ratings. The home visiting working alliance was measured by parent and home visitor ratings of their working alliance, using parallel self-report short forms of the Working Alliance Inventory (WAI) (Horvath, 1994; Tracey & Kokotovic, 1989; see also Busseri & Tyler, 2003), modified for use in home visiting with the permission of the author (A. O. Horvath, personal communication, July 24, 1998). The Working Alliance Inventory (WAI) comprises three subscales (12 items each for a total of 36 items) measuring the three components of Bordin's trans-theoretical model, namely goals, tasks, and bond. Each item is rated for the degree to which it applies to the alliance on a 7-point scale, from

never (1) to *always* (7). Subscale scores are computed by averaging their respective 12 items. The general alliance score is computed by averaging all 36 items. The WAI can be completed from three perspectives: client, therapist, and independent observer.

Reliability and validity of the Working Alliance Inventory (WAI). In early reliability studies, internal consistency estimates of the three WAI subscales ranged from .85 to .92, with the general score (i.e., the average of the three scales) attaining a coefficient alpha of .93 (Horvath & Greenberg, 1989). Tracey and Kokotovic (1989) selected the four items with the highest factor loadings from each of the three WAI subscales to form a short version of the WAI. Their confirmatory factor analysis of this short version of the WAI revealed a hierarchical, bilevel factor structure with three first-order factors (goals, tasks, and bond) and a second-order general alliance factor. The short version of the WAI has high internal consistency reliability, with a Cronbach's alpha of .98 for client-rated general alliance (.90, .90, .92 for goals, tasks, bond) and .95 for therapist-rated general alliance (.88, .83, .91 for goals, tasks, bond) (Tracey & Kokotovic, 1989). An early multitrait-multimethod analysis demonstrated the construct validity of the original WAI (Horvath & Greenberg, 1986) according to the convergent and discriminant validation criteria of Campbell and Fiske (1959), including significant correlations between the WAI and client characteristics, therapeutic activities, and therapeutic outcome (reviewed in Horvath, 1994). A recent study confirmed the interchangeability of the original WAI and the short form of the WAI (Busseri & Tyler, 2003).

In a recent meta-analysis of alliance measures (Martin et al., 2000), the reliability of the WAI matched or exceeded all other available alliance measures, with an overall reliability of .84 across 15 studies ($SD = .11$). When Cronbach's alpha was used, the average internal consistency was .90 ($n = 8$ studies, $SD = .05$); when inter-rater reliability was reported, the average reliability was .92 ($n = 2$ studies, $SD = .01$); and when a test-retest correlation was calculated, the average correlation was .73 ($n = 5$ studies, $SD = .11$) (Martin et al., 2000).

As described in an earlier section, the average weighted alliance-outcome correlation of the WAI was .24 across 22 studies (Martin et al., 2000).¹³ This was similar to the average r for all alliance measures (.22, $SD = .12$) across the 79 studies included in the full meta-analysis (Martin et al., 2000). The authors also conducted a file drawer analysis (Rosenthal, 1979, 1995)¹⁴ and showed that it would take 331 studies averaging null findings to reduce the correlation from .22 to .05. Thus, it is highly unlikely that a sufficient number of unreported studies exist to substantively reduce the overall correlation. Martin et al. (2000) concluded that "the WAI is likely to be appropriate for most research projects" (p. 447).

Age 1 Outcome Variables

Within approximately 2 months after the child's first birthday, families were re-assessed with the same evaluation measures administered at baseline. Measures included in the present study were selected from the larger set of measures collected in the BabyFirst program evaluation. As discussed above in the *Program Theoretical Model*, six outcomes in three ecological domains (family, parent, and parent-child) were included in the present study.

Family outcome. Family outcome consisted of overall social support, measured by the parent-rated NLSCY Social Support scale (see *Baseline Variables* section above).

Parent outcomes. Self-rated parent outcomes included parental depression, measured by a section of the NLSCY Adult Health scale (total score on the shortened 12-item CES-D); and parental psychological well-being (overall PWB score, see *Baseline Variables* section above). Follow-up RDD analyses for each of the six PWB factor scores (Autonomy, Environmental Mastery, Personal Growth, Positive Relations with Others, Purpose in Life, and Self-Acceptance;

¹³ This was lower than the findings in the Horvath (1994) meta-analysis (.33, 95% confidence interval of .23 to .43). Three of the 8 studies in Horvath (1994) were not included in the Martin et al. (2000) meta-analysis, although they both used the criteria in the Horvath and Symonds (1991) meta-analysis.

¹⁴ Coined by Rosenthal (1979), the *file drawer problem* refers to "the well-supported suspicion that the studies retrievable in a meta-analysis are not likely to be a random sample of all studies actually conducted" (Rosenthal, 1995, p. 189). A file drawer analysis calculates the number of unlocated studies with null results that would need to exist in file drawers to substantively reduce the effect size. If the overall effect size of the meta-analysis is reduced by the addition of just a few more null results, the finding is not considered resistant to the file drawer problem (Rosenthal, 1995; see also Becker, 1994; Begg, 1994; Greenhouse & Iyengar, 1994).

see Table 2) were contingent on results for the overall PWB score. If results indicated program effects on overall PWB, then further RDD analyses were conducted on the six PWB factors.

Parent-child outcomes. Parent-child outcomes included the Positive Parenting and Hostile/Ineffective Parenting factor scores on the NLSCY Parenting scale as well as scores on the NLSCY Literacy scale, which measured the frequency of parent-child literacy activities, such as reading together. Both scales were completed by parents.

The NLSCY Parenting scale is a modified version of the Parent Practices Scale (PPS), a self-report instrument on parents' patterns of interaction with their preschool children (Strayhorn & Weidman, 1988). The original PPS comprised 34 items and demonstrated good internal consistency (.78), 6-month test-retest reliability (.79), and excellent construct validity in multitrait-multimethod analyses (Campbell & Fiske, 1959) conducted by the PPS authors (Strayhorn & Weidman, 1988). These included significant correlations with parenting measures (blind independent ratings of videotaped interactional behaviour of parents and children) and child outcomes (e.g., parent ratings of children's home behaviour, blind classroom observational ratings of children's classroom behaviour, blind tests of children's verbal ability) (Strayhorn & Weidman, 1988). In the NLSCY, seven items from the PPS are used for parents with children aged 0 to 23 months. Items are rated on a 5-point scale describing the frequency of the parenting behaviour, ranging from *never* (1) to *many times a day* (5). Factor analysis of NLSCY data for this age group ($n = 4,696$) confirmed the presence of two factors: Positive Parenting and Hostile/Ineffective Parenting (Statistics Canada & HRDC, 1995). Positive Parenting includes providing praise to the child, talking, playing, laughing, and engaging in special activities with the child. This factor comprises five items and showed good internal consistency (.73) in the NLSCY. The Hostile/Ineffective Parenting factor includes expressing annoyance with the child and expressing that the child is not as good as others. As this factor comprises only two items, its

internal consistency was low (.39) in the NLSCY.¹⁵ The NLSCY Parenting scale has demonstrated excellent construct validity across age groups (0 – 11 years) in the NLSCY. In multivariate analyses of NLSCY data, the Positive Interaction and Hostile/Ineffective factors emerged as the most powerful predictors of children's learning and behaviour outcomes, overshadowing other well-established risk factor variables such as low income, single parenthood, and adolescent parenthood (Statistics Canada, 1998). For example, hostile/ineffective parenting was associated with a seven-fold increase in risk of aggressive behaviour in 2- and 3-year-olds (Statistics Canada, 1998). In multivariate analyses of NLSCY data for children ages 6 to 11 years in single-mother families ($n = 1,286$) and two-parent families ($n = 8,112$), hostile parenting and maternal depression were most strongly associated with children's social impairment, and with emotional and behavioural problems (Lipman et al., 2002). This effect was found for both types of families, suggesting that children from single-mother families develop difficulties for the same reasons as children in two-parent families. A statistically significant interaction was found between single-mother family status and hostile parenting, demonstrating a risk effect beyond that found in two-parent families (Lipman et al., 2002).

The NLSCY Literacy scale (for children ages 0 to 23 months) includes an item that inquires about the frequency of parent-child literacy activities (e.g., reading to the child), rated on an 8-point scale ranging from *rarely* (1) to *many times a day* (8). The NLSCY Literacy Scale items were originally developed by DeBaryshe for the U.S. National Assessment of Educational Progress (NAEP) (Statistics Canada & HRDC, 1995; see also DeBaryshe, 1992; NAEP, 1997). The NAEP is a congressionally-mandated study funded by the Office of Educational Research and Improvement, U.S. Department of Education, to determine the nation's progress in education (see <http://nces.ed.gov/nationsreportcard/about/>).

¹⁵ A factor analysis of the NLSCY Parenting Scale for 2- to 11-year-olds in the NLSCY ($n = 18,135$) revealed three factors: Positive Interaction (5 items), Hostile/Ineffective (7 items), and Consistency (5 items), with internal consistency coefficients of .81, .71, and .66 respectively (Statistics Canada & HRDC, 1995). Note that the internal consistency of the Hostile/Ineffective factor has increased as the number of scale items increased.

Summary and Research Framework

As described in an earlier section (see *What We Don't Know: Effective Ingredients of Early Childhood Intervention* above), Berlin et al. (1998) proposed a tripartite framework for understanding what makes early childhood interventions work: the *program* (program characteristics, program staff characteristics, and cultural relevance), its *participants* (child characteristics, family characteristics, community characteristics, and cumulative risk), and their *interaction* (participant-staff relationship and match of service with need) (see Table 1). As described in the foregoing sections, data from the BabyFirst program were available for two of the three principal dimensions of this framework: participants and interaction. Specifically, these data represent five of the nine framework subcomponents: child characteristics, family characteristics, community characteristics, and cumulative risk (from the *participant* dimension); and participant-staff relationship (from the *interaction* dimension).

Statistical Considerations: Significance Tests, Effect Sizes, and Confidence Intervals

Before discussing the approach to statistical inference used in the present study, it is important to consider current discussions and controversies regarding significance testing, the most common approach to statistical inference used in psychological research. The latest version of the American Psychological Association (APA) *Publication Manual* offers a starting point:

The field of psychology is not of a single mind on a number of issues surrounding the conduct and reporting of what is commonly known as *null hypothesis significance testing*. These issues include, but are not limited to, the reporting and interpretation of results of hypothesis tests, the selection of effect size indicators . . . A discussion of these and other issues can be found in Wilkinson and the Task Force on Statistical Inference (1999). It is not the role of the *Publication Manual* to resolve these issues (APA, 2001, pp. 21-22, italics in original).

Null hypothesis significance testing (NHST) has long been controversial, with criticisms as old as the method itself (Kline, 2004; B. Thompson, 2002a), dating back nearly a century (Boring, 1919), recurring in the 1940s and 1950s (Berkson, 1938, 1942; Hogben, 1957), and resurging in the 1960s and 1970s (Bakan, 1966; Carver, 1978; Lykken, 1968; Meehl, 1967, 1978; Morrison & Henkel, 1970; Rozeboom, 1960). Kaufman (1998) called the NHST controversy "the major

methodological issue of our generation" (p. 1). In the quotation above, the *APA Publication Manual* references the final report of the APA Task Force on Statistical Inference (TFSI):

In the light of continuing debate over the applications of significance testing in psychology journals and following the publication of Cohen's (1994) article, the Board of Scientific Affairs (BSA) of the American Psychological Association (APA) convened a committee called the Task Force in Statistical Inference (TFSI) whose charge was "to elucidate some of the controversial issues surrounding applications of statistics including significance testing and its alternatives" (Wilkinson & TFSI, 1999, p. 594)

The TFSI was co-chaired by three statistical luminaries in the field of psychology (Robert Rosenthal, Robert Abelson, and Jacob Cohen), and included nine other leading statisticians, with support from four methodological giants in the field (Lee Cronbach, Paul Meehl, Frederick Mosteller, and John Tukey), who served as senior advisors to the committee. Regarding NHST, the report indicated that

It is hard to imagine a situation in which a dichotomous accept-reject decision is better than reporting an actual p value or, better still, a confidence interval. Never use the unfortunate expression "accept the null hypothesis." Always provide some effect-size estimate when reporting a p value. Cohen (1994) has written on this subject in this journal. All psychologists would benefit from reading his insightful article (p. 599, italics in original).

In order to "initiate discussion in the field about changes in current practices of data analysis and reporting" (Wilkinson & TFSI, 1999, p. 594), the final report offers important guidelines regarding method (design, population, sample, assignment, measurement), results (complications, analysis), and discussion (interpretation, conclusions).¹⁶ In its subsequent edition, the APA (2001) *Publication Manual* went on to strongly recommend the reporting of effect sizes, confidence intervals, statistical power analyses, and, if significance tests are still used, the exact probabilities (*p* values) of these tests.

The NHST controversy is due "in large part from the realization that conventional significance-testing procedures provide an impoverished and possibly even misleading view of how seriously to take any particular result" (Prentice & Miller, 1992, p., 160). This "misleading"

¹⁶ Not everyone was satisfied with the final report: "I don't wish to be invidious, but I am afraid that the APA committee has labored to bring forth a mouse. The report reads like a politician's "blue-ribbon" committee, coming out in favor of motherhood, the flag, and apple pie; and it has no teeth in it. It does not require or forbid anything, including the most irrational current practices. I was one of the four outside experts, named as consultants--the others being Mosteller, Tukey, and Cronbach--and I would be curious to know whether the committee paid as little attention to the other three as they did to me" (Meehl, 1998, p. 5).

nature of NHST is related to the general relationship between the *p* value and the effect size (Rosnow & Rosenthal, 2003):

$$\text{significance test} = \text{effect size} \times \text{study size}$$

Because the results of significance tests are a function of both effect size and sample size, even the most trivially small effects will be statistically significant given sufficiently large sample sizes, whereas meaningfully large effects may fail to reach statistical significance given inadequately large sample sizes.

Alternatives to NHST have been suggested and debated (Denis, 2003; Harlow et al., 1997; Kline, 2004; McCartney & Rosenthal, 2000; Nickerson, 2000; Shrout, 1997; B. Thompson, 1998), including the reporting of effect sizes alongside *p* values (Cohen, 1990, 1994) and the outright replacement of significance tests with effect sizes and confidence intervals (Schmidt, 1992, 1996). The use of effect sizes itself is over a century old (reviewed in Kline, 2004), but their increased prominence is generally attributed to the rise of modern meta-analysis and “meta-analytic thinking” led by Glass (1976) and Rosenthal (1976). Meta-analytic thinking, according to Kline (2004), requires (a) an accurate appreciation of the results of prior studies, (b) a view of one's own study as making a modest contribution, (c) reporting of results so they can be included in future meta-analyses, and (d) interpretation of new results via comparison with prior effect sizes, a view incompatible with the use of NHST as the sole way to test hypotheses.

Meehl (1997) recommended several resolutions to the NHST debate: (a) state confidence intervals before mentioning null hypothesis significance tests; (b) avoid the misleading term *significant* and indicate instead that, because the result falls outside of the confidence interval, “it is unpleasurable to explain the observed difference as arising from random sampling error” (p. 421); (c) state statistical power values if inferences are made by contrasting significant and

nonsignificant differences¹⁷; (d) compute a suitable measure of overlap (e.g., Cohen's *U3*, the percentile of the intervention group corresponding to the 50th percentile of the comparison group, Cohen, 1988; considered the most useful of overlap measures, Kline, 2004); and (e) explicitly state in the discussion section that proving the counter-null hypothesis by refuting the null hypothesis provides only weak corroboration of the theory. Finally, Meehl (1997) reminded us that "the most important property of an empirical finding is intersubjective replicability, that other investigators, relying on the description of what was done, will (almost always) make the same (or closely similar) observations" (p. 423).

Based on these and similar methodological recommendations (APA, 2001; Cohen, 1988; 1990, 1994; Dar, Serlin, & Omer, 1994; Hunter & Schmidt, 1996; Kirk, 1996; Kline, 2004; Meehl, 1990a; Rosnow & Rosenthal, 2003; Schmidt, 1992, 1996; B. Thompson, 1998, 2002a, 2002b; Valentine & Cooper, 2003; Wilkinson & TFSI, 1999), the present study focused on unstandardized effect sizes and their confidence intervals, standardized effect sizes in the context of meta-analytic findings, statistical power, distribution overlap in between-group comparisons, and replication of previous research findings, particularly from meta-analyses.

Statistical Considerations: Type I and Type II Error Rates

It is common for researchers to set the Type I error rate (or alpha) to .05 and the Type II error rate (or beta) to .20 (power = .80). This statistical convention suggests that Type I errors (false positives) are four times as serious as Type II errors (false negatives). However, it is important that each study consider the appropriate balance of Type I and Type II errors:

Recognition of the often overlooked cost of Type II errors would seem to dictate the use of criteria that are not so extremely protective against errors of Type I. More generally, decision criteria should be set so as to take into account the consequences of the various possible ways of being wrong and with recognition that the relative costs of Type I and Type II errors can vary considerably from situation to situation (Nickerson, 2000, p. 273).

¹⁷ "Failing to do so is not a minor weakness; it makes the manuscript unacceptable. Leaving out the value of the power function is as bad scientific reporting as not telling the sample size, or where you got your subjects, or which measuring instrument you used" (Meehl, 1998, p. 6).

Nickerson (2000) further notes that the use of statistical convention "can be especially problematic in situations in which a Type II error is likely to be as costly as, or even more costly than, a Type I error, and this may often be the case in applied settings" (p. 272), including those in effectiveness studies of prevention and early childhood intervention (Collins, Murphy, Nair, & Strecher, in press; Huston, 1994, 2002; Korfmacher, 2002).

This perspective is elaborated in a recent strategy for optimizing and evaluating interventions (Collins et al., in press). This strategy includes three phases: (a) the *screening* phase assesses an array of intervention components and selects those that merit further investigation; (b) the *refining* phase analyzes interactions among components identified in the screening phase, and their interrelationships with covariates. Finally, (c) the *confirming* phase evaluates the optimized intervention that is developed in the first two phases. The screening and refining phases may be alternated several times, particularly if development is taking place in a relatively uncharted area. The information gained from the activities of the screening and refining phases is then used to formulate the intervention (Collins et al., in press). As discussed in the introduction to the present study, the field of early childhood intervention research has only recently entered the uncharted areas of effectiveness and effective ingredients (or components) of intervention. Therefore, it can be considered to be in the screening and refining phases of the foregoing strategy.

In the screening and refining phases, a Type II error may be considered at least as serious as a Type I error (Collins et al., in press). That is, overlooking an effective intervention (or an active ingredient of intervention) is at least as serious as mistakenly concluding that an ineffective intervention is effective (or an inactive ingredient is active). Thus, a larger Type I error rate (e.g., .10 or .20) may be used (Collins et al., in press). Similarly, regarding harmful effects, a Type II error may be considered more serious than a Type I error. That is, failing to detect a harmful intervention has more serious consequences than falsely concluding that a benign intervention is harmful, given the ethical imperative to minimize harm. Later, when the optimized intervention is

tested in the confirming phase¹⁸, a more traditional Type I error rate (e.g., .05) may be used (Collins et al., in press).

Given the focus of the present study on the effectiveness and effective ingredients of early childhood intervention, it was considered essential to minimize Type II error rates as much as possible to detect potential effects, both positive and negative. Type II errors were considered to be at least as serious as Type I errors. Setting the Type I error rate to .20, generally considered to be the largest acceptable value, indicates a relatively high tolerance for incorrectly rejecting the null hypothesis (i.e., falsely concluding that there is an intervention effect). Based on the foregoing considerations, in the present study, the Type I error rate (alpha) and the Type II error rate (beta) were both set to values of .20, reflecting 80% confidence intervals¹⁹ (Cohen, 1990; Denis, 2003) and 80% statistical power (Cohen, 1988).

Statistical Power Analysis

Statistical power analyses were computed prior to conducting the data analyses for the present study, setting alpha = .20 (80% confidence intervals) and beta = .20 (80% power). Effect size selection for power analyses was guided by effect sizes from the literature, particularly results from meta-analyses (i.e., the “empirical literature context” for effect sizes; McCartney & Rosenthal, 2000, p. 175). The present study also used available methodological tactics to increase statistical power. According to Lipsey (1990), “in most circumstances, the first tactics a researcher should use to attain adequate statistical power are those that increase the effect size” (p. 138). These include (a) the selection of reliable and valid measures, (b) the use of statistical control of variance, and (c) maintaining the strength and integrity of the independent variable.

¹⁸ A full confirmatory intervention trial is mounted only when an optimized intervention has been reached, and only when there is sufficient potential for efficacy or effectiveness, based on information gathered in the screening and refining phases; i.e., program evaluation resources are reserved for interventions that have a demonstrably high probability of success (Collins et al., in press).

¹⁹ “Then, having found the sample effect size, you can attach a *p* value to it, but it is far more informative to provide a confidence interval. . . . (By the way, I don’t think that we should routinely use 95% intervals: Our interests are often better served by more tolerant 80% intervals.)” (Cohen, 1990, p. 1310).

"Only when these factors are fine-tuned to maximize the operative effect size will it generally be profitable to consider sample size" (Lipsey, 1990, p. 138).

For the present study, the first and second tactics were the most promising (the third tactic concerns the application of the intervention, which was not under the control of the researcher). First, reliable and valid measures, predominantly from the National Longitudinal Survey of Children and Youth (NLSCY) (HRDC & Statistics Canada, 1996), were selected for use as pretests and outcome measures. Second, the present study used statistical controls of variance to increase the effect size by (a) using pretest scores as covariates and (b) using unequal sample sizes, thereby optimizing the statistical power of the available sample sizes in the intervention and control groups.

Using pretests as covariates. "If we know the correlation between the covariate and the dependent variable, or the multiple correlation between a set of covariates and the dependent variable, it is easy to determine the effect [this] will have upon statistical power" (Lipsey, 1990, p. 131). In fact, the use of such a pretest-posttest analysis of covariance (ANCOVA) design "is so powerful and so readily attainable in most instances of treatment effectiveness research that it should be taken as the standard to be used routinely unless there are good reasons to the contrary" (Lipsey, 1990, p. 132). According to Table 6.2 in Lipsey (1990, p. 127), there are at least three ANCOVA possibilities for the RDD. First, using a set of pretest covariates that account for at least 50% of the variance in outcomes would multiply the effect size by 1.41. Second, pretest covariates accounting for 85% of the variance in outcomes would multiply the effect size by 2.57. Third, pretest covariates accounting for 90% or more of the variance in outcomes would multiply the effect size by 3.64. The present study attempted to select pretest covariates that accounted for at least 50% of the variance in outcomes in order to increase statistical power.

Using unequal sample sizes. As described above, in the BabyFirst program, on average, about 75% of families score above the cutoff in the in-depth assessment, creating unequal sample sizes between intervention and control groups in the RDD. At any given time, there will be at

least three times as many participants in the intervention group than the control group. This created a further opportunity to increase the statistical power of the present study: "If there are limits on the numbers of subjects available or appropriate for one experimental condition, power may be increased by adding more subjects to the other condition" (Lipsey, 1990, p. 170).

According to Table 6.6 in Lipsey (1990, p. 141), if one group is reduced to about 0.5 of the equal n value required, the other group must increase by about 2.9 times the equal n value. For example, to detect large effect sizes in the RDD, where $n = 68$ is required in each sample, the researcher could maintain similar power with one group of 37 by increasing the other group to 163.²⁰ Combined with pretest covariates that account for at least 50% of the outcome variance, these unequal ns would be sufficiently powerful to detect medium effect sizes. The present study used the unequal ns arising from the BabyFirst assessment process to increase statistical power.

Statistical power for effectiveness analyses. Statistical power analysis for RDD analyses began with RDD effect size classifications from Cappelleri et al. (1994), calculated using Cohen's (1988) general guidelines²¹ for small, medium, and large effects, namely partial r values of .14, .36, and .51, respectively.

Four recent meta-analyses provide an empirical context (McCartney & Rosenthal, 2000) for estimating the effect sizes (in terms of the standardized mean difference, or d) for the effectiveness of early childhood intervention.²² The Bakermans-Kranenburg et al. (2003) meta-analysis of 70 early childhood intervention studies found $d = .33$ for insensitive parenting and $d = .20$ for infant attachment insecurity (equivalent to rs of .17 and .10, respectively). The Macleod and Nelson (2000) meta-analysis of 47 studies found that home visiting had an average effect of $d = .38$ across outcomes and $d = .24$ for parent behaviour (equivalent to rs of .19 and .12,

²⁰ The total N , however, has increased 47%, from 136 to 200, showing the inefficiency of unequal sample sizes (Lipsey, 1990).

²¹ It should be noted that Cohen (1988) cautioned against application of these effect size conventions to particular social science disciplines or topic areas. Such uncritical application can lead to highly misleading conclusions regarding the importance of effects (Prentice & Miller, 1992, Valentine & Cooper, 2003).

²² According to Meehl (1997), "significance tests are appropriate in technological contexts, such as evaluating the efficacy of interventions, but setting up confidence intervals is preferable" (p. 397).

respectively). The Nelson et al. (2003) meta-analysis of 34 long-term studies found that home visiting had an average effect of $d = .33$ on parent-family wellness, with an average effect of $d = .19$ across preschool outcomes (equivalent to rs of .17 and .10, respectively). Most recently, the Sweet and Appelbaum (2004) meta-analysis of 60 home visiting programs found $d = .14$ for parent behaviour and $d = .11$ for parent attitudes (equivalent to rs of .07 and .06 respectively).

Taken together, for early childhood interventions similar to the BabyFirst program, these four meta-analyses found average effects ranging, in r terms, from .06 to .19, with an overall average effect size of $r = .12$ on family, parent, and parent-child outcomes similar to those included in the present study. In the first report from the randomized effectiveness trial of Hawaii Healthy Start (Duggan et al., 1999), statistically significant outcome findings were equivalent to r values ranging from .02 to .11. Overall, $r \geq .10$ can be regarded as a meaningful expected effect size regarding early childhood interventions in general and home visiting in particular.²³

Based on this empirical context, using *SamplePower* software (Borenstein, Rothstein, & Cohen, 2001) calculations for multiple regression analysis, it was determined that the minimum sample size required for RDD analyses to detect intervention effects with partial r values of .10 or greater (equivalent to Cohen's $U3$ values of 58% or greater) with 80% power and 80% confidence intervals ($p = .20$) was $N = 200$ (representing unequal ns of 163 for the intervention group and 37 for the comparison group, following Table 6.6 in Lipsey, 1990, p. 141, described in the example above), provided that a set of pretest covariates accounts for at least 57% of the outcome variance. As described above in the *Participants* section, data for 250 families (187 BabyFirst families and 63 Comparison families) were available for the present study.

Statistical power for alliance-outcome analyses. Three meta-analyses provide an empirical context for estimating the effect sizes (in r terms) for the theorized relationship between

²³ Given the relatively small magnitude, in statistical terms, of early childhood intervention effects, conducting statistically powerful effectiveness studies is an ongoing challenge for the field. A similar challenge was reported in a recent meta-analysis of 46 health psychology intervention studies (Maddock & Rossi, 2001), most studies had sufficient power (.81) to detect large effects, weak power (.57) to detect medium effects, and were significantly underpowered (.17) to detect small effects (Maddock & Rossi, 2001).

the working alliance and intervention outcome.²⁴ In Horvath's (1994) meta-analysis of eight WAI studies relating the working alliance and therapeutic outcome, the average effect size was $r = .33$. In the more recent Martin et al. (2000) meta-analysis, the average WAI-outcome correlation across 80 effect sizes in 22 studies was $r = .24$. Most recently, Horvath (2001; Horvath & Bedi, 2002) updated the Martin et al. (2000) meta-analysis to include 10 additional studies from 1997 to 2000 and found an average r of .21 (median r of .25). Two published studies have linked alliance-related variables to outcome in early childhood intervention, both using multivariate predictor regression models consisting of both participant and interaction variables. In the first study, multiple R values ranged from .26 to .53 across outcomes; standardized beta coefficients for alliance-related variables in the regression analyses ranged from .13 to .20, with an average of .17 (Korfmacher, Kitzman, & Olds, 1998). In the second study, partial correlation coefficients for alliance-rated variables in the regression model ranged from .05 to .67, with an average of .20 (Heinicke et al., 2000). Overall, $r \ge .20$ can be regarded as a meaningful expected effect size for the relationship between alliance and outcome.

Based on this empirical context, using *SamplePower* software calculations for multiple regression analysis, it was determined that the minimum sample size required to detect alliance-outcome partial r values of .20 or greater with 80% power and 80% confidence intervals ($p = .20$) was $n = 50$, provided that a set of covariates accounts for at least 54% of the outcome variance.

Data Analyses

To test Hypothesis 1 (the *effectiveness hypothesis*), BabyFirst families and comparison families were compared on each of the six outcomes using RDD analysis. Baseline risk, baseline outcome status, and six baseline variables (child temperament, parental depression, parental psychological well-being, parental Aboriginal status, family social support, and neighbourhood

²⁴ It should be noted that Meehl (1990b) did not regard meta-analysis as a tool for theory testing. However, Schmidt (1992, 1996) has demonstrated the theory-building role of meta-analysis (see also Hunter & Schmidt, 1996). For more detailed discussion regarding theory testing, see Serlin and Lapsley (1985) and responses by Dar (1987) and Meehl (1990a, 1998).

safety) were entered as covariates in the regression model to increase the statistical power of the RDD (Judd & Kenny, 1981; Lipsey, 1990).

To test Hypothesis 2 (the *internal working model-early alliance hypothesis*), parents' baseline ratings of positive relationships with others and family social support were correlated with parents' subsequent ratings of the early alliance (after 6 home visits).

To test Hypothesis 3 (the *early alliance-outcome hypothesis*) and Hypothesis 4 (the *alliance improvement-outcome hypothesis*), multiple regression analyses were conducted, with early alliance and alliance change (over the first year of the program) entered together as predictor variables for outcomes of the BabyFirst group, controlling for baseline outcome status in each case. These multiple regression analyses were conducted separately for parent and home visitor alliance ratings, and only for outcomes impacted by the BabyFirst program, based on the results of the foregoing effectiveness analyses.

Results

Group Comparisons Regarding Sociodemographic Characteristics at Baseline

Baseline sociodemographic characteristics of BabyFirst families and Comparison families are presented for continuous variables in Table 3. On average, BabyFirst children and Comparison children were similar in age. However, BabyFirst parents were 2.62 years younger and 0.70 years less educated than comparison parents. Further, BabyFirst families had not lived as long at their current residence (1.50 years less). As shown in Table 3, the most dramatic difference between the two groups was in terms of annual family income, with BabyFirst families earning an average of \$24,000 less per year than Comparison families (see Table 3), a difference of 1.08 standard deviations ($p \leq .001$).

Turning to dichotomous variables, statistically similar majorities of both BabyFirst and Comparison parents were birth parents (97.9% and 100%, respectively) and female (97.8% and 98.4%, respectively). Each group had statistically similar percentages of male and female children. Of the BabyFirst families, 51.1% had boys and 48.9% had girls. Of the Comparison families, 52.4% had boys and 47.6% had girls. However, at baseline almost twice as many BabyFirst parents were teenagers (age ≤ 19 years) (28.8%), contrasted with Comparison parents (15.9%), $\chi^2 (1, N = 247) = 4.13, \phi = .13, p = .04$. Further, a smaller percentage of BabyFirst parents had completed high school (45.9%), contrasted with Comparison parents (77.8%), $\chi^2 (1, N = 246) = 19.18, \phi = -.28, p = .0001$.

Table 3

*Sociodemographic Characteristics of BabyFirst and Comparison Families at Baseline
(Continuous Variables)*

Sociodemographic variable	BabyFirst families (n = 187)		Comparison families (n = 63)		df	t
	M	SD	M	SD		
Child age (months)	4.50	1.52	4.25	1.69	240	1.06
Parent age (years)	24.38	6.06	27.00	6.03	245	-2.97**
Parental education (years)	10.62	2.04	11.32	1.67	240	-2.68**
Family income (dollars)	24,300	16,900	48,300	27,900	216	-5.97***
Years at current residence	1.69	3.13	3.19	3.57	246	-2.98**

** $p \leq .01$. *** $p \leq .001$.

The considerable difference in annual family income between the two groups at baseline may be explained, in part, by differences in family structure. On average, two-parent families earn more income than one-parent families. In the present study, single parent families comprised a larger percentage of the BabyFirst families (43.2%) than the Comparison families (25.4%), $\chi^2(1, N = 248) = 6.31, \phi = .16, p = .01$. Specifically, of the BabyFirst parents, 67 were single (never married) (36.2%) and 13 were either separated or divorced (7.0%). The remaining BabyFirst parents were from two-parent families: 61 were either common-law or living with a partner (33.0%), and 44 were married (23.8%). In contrast, the majority (74.6%) of Comparison parents were from two-parent families: 39 were married (61.9%) and 8 were either common-law or living with a partner (12.7%). The remaining Comparison parents were from one-parent families: 14 were single (never married) (22.2%) and 2 were either separated or divorced (3.2%).

BabyFirst parents and Comparison parents were statistically similar in terms of labour force participation. At baseline, the majority of both BabyFirst and Comparison parents reported that their current primary activity was caring for their family (85.2% and 79.4%, respectively). Smaller percentages of BabyFirst parents and Comparison parents reported that they were both working and caring for their family (9.3% and 11.1%). The remaining BabyFirst parents and Comparison parents indicated that they were going to school (3.3% and 7.9%), working (0.5% and 1.6%), looking for work (1.1% and 0.0%), or other (0.5% and 0.0%). Overall, these between-group differences did not appear practically important and were not statistically significant.

BabyFirst families and Comparison families were statistically different in terms of primary source of family income. At baseline, 44.5% of BabyFirst families reported that their primary source of family income came from wages and salaries, contrasted with 67.2% of Comparison families. Further, 4.0% of BabyFirst families indicated income from self-employment as their primary source, contrasted with 13.1% of Comparison families. Nearly six times as many BabyFirst families reported that their primary source of income was social assistance, contrasted with Comparison families (39.3% vs. 6.6%). Of the remaining families,

similar proportions of both BabyFirst families and Comparison families reported employment insurance (6.4% and 6.6%), worker's compensation (0% and 1.6%), benefits from the Canada Pension Plan (0.6% and 0%), retirement pension (0.6% and 1.6%), the Canada Child Tax Benefit (4.0% and 0%), or other sources (0.6% and 3.3%) as their primary income source. Overall, these differences in primary source of income between BabyFirst and Comparison families appear practically meaningful and statistically significant, $\chi^2 (8, N = 234) = 34.94, \phi = .39, p = .0001$.

BabyFirst parents and Comparison parents were statistically similar in terms of their country of birth. The majority of both groups were born in Canada (93.4% and 81.0%). Although the majority of both groups were White (74.9% and 90.5%), over twice as many BabyFirst parents identified themselves as Aboriginal (25.1%), contrasted with Comparison parents (9.5%), $\chi^2 (1, N = 250) = 6.87, \phi = .17, p = .009$.

The meaningful differences between BabyFirst and Comparison families on a range of sociodemographic risk factors (parental age, parental education, family income, residential stability) were expected, given that assignment to intervention in the regression-discontinuity design (RDD) was based on the Family Stress Checklist (FSC) assessment process described in the *Method* section above. On average, BabyFirst families scored 30.83 points higher on the 100-point FSC than Comparison families, a difference of 1.64 standard deviations ($p \leq .001$). These higher baseline risk scores are also reflected in meaningful differences between the two groups on several of the child, parent, family, and community measures at baseline (see *Group Comparisons Regarding Participant Measures at Baseline* section below).

Reliability of Participant Measures at Baseline and 1 Year

Internal consistency reliability coefficients (Cronbach's alpha) were computed for multiple-item participant measures (child difficult temperament, parental depression, parental psychological well-being, positive parenting, hostile/ineffective parenting, family social support, and neighbourhood safety) at baseline and 1 year, and are presented in Table 4.

Table 4*Internal Consistency Reliability Coefficients for Participant Measures at Baseline and 1 Year*

Measure	Baseline		1 Year		Number of scale items
	N	α	N	α	
Child difficult temperament	238	.84	242	.82	10
Parental depression	234	.81	243	.82	12
Parental PWB	230	.81	228	.86	18
Autonomy	239	.52	238	.55	3
Environmental mastery	244	.47	245	.60	3
Personal growth	243	.46	244	.48	3
Positive relations with others	243	.51	245	.62	3
Purpose in life	243	.37	246	.47	3
Self-acceptance	244	.64	241	.68	3
Positive parenting	244	.71	244	.77	5
Hostile/ineffective parenting	238	.48	246	.52	2
Family social support	239	.79	248	.82	6
Neighbourhood safety	247	.60	246	.60	3

Note. α = Cronbach's alpha, PWB = psychological well-being.

Coefficient alpha is not a property of instruments but of scores for a particular population (Wilkinson & TFSI, 1999). Coefficient alpha values greater than or equal to .70 (Cortina, 1993) or .80 (Streiner, 2003) generally indicate high levels of internal consistency. The value of coefficient alpha is related to the number of items in a scale; the fewer the items, the lower the value (Cortina, 1993; Schmitt, 1996; Streiner, 2003). In the present study, for measures with five or more items, coefficients alphas were high, ranging from .71 to .84 at baseline and from .77 to .86 at 1 year (see Table 4). For measures with fewer than five items, coefficient alphas were lower, ranging from .37 to .64 at baseline and from .47 to .68 at 1 year (see Table 4). It is important to note that even relatively low (e.g., .50) levels of coefficient alpha do not seriously attenuate validity coefficients (Schmitt, 1996) and that coefficient alpha may underestimate reliability (Osburn, 2000). Coefficient alphas for NLSCY measures used in the present study were very similar to those reported in the national probability sample for the NLSCY (Statistics Canada & HRDC, 1995). Coefficient alphas for the six psychological well-being (PWB) factors were very similar to those reported for national probability samples from several large-scale surveys in the U.S. (Keyes et al., 2002; Ryff & Keyes, 1995) and Canada (Clarke et al., 2000). In summary, measures in the present study showed acceptable levels of internal consistency at baseline and 1 year.

Group Comparisons Regarding Participant Measures at Baseline

Means and standard deviations for measures at baseline of child difficult temperament, parental depression, parental psychological well-being (PWB), positive parenting, hostile/ineffective parenting, frequency of parental reading with children, family social support, and neighbourhood safety were contrasted in the BabyFirst and Comparison families with *t* tests, and are presented in Table 5. At baseline, BabyFirst children and Comparison children were similar in terms of difficult temperament. However, several statistically significant differences emerged. BabyFirst parents evidenced higher depression and lower psychological well-being (PWB). The latter difference reflected differences on three PWB factors: positive relations with others, purpose in life, and self-acceptance. BabyFirst parents and Comparison parents were similar in terms of the other three PWB factors (autonomy, environmental mastery, personal growth). Statistically similar (and high) percentages of BabyFirst parents and Comparison parents were already reading with their children at baseline (89.2% and 90.5% respectively). On average, the two groups of parents were also similar in terms of positive parenting and frequency of reading with their children, as shown in Table 5. However, BabyFirst parents reported more hostile/ineffective parenting than Comparison parents. Finally, BabyFirst parents reported less family social support and lower levels of neighbourhood safety than Comparison parents.

Intercorrelations of baseline variables are presented in Table 6. These correlations range in absolute value from .01 to .52. Only two correlations were greater than .33. Both of these were between parental PWB and closely related constructs (parental depression and family social support). Thus, these correlations are neither surprising nor problematic. Correlation results suggest that multicollinearity (i.e., very high correlations between predictor variables in a multiple regression model) was not a problem in the regression-discontinuity design analyses.

Table 5*Baseline Participant Measures for BabyFirst Families and Comparison Families*

Variable	BabyFirst families (n = 187)		Comparison families (n = 63)		df	t
	M	SD	M	SD		
Family Stress Checklist score	43.29	14.80	12.46	6.71	248	22.44***
Child difficult temperament	3.25	0.68	3.10	0.93	236	1.08
Parental depression	1.75	0.51	1.52	0.40	232	3.53***
Parental PWB	4.54	0.64	4.83	0.56	228	-3.09**
Autonomy	4.45	0.85	4.44	0.73	237	0.06
Environmental mastery	4.63	0.87	4.79	0.70	242	-1.42
Personal growth	4.94	0.84	5.02	0.70	241	-0.72
Positive relations with others	4.35	1.07	4.93	1.01	241	-3.74***
Purpose in life	4.38	0.98	4.89	0.83	241	-3.67***
Self-acceptance	4.54	0.99	4.96	0.83	242	-2.97**
Positive parenting	4.55	0.50	4.64	0.39	242	-1.21
Hostile/ineffective parenting	1.62	0.75	1.40	0.51	236	2.11*
Parental reading with children	6.20	1.27	6.09	1.20	221	0.61
Family social support	3.52	0.45	3.70	0.39	237	-3.01**
Neighbourhood safety	3.05	0.58	3.33	0.55	245	-3.34***

Note. PWB = psychological well-being.

* $p \leq .05$. ** $p \leq .01$. *** $p \leq .001$.

Table 6*Intercorrelations for Baseline Child, Parent, Parent-Child, Family, and Community Variables*

Variable	1	2	3	4	5	6	7	8	9	10
1. Family Stress Checklist score	—	.03	.21	-.19	.12	-.04	.12	-.01	-.17	-.23
2. Child difficult temperament		—	.11	-.20	.07	-.14	.20	.07	-.08	-.08
3. Parental depression			—	-.50	.11	-.25	.07	-.07	-.33	-.17
4. Parental PWB				—	-.11	.33	-.08	.16	.52	.13
5. Parental Aboriginal status					—	-.11	.01	.05	-.13	-.13
6. Positive parenting						—	-.04	.20	.17	.29
7. Hostile/ineffective parenting							—		-.08	-.06
8. Parental reading with child								—	.02	.03
9. Family social support									—	.20
10. Neighbourhood safety										—

Note. Intercorrelation sample sizes range from 223 to 250. PWB = psychological well-being. Correlations in **bold** typeface (i.e., $r \geq .10$ in absolute value) fall within 80% confidence intervals ($p \leq .20$). Lower correlations may be due to low power and/or sampling error.

Testing Hypothesis 1: Does Participation in the BabyFirst Home Visiting Program Improve Family, Parent, and Parent-Child Outcomes at 1 Year?

To test Hypothesis 1 (the *effectiveness hypothesis*), regression-discontinuity design (RDD) analyses were conducted for 1 year outcomes and are presented in Tables 7 through 12. These tables are organized by outcome, as illustrated in Figure 3. In the RDD analyses, partial correlations (pr_s) $> .20$ represent strong corroboration, partial rs between .10 and .20 represent moderate corroboration, and partial $rs < .10$ represent weak corroboration of the effectiveness hypothesis. Using 80% confidence intervals (CIs) at 80% power, all RDD analyses controlled for the following baseline variables: cumulative risk status, child difficult temperament, parental depression, parental psychological well-being, parental Aboriginal status, family social support, neighbourhood safety, and pre-intervention outcome status.

Effect of home visiting on parental psychological well-being (PWB) at 1 year. As shown in Table 7, the home visiting program had the intended positive effect of improving parental psychological well-being (PWB) at 1 year ($B = .20$, 80% CI = .06 to .34, $pr = .14$, $p = .06$). Table 7 also shows that baseline parental depression ($pr = -.22$, $p = .002$) and parental Aboriginal status ($pr = -.10$, $p = .19$) were both negative predictors of parental PWB at 1 year. These results suggest that parental depression and parental Aboriginal status may limit the positive effects of home visiting on parental PWB after 1 year.

Given the positive effect of home visiting on overall parental PWB, follow-up RDD analyses of the six PWB factors were conducted. These analyses indicated that the positive program effect on overall parental PWB was largely due to positive program effects on three of the six PWB factors: autonomy ($pr = .10$, $p = .16$), environmental mastery ($pr = .11$, $p = .13$), and self-acceptance ($pr = .19$, $p = .008$). Home visiting did not appear to have any effect on personal growth, positive relations with others, or purpose in life after 1 year.

Table 7*Predictors of Parental Psychological Well-Being at 1 Year*

Baseline predictor variable	<i>B</i>	<i>SE B</i>	β	<i>pr</i>	<i>p</i>
Intervention group	.20	.11	.13	.14	.06
Family Stress Checklist score	.00	.00	-.08	-.09	.22
Child difficult temperament	.01	.04	.01	.02	.80
Parental depression	-.25	.08	-.17	-.22	.002
Parental Aboriginal status	-.11	.08	-.06	-.10	.19
Family social support	.07	.09	.05	.06	.42
Neighbourhood safety	.05	.06	.04	.06	.40

Note. Intervention group was coded 0 for comparison group and 1 for BabyFirst program group. Parental Aboriginal status was coded 0 for non-Aboriginal and 1 for Aboriginal. $R = .77$, adj. $R^2 = .57$ ($N = 189$, $p = .0001$).

Effect of home visiting on positive parenting at 1 year. As shown in Table 8, the home visiting program had the intended positive effect of increasing positive parenting ($B = .13$, 80% CI = .01 to .21, $pr = .12$, $p = .11$). Table 8 also shows that baseline neighbourhood safety was a positive predictor of positive parenting at 1 year ($pr = .21$, $p = .004$). This result suggests that safer neighbourhoods may also contribute to improvements in positive parenting after 1 year.

Table 8*Predictors of Positive Parenting at 1 Year*

Baseline predictor variable	<i>B</i>	<i>SE B</i>	β	<i>pr</i>	<i>p</i>
Intervention group	.13	.08	.14	.12	.11
Family Stress Checklist score	.00	.00	-.05	-.04	.57
Child difficult temperament	.02	.03	.04	.05	.53
Parental depression	-.06	.06	-.07	-.08	.31
Parental PWB	.06	.05	.10	.09	.23
Parental Aboriginal status	-.02	.06	-.02	-.03	.70
Family social support	-.04	.07	-.04	-.05	.54
Neighbourhood safety	.13	.04	.19	.21	.004

Note. Intervention group was coded 0 for comparison group and 1 for BabyFirst program group. Parental Aboriginal status was coded 0 for non-Aboriginal and 1 for Aboriginal. PWB = psychological well-being.

R = .59, adj. *R*² = .31 (*N* = 195, *p* = .0001).

Effect of home visiting on parental depression at 1 year. As shown in Table 9, the home visiting program had virtually no intended effect on parental depression at 1 year ($pr = -.02, p = .78$). Table 9 also shows that higher levels of cumulative family risk at baseline, as measured by the Family Stress Checklist, predicted higher levels of depression at 1 year ($pr = .13, p = .08$). Similarly, parental Aboriginal status also predicted higher levels of depression ($pr = .13, p = .08$). As would be expected, Table 9 shows that parental PWB was inversely related to parental depression. Higher levels of parental PWB at baseline predicted lower levels of parental depression at 1 year ($pr = -.19, p = .01$). This result was expected given that these constructs are closely related theoretically, and because they were negatively correlated at baseline ($r = -.50$), as shown in Table 6.

Table 9*Predictors of Parental Depression at 1 Year*

Baseline predictor variable	B	SE B	β	pr	p
Intervention group	-.03	.10	-.02	-.02	.78
Family Stress Checklist score	.00	.00	.15	.13	.08
Child difficult temperament	.02	.03	.04	.05	.53
Parental PWB	-.16	.06	-.20	-.19	.01
Parental Aboriginal status	.14	.08	.11	.13	.08
Family social support	.01	.08	.01	.01	.86
Neighbourhood safety	-.07	.06	-.08	-.09	.22

Note. Intervention group was coded 0 for comparison group and 1 for BabyFirst program group. Parental Aboriginal status was coded 0 for non-Aboriginal and 1 for Aboriginal. PWB = psychological well-being.

$R = .59$, adj. $R^2 = .32$ ($N = 197$, $p = .0001$).

Effect of home visiting on parental reading with children at 1 year. As shown in Table 10, the home visiting program had virtually no intended effect on parents' frequency of reading with their children ($pr = -.08, p = .30$). This may have been due to ceiling effects; about 90% of parents in both program and comparison groups reported reading with their children at baseline. Further, Table 10 shows that baseline parental depression ($pr = -.11, p = .14$) and parental Aboriginal status ($pr = -.15, p = .05$) were both negative predictors of parental reading with children at 1 year. More depressed parents read with their children less frequently, as did Aboriginal parents. Surprisingly, cumulative family risk at baseline was positively related to parental reading with children. That is, higher Family Stress Checklist scores at baseline predicted more frequent parental reading with children at 1 year ($pr = .12, p = .13$).

Table 10*Predictors of Parental Reading with Children at 1 Year*

Baseline predictor variable	B	SE B	β	pr	p
Intervention group	-.25	.24	-.10	-.08	.30
Family Stress Checklist score	.01	.01	.15	.12	.13
Child difficult temperament	.04	.08	.03	.03	.67
Parental depression	-.25	.17	-.12	-.11	.14
Parental PWB	-.09	.15	-.05	-.05	.54
Parental Aboriginal status	-.37	.19	-.14	-.15	.05
Family social support	.21	.20	.09	.08	.30
Neighbourhood safety	.07	.13	.04	.04	.57

Note. Intervention group was coded 0 for comparison group and 1 for BabyFirst program group. Parental

Aboriginal status was coded 0 for non-Aboriginal and 1 for Aboriginal. PWB = psychological well-being.

$R = .47$, adj. $R^2 = .18$ ($N = 178$, $p = .0001$).

Effect of home visiting on family social support at 1 year. As shown in Table 11, the home visiting program had virtually no intended effect on family social support ($pr = .05, p = .35$). Table 11 also shows two positive baseline predictors of family social support at 1 year: child difficult temperament ($pr = .11, p = .12$) and parental PWB ($pr = .14, p = .05$). Parents with more temperamentally difficult children may be more likely to seek out additional social support. Parental PWB at baseline was expected to predict family social support at 1 year because these constructs are closely related theoretically, and because they were positively correlated at baseline ($r = .52$), as shown in Table 6.

Table 11 also shows two negative predictors of family social support at 1 year: cumulative family risk (i.e., Family Stress Checklist scores) ($pr = -.14, p = .05$) and parental Aboriginal status ($pr = -.25, p = .0001$). Families living with higher levels of cumulative family risk, especially Aboriginal families, had lower levels of social support.

Table 11*Predictors of Family Social Support at 1 Year*

Baseline predictor variable	<i>B</i>	<i>SE B</i>	β	<i>pr</i>	<i>p</i>
Intervention group	.08	.08	.07	.07	.35
Family Stress Checklist score	.00	.00	-.15	-.14	.05
Child difficult temperament	.04	.03	.09	.11	.12
Parental depression	.02	.06	.02	.02	.79
Parental PWB	.10	.05	.14	.14	.05
Parental Aboriginal status	-.22	.06	-.20	-.25	.0001
Neighbourhood safety	.05	.04	.07	.08	.25

Note. Intervention group was coded 0 for comparison group and 1 for BabyFirst program group. Parental Aboriginal status was coded 0 for non-Aboriginal and 1 for Aboriginal. PWB = psychological well-being.
R = .66, adj. *R*² = .42 (*N* = 202, *p* = .0001).

Effect of home visiting on hostile/ineffective parenting at 1 year. As shown in Table 12, the home visiting program had the unintended negative effect of increasing hostile/ineffective parenting at 1 year ($B = .21$, 80% CI = .01 to .41, $pr = .10$, $p = .18$). Moreover, higher levels of child difficult temperament ($pr = .17$, $p = .02$) and parental Aboriginal status ($pr = .11$, $p = .15$) predicted higher levels of hostile/ineffective parenting at 1 year. Children who are more temperamentally difficult may elicit more hostile/ineffective parenting. Parental Aboriginal status is likely a proxy for greater socioeconomic disadvantage (Hallett et al., 2000; Martens et al., 2002), which is strongly associated with hostile/ineffective parenting (Grant et al., 2003). Higher levels of parental PWB predicted lower levels of hostile/ineffective parenting at 1 year ($pr = -.16$, $p = .04$).

Table 12*Predictors of Hostile/Ineffective Parenting at 1 Year*

Baseline predictor variable	B	SE B	β	pr	p
Intervention group	.21	.16	.12	.10	.18
Family Stress Checklist score	.00	.00	-.01	-.01	.88
Child difficult temperament	.13	.06	.16	.17	.02
Parental depression	.04	.11	.03	.03	.71
Parental PWB	-.21	.10	-.17	-.16	.04
Parental Aboriginal status	.17	.12	.09	.11	.15
Family social support	.13	.13	.07	.07	.34
Neighbourhood safety	-.06	.09	-.04	-.05	.52

Note. Intervention group was coded 0 for comparison group and 1 for BabyFirst program group. Parental Aboriginal status was coded 0 for non-Aboriginal and 1 for Aboriginal. PWB = psychological well-being.

R = .57, adj. R² = .29 (N = 192, p = .0001).

Taken together, these results provide support for Hypothesis 1 (the *effectiveness hypothesis*). After 1 year, home visiting had beneficial effects on (a) parental psychological well-being (PWB), especially parental autonomy, environmental mastery, and self-acceptance; and (b) positive parenting. These effects can also be expressed in terms of the overlap between two distributions, using Cohen's *U3*, the percentile of the intervention group corresponding to the 50th percentile of the comparison group (Cohen, 1988). These percentiles were as follows: parental PWB = 61% (parental self-acceptance = 65%, parental environmental mastery = 59%, and parental autonomy = 58%); and positive parenting = 60%. These results are comparable in magnitude to the *U3* values (55% to 65%) computed from statistics reported in recent meta-analytic reviews of the effects of home visiting on parent and parent-child outcomes (Bakermans-Kranenburg et al., 2003; MacLeod & Nelson, 2000; Nelson et al., 2003; Sweet & Appelbaum, 2004). However, the BabyFirst home visiting program also had an unfavourable effect of increasing hostile/ineffective parenting at 1 year (*U3* = 58%). Possible pathways for these positive and negative program effects were explored further in analyses of the working alliance.

Parent and Home Visitor Ratings of the Working Alliance

Using the short form of the Working Alliance Inventory (WAI), BabyFirst parents and home visitors each rated the working alliance at two times: early in the program (i.e., after six home visits, or about 1.5 months after baseline) and after 1 year of home visits. Of the 187 BabyFirst parents, Working Alliance Inventory (WAI) data were available for 97 parents (23 for early WAI only, 40 for 1-year WAI only, and 34 for both early and 1-year WAI) and 100 home visitors (28 for early WAI only, 46 for 1-year WAI only, and 26 for both early and 1-year WAI). Correlation analyses for testing Hypothesis 2 used the subsamples with early WAI data ($n = 57$ for BabyFirst parents, $n = 54$ for home visitors). Regression analyses for testing Hypotheses 3 and 4 used the subsamples with both early and 1-year WAI data ($n = 34$ for BabyFirst parents, $n = 26$ for home visitors).

Analyses of available and missing working alliance data. To examine differences between BabyFirst parents with and without WAI data, a series of logistic regression analyses were computed to identify predictors of missing data (Tabachnick & Fidell, 2001). The outcome variable was a group membership dummy variable indicating the presence (coded 1) or absence (coded 0) of WAI data. It should be noted that the nature of the missing WAI data is unknown. For example, it is unclear whether WAI data were unavailable because they were not collected, completed, or submitted by BabyFirst parents or home visitors, or whether they were due to BabyFirst parent attrition from the program. This information was not available for the present study and is reportedly not available from the BabyFirst program (M. Chartier, personal communication, December 2004). Nonetheless, because baseline data was available on the 187 BabyFirst parents included in the present study, it was possible to examine baseline differences between BabyFirst parents with and without WAI data.

In the first logistic regression analysis, five sociodemographic baseline variables, namely child gender, parental age, family income, family structure (i.e., single parent or two parent), and parental education were entered as predictors of missing data group membership. This initial analysis found that the only sociodemographic predictor of missing early WAI data was parental education. The odds ratio (OR) indicated that for every additional year of education, BabyFirst parents were 1.21 times more likely ($p = .07$) to have early WAI data available. A follow-up logistic regression analysis was then conducted, using the same sociodemographic variables above, but replacing the continuous parental education variable with the dichotomous parental high school completion variable. This analysis found this dichotomous parental education variable to be the sole, and much stronger, sociodemographic predictor of missing WAI data. BabyFirst parents who had completed high school were 2.67 times more likely ($p = .02$) to have early WAI data available, compared to BabyFirst parents who had not completed high school.

In the next set of logistic regression analyses, based on the preceding results, parental high school completion was included in the model with the seven baseline child, parent, family,

and community variables from the preceding RDD analyses: child difficult temperament, parental depression, parental PWB, parental Aboriginal status, cumulative family risk (Family Stress Checklist score), family social support, and neighbourhood safety. Outcome variables reflected the presence or absence of (a) parent-rated early WAI data, (b) parent-rated 1-year WAI data, (c) home visitor-rated early WAI data, and (d) home visitor-rated 1-year WAI data.

Results indicated the following: (a) parents who had completed high school ($OR = 2.04$, $p = .06$) and parents with more family social support ($OR = 2.59$, $p = .08$) were over twice as likely to have early WAI data available, compared to BabyFirst parents who had not completed high school and parents with less family social support; (b) parents with more temperamentally difficult children ($OR = .73$, $p = .14$) and Aboriginal parents ($OR = .32$, $p = .01$) were 27% to 68% less likely to have 1-year WAI data available; (c) home visitors of parents with more temperamentally difficult children ($OR = .61$, $p = .03$) were 39% less likely to have early WAI data available; and (d) home visitors of parents with higher parental PWB ($OR = .51$, $p = .07$) were 49% less likely to have 1-year WAI data available, whereas home visitors of parents with higher family social support ($OR = 2.71$, $p = .06$) were 2.71 times more likely to have 1-year WAI data available.

In summary, missing parent-rated WAI data was associated with parental high school non-completion, more difficult child temperament, less family social support, and parental Aboriginal status. Similarly, missing home visitor-rated WAI data was also associated with more difficult child temperament and less family social support. Parents with the latter characteristics may have been more likely to drop out of the program. Missing home visitor-rated 1-year WAI data was also associated, somewhat unexpectedly, with higher baseline parental PWB. It is possible that these higher-functioning parents graduated from the program before home visitors completed their 1-year WAI ratings.

The foregoing analyses of missing WAI data in the present study suggest that WAI data were more likely to be available from (or regarding) parents who had completed high school, had

children with less difficult temperament, had more family social support, and who were non-Aboriginal. These potential limits to generalizability should be considered when interpreting the correlation and multiple regression analyses using parent and home visitor alliance ratings in the sections below. Notwithstanding these potential limitations, BabyFirst parents for whom WAI data were available nonetheless met the inclusion criteria for home visiting, and were similar to BabyFirst parents for whom WAI were not available in terms of cumulative family risk (Family Stress Checklist score), parental depression, and neighbourhood safety. In other words, none of these baseline variables were statistically significant predictors of missing WAI data in the foregoing logistic regression analyses, and provide a measure of confidence in conducting and interpreting analyses of available WAI data.

Reliability of Early and 1 Year Working Alliance Ratings

Internal consistency coefficients (Cronbach's alpha) were computed for early and 1 year working alliance ratings of parents and home visitors. Coefficient alphas for parent ratings of the early alliance ($n = 57$) were as follows: general alliance (.68), goals (.41), tasks (.63), and bond (.71). Coefficient alphas for parent ratings of the 1 year alliance ($n = 74$) were as follows: general alliance (.83), goals (.65), tasks (.84), and bond (.84). As these results demonstrate, reliabilities of parent alliance ratings in the present study were moderate to high. It should be noted that parent alliance ratings were relatively lower than those reported for psychotherapy clients (from university counseling centres) using the short form of the Working Alliance Inventory (WAI), particularly for goals (Busseri & Tyler, 2003; Tracey & Kokotovic, 1989). Alpha coefficients for client ratings from these two studies were as follows: general alliance (.91, .98), goals (.73, .90), tasks (.86, .89), and bond (.80, .92). It is possible that the present modifications to the wording of the WAI for use in home visiting did not adequately reflect important differences between these two service populations (parents in home visiting and clients in psychotherapy) in terms of their respective intervention goals and tasks.

Coefficient alphas for home visitor ratings of the early alliance ($n = 54$) were as follows: general alliance (.93), goals (.79), tasks (.92), and bond (.81). Coefficient alphas for home visitor ratings of the 1 year alliance ($n = 72$) were as follows: general alliance (.93), goals (.82), tasks (.88), and bond (.81). Reliabilities of home visitor alliance ratings in the present study were high, and similar to those reported for therapists using the short form of the WAI (Busseri & Tyler, 2003; Tracey & Kokotovic, 1989). Alpha coefficients for therapist ratings from these two studies were as follows: general alliance (.91, .95), goals (.81, .88), tasks (.89, .83), and bond (.77, .91).

Overall, in the present study, the reliability of home visitor alliance ratings was higher than parent alliance ratings. However, the reliability of parent ratings did improve over time, from a range of .41 to .71 (early) to a range of .65 to .83 (1 year). To improve statistical power, only general alliance ratings were used in the following analyses when testing the alliance hypotheses.

Testing Hypothesis 2: Do More Positive Internal Working Models of Parents Predict More Positive Parent Ratings of the Early Alliance in Home Visiting?

Correlations of baseline family and parent variables with parent-rated and home visitor-rated early alliance are presented in Table 13. Hypothesis 2 was tested through convergent and discriminant validation approaches (Campbell & Fiske, 1959). Convergent validity refers to empirical associations between theoretically associated constructs. Discriminant validity refers to the absence of empirical association between constructs that are not theoretically associated. Parents' internal working models should predict their early alliance ratings, but would not be expected theoretically to predict the early alliance ratings of their home visitors. The present results provide both convergent and discriminant validation for Hypothesis 2. Higher levels of parents' positive relations with others ($r = .23, p = .09$) and family social support ($r = .29, p = .03$) at baseline each predicted higher levels of subsequent parent-rated early alliance (convergent validation). However, these parent-rated interpersonal variables were poor predictors of home visitor-rated early alliance ($rs = .02$ and $.08$ respectively) (discriminant validation). These results provide partial support for Hypothesis 2 (the *internal working model-early alliance hypothesis*).

Table 13*Correlations of Family and Parent Variables at Baseline with Early Alliance Ratings*

Variable	Parent-rated early alliance	Home visitor-rated early alliance
Family Stress Checklist (FSC) score	.21	-.14
Family social support	.29	.08
Parental depression	-.25	.23
Parental PWB (overall score)	.18	-.32
Parental autonomy	-.07	-.32
Parental environmental mastery	.27	-.16
Parental personal growth	.12	-.07
Parental positive relations with others	.23	.02
Parental self-acceptance	.15	-.27
Parental Aboriginal status	.04	-.11

Note. Sample sizes for (a) parent-rated early alliance range from 51 to 57 and (b) home visitor-rated early alliance range from 48 to 54. Sample sizes provided sufficient power to detect correlations greater than or equal to approximately .20. PWB = psychological well-being. Correlations in **bold** typeface fall within 80% confidence intervals ($p \leq .20$). Lower correlations may be due to low power and/or sampling error.

What other baseline family and parent variables predict the early alliance? Some other baseline variables predicted parent-rated vs. home visitor-rated early alliance, but in opposite directions (see Table 13). For example, higher levels of parental PWB ($r = .18, p = .20$) and lower levels of parental depression ($r = -.25, p = .06$) each predicted higher levels of parent-rated early alliance, whereas lower levels of parental PWB ($r = -.32, p = .03$) and higher levels of parental depression ($r = .23, p = .11$) each predicted higher levels of home visitor-rated early alliance. Further, higher levels of parental environmental mastery predicted higher levels of parent-rated early alliance ($r = .27, p = .05$). In contrast, lower levels of parental environmental mastery predicted higher levels of home visitor-rated early alliance ($r = -.16, p = .26$). Higher levels of parental self-acceptance predicted higher parent-rated early alliance ($r = .15, p = .29$) but lower levels of home visitor-rated early alliance ($r = -.27, p = .05$). These results suggest that parents' personal strengths (high total PWB, high environmental mastery, high self-acceptance, low depression) are especially important to parents in developing a strong early alliance. In contrast, it appears that parents' personal limitations (low overall PWB, low environmental mastery, low self-acceptance, high depression) are especially important to home visitors in developing a strong early alliance. Home visitors may be more strongly motivated to establish positive working alliances with parents who have greater levels of need.

Results for cumulative family risk, as measured by the Family Stress Checklist (FSC), however, suggest an opposite pattern of effects to the foregoing individual parent characteristics. As shown in Table 13, higher levels of cumulative family risk predicted higher parent-rated early alliance ($r = .21, p = .12$), whereas higher levels of cumulative family risk predicted lower home visitor-rated early alliance ($r = -.14, p = .30$). Thus, it appears that higher levels of cumulative family risk in the lives of parents may motivate them to develop stronger early alliances. In contrast, home visitors may be limited by high levels of family risk in establishing strong early alliances with the parents they serve under these risky conditions.

Home visitors may be further limited in establishing a strong early alliance by (a) higher levels of parental autonomy at baseline, which predicted lower home visitor-rated early alliance ($r = -.32, p = .03$), but was not strongly related to parent-rated early alliance ($r = -.07, p = .65$); and (b) parental Aboriginal status, which was negatively related to home visitor-rated early alliance ($r = -.11, p = .44$), but was not strongly related to parent-rated early alliance ($r = .04, p = .76$), as shown in Table 13. This may account, in part, for the aforementioned limiting effect of parental Aboriginal status on all program outcomes except positive parenting, as shown in Tables 7-12. Home visitors may feel slightly more challenged in developing early working alliances with Aboriginal parents. Based on the correlations between parental Aboriginal status and other baseline variables shown in Table 6, it is possible that home visitors may also feel somewhat more challenged by the higher levels of cumulative family risk ($r = .12, p = .06$) and the lower levels of neighbourhood safety ($r = -.13, p = .04$) associated with being Aboriginal parents. Descriptive data (e.g., age, gender, Aboriginal status) about home visitors would have been useful to further interpret these findings, but such data were not routinely collected by the program.

Testing Hypotheses 3 and 4: The Early Alliance-Outcome Hypothesis and the Alliance Improvement-Outcome Hypothesis

Improvement-Outcome Hypothesis

To test Hypothesis 3 (the *early alliance-outcome hypothesis*) and Hypothesis 4 (the *alliance improvement-outcome hypothesis*), early alliance scores and alliance change scores were entered together as predictors to improve statistical power and examine their unique contributions to outcome. Only 1-year outcomes that were influenced by the program (as shown by the RDD analyses above) were tested, controlling for baseline outcome scores.

Hypothesis 3: Does a stronger early alliance between parents and home visitors predict better family, parent, and parent-child outcomes at 1 year? Hypothesis 4: Does a home visiting alliance that improves over the first year predict better family, parent, and parent-child outcomes at 1 year? Given that the BabyFirst home visiting program did not have an effect on family social support, parental depression, or parental reading with children, only parental PWB,

positive parenting, and hostile/ineffective parenting outcomes were included in the analyses. Parent alliance ratings were analyzed separately from home visitor alliance ratings. Alliance change was measured by subtracting 1-year alliance scores from early alliance scores (i.e., change scores, also known as difference scores or gain scores).²⁵ Given the range of average effect sizes ($r_s = .21$ to $.33$) reported in meta-analyses of alliance-outcome studies (Horvath & Symonds, 1991; Horvath, 2001; Horvath & Bedi, 2002; Martin et al., 2000), partial correlations (pr_s) in multiple regression analyses of alliance and outcomes were designated as weak ($<.10$), moderate (.10 to .20), or strong ($>.20$) corroboration of the hypotheses.

Alliance and parental psychological well-being (PWB) at 1 year. Parent-rated early alliance ($pr = .11, p = .87$) and parent-rated alliance improvement ($pr = .16, p = .80$) were moderate and positive predictors of parental psychological well-being (PWB) at 1 year. Home visitor-rated early alliance ($pr = .45, p = .06$) and alliance improvement ($pr = .43, p = .07$) were both strong and positive predictors of parental PWB at 1 year. These results are generally corroborated by similar results for the three PWB factors (parental autonomy, parental environmental mastery, and parental self-acceptance) described next.

Alliance and parental autonomy at 1 year. Parent-rated early alliance was a moderate and unexpectedly negative predictor ($pr = -.14, p = .46$) and parent-rated alliance improvement was a moderate and positive predictor ($pr = .14, p = .47$) of parental autonomy at 1 year. Home visitor-rated early alliance ($pr = .21, p = .39$) and home visitor-rated alliance improvement ($pr = .21, p = .37$) were both strong and positive predictors of parental autonomy at 1 year. Overall,

²⁵ Change scores have long been controversial, particularly since an influential paper by Cronbach and Furby (1970) that advocated against their use. Yet for over two decades it has been established that change scores provide unbiased estimates of change, regardless of the magnitude of measurement error (Rogosa, Brandt, & Zumvich, 1982). Because the controversy has persisted, residualized change scores have been widely used in place of change scores, following concerns about the unreliability of change scores and their proneness to regression to the mean. However, Rogosa (1988) and Willett (1988) have shown these concerns are largely unfounded. They have demonstrated that residualized change scores have serious logical, substantive, and technical flaws, and have strongly advised against their use, while confirming that change scores are moderately successful estimates of change (cf. Clarke, 2004; Maris, 1998). Therefore, the present study used change scores, rather than residualized change scores. Nonetheless, it is widely acknowledged that two-wave designs are very limited in studying individual change (Gottman & Rushe, 1993; Rogosa, 1988, 1995; Rogosa & Willett, 1983, 1985; Willett, 1988, 1997). Additional waves of data make possible the use of newer, more powerful analyses of change (Collins & Horn, 1991; Collins & Sayer, 2001; Willett, Singer, & Martin, 1998).

home visitor ratings were better predictors of parental autonomy than were parent ratings.

Increased alliance strength and improvement predicted increased parental autonomy at 1 year.

Alliance and parental environmental mastery at 1 year. Parent-rated early alliance was a strong and positive predictor ($pr = .22, p = .23$) of parental environmental mastery at 1 year. Similarly, parent-rated alliance improvement was a moderate and positive predictor ($pr = .15, p = .41$). In contrast, home visitor-rated early alliance was a weak predictor ($pr = .03, p = .91$), whereas home visitor-rated and alliance improvement was a strong and positive predictor ($pr = .49, p = .03$) of parental environmental mastery at 1 year. Taken together, both parent and home visitor ratings were predictors of parental environmental mastery, with home visitor-rated alliance improvement being the strongest predictor. Increased alliance strength and improvement predicted increased parental environmental mastery at 1 year.

Alliance and parental self-acceptance at 1 year. Parent-rated early alliance ($pr = .14, p = .45$) and parent-rated alliance improvement ($pr = .14, p = .45$) were both moderate and positive predictors of parental self-acceptance at 1 year. Home visitor-rated early alliance ($pr = .29, p = .20$) and home visitor-rated alliance improvement ($pr = .37, p = .10$) were both strong and positive predictors of parental self-acceptance at 1 year. Overall, home visitor ratings were better predictors of parental self-acceptance than were parent ratings. Increased alliance strength and improvement predicted increased parental self-acceptance at 1 year.

In summary, better home visiting alliances predicted improvements in parental PWB. Increases in alliance strength and improvement were associated with increases in parental PWB. Further, home visitor ratings were better predictors of parental PWB and the three PWB factors than were parent ratings. Home visitor perspectives on the alliance may be more important than parent perspectives in improving parental PWB.

Alliance and positive parenting at 1 year. Parent-rated early alliance ($pr = .33, p = .07$) and parent-rated alliance improvement ($pr = .47, p = .007$) were both strong and predictors of positive parenting at 1 year. Home visitor-rated early alliance was a weak predictor ($pr = .09, p = .49$).

.71) whereas home visitor-rated alliance improvement was a moderate and positive predictor ($pr = .16, p = .50$) of positive parenting at 1 year. Increases in alliance strength and improvement predicted increases in positive parenting at 1 year. Overall, parent ratings were generally better predictors of positive parenting than were home visitor ratings. Parent perspectives on the alliance may be more important than home visitor perspectives in improving positive parenting.

Alliance and hostile/ineffective parenting at 1 year. Parent-rated early alliance was a weak predictor ($pr = -.06, p = .75$), whereas parent-rated alliance improvement was a strong and negative predictor ($pr = -.33, p = .07$) of hostile/ineffective parenting at 1 year. Home visitor-rated early alliance ($pr = -.30, p = .21$) and home visitor-rated alliance change ($pr = -.52, p = .02$) were both strong and negative predictors of hostile/ineffective parenting at 1 year. As alliance strength and improvement increased, hostile/ineffective parenting decreased. Conversely, as alliance strength and improvement decreased, hostile/ineffective parenting increased. From a parent's perspective, an alliance that worsened over time was associated with an increase in hostile/ineffective parenting. From a home visitor's perspective, both a weak early alliance and an alliance that worsened over time were associated with an increase in hostile/ineffective parenting.

Summary of alliance-outcome correlations. Results were consistent with both Hypotheses 3 and 4. As alliance strength and improvement increased, desired program outcomes also increased (i.e., improvements in PWB and positive parenting). Where the alliance was initially weak or worsened over time, unfavourable outcomes increased (i.e., increase in hostile/ineffective parenting).

In terms of predictive power, absolute values for 11 of the 12 parent-rated alliance-outcome partial correlations were moderate to strong in magnitude ($pr \geq .10$). However, one of these moderate correlations was in the opposite direction predicted by Hypothesis 3 (the *early alliance-outcome hypothesis*). Absolute values for 10 of the 12 home visitor-rated alliance-outcome partial correlations were moderate to strong in magnitude ($pr \geq .10$).

In terms of rater perspective, home visitor ratings of the working alliance appear to be somewhat better predictors of differential 1-year outcome (mean $pr = .30$ across early alliance and alliance change ratings, using absolute values), compared to parent ratings (mean $pr = .20$ across early alliance and alliance change ratings, using absolute values).

In terms of outcome, home visitor alliance ratings were generally better predictors of parental psychological well-being than were parent ratings. Parent alliance ratings were better predictors of positive parenting than were home visitor ratings. Parent alliance ratings and home visitor alliance ratings were both predictors of hostile/ineffective parenting.

Alliance improvement appears to be a somewhat better predictor of 1-year outcome (mean $pr = .30$ across parent and home visitor ratings, using absolute values) than the early alliance (mean $pr = .20$ across parent and home visitor ratings, using absolute values). These effect sizes are comparable in magnitude to the range of effect sizes ($rs = .21$ to $.33$) reported in meta-analyses of alliance-outcome studies (Horvath & Symonds, 1991; Horvath & Bedi, 2002; Martin et al., 2000).

Taken together, these results provide empirical support for Hypothesis 3 (the *early alliance-outcome hypothesis*) and Hypothesis 4 (the *alliance improvement-outcome hypothesis*). Overall, in the BabyFirst home visiting program, both a stronger early working alliance and improvement in the alliance predicted improvements in parental PWB and positive parenting at 1 year, from both parent and home visitor perspectives. Further, as the strength of the alliance decreased, hostile/ineffective parenting increased.

Comparison of results using available data with results using imputed data. The potential limitations to the generalizability of alliance results were discussed in an earlier section (see *Parent and Home Visitor Ratings of the Working Alliance* above), following analyses of available and missing working alliance data. Given the small subsamples with both early and 1-year WAI data available ($n = 34$ parents, $n = 26$ home visitors) for testing Hypotheses 3 and 4, the substantial amount of missing data were imputed and compared with results from available data.

According to Tabachnick and Fidell (2001):

In the absence of all other information, the mean is the best guess about the value of a variable. Part of the attraction of the procedure is that it is conservative; the mean for the distribution as a whole does not change and the researcher is not required to guess at missing values. On the other hand, the variance of a variable is reduced because the mean is closer to itself than to the missing value it replaces, and the correlation the variable has with other variables is reduced because of the reduction in variance. The extent of loss in variance depends on the amount of missing data and on the actual values that are missing (p. 62).

If you use some method of estimating missing values . . . consider repeating your analyses using only complete cases. This is particularly important if the data set is small, the proportion of missing values high, or data are missing in a nonrandom pattern. If the results are similar, you can have confidence in them. If they are different, however, you need to investigate the reasons for the difference, and either evaluate which result more nearly approximates "reality" or report both sets of results (p. 65).

Missing data were imputed using mean substitution, providing parent-rated WAI data from, and home visitor-rated WAI data regarding, all 187 BabyFirst parents. Multiple regression analyses for Hypotheses 3 and 4 were repeated using this imputed data set. In Table 14, results are compared to results from the first set of regression analyses using only complete cases (reported above). As expected, the absolute values of the partial correlation coefficients are dramatically reduced in the imputed analyses, because of the reduction in variance resulting from mean substitution. Despite this conservative analytic approach, 10 of the 24 alliance variables retain moderate to strong effect sizes in absolute terms ($prs = .10$ to $.26$) within the 80% confidence intervals ($p \leq .20$) of the present study (see Table 14). Further, a comparison of the correlation patterns in the two sets of analyses shows that both sets generally yield similar conclusions about (a) the *direction* of statistical relationships between alliance and outcome variables and (b) the *relative predictive power* of parent versus home visitor ratings of the alliance. That is, increasingly strong early alliances and increasing alliance improvements predict increases in parental PWB and positive parenting. Similarly, increasingly weak early alliances and increasing alliance deterioration predict increases in hostile/ineffective parenting. Moreover, home visitor alliance ratings are relatively better predictors of parental PWB than parent ratings, parent alliance ratings are better predictors of positive parenting than home visitor ratings, and parent and home visitor ratings are generally comparable predictors of hostile/ineffective parenting. These findings from imputed analyses increase the level of confidence in the study findings.

Table 14

*Early Alliance and Alliance Improvement as Predictors of Parent and Parent-Child Outcomes
at 1 Year: Comparison of Results Using Available Data and Imputed Data*

Predictor variable	Parent outcomes				Parent-child outcomes	
	PWB	AU	EM	SA	POS	HOS
Available data						
Parent-rated early alliance	.11	-.14	.22	.14	.33	-.06
Parent-rated alliance improvement	.16	.14	.15	.14	.47	-.33
Home visitor-rated early alliance	.45	.21	.03	.29	.09	-.30
Home visitor-rated alliance improvement	.43	.21	.49	.37	.16	-.52
Imputed data						
Parent-rated early alliance	.02	-.02	.02	.05	.14	.06
Parent-rated alliance improvement	.03	.07	.03	.06	.26	-.12
Home visitor-rated early alliance	.14	.10	.11	.09	.09	-.09
Home visitor-rated alliance improvement	.11	.05	.20	.11	.09	-.16

Note. Sample sizes for (a) available data are 34 parent ratings and 26 home visitor ratings; and (b) imputed data are 187 parent ratings and 187 home visitor ratings. PWB = psychological well-being, AU = autonomy, EM = environmental mastery, SA = self-acceptance, POS = positive parenting, and HOS = hostile/ineffective parenting. Parent-rated alliance data were analyzed separately from home visitor-rated alliance data. Early alliance and alliance improvement were entered together in the regression model, controlling for baseline outcome scores. Partial correlations in **bold** typeface fall within 80% confidence intervals ($p \leq .20$). Lower partial correlations may be due to low power and/or sampling error.

Discussion

There are now home visiting programs in virtually every jurisdiction across Canada (Government of Canada, 2003). These programs are serving thousands of young children and their families, most of whom are living in conditions of significant socioeconomic disadvantage and risk. Yet few of these programs have been rigorously evaluated for their effectiveness. The present results supported the hypothesis that, under real-world conditions of service delivery, home visiting can have beneficial effects on parental psychological well-being (PWB) and positive parenting after 1 year. These effects are notable because few prevention studies have reported effects as early as 12 months (Culp et al., 2004). However, the present results also suggest that home visiting can have an unintended and unfavourable effect on hostile/ineffective parenting. It is imperative that we determine the pathways through which home visiting actually exerts these positive and negative effects (cf. Brooks-Gunn et al., 2000).

The present study represents one of the first empirical tests of the theoretically crucial²⁶ hypotheses that the early working alliance between parents and home visitors, as well as improvements in the alliance, are key ingredients in the effectiveness of early childhood intervention (Emde, 1988b; Emde et al., 2000; Emde et al., 2004). Thus, it adds to a small but growing number of studies relating the working alliance to early childhood intervention processes and outcomes (Heinicke et al., 2000; Korfmacher et al., 1998; Sharp, Ispa, Thornburg, & Lane, 2003; Whipple, 2004).

The present results provide support for the hypothesis that more positive relationship histories of parents enhance their capacity to develop a strong early alliance with home visitors, whereas more negative relationship histories appear to limit this capacity. Results further support the hypotheses that strong alliances established early in home visiting, and that alliance improvement over the first year of intervention, both predict improvement in parental PWB and

²⁶ It should be noted that theoretical development regarding the alliance, as in "most of psychology is nowhere near [the ideal] stage of theory testing yet" (Meehl, 1997, p. 415) and so Meehl's (1967, 1990a, 1990b) strong use of significance testing cannot (yet) be used.

positive parenting (see Figure 3, C). Results also indicate that weak early alliances and alliances that worsened over the first year of home visiting both predict increases in hostile/ineffective parenting. Taken together, the results of the present study strongly suggest that the working alliance is a central pathway of success or failure in home visiting. More generally, the alliance appears to be a core change process common to all interventions, both preventive and clinical, consistent with relationship-based theories of early childhood intervention (Emde et al., 2000, 2004) and clinical intervention (Henry & Strupp, 1994).

The present study provides new evidence that supports important theoretical linkages between parents' past relationships, their intervention relationship, and their subsequent relationships. Parents' interpersonal histories (positive relationships with others, social support) predict parents' development of a new relationship with their home visitors (early alliance). This intervention relationship, in turn, predicts parents' relationships with themselves (increased self-acceptance), parents' relationships with their environments (increased autonomy and environmental mastery) and, perhaps most importantly, parents' relationships with their newborn children through the first year of life, for better (increased positive parenting) and for worse (increased hostile/ineffective parenting).

What Influences the Development of the Working Alliance in Home Visiting?

Results suggest that parents with more negative relationship histories may be limited in their capacity to develop a strong alliance in home visiting. Parent-home visitor alliances appear to develop early on (after six home visits, in the current context) and the success or failure of home visiting may depend on their initial strength and capacity to improve over time. These results replicate emerging evidence from early childhood intervention research linking alliance-related variables to intervention outcomes (Heinicke et al., 2000; Korfmacher et al., 1998).

Results also suggest that home visitors may find it more challenging to establish strong early alliances with parents who are more autonomous at baseline. However, baseline parental autonomy does not appear to be related to parents' ratings of the development of the early

alliance. Further, stronger parent ratings of the early alliance predicted less favourable program effects on parental autonomy at 1 year. If these findings are replicated, they may suggest that parents rely more heavily on home visitors during the early phase of the alliance, as represented by the negative correlation between early alliance and parental autonomy at 1 year. Over time, further improvements in a positive early alliance may lead to improvements in parental autonomy, as represented by the positive correlation between alliance change and parental autonomy at 1 year.

Parents' personal strengths and resources (high overall PWB, high environmental mastery, high self-acceptance, low depression), in the face of high levels of cumulative family risk (e.g., poor emotional functioning, parental stress), appear to be especially important to parents in developing a strong early alliance. In contrast, it appears that, for home visitors, parents' personal limitations (low overall PWB, low environmental mastery, low self-acceptance, high depression) are especially important in developing a strong early alliance. In the present study, home visitors seemed to be more strongly motivated to establish positive working alliances with more personally needful parents living in less risky family environments.

In contrast, home visitors may be particularly limited in establishing strong working alliances with Aboriginal parents, which may account, in part, for the finding that parental Aboriginal status was a limiting factor on the effectiveness of home visiting for all outcomes except positive parenting, even after controlling for other established risk factors for healthy child development (cumulative risk, child difficult temperament, and parental depression). The risky family environments and unsafe neighbourhoods associated with being an Aboriginal parent may partially account for home visitors' limitations in developing effective early alliances with Aboriginal parents. These difficult conditions may distract or deter home visitors from establishing trusting relationships with Aboriginal parents, due to home visitor concerns for their own personal safety or well-being.

How Does Home Visiting Improve Parental Psychological Well-Being and Positive Parenting?

The present findings indicate that a strong alliance early on in home visiting, especially one that improves over the first year, is a central pathway by which home visiting improves (a) parents' overall psychological well-being (PWB), particularly their self-acceptance, their sense of mastery over their environments, and their sense of autonomy, as well as their (b) positive interactions with their children. Results indicate that home visitor ratings are better predictors of parental PWB, whereas parent ratings are better predictors of positive parenting. These results make sense, given the different emphases that home visitors and parents may place on the interpersonal process of intervention. Home visitors, given their primary focus on parents as their clientele, may emphasize (more than parents) actions on improving parental PWB. Home visitors may use their sense of a strong and improving alliance as leverage for facilitating improvements in parental self-acceptance, environmental mastery, and autonomy. In contrast, parents, given their primary focus on their newborn children, may emphasize (more than home visitors) actions on improving positive parenting. Parents may use their sense of a strong and improving alliance as a secure base for learning and improving their positive parenting skills.

Why Did Home Visiting Increase Hostile/Ineffective Parenting?

The working alliance may also account, in part, for the unintended and unfavourable effect of home visiting on hostile/inffective parenting. Parents who observed deterioration in their relationships with their home visitors reported a subsequent deterioration in their relationships with their newborn children. Home visitors who observed weaker early alliances, along with deterioration in their relationships with parents, may also have contributed to this negative pattern in parents' relationships with their newborn children. Henry and Strupp (1994) have provided evidence from studies of interpersonal process in therapy that people learn to treat themselves and others as they have been treated by significant others in their life. As the therapist becomes a significant other to their clients, these clients often go on to treat themselves and others as they have been treated by their therapist. Further, the interpersonal process underlying

therapeutic interventions "either ameliorates or entrenches" the client's negative internal working models of previous relationships (Henry & Strupp, 1994, p. 64). A similar dynamic may be operating in the home visiting relationship. A poor home visiting alliance, wherein parents do not feel accepted, and perhaps feel rejected, may in turn act in a less accepting and more rejecting manner toward their children.

What happened (or didn't happen) between home visitors and parents that made for success or failure in this important parent-child outcome? The present results show how negative relationship histories of parents predict negative early relationships with their home visitors. Parents with more negative internal working models appear to be more prone to poor early alliances in home visiting. This proneness would tend to produce feelings of low self-esteem, mistrust, and sensitivity to criticism (real or perceived). A history of dealing with prejudice, discrimination, or racism (e.g., common experiences for many Aboriginal parents) would tend to magnify all of these feelings. Parents with negative internal working models may be predisposed to future negative interpersonal experiences which are, in part, elicited by their own negative interpersonal expectations and actions, thereby confirming their negative internal working models; an interpersonal case of the vicious cycle or self-fulfilling prophecy. There is extensive evidence that, to varying degrees, individuals shape their interpersonal environments, wittingly or unwittingly, for better or worse (Rutter, 2005). Together with an increasingly negative relationship over the first year of home visiting, a negative early alliance predicts an increasingly negative parent-child relationship, manifested by increases in hostile/ineffective parenting.

By implication, two important tasks of the home visitor are to recognize this negative interpersonal process in the home visiting alliance and repair the ruptures in the alliance (Safran, Muran, Samstag, & Stevens, 2001, 2002) that it represents. The present results appear to suggest that in early childhood intervention, as in clinical intervention (Henry & Strupp, 1994), a little bit of negative interpersonal process can go a long way in producing negative outcomes.

The above conclusion is consistent with another, based on a large body of well-replicated findings across diverse areas of research, namely that “bad is stronger than good” (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001, p. 323). In close relationships, such as the home visiting relationship, this conclusion means that “increasing positive behaviours in a relationship will not affect the relationship as much as decreasing negative behaviours” (Baumeister et al., 2001, p. 328). Positive behaviours may include warmth, praise, or acceptance. Negative behaviours may include harsh criticism or blatant rejection. Gottman (1994) suggested that, in relationships, good must outweigh bad by at least 5 to 1 (cited in Baumeister et al., 2001, p. 329), and some evidence from the NLSCY suggests that a similar ratio may apply to the relationship between parenting and children’s behaviour (Statistics Canada, 1998, 2004).²⁷ In short, “the implication is that the long-term success of a relationship depends more on not doing bad things than on doing good things” (Baumeister et al., 2001, p. 329). In terms of parenting, “bad things” include a lack of warmth and consistency, along with harsh, hostile, and rejecting interactions (Collins et al., 2000; O’Connor, 2002; Rohner, 2004; Repetti et al., 2002; Rutter, 2005) that make a child feel unloved and unwanted. This body of evidence strongly indicates that although negative interpersonal interactions may begin on a small scale, the accumulation of these negative interactions over time can lead to deleterious effects. The present findings suggest something similar may occur in home visiting. An initially weak or negative working alliance may begin on a small scale in the early phase of intervention. Over time, negative experiences accumulate, as reflected in a deteriorating alliance across the first year of home visiting. As parents internalize this increasingly negative alliance, deleterious effects such as increased hostile/ineffective parenting may follow, which, in turn, leads to deleterious effects on their children’s development. As discussed in an earlier section (see *Evaluation Logic of Prevention and Early Intervention Programs* above), a major

²⁷ In terms of child development, “fairly little work is available to examine the relative power of good versus bad events on child development” (Baumeister et al., 2001, p. 337) except IQ: “the effects of parenting and other environmental experiences on children are mainly negative: Bad family environments override the effect of genetic heredity on intelligence, but good environments do not” (Baumeister et al., 2001, p. 338).

goal of early intervention is to prevent proximal negative outcomes that lead to distal negative outcomes. Preventing such “chains of risk” is a major intervention priority (Rutter, 2005).

Other recent evidence indicates that the intergenerational transmission of antisocial behaviour (Serbin & Karp, 2004) probably begins as early as infancy. In a recent study, hostile/ineffective parenting at age 5 months was one of the strongest predictors of high physical aggression 12 months later (Tremblay et al., 2004). In this study, difficult child temperament and hostile/ineffective parenting were positively correlated ($r = .30$), similar to the present findings ($r = .20$ at baseline), suggesting a reciprocal effect of temperament on hostile/ineffective parenting. In the Hawaii Healthy Start (HHS) effectiveness trial, difficult child temperament predicted severe physical assault in the context of maternal substance abuse, and predicted “assault on the child’s self-esteem” in the *absence* of maternal depression (Windham et al., 2004). That is, the presence of maternal depression protected children with difficult temperaments from parental assaults to the children’s self-esteem, primarily due to a decrease in parent-child interaction brought on by parents’ depression. In the present study, difficult child temperament remained an important predictor of hostile/ineffective parenting at 1 year ($pr = .17, p = .02$), even after controlling for the effects of the intervention and other baseline variables (see Table 12).

The reciprocal nature of parent-child interaction (Bell, 1968) is now well-established, particularly by recent genetically-sensitive research studies (i.e., studies that measure the effects of both genetic and environmental factors, e.g., twin and adoption studies) (Collins et al., 2000; Maccoby, 2000; O’Connor, 2002). Child characteristics such as difficult temperament, through their risk effects on the environment (Rutter, 2005), may impose limits on the effectiveness of early childhood intervention (Tremblay et al., 2004; Windham et al., 2004). However, a recent genetically-sensitive study found that environmental factors account for most of the variation in corporal punishment and physical maltreatment (Jaffee et al., 2004), parenting behaviour considerably more harmful than the hostile/ineffective parenting measured in the present study. Although “difficult children may provoke corporal punishment . . . the factors causing child abuse

are more likely to be found within the family environment or the adult abuser" (Jaffee et al., 2004, p. 1056). Another recent genetically-sensitive study showed that parents' negative emotional attitudes toward their children may be an environmental cause of children's early emerging antisocial behavior problems (Caspi et al., 2004).

For local practitioners and policymakers, it will be important to address and monitor the observed unintended effect on hostile/ineffective parenting over the three-year course of the BabyFirst program.²⁸ Intervention effects should continue to be followed through the second and third years of program delivery, and efforts should be made to improve the effectiveness of home visiting in reducing hostile/ineffective parenting. This is discussed further in a later section (see *How Can Home Visiting Programs Be Strengthened?* below). Longitudinal outcome studies, before and after program completion at age 3, to establish causal linkages between parent-child outcomes and child outcomes²⁹, are also strongly recommended (see Figure 3, C).

What Are the Limitations of Home Visiting Programs?

The present findings indicate that early childhood intervention can have positive effects on personal qualities that may be necessary to survive, if not overcome, socioeconomic hardship: autonomy, environmental mastery, and self-acceptance. It may be noted that these PWB variables are core tenets of self-efficacy theory (Bandura, 1977, 2001), which is based on a fundamental belief in the "capacity to exercise control over the nature and quality of one's life [as] the essence of humanness" (Bandura, 2001, p. 1). In addition to these direct benefits for parents, the positive program effects on parental psychological well-being (PWB) in the present study are noteworthy.

As described in an earlier section (see *Program Theoretical Model* above), parents' PWB has

²⁸ Measures for estimating the clinical or policy significance of risk and protective factors are an important and recent methodological innovation, but the final decisions must still be made by "clinicians and policy makers who are aware of the nature of the population and of the clinical consequences of false-positive and false-negative judgements in that population, not by methodologists without such knowledge" (Kraemer et al., 1999, p. 270).

²⁹ This represents an important next step in the research program: testing hypothesized links between direct and indirect outcomes in the program theoretical model (see Figure 3, C). In 2000, Brooks-Gunn et al. noted that "to date, no evidence exists as to the effects of home visiting on parents acting as a pathway through which home visiting influences children" (Brooks-Gunn et al., 2000, p. 570). A new study provides the first evidence linking the effects of home visiting on parenting to longer-term effects on children's behaviour outcomes (Lyons-Ruth & Melnick, 2004).

several indirect benefits for their children (Bugental & Johnston, 2000), because parental PWB appears to mediate the adverse effects of socioeconomic stress and disadvantage on parenting (Mistry et al., 2002; Taylor et al., 1997). Children are shielded from the risk effects of socioeconomic disadvantage and benefit from the improved parenting fostered by parental PWB.

Extant research suggests, however, that home visiting alone may be insufficient to meaningfully improve parental depression or family social support. The magnitude and pattern of effects in the present study of the BabyFirst program are similar to those reported in effectiveness studies of its forebears, Hawaii Healthy Start (HHS) and Healthy Families America (HFA) (see also Gomby, 2003). Both HHS and HFA had intended effects on parent-child outcomes but virtually no intended effects on parental mental health (Daro & Harding, 1999; Duggan, Fuddy, et al., 2004; Landsverk et al., 2002) or family social support (Daro & Harding, 1999; Duggan et al., 1999; Landsverk et al., 2002; McCurdy, 2001). These minimal effects of home visiting on parental depression and social support are concerning, because of their importance to parental well-being and in influencing children's development over time. Improving these two outcomes, therefore, remains an important but largely unrealized goal for home visiting (Barnet et al., 2002; Duggan, Fuddy, et al., 2004; Stevens, Ammerman, Putnam, & Van Ginkel, 2002). Possibilities for improving intervention effects on depression and social support are discussed below.

The limitations of home visiting programs reflect clinical concerns and call for high levels of ethical and professional responsibility but, as Musick & Stott (2000) have emphasized, home visitors are not clinicians, and are usually not trained to address clinical problems such as depression. This reality presents major training and service challenges for home visiting programs that seek to prevent child maltreatment, because clinical problems are the major risk factors for child maltreatment (Duggan, Fuddy, et al., 2004; NRC & IOM, 1999; Windham et al., 2004). Home visitors could, however, be trained to recognize clinical problems (or systematically screen for them), and then make appropriate referrals. The success of such referrals, of course, depends greatly on the availability of effective supports and resources in the community.

The present study calls much-needed attention to the issue of negative effects of early childhood intervention. Despite Durlak and Wells' (1997) exhortation to continue monitoring interventions to detect any possible negative side effects, there has been very little discussion in the literature regarding the negative effects of early childhood intervention (for one exception, see Shonkoff, 2004).³⁰ Negative effects were found for the highest-risk families in the randomized Infant Health and Development Program (IHDP) efficacy trial (Brooks-Gunn et al., 1994; see also Berlin et al., 1998), the highest-risk families in the randomized Early Head Start (EHS) effectiveness trial (ACYF, 2002), and the high-risk families receiving the highest service dose in the randomized Hawaii Healthy Start (HHS) effectiveness trial (Duggan, McFarlane, et al., 2004). As described in the introduction to the present study, the EHS authors posited a "major life change hypothesis" as an explanation for these unfavourable impacts:

low-income families who have experienced high levels of instability, change, and risk may be overwhelmed by the changes that a new program introduces into their lives, even though the program is designed to help. As a result, the program requirements may create unintended negative consequences for these families (ACYF, 2002, p. 355).

Home visitors are the concrete manifestation of the program. A strong alliance can bear the brunt of parents feeling frustrated, overwhelmed, and/or criticized. However, timing is everything. Efforts to help that are too much, too soon, will backfire. After forty years in primary prevention, Caplan and Caplan (2000) provided a model of self-reflection for the field:

We are beginning to realize that the methods and techniques that we are developing in the field of primary prevention may cause harm rather than the good we hope to achieve. Our well-meant actions must be carefully controlled and limited by sensitivity to the rights of people for privacy and autonomy in deciding about intimate aspects of their own lives. It may be that we pioneers are sufficiently experienced and interpersonally sensitive to be able to recognize the early signs of potential harm when we cross the boundaries of privacy, but it remains for the future to determine how, if at all, it is possible to recruit the large numbers of line workers who will intercede with our level of sensitivity and skill into the intimate lives of a whole population. *For instance, many people who are grappling with current situations of adversity will be weakened, rather than strengthened, by well-meaning intervention that is aimed at supporting them. Such people, if asked, are likely to say that they wish to be left alone to struggle with their difficulties at their own pace, rather than being forced by the pressure of others to deal with the issues according to some universal timetable* (p. 134, emphasis added).

³⁰ It should be noted that short- and long-term negative effects have also been found in clinical interventions delivered during later periods of the life course, including adolescence (Dishion, McCord, & Poulin, 1999) and adulthood (D. C. Mohr, 1995; Strupp, Hadley, & Gomes-Schwartz, 1977, 1994).

It is unfortunately true that many, if not all, caregivers may inadvertently and on occasion harm the people they are trying to help. Combating this hazard should have high priority in all our programs of primary prevention (Caplan & Caplan, 2000, p. 135).

In the EHS trial, the authors suggested that it was

possible that the services provided by the Early Head Start programs--primarily weekly home visits or regular attendance at centers--were not sufficient to meet the needs of these families, and program expectations for participation may have added to the challenges these parents faced (ACYF, 2002, pp. 354-355).

In other words, administrative pressures for program recruitment and delivery may have led program providers to minimize or overlook families' needs to "struggle with their difficulties at their own pace" as mentioned by Caplan and Caplan in the quotation above (2000, p. 134). As Halpern (2000) warned, "deliberately setting out to alter parenting in a significant way is an uncertain proposition" (p. 376). Musick and Stott (2000) have further cautioned:

intervening in the parenting process is a complex, fundamentally therapeutic endeavor. Such an endeavor calls for well-trained and sophisticated workers with theoretical as well as practical knowledge of the parenting process and with an understanding of how, and under what circumstances, people change. It calls for excellent supervision as well (p. 442).

Recent developments in the BabyFirst program to facilitate quality management and assurance, enhance the supervisory skills of nurses, and provide professional development opportunities for home visitors (D. Girard, personal communication, January 2005), all suggest that the program is steadily improving its sensitivity to these perennial service delivery challenges.

Limitations on the effectiveness of early childhood intervention. When appraising the results of the present study, it is important to consider the context of disadvantage for families in the intervention group. On average, these families lived on poverty levels of annual income, with less social support, greater parental depression, lower parental psychological well-being, greater residential instability, and lower levels of neighbourhood safety than did parents in the comparison group and in the general population. This context of disadvantage was described in a recent review of research regarding the "environment of childhood poverty":

Poor children confront widespread environmental inequalities. Compared to economically advantaged children, they are exposed to more family turmoil, violence, separation from their families, instability, and chaotic households. Poor children experience less social support, and their parents are less responsive and more authoritarian. Low-income children are read to relatively infrequently, watch more TV, and have less access to books and computers. Low-income parents are less involved in their children's school activities. The air and water poor children consume are more polluted. Their homes are more crowded, noisier, and of lower quality. Low-income neighborhoods are more dangerous, offer poorer municipal services, and suffer greater physical deterioration. Predominantly low-income schools and day care are inferior. The accumulation of multiple environmental risks rather than singular risk exposure may be an especially pathogenic aspect of childhood poverty (Evans, 2004, p. 77).

As noted in the introduction of the present study, given these conditions of cumulative risk, it is remarkable that home visits for a few hours per month over a few years have any positive or lasting effects at all (Bryant, 2003; Gomby, 2000; NRC & IOM, 1999).

Given the environment of childhood poverty, some leading thinkers in early childhood intervention have argued that

goals such as "improved parent-child relationships" . . . and "strengthening families" are premised on naïve notions about social services. A focus on family relationships and individual mental health cannot by itself prevent or cure what are essentially societal problems such as poverty, unemployment, and crime (Music & Stott, 2000, p. 452).

Farran (2000) expressed similar concerns regarding early childhood intervention for children with disabilities: "It is possible that the primary problem with the intervention efforts mounted so far is that they do not take a contextual focus . . . None of the programs reviewed have made any difference to the income, housing conditions, or employment of the parents involved" (p. 525)³¹ and "there continues to be a persistent lack of recognition of the parenting context of poverty" (p. 542). Given this context, Halpern (2000) has suggested an "urgent need to lower expectations" regarding early childhood intervention (p. 375).

³¹ Recent findings from systematic efforts to intervene in these social conditions, such as the Moving to Opportunity (Leventhal & Brooks-Gunn, 2004), and New Hope (Huston et al., 2001, 2003) experiments suggest that such changes are indeed possible.

The present results lend support to recent recommendations to develop more modest, balanced, and realistic expectations³² of home visiting (Gomby et al., 1999; NRC & IOM, 1999), and to focus on program quality and implementation (Duggan, Fuddy, et al., 2004; Gomby, 2000; Leventhal, 2001; NRC & IOM, 1999; Olds, 2003), particularly the quality of the relationship between parent and home visitor.

It is also important to emphasize that no single program, home visiting or otherwise, however effective, should be expected to substantially reverse the deleterious effects of poverty, unemployment, and social disorganization (Weiss, 1993). Nor can even the most effective early childhood intervention programs be expected to inoculate against the effects of inadequate schools and impoverished opportunities for education later in life (Zigler & Styfco, 2001). It is clear from the available evidence that a comprehensive and integrated system of multiyear, multicomponent programs and services, including home visiting, lasting from early childhood through middle childhood and adolescence, is needed to facilitate successful life course development, providing "safe passage" (Dryfoos, 1998, p. 4) for children from conception to adulthood (Catalano et al., 2002; Earls, 1998; Greenberg et al., 2001, 2003; NRC & IOM, 2000, 2002; Weissberg, Kumpfer, & Seligman., 2003). More broadly, we must infuse society with a sense of shared responsibility for all its children (Daro & Donnelly, 2002; NRC & IOM, 2000).

How Can Home Visiting Programs Be Strengthened?

The present findings suggest that in order to improve the effects of home visiting on hostile/ineffective parenting, home visiting programs must first understand the strong relationship between socioeconomic disadvantage and hostile/ineffective parenting (Grant et al., 2003) and how this disadvantage exerts its negative effects on children's development via hostile/ineffective parenting (Grant et al., 2003; Repetti et al., 2002; Rutter, 2005). Hostile/ineffective parenting appears to be transmitted intergenerationally (Capaldi, Pears, et al., 2003; Newcomb & Locke,

³² "It seems fair to say that early intervention works, but not as consistently as hoped, not as strongly as hoped, and not across as many areas of the parent's and child's development as hoped. Answering *why* this is so is more difficult" (Korfmacher, 2002, p. 277).

2001; Serbin & Karp, 2004; Tremblay et al., 2004), so parents with childhood histories of hostile/ineffective parenting are more likely to be similarly hostile with their own children.

The daunting challenge for home visiting programs targeting parents and families living with significant socioeconomic disadvantage is manifold. These parents are (a) more likely to use more hostile/ineffective parenting, (b) more likely to have had negative relationships and, in turn, are (c) more predisposed to misinterpreting new relationships in a negative manner, particularly in the context of interpersonally “complex” communications that convey mixed messages (Henry & Strupp, 1994) or in the context of support provided under interpersonally ambiguous conditions. According to Collins and Feeney (2004):

Support attempts may be ambiguous either because support providers are unskilled at providing effective support—because their behavior contains a mixture of helpful and unhelpful responses—or because they misunderstand the type (or amount) of support that is wanted or needed by the support recipient (p. 364).

In these interpersonal contexts, disadvantaged parents may be (d) less likely to successfully establish a positive new relationship, the foundation of all successful interventions. “In such cases, the support recipient's subjective sense of feeling supported may hinge greatly on his or her appraisal of the support provider's good will and benevolent intent” (Collins & Feeney, 2004, p. 364). In other words, much depends on parents' perception of the early working alliance.

In short, home visiting programs must improve their capacity to recognize recurrent maladaptive patterns in parents' relationships and provide new, more positive interpersonal experiences in home visiting which, in turn, can help parents create new and more positive relationship patterns with their young children. Alongside the extensive literature, the present findings suggest that the initial interactions and communications from home visitor to parent (i.e., the home visitor's contribution to the interpersonal process of the early alliance), must not only be warm, accepting, consistent, and supportive, but also clear, direct, and unambiguous in order to minimize negative misinterpretation or misattribution on the part of vulnerable parents:

When the support message is somewhat ambiguous and more open to construal . . . insecure working models appear to be a cognitive liability primarily when ambiguous or potentially negative events . . . activate doubts or vulnerabilities, which then distort perception and interfere with one's ability to engage (Collins & Feeney, 2004, p. 379)

For example, a given parent may misinterpret her home visitor's lighthearted or humorous attempts to address a stressful situation as minimizing the seriousness of the situation or being insensitive to the parent's feelings. The parent may in turn be less reluctant to engage in home visiting interventions, or even drop out of the program entirely. Such misinterpretations are likely to be common among families targeted for home visiting, and may contribute to the high attrition rates common to most home visiting programs (Duggan et al., 2000; NRC & IOM, 1999).

The present results, therefore, suggest that important characteristics to consider in the recruitment, training, and supervision of home visitors include their capacity and skills to develop an alliance and their skills in repairing ruptures in the alliance. This is consistent with research from both early childhood intervention (Berlin et al., 1998; Daro, McCurdy, Falconnier, & Stojanovic, 2003; McCurdy & Daro, 2001; McCurdy, Gannon, & Daro, 2003; Wasik, 1993; Wasik & Roberts, 1994) and clinical intervention (Henry & Strupp, 1994; Horvath & Bedi, 2002; Safran et al., 2001, 2002). The present interpersonal process view of the alliance in early childhood intervention suggests that future training and supervision of early childhood intervention providers should facilitate their recognition of the subtle interactions between themselves and the parents they serve, with special attention to negative interpersonal process. Indeed, these interpersonal qualities and skills may be essential for effective parent recruitment, participation, engagement, and retention in early childhood intervention (Berlin et al., 1998; Daro et al., 2003; Korfomacher, O'Brien, Hiatt, & Olds, 1999; McCurdy & Daro, 2001; McCurdy et al., 2003; Navaie-Waliser et al., 2000). The present findings indicate that home visitors' ability to establishing a strong alliance early on in home visiting is essential for accomplishing the longer-term goals of home visiting, including enhanced parenting skills.

Improving the effectiveness of home visiting on hostile/ineffective parenting.

New research suggests that multimodal early childhood interventions (e.g., including group therapy, home practice of group therapy activities, and home visiting; Peterson, Tremblay, Ewigman, & Saldana, 2003) and home visiting enhanced with cognitive retraining (Bugental et al., 2002) can reduce hostile/ineffective parenting. Cognitive retraining involves a two-part intervention. First, parents are “assisted in acquiring skills in reading children’s cues to distress and in countering misattributional processes (e.g., countering the view that infants or very young children can read parents’ minds, are behaving with negative intent, or are challenging parental power)” (Bugental et al., 2002, p. 245). Second, parents receive “problem-solving training in which they define the problem, brainstorm possible solutions, evaluate possible consequences, develop an action plan, and observe and evaluate the success of their children” (Bugental et al., 2002, p. 245). Results from a recent prospective study of child maltreatment during the first year of life showed that a lack of perceived parental power interacts with child risk to predict maltreatment in low-income families (Bugental & Happaney, 2004), supporting the theorized mechanism for the superior efficacy of home visiting enhanced with cognitive retraining (Bugental et al., 2002).

Improving the effectiveness of home visiting on parental depression.

Effective prevention approaches are available for adult depression (e.g., cognitive-behavioural mood management courses; Muñoz, 2001). There are also effective approaches for preventing the negative effects of parental depression on infants and children (e.g., programs promoting massage, stimulation, and other forms of early parent-child interaction; Field, Hernandez-Reif, & Freedman, 2004; Gladstone & Beardslee, 2002). These approaches may represent service integration opportunities for improving the effects of home visiting on parental depression.

Improving the effectiveness of home visiting on family social support.

To date, most interventions intended to improve social support have been disappointing (Cohen, 2004). Social support may be one of the mechanisms through which neighbourhoods influence children’s development (Leventhal & Brooks-Gunn, 2000). This suggests that population-, community-, and

neighbourhood-level interventions (Reppucci, Woolard, & Fried, 1999) may be essential to directly improve social support for families and indirectly improve the effectiveness of home visiting (Duggan, McFarlane, et al., 2004; McCurdy, 2001). Studies of neighbourhood variables as mediators and moderators of parenting are emerging (Hill & Herman-Stahl, 2002; Silk, Sessa, Morris, Steinberg, & Avenevoli, 2004). In the present study, higher levels of baseline neighbourhood safety predicted higher levels of positive parenting at 1 year ($pr = .21, p = .004$), even after controlling for the effects of the intervention and other baseline variables.

Neighbourhood variables such as safety and social cohesion (Leventhal & Brooks-Gunn, 2000) may represent important targets or mechanisms for effective early childhood intervention.

Strengths and Limitations of the Present Study

Strengths of the present study include the following: (a) its high level of internal validity through the use of the regression-discontinuity design (RDD); (b) its use of reliable and valid alliance and outcome measures; (c) its study of intervention effects under ecologically valid, real-world conditions; with (d) a sample that was sociodemographically comparable to the population served in the EHS effectiveness trial (Duggan, McFarlane, et al., 2004) and to targeted populations served by most home visiting programs (Sweet & Appelbaum, 2004); (e) its difficult context of intervention (i.e., conditions of socioeconomic disadvantage), which may magnify the importance of obtained positive effects (Prentice & Miller, 1992); and (f) its replication of the meta-analytic direction and magnitude of program effects (Bakermans-Kranenburg et al., 2003; MacLeod & Nelson, 2000; Nelson et al., 2003; Sweet & Appelbaum, 2004) and alliance-outcome effects (Horvath & Bedi, 2002; Horvath & Symonds, 1991; Martin et al., 2003).

Limitations of the present study include the following: (a) the preliminary nature of its effectiveness results (i.e., 1-year outcomes in a 3-year program), which indicates a need for continued longitudinal follow-up; (b) the lack of available data on program implementation (e.g., data on actual number of home visits) and actual home visit processes to test for fidelity to the program model; (c) missing alliance data and attendant potential selection effects, which may

limit the generalizability and replicability of alliance results (as discussed in the *Results* section above); (d) the lack of outcome measurement perspectives apart from parent reports (e.g., the lack of observation or testing by independent observers); and (e) the extensive time required by the program to accrue baseline and 1-year data on a sufficiently large sample (a duration of nearly five years from 2000 to 2004), which may have led to unmeasured participant attrition, given the relatively high attrition rates found in most home visiting programs during the first year of service delivery (Duggan et al., 2000; NRC & IOM, 1999). With the exception of the first limitation above, all of the foregoing limitations are directly related to the fact that these data were not routinely collected by the program during its initial years of operation, and were not under the direct control of the present researcher. Most public programs have insufficient resources for collecting even the most basic management information, much less systematic measures of process and outcome. However, several improvements to the data collection protocol, including detailed and routinized measures of interventions during each home visit, have been recently initiated (M. Chartier, personal communication, December 2004). The above limitations highlight both the potentials and pitfalls of data collection in the non-research context of public programs, and the subsequent limitations that ensue for secondary analyses of these data (Brooks-Gunn, Berlin, Leventhal, & Fuligni, 2000; Brooks-Gunn, Phelps, & Elder, 1991).

Implications and Future Directions for Research

The science of early childhood intervention tells us what is possible. The translation of that knowledge into policies and programs demonstrates what we are willing to do with what we know. The politics of resource allocation would be guided more constructively by empirical research if we moved beyond the basic question of *whether* early childhood interventions work and began to seriously address the more compelling challenge of *how* to achieve a maximum return on our early childhood investments. For the intervention research community, this underscores the importance of a more explicit focus on understanding what does and what does not improve the odds of more favorable outcomes for children with different needs, particularly for those who are the most difficult to reach (Shonkoff, 2003, pp. 4-5).

Measuring positive outcomes. The present study may be one of the first effectiveness studies of early childhood intervention that has included positive outcome measures of parental psychological well-being. Most efficacy and effectiveness studies have included only negative

outcome measures of parental psychological well-being (e.g., parental stress, parental depression). The present PWB findings serve as a reminder that families identified as being "at risk" also have strengths. For example, at baseline, high-risk families in the intervention group were similar to the low-risk comparison families in terms of three parental PWB factors: autonomy, environmental mastery, and personal growth, strengths that may have been otherwise overlooked. Given that almost no early childhood intervention studies have used positive outcome measures³³, and none appear to have used the PWB measure used in the present study, the effects of early childhood intervention on parental PWB require further investigation and the present findings require replication. A recent exploratory study (Kovach, Becker, & Worley, 2004) suggests that parents participating in early childhood intervention show improvements on variables (self-determination, self-sufficiency, and decision-making) that appear conceptually related to two of the three PWB factors improved by the BabyFirst home visiting program (parental autonomy and parental environmental mastery) and warrants more rigorous investigation (e.g., use of a comparison group, random assignment to intervention).

Pursuing the program, its participants, and their interaction. The tripartite framework (Berlin et al., 1998) of program, participants, and interaction should continue to guide future early childhood intervention research. The present study provides quantitative estimates for five of the nine framework components (included in two of the three principal dimensions; see Table 1): child characteristics, family characteristics, community characteristics, cumulative risk (*participants dimension*); and participant-staff relationship (*interaction dimension*). Future studies should incorporate data on the *program dimension* (program characteristics, program staff characteristics, and cultural relevance) and the other component of the *interaction dimension* (match of service with need). Specifically, future studies should investigate (a) the influence of the home visitor's relationship history on the home visiting relationship and the parent's internal

³³ For example, the recent positive psychology movement (Aspinwall & Staudinger, 2003; Keyes & Haidt, 2003; Seligman & Csikszentmihalyi, 2000) offers extensive theoretical, methodological, and empirical resources.

working models; and (b) the influence of the parent-child relationship on the child's internal working models. Based on the few studies available to date, Korfomacher (2002) suggested that "participants respond to interventions in unique ways based upon personal histories and characteristics" (p. 284). Future research needs to determine whether such responses are indeed unique, or whether common patterns can be discerned, in order to determine a more appropriate balance of ideographic and nomothetic approaches to future studies (more on this issue later).

Early childhood intervention provides a good "natural laboratory" for studying the development of attachment relationships in adulthood (Main, 1999, p. 861). Future research should study the influence of the parent-home visitor alliance on other relationships, the role of the home visitor as an attachment figure for parents (as suggested by Berlin & Cassidy, 1999), and the alliance as a source of leverage in the larger system of relationships surrounding parents (as suggested by Berlin & Cassidy, 1999; Emde et al., 2004; Emde & Robinson, 2000).

Learning what works in the real world. The regression-discontinuity design (RDD) has proven useful in the present study for evaluating the effectiveness of home visiting, when families are assigned to the program on the basis of risk or need. Such targeting is the most common approach (75%) in home visiting programs (Sweet & Appelbaum, 2004) and most early childhood interventions (Nelson et al., 2003). Trochim (2000) has suggested that "the real allure of the RDD is that it allows us to assign the treatment or program to those who most need or deserve it. Thus, the real attractiveness of the design is ethical – we don't have to deny the program or treatment to participants who might need it as we do in randomized studies."³⁴ He has suggested that the scientific community should facilitate the use of high quality methodologies such as the randomized experimental design or the RDD by encouraging "administrators and legislators who wish their programs to be accountable to make explicit their criteria for program eligibility by either using probabilistically based lotteries or by relying on quantitative eligibility

³⁴ The ethical advantages of the RDD may be offset by its sample size needs, given that 2.75 times as many participants will not receive the intervention (Senn, 1995, cited in Cappelleri & Trochim, 2003). The use of unequal n (Lipsey, 1990) may be a solution.

ratings and cutoff values" (Trochim, 2000) in order to improve accountability for public investments in programs through rigorous evaluation of effectiveness.

Understanding neighbourhood and community influences. There is almost no research on neighbourhood or community determinants of social support, social support as a mediator of community structures and processes, or on effective social support interventions (Barrera, 2000; Cohen, 2004; Earls & Carlson, 2001; Hogan, Linden, & Najarian, 2002; Sampson et al., 2002; Shinn & Toohey, 2003). These are scientific priorities for improving the effectiveness of early childhood intervention (Leventhal & Brooks-Gunn, 2000).

Studying and valuing culture, race, and ethnicity. Additional research is also needed on the influences of parental culture, race, and ethnicity on the development of the working alliance in home visiting (cf. McCurdy et al., 2003). The present study found a small negative correlation ($r = -.11$) between parental Aboriginal status and home visitor-rated early alliance. Future studies need to include larger samples in order to determine whether this result is valid and replicable, or due to low power and/or sampling error. If the former, future studies should explore the role of Aboriginal status in determining the effectiveness of early childhood intervention in Canada. Given that Aboriginal people are more likely to live in conditions of disadvantage (Hallett et al., 2000; MacMillan et al., 1996; Martens et al., 2002; Young, 2003), and Aboriginal children are more likely to grow up in the environment of childhood poverty (CICH, 2000; Hallett et al., 2000), it is essential that early childhood intervention researchers collaborate with Aboriginal communities to learn how better to improve the life chances and life quality of their children.

Moving toward meta-analysis. Sufficient studies relating working alliance to early childhood intervention outcomes are needed to permit meta-analytic review. This is necessary to test whether the direction and magnitude of alliance-outcome relationships in psychotherapy studies are replicated across early childhood intervention studies. The large sample in the national Early Head Start (EHS) evaluation (ACYF, 2002) promises to be a major contributor in this regard, as the Working Alliance Inventory (WAI) was included in the EHS assessment protocol

(Brooks-Gunn et al., 2000). However, to date, preliminary WAI results have been published for only a single EHS site (Sharp et al., 2003). A recent, unpublished study of a single Healthy Families America (HFA) site has reported links between the alliance, as measured by the WAI, and outcome (Whipple, 2004).

Charting the dimensions of the alliance. Future studies should elaborate the alliance and its relationship to outcome in early childhood intervention. The as-yet unspecified critical dimensions of the helping relationship (Emde et al., 2000; Korfsmacher, 2002) may include those posited by Bordin (1975, 1979, 1994), namely goals, tasks, and bond. These studies should follow the lead of research using the Working Alliance Inventory (WAI) (Horvath, 1994). An interpersonal process view of early childhood intervention also indicates a need for finer-grained process analyses of the alliance, following the lead of research using the Structural Analysis of Social Behavior (SASB; Benjamin, 1974) in psychotherapy (Henry & Strupp, 1994; Hilliard, Henry, & Strupp, 2000).

Looking at the alliance across time. The two-wave design in the present study needs to be improved with additional waves of data after subsequent years of intervention, to provide a more rigorous analysis of continuity and change in the alliance over time, and their relation to early childhood intervention outcome (cf. Korfsmacher, 2002). Such an undertaking would follow the lead of emerging research on the patterns of alliance development over time in counseling and psychotherapy (Kivlighan & Shaughnessy, 1995, 2000; Stiles et al., 2004).

Looking across generations. Notwithstanding their inherent costs, intergenerational studies of early childhood intervention represent a pressing priority for studying the effectiveness of early childhood intervention. Some key questions for future research include the following: How do first-generation changes in parent-child interaction created through early childhood intervention relate to subsequent patterns of second-generation parenting by those children as adults? How do these changes influence the third generation of children? Many of the leading influences of healthy development over the life course appear to begin very early in childhood

and carry forward into future generations, presenting significant ethical, methodological, and practical challenges, as well as opportunities, to literally change the future for a new generation of children (Capaldi, Pears, et al., 2003; Serbin & Karp, 2004; Tremblay et al., 2004).

Similarly, intergenerational studies (Dubow, Huessmann, & Boxer, 2003; Shaw, 2003; Thornberry, Hops, Conger, & Capaldi, 2003) are also a priority for the process-outcome study of effective ingredients. Intergenerational studies are essential to test hypotheses generated by relationship-based theories of early childhood intervention, including the influence of internalized relationship experiences across generations (Emde et al., 2000, 2004).

Lessons from the litany in psychotherapy research. As noted above, Shonkoff (2003) has urged the field of early childhood intervention to move beyond asking *whether* early childhood intervention works to asking *how* its effects can be maximized. Others share this perspective: “Instead of the typical outcome question “Does the program work?”, a more important set of questions to ask is “How did the intervention best work? For whom? Under what circumstances?” (Korfmacher, 2002, p. 283). “Accordingly, the key question for evaluators and for consumers of evaluations—especially policy makers—may be not “did an intervention work?” but “for whom and under what circumstances are particular services most effective?”” (Berlin et al., 1998, p. 13).

These relatively recent recommendations from leading early childhood intervention researchers are reminiscent of exhortations to psychotherapy researchers that date back at least five decades:

From the point of view of science, the question ‘Does psychotherapy do any good?’ has little interest because it is virtually meaningless . . . The question is which people, in what circumstances, responding to what psychotherapeutic stimuli (Sanford, 1953, pp. 335-336).

. . . which therapist behaviors are most effective with which types of patients? (Kiesler, 1966, p. 113)

. . . which procedures and techniques, when used to accomplish what kinds of behavior change, are most effective with what kinds of clients when applied by what kind of counselor? (Krumboltz, 1966, p. 22)

What treatment, by whom, is most effective for this individual with that specific problem, under which set of circumstances? (Paul, 1967, p. 111, italics in original).

These “calls to arms” (Kiesler, 2004), especially the call from Gordon Paul (1967)—“one of the most quoted statements in the history of psychotherapy research” (Omer & Dar, 1992, p. 91) and eventually christened as the *litany* (Stiles, Shapiro, & Elliott, 1986)³⁵—became especially pronounced during the 1970s and influenced three decades of subsequent research. However, it eventually became all too clear that this atheoretical framework was fatally flawed. Over a decade ago, in 1992, Omer and Dar considered Paul’s “matrix of all possible combinations of treatments, therapists, clients, problems, and sets of circumstances” (Omer & Dar, 1992, p. 92). Based on a review of 252 studies published between 1967 and 1987 in a leading journal, they estimated that

the overall number of studies of this kind published in various journals in the last 2 decades reaches the count of thousands. Yet after all this labor, there is no clear agreement about even a single cell in this matrix. Clearly, a blind accumulation of comparisons, unrelated to or unanchored in theories, teaches us little. . . . Filling Paul’s matrix, therefore, is a Sisyphean task (Omer & Dar, 1992, p. 92).

This “flight from theory to pragmatics” in psychotherapy research (Omer & Dar, 1992) provides important lessons for early childhood intervention researchers as they now sound the same call to arms. Omer and Dar (1992) recommended that future research strive to balance practical relevance with theoretical relevance.³⁶ This “new research balance” promises to bridge research and real-world practice:

The best process-outcome studies combine sturdy clinical parameters with a detailed attention to questions of mechanism . . . clinical validity and theoretical relevance may not be mutually exclusive. . . . The challenge for modern research is whether theoretical studies can be kept close enough to clinical conditions and whether pragmatic clinical research can increase its theoretical relevance. At least two kinds of studies could provide the desired link between the two: (a) process-outcome studies combining credible outcome criteria with theory-guided process measures and (b) the almost abandoned case study with its possibilities for detailed investigations of mechanism (Omer & Dar, 1992, p. 92).

³⁵ Kiesler recently offered a more distilled version: “What package of which specific therapist miniature interventions, in what optimal sequence, applied by therapists demonstrating what level of expertise within what relational context, will successfully engage which proximate and ultimate mechanisms of patient change toward the alleviation of what specific pattern of DSM patient adjustment?” (Kiesler, 1995, p. 100; reiterated in Kiesler, 2004).

³⁶ As Omer and Dar (1992) note, like the *litany*, this debate between theoretical relevance and practical relevance also dates back at least five decades (see Brunswick, 1955; and Postman, 1955).

In this way, research frameworks that set theory-based parameters on a potentially infinite set of combinations are a necessity for future studies of early childhood intervention. The present process-outcome study has attempted to find a balance by applying theory-guided concepts and measures from early childhood intervention (Emde et al., 2000, 2004) and clinical intervention (Bordin, 1994; Horvath, 1994), anchored within a more general, theoretical framework for understanding what makes early childhood interventions work (Berlin et al., 1998), to a large-scale early childhood intervention program delivered under real-world conditions of practice.

In summary, future research must be theory-based, and would do well to (a) elucidate the direct and indirect effects of the alliance on parent and child development through process-outcome studies of relationship-based early childhood intervention (Emde et al., 2000, 2004) using the tripartite framework (Berlin et al., 1998); (b) examine the influence of home visitors' own personal qualities and interpersonal histories ("It is surprising—given how obviously crucial they are to any early intervention program—how little we know about what goes into good interventionists" Korfsmacher, 2002, p. 287) in the development of the working alliance with parents; (c) elaborate more fine-grained analysis of the moment-to-moment interpersonal process of the alliance (including goals, tasks, and bond), in order to improve training, service delivery, and supervision; (d) follow patterns of alliance continuity and change over time; and (e) monitor the longitudinal effects of these interventions across generations.

This proposed ensemble of studies is necessary to bridge the gaps between research, policy and practice (Shonkoff, 2000) in order to improve the real-world effectiveness of early childhood intervention. Psychotherapy has been defined as a complex human relationship in a specialized context (Henry & Strupp, 1994). The present study suggests that this interpersonal definition appears to apply as well to early childhood intervention. The personal qualities of its participants, and the relationship between participants, may be the most important ingredients of effective early childhood intervention.

References

- Administration on Children, Youth and Families. (2002, June). *Making a difference in the lives of infants and toddlers and their families: The impacts of Early Head Start. Vol. I: Final technical report*. Washington, DC: U.S. Department of Health and Human Services.
- Ainsworth, M. D. S. (1989). Attachments beyond infancy. *American Psychologist*, 44, 709-716.
- Albee, G. W., & Gullotta, T. P. (Eds.). (1997). *Primary prevention works*. Thousands Oaks, CA: Sage.
- Alper, J. (2002). The nurse home visitation program. In S. L. Isaacs & J. R. Knickman (Eds.), *To improve health and health care* (vol. 5). San Francisco: Jossey-Bass. Retrieved August 7, 2002, from <http://www.rwjf.org/publicationsPdfs/anthology2002/chapt1.htm>
- American Academy of Pediatrics Committee on Children with Disabilities. (2001). Role of the pediatrician in family-centred early intervention services. *Pediatrics*, 107, 1155-1157.
- American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (Rev. 3rd ed.). Washington, DC: Author.
- American Psychological Association. (2001). *Publication manual of the American Psychological Association* (5th ed.). Washington, DC: Author.
- Andersen, S. M., & Chen, S. (2002). The relational self: An interpersonal social-cognitive theory. *Psychological Review*, 109, 619-645.
- Anderson, L. M., Shinn, C., Fullilove, M. T., Scrimshaw, S. C., Fielding, J. E., Normand, J., Carande-Kulis, V. G., & the Task Force on Community Preventive Services. (2003). The effectiveness of early childhood development programs: A systematic review. *American Journal of Preventive Medicine*, 24(3S), 32-46.
- Armstrong, K. L., Fraser, J. A., Dadds, M. R., & Morris, J. (1999). A randomized, controlled trial of nurse home visiting to vulnerable families with newborns. *Journal of Paediatrics and Child Health*, 35, 237-244.
- Armstrong, K. L., Fraser, J. A., Dadds, M. R., & Morris, J. (2000). Promoting secure attachment, maternal mood and child health in a vulnerable population: A randomized controlled trial. *Journal of Paediatrics and Child Health*, 36, 555-562.
- Arnold, D. H., & Doctoroff, G. L. (2003). The early education of socioeconomically disadvantaged children. *Annual Review of Psychology*, 54, 517-545.
- Aspinwall, L. G., & Staudinger, U. M. (Eds.). (2003). *A psychology of human strengths: Fundamental questions and future directions for a positive psychology*. Washington, DC: American Psychological Association Press.
- Atkinson, L., Paglia, A., Coolbear, J., Niccols, A., Parker, K. C. H., & Guger, S. (2000). Attachment security: A meta-analysis of maternal mental health correlates. *Clinical Psychology Review*, 20, 1019-1040.

- Bakan, D. (1966). The test of significance in psychological research. *Psychological Bulletin, 66*, 423-437.
- Baker, A. J. L., Piotrkowski, C. S., & Brooks-Gunn, J. (1999). The Home Instruction Program for Preschool Youngsters (HIPPY). *The Future of Children, 9*, 116-133.
- Bakermans-Kranenburg, M. J., van IJzendoorn, M. H., & Juffer, F. (2003). Less is more: Meta-analyses of sensitivity and attachment interventions in early childhood. *Psychological Bulletin, 129*, 195-215.
- Baldwin, M. W. (1992). Relational schemas and the processing of social information. *Psychological Bulletin, 112*, 461-484.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*, 191-215.
- Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology, 52*, 1-26.
- Barnard, K. E., Magary, D., Sumner, G., Booth, C. L., Mitchell, S. K., & Spieker, S. (1988). Prevention of parenting alterations for women with low social support. *Psychiatry, 51*, 248-253.
- Barnes McGuire, J. (1997). The reliability and validity of a questionnaire describing neighborhood characteristics relevant to families and young children living in urban areas. *Journal of Community Psychology, 25*, 551-566.
- Barnet, B., Duggan, A. K., Devoe, M., & Burrell, L. (2002). The effect of volunteer home visitation for adolescent mothers on parenting and mental health outcomes. *Archives of Pediatric and Adolescent Medicine, 156*, 1216-1222.
- Barrera, M., Jr. (2000). Social support research in community psychology. In J. Rappaport & E. Seidman (Eds.), *Handbook of community psychology* (pp. 215-245). New York: Kluwer Academic/Plenum.
- Bates, J. E. (1984). *The infant characteristics questionnaire*. Bloomington, IN: Indiana University Department of Psychology.
- Bates, J. E., & Bayles, K. (1984). Objective and subjective components in mothers' perceptions of their children from age 6 months to 3 years. *Merrill-Palmer Quarterly, 30*, 111-130.
- Bates, J. E., Freeland, C. A. B., & Lounsbury, M. L. (1979). Measurement of infant difficultness. *Child Development, 50*, 794-803.
- Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology, 5*, 323-370.
- Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin, 117*, 497-529.

- Beardslee, W. R., Versage, E. M., & Gladstone, T. R. G. (1998). Children of affectively ill parents: A review of the past 10 years. *Journal of the American Academy of Child and Adolescent Psychiatry*, 37, 1134-1141.
- Beck, C. T. (1999). Maternal depression and child behaviour problems: A meta-analysis. *Journal of Advanced Nursing*, 29, 623-629.
- Becker, B. J. (1994). Combining significance levels. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 215-230). New York: Russell Sage Foundation.
- Beckwith, L. (1988). Intervention with disadvantaged parents of sick preterm infants. *Psychiatry*, 51, 242-247.
- Beckwith, L. (2000). Prevention science and prevention programs. In C. H. Zeanah, Jr. (Ed.), *Handbook of infant mental health* (pp. 439-456). New York: Guilford Press.
- Begg, C. (1994). Publication bias. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 399-409). New York: Russell Sage Foundation.
- Bell, R. Q. (1968). A reinterpretation of the direction of effects in studies of socialization. *Psychological Review*, 75, 81-95.
- Benard, B. (1999). Applications of resilience: Possibilities and promise. In M. D. Glantz & J. L. Johnson (Eds.) *Resilience and development: Positive life adaptations* (pp. 269-277). New York: Kluwer Academic/Plenum Publishers.
- Benjamin, L. S. (1974). Structural analysis of social behavior. *Psychological Review*, 81, 392-425.
- Berkman, L. F., & Glass, T. (2000). Social integration, social networks, social support, and health. In L. F. Berkman & I. Kawachi (Eds.), *Social epidemiology* (pp. 137-173). New York: Oxford University Press.
- Berkson, J. (1938). Some difficulties of interpretation encountered in the application of the chi-square test. *Journal of the American Statistical Association*, 33, 526-536.
- Berkson, J. (1942). Tests of significance considered as evidence. *Journal of the American Statistical Association*, 37, 325-335.
- Berlin, L. J., & Cassidy, J. (1999). Relations among relationships: Contributions from attachment theory and research. In J. Cassidy & P. R. Shaver (Eds.), *Handbook of attachment: Theory, research, and clinical applications* (pp. 688-712). New York: Guilford Press.
- Berlin, L. J., O'Neal, C. R., & Brooks-Gunn, J. (1998). What makes early intervention programs work? The program, its participants, and their interaction. *Zero to Three*, 18, 4-15.
- Berrueta-Clement, J. R., Schweinhart, L. J., Barnett, W. S., Epstein, A. S., & Weikart, D. P. (1984). *Changed lives: The effects of the Perry Preschool Program on youths through age 19* (Monographs of the High/Scope Educational Research Foundation, no. 8). Ypsilanti, MI: High/Scope Press.

- Berscheid, E. (1994). Interpersonal relationships. *Annual Review of Psychology, 45*, 79-129.
- Berscheid, E. (1999). The greening of relationship science. *American Psychologist, 54*, 260-266.
- Berscheid, E., & Reis, H. T. (1998). Attraction and close relationships. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology: Vol. 2* (4th ed., pp. 193-281). New York: McGraw-Hill.
- Black, M. M. (2000). Prevention: For infants and young children. In A. E. Kazdin (Ed.), *Encyclopedia of psychology: Vol. 6* (pp. 290-293). American Psychological Association and Oxford University Press.
- Bordin, E. S. (1975, September). *The working alliance: Basis for a general theory of psychotherapy*. Paper presented at the annual meeting of the American Psychological Association. Washington, DC.
- Bordin, E. S. (1979). The generalizability of the psychoanalytic concept of the working alliance. *Psychotherapy: Theory, Research and Practice, 16*, 252-260.
- Bordin, E. S. (1994). Theory and research on the therapeutic working alliance: New directions. In A. O. Horvath & L. S. Greenberg (Eds.), *The working alliance: Theory, research, and practice* (pp. 13-37). New York: John Wiley & Sons.
- Borenstein, M., Rothstein, H., & Cohen, J. (2001). *SamplePower 2.0* [Computer manual]. Chicago: SPSS Inc.
- Boring, E. G. (1919). Mathematical vs. scientific importance. *Psychological Bulletin, 16*, 335-338.
- Bowlby, J. (1969). *Attachment and loss: Vol. 1: Attachment*. New York: Basic Books.
- Bowlby, J. (1988). *A secure base*. New York: Basic Books.
- Bradley, R. H., & Corwyn, R. F. (2002). Socioeconomic status and child development. *Annual Review of Psychology, 53*, 371-399.
- Bringing up baby – better. (2002, Winter). *Harvard Public Health Review*. Retrieved August 25, 2002, from <http://www.hsph.harvard.edu/review/677baby.html>
- Brink, S., & McKellar, S. (2000). NLSCY: A unique Canadian survey. *ISUMA: Canadian Journal of Policy Research, 1(2)*, 111-113.
- Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Cambridge, MA: Harvard University Press.
- Brooks-Gunn, J. (2000, May). *Do you believe in magic? What we can expect from early childhood intervention programs*. Paper presented at a research briefing sponsored by the Subcommittee on Human Resources of the U.S. House of Representatives Committee on Ways and Means and the Joint Center for Poverty Research (University of Chicago and Northwestern University), Washington, DC.

- Brooks-Gunn, J. (2003). Do you believe in magic? What we can expect from early childhood intervention programs. *Social Policy Report*, 18, 1-16.
- Brooks-Gunn, J., Berlin, L. J., & Fuligni, A. S. (2000). Early childhood intervention programs: What about the family? In J. P. Shonkoff & S. J. Meisels (Eds.), *Handbook of early childhood intervention* (2nd ed., pp. 549-588). New York: Cambridge University Press.
- Brooks-Gunn, J., Berlin, L. J., Leventhal, T., & Fuligni, A. S. (2000). Depending on the kindness of strangers: Current national data initiatives and developmental research. *Child Development*, 71, 257-268.
- Brooks-Gunn, J., McCarton, C., Casey, P., McCormick, M., Bauer, C., Bernbaum, J., Tyson, J., Swanson, M., Bennett, F., Scott, D., Tonascia, J., & Meinert, C. (1994). Early intervention in low birthweight, premature infants: Results through age 5 years from the Infant Health and Development Program. *Journal of the American Medical Association*, 272, 1257-1262.
- Brooks-Gunn, J., Phelps, E., & Elder, G. H. (1991). Studying lives through time: Secondary data analyses in developmental psychology. *Developmental Psychology*, 27, 899-910.
- Brunswick, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, 62, 193-217.
- Bryant, D. (2003, March 17). *Are we at home yet with home visiting?* Paper presented at the 35th Annual Banff International Conference on Behavioural Science, Banff, Alberta, Canada.
- Bugental, D. B., Ellerson, P. C., Lin, E. K., Rainey, B., Kokotovic, A., & O'Hara, N. (2002). A cognitive approach to child abuse prevention. *Journal of Family Psychology*, 16, 243-258.
- Bugental, D. B., & Happaney, K. (2004). Predicting infant maltreatment in low-income families: The interactive effects of maternal attributions and child status at birth. *Developmental Psychology*, 40, 234-243.
- Bugental, D. B., & Johnston, C. (2000). Parental and child cognitions in the context of the family. *Annual Review of Psychology*, 51, 315-344.
- Bus, A. G., van IJzendoorn, M. H., & Pellegrini, A. (1995). Joint book reading makes for success in learning to read: A meta-analysis on the intergenerational transmission of literacy. *Review of Educational Research*, 65, 1-21.
- Busseri, M. A., & Tyler, J. D. (2003). Interchangeability of the Working Alliance Inventory and Working Alliance Inventory, Short Form. *Psychological Assessment*, 15, 193-197.
- Butler, S. F., & Strupp, H. H. (1986). Specific and nonspecific factors in psychotherapy: A problematic paradigm for psychotherapy research. *Psychotherapy*, 23, 30-40.
- Campbell, D. T. (1969). Reforms as experiments. *American Psychologist*, 24, 409-429.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81-105.

- Campbell, D. T., & Stanley, J. C. (1966). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.
- Campbell, F. A., Pungello, E. P., Miller-Johnson, S., Burchinal, M., & Ramey, C. T. (2001). The development of cognitive and academic abilities: Growth curves from an early childhood educational experiment. *Developmental Psychology, 37*, 231-242.
- Campbell, F. A., & Ramey, C. T. (1994). Effects of early intervention on intellectual and academic achievement: A follow-up study of children from low-income families. *Child Development, 65*, 684-698.
- Campbell, F. A., & Ramey, C. T. (1995). Cognitive and school outcomes for high-risk African American students into middle adolescence: Positive effects of early intervention. *American Educational Research Journal, 32*, 743-772.
- Campbell, F. A., Ramey, C. T., Pungello, E., Sparling, J., & Miller-Johnson, S. (2002). Early childhood education: Outcomes as a function of different treatments. *Applied Developmental Science, 6*, 42-57.
- Canadian Institute of Child Health. (2000). *The health of Canada's children: A CICH profile* (3rd ed.). Ottawa, ON: Author.
- Capaldi, D. M., Conger, R. D., Hops, H., & Thornberry, T. P. (2003). Introduction to special section on three-generation studies. *Journal of Abnormal Child Psychology, 31*, 123-125.
- Capaldi, D. M., Pears, K. C., Patterson, G. R., & Owen, L. D. (2003). Continuity of parenting practices across generations in an at-risk sample: A prospective comparison of direct and mediated associations. *Journal of Abnormal Child Psychology, 31*, 127-142.
- Caplan, G., & Caplan, R. B. (2000). The future of primary prevention. *Journal of Primary Prevention, 21*, 131-136.
- Cappelleri, J. C., Darlington, R. B., & Trochim, W. M. K. (1994). Power analysis of cutoff-based randomized clinical trials. *Evaluation Review, 18*, 141-152.
- Cappelleri, J. C., & Trochim, W. M. K. (2003). Cutoff designs. In S. C. Chow (Ed.), *Encyclopedia of biopharmaceutical statistics* (2nd ed., pp. 263-269). New York: Marcel Dekker.
- Carver, R. P. (1978). The case against significance testing. *Harvard Educational Review, 48*, 378-399.
- Caspi, A. (1998). Personality development across the life course. In W. Damon (Series Ed.) & N. Eisenberg (Vol. Ed.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (5th ed., pp. 311-388). New York: John Wiley & Sons.

- Caspi, A., Moffitt, T. E., Morgan, J., Rutter, M., Taylor, A., Arsenault, L., Tully, L., Jacobs, C., Kim-Cohen, J., & Polo-Tomas, M. (2004). Maternal expressed emotion predicts children's antisocial behavior problems: Using monozygotic-twin differences to identify environmental effects on behavioral development. *Developmental Psychology, 40*, 149-161.
- Catalano, R. F., Berglund, M. L., Ryan, J. A. M., Lonczak, H. S., & Hawkins, J. D. (2002, June 24). Positive youth development in the United States: Research findings on evaluations of positive youth development programs. *Prevention and Treatment, 5*, Article 15. Retrieved July 16, 2002, from <http://journals.apa.org/prevention/volume5/pre0050015a.html>
- Chaffin, M. (2004). Is it time to rethink Healthy Start/Healthy Families? *Child Abuse and Neglect, 28*, 589-595.
- Chen, E., Matthews, K. A., & Boyce, W. T. (2002). Socioeconomic differences in children's health: How and why do these relationships change with age? *Psychological Bulletin, 128*, 295-329.
- Clarke, P. J., Marshall, V. W., Ryff, C. D., & Rosenthal, C. J. (2000). Well-being in Canadian seniors: Findings from the Canadian Study of Health and Aging. *Canadian Journal on Aging, 19*, 139-159.
- Clarke, P. S. (2004). Causal analysis of individual change using the difference score. *Epidemiology, 15*, 414-421.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Cohen, J. (1990). Things I have learned (so far). *American Psychologist, 45*, 1304-1312.
- Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159.
- Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist, 49*, 997-1003.
- Cohen, S. (2004). Social relationships and health. *American Psychologist, 59*, 676-684.
- Collins, L. M., & Horn, J. L. (1991). *Best methods for the analysis of change: Recent advances, unanswered questions, future directions*. Washington, DC: American Psychological Association.
- Collins, L. M., Murphy, S. A., Nair, V. N., & Strecher, V. J. (In press). A strategy for optimizing and evaluating behavioral interventions. *Annals of Behavioral Medicine*.
- Collins, L. M., & Sayer, A. G. (Eds.). (2001). *New methods for the analysis of change*. Washington, DC: American Psychological Association.
- Collins, N. L., & Feeney, B. C. (2004). Working models of attachment shape perceptions of social support: Evidence from experimental and observational studies. *Journal of Personality and Social Psychology, 87*, 363-383.

- Collins, W. A., Maccoby, E. E., Steinberg, L., Hetherington, E. M., & Bornstein, M. H. (2000). Contemporary research on parenting: The case for nature and nurture. *American Psychologist*, 55, 218-232.
- Conger, R. D., Neppl, T., Kim, K. J., & Scaramella, L. (2003). Angry and aggressive behavior across three generations: A prospective, longitudinal study of parents and children. *Journal of Abnormal Child Psychology*, 31, 143-160.
- Connell, A. M., & Goodman, S. H. (2002). The association between psychopathology in fathers versus mothers and children's internalizing and externalizing behavior problems: A meta-analysis. *Psychological Bulletin*, 128, 746-773.
- Cook, C., & Willms, J. D. (2002). Balancing work and family life. In J. D. Willms (Ed.), *Vulnerable children: Findings from Canada's National Longitudinal Survey of Children and Youth* (pp. 183-197). Edmonton, AB: University of Alberta Press and Human Resources Development Canada-Applied Research Branch.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. New York: Springer-Verlag.
- Cook, T. D., & Shadish, W. R. (1994). Social experiments: Some developments over the past fifteen years. *Annual Review of Psychology*, 45, 545-580.
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78, 98-104.
- Cowan, P. A., Powell, D., & Cowan, C. P. (1998). Parenting interventions: A family systems perspective. In W. Damon (Series Ed.) & I. E. Siegel & K. A. Renninger (Vol. Eds.), *Handbook of child psychology: Vol. 4. Child psychology in practice* (5th ed., pp. 3-72). New York: John Wiley & Sons.
- Cowen, E. L. (1997). The coming of age of primary prevention: Comments on Durlak and Wells' meta-analysis. *American Journal of Community Psychology*, 25, 153-167.
- Cowen, E. L. (2000a). Community psychology and routes to psychological wellness. In J. Rappaport & E. Seidman (Eds.), *Handbook of community psychology* (pp. 79-99). New York: Kluwer Academic/Plenum Publishers.
- Cowen, E. L. (2000b). Now that we all know that primary prevention in mental health is great, what is it? *Journal of Community Psychology*, 28, 5-16.
- Coyne, J. C., Thompson, R., Klinkman, M. S., & Nease, D. E., Jr. (2002). Emotional disorders in primary care. *Journal of Consulting and Clinical Psychology*, 70, 798-809.
- Crockenberg, S. C., & Leerkes, E. M. (2003). Parental acceptance, postpartum depression, and maternal sensitivity: Mediating and moderating processes. *Journal of Family Psychology*, 17, 80-93.
- Cronbach, L. J., & Furby, L. (1970). How should we measure "change"—or should we? *Psychological Bulletin*, 74, 68-80.

- Culp, A. M., Culp, R. E., Hechtner-Galvin, T., Howell, C. S., Saathoff-Wells, T., & Marr, P. (2004). First-time mothers in home visitation services utilizing child development specialists. *Infant Mental Health Journal*, 25, 1-15.
- Cutrona, C. E. (1986). Objective determinants of perceived social support. *Journal of Personality and Social Psychology*, 50, 349-355.
- Cutrona, C. E., & Russell, D. C. (1987). The provisions of social relationships and adaptation to stress. In W. H. Jones & D. Perlman (Eds.), *Advances in personal relationships* (pp. 33-67). Greenwich, CT: JAI Press.
- Dadds, M. R. (2001, March 30). Fads, politics, and research: Keeping prevention on the mental health agenda. *Prevention & Treatment*, 4, Article 6. Retrieved July 16, 2002, from <http://journals.apa.org/prevention/volume4/pre0040006c.html>
- Dar, R. (1987). Another look at Meehl, Lakatos, and the scientific practices of psychologists. *American Psychologist*, 42, 145-151.
- Dar, R., Serlin, R. C., & Omer, H. (1994). Misuse of statistical tests in three decades of psychotherapy research. *Journal of Consulting and Clinical Psychology*, 62, 75-82.
- Daro, D., & Donnelly, A. C. (2002). Charting the waves of prevention: Two steps forward, one step back. *Child Abuse and Neglect*, 26, 731-742.
- Daro, D. A., & Harding, K. A. (1999). Healthy Families America: Using research to enhance practice. *The Future of Children*, 9, 152-176.
- Daro, D., McCurdy, K., Falconnier, L., & Stojanovic, D. (2003). Sustaining new parents in home visitation services: Key participant and program factors. *Child Abuse and Neglect*, 27, 1101-1125.
- DeBaryshe, B. D. (1992). *Early language and literacy activities in the home: Final report*. Washington, DC: U.S. Department of Education Field Initiated Studies Program.
- Denis, D. J. (2003). Alternatives to null hypothesis significance testing. *Theory and Science*, 4(1). Retrieved December 5, 2004, from: http://theoryandscience.icaap.org/content/vol4.1/02_denis.html
- Dishion, T. J., McCord, J., & Poulin, F. (1999). When interventions harm: Peer groups and problem behavior. *American Psychologist*, 54, 755-764.
- Dryfoos, J. G. (1998). *Safe passage: Making it through adolescence in a risky society: What parents, school, and communities can do*. New York: Oxford University Press.
- Dubow, E. F., & Huessmann, L. R., & Boxer, P. (2003). Theoretical and methodological considerations in cross-generational research on parenting and child aggressive behavior. *Journal of Abnormal Child Psychology*, 31, 185-192.
- Duggan, A., Fuddy, L., Burrell, L., Higman, S. M., McFarlane, E., Windham, A., & Sia, C. (2004). Randomized trial of a statewide home visiting program to prevent child abuse: Impact in reducing parental risk factors. *Child Abuse and Neglect*, 28, 623-643.

- Duggan, A., McFarlane, E., Fuddy, L., Burrell, L., Higman, S. M., Windham, A., & Sia, C. (2004). Randomized trial of a statewide home visiting program: Impact in preventing child abuse and neglect. *Child Abuse and Neglect, 28*, 597-622.
- Duggan, A. K., McFarlane, E. C., Windham, A. M., Rohde, C. A., Salkever, D. S., Fuddy, L., Rosenberg, L. A., Buchbinder, S. B., & Sia, C. C. J. (1999). Evaluation of Hawaii's Healthy Start program. *The Future of Children, 9*, 66-90.
- Duggan, A., Windham, A., McFarlane, E., Fuddy, L., Rohde, C., Buchbinder, S., & Sia, C. (2000). Hawaii's Healthy Start program of home visiting for at-risk families: Evaluation of family identification, family engagement, and service delivery. *Pediatrics, 105*, 250-259.
- Duncan, G. J., & Brooks-Gunn, J. (2000). Family poverty, welfare reform, and child development. *Child Development, 71*, 188-196.
- Durlak, J. A. (2003, December 18). The long-term impact of preschool prevention programs: A commentary. *Prevention & Treatment, 6*, Article 32. Retrieved February 16, 2004, from <http://journals.apa.org/prevention/volume6/pre0060032c.html>
- Durlak, J. A., & Wells, A. M. (1997). Primary prevention mental health programs for children and adolescents: A meta-analytic review. *American Journal of Community Psychology, 25*, 115-152.
- Earls, F. (1998). Positive effects of prenatal and early childhood intervention. *Journal of the American Medical Association, 280*, 1271-1273.
- Earls, F., & Carlson, M. (2001). The social ecology of child health and well-being. *Annual Review of Public Health, 22*, 143-166.
- Eckenrode, J., Ganzel, B., Henderson, C. R., Jr., Smith, E., Olds, D. L., Powers, J., Cole, R., Kitzman, H., & Sidora, K. (2000). Preventing child abuse and neglect with a program of nurse home visitation: The limiting effects of domestic violence. *Journal of the American Medical Association, 284*, 1385-1391.
- Emde, R. N. (1988a). Research forum: Preventive intervention and early development. *Psychiatry, 51*, 235.
- Emde, R. N. (1988b). Risk, intervention and meaning. *Psychiatry, 51*, 254-259.
- Emde, R.N., Everhart, K.D., & Wise, B. K. (2004). Therapeutic relationships in infant mental health and the concept of leverage. In A.J. Sameroff, S.C. McDonough, K.L. Rosenblum (Eds.) *Treating parent-infant relationship problems* (pp. 267-292). New York: Guilford.
- Emde, R. N., Korfmacher, J., & Kubicek, L. F. (2000). Toward a theory of early relationship-based intervention. In J. D. Osofsky & H. E. Fitzgerald (Eds.), *WAIMH handbook of infant mental health: Vol. 2: Early intervention, evaluation, and assessment* (pp. 1-32). New York: John Wiley & Sons.

- Emde, R. N., & Robinson, J. (2000). Guiding principles for a theory of early intervention: A developmental-psychoanalytic perspective. In J. P. Shonkoff & S. J. Meisels (Eds.), *Handbook of early childhood intervention* (2nd ed., pp. 160-178). New York: Cambridge University Press.
- Evans, G. W. (2004). The environment of childhood poverty. *American Psychologist*, 59, 77-92.
- Farran, D. C. (2000). Another decade of early intervention for children who are low income or disabled: What do we know now? In J. P. Shonkoff & S. J. Meisels (Eds.), *Handbook of early childhood intervention* (2nd ed., pp. 510-548). New York: Cambridge University Press.
- Fechner-Bates, S., Coyne, J. C., & Schwenk, T. L. (1994). The relationship of self-reported distress to depressive disorders and other psychopathology. *Journal of Consulting and Clinical Psychology*, 62, 550-559.
- Field, T., Hernandez-Reif, M., & Freedman, J. (2004). Stimulation programs for preterm infants. *Social Policy Report*, 18, 1-20.
- Flay, B. R. (1986). Efficacy and effectiveness trials (and other phases of research) in the development of health promotion programs. *Preventive Medicine*, 15, 451-474.
- Fraser, J. A., Armstrong, K. L., Morris, J. P., & Dadds, M. R. (2000). Home visiting intervention for vulnerable families with newborns: Follow-up results of a randomized controlled trial. *Child Abuse and Neglect*, 24, 1399-1429.
- Gelso, C. J. (2000). Working alliance. In A. E. Kazdin (Ed.), *Encyclopedia of psychology: Vol. 8* (pp. 274-276). American Psychological Association and Oxford University Press.
- Gillham, J. E., Shatté, A. J., & Reivich, K. (2001, March 30). Needed for prevention research: Long-term follow up and the evaluation of mediators, moderators, and lay providers. *Prevention & Treatment*, 4, Article 9. Retrieved July 16, 2002, from <http://journals.apa.org/prevention/volume4/pre0040009c.html>
- Gladstone, T. R. G., & Beardslee, W. R. (2002). Treatment, intervention, and prevention with children of depressed parents: A developmental perspective. In S. H. Goodman & I. H. Gotlib (Eds.), *Children of depressed parents: Mechanisms of risk and implications for treatment* (pp. 277-305). Washington, DC: American Psychological Association.
- Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, 10, 3-8.
- Glass, T. A. (2000). Psychosocial intervention. In L. F. Berkman & I. Kawachi (Eds.), *Social epidemiology* (pp. 267-305). New York: Oxford University Press.
- Goldfried, M. R. (2000). Editorial: Psychotherapy in the twenty-first century. *Journal of Clinical Psychology/In Session: Psychotherapy in Practice*, 56, 157-158.
- Gomby, D. S. (1999). Understanding evaluations of home visiting programs. *The Future of Children*, 9, 27-43.

- Gomby, D. S. (2000). Promise and limitations of home visitation. *Journal of the American Medical Association, 284*, 1430-1431.
- Gomby, D. S. (2003, January 29). Meta-analyses, literature reviews, and recent studies of home visiting programs. In First 5 California Children and Families Commission, *Building school readiness through home visitation* (Appendix B). Sacramento, CA: First 5 California Children and Families Commission.
- Gomby, D. S., Culross, P. L., & Behrman, R. E. (1999). Home visiting: Recent program evaluations—analysis and recommendations. *The Future of Children, 9*, 4-26.
- Gomby, D. S., Larson, C. S., Lewit, E. M., & Behrman, R. E. (1993). Home visiting: Analysis and recommendations. *The Future of Children, 3*, 6-22.
- Goodman, S. H., & Gotlib, I. H. (1999). Risk for psychopathology in the children of depressed mothers: A developmental model for understanding mechanisms of transmission. *Psychological Review, 106*, 458-490.
- Goodman, S. H., & Gotlib, I. H. (Eds.). (2002). *Children of depressed parents: Mechanisms of risk and implications for treatment*. Washington, DC: American Psychological Association.
- Gopnik, A., Meltzoff, A. N., & Kuhl, P. K. (1999). *The scientist in the crib: Minds, brains, and how children learn*. New York: William Morrow & Company.
- Gottman, J. M., & Rushe, R. H. (1993). The analysis of change: Issues, fallacies, and new ideas. *Journal of Consulting and Clinical Psychology, 61*, 907-910.
- Government of Canada. (2003). *Early childhood development activities and expenditures: Government of Canada report 2002-2003*. Ottawa, ON: Human Resources Development Canada, Health Canada, and Indian and Northern Affairs Canada.
- Government of Manitoba. (1998). *Children and Youth Secretariat: Annual report 1997 - 1998*. Winnipeg, MB: Author.
- Government of Manitoba. (2001). *Healthy Child Manitoba: Annual report 2000 - 2001*. Winnipeg, MB: Author.
- Government of Manitoba. (2004). *Healthy Child Manitoba: Annual report 2003 - 2004*. Winnipeg, MB: Author.
- Grant given to study low birth-weight children as teens. (2001, October 19). *Harvard Public Health NOW*. Retrieved August 25, 2002, from <http://www.hsph.harvard.edu/now/oct19/ats.html>
- Grant, K. E., Compas, B. E., Stuhlmacher, A. F., Thurm, A. E., McMahon, S. D., & Halpert, J. A. (2003). Stressors and child and adolescent psychopathology: Moving from markers to mechanisms of risk. *Psychological Bulletin, 129*, 447-466.

- Great Kids, Inc. (2000). *Growing great kids: Prenatal to 36 months: An interactive parenting and child development curriculum and a staff development program*. San Angelo, TX: Author.
- Greenberg, M. T., Domitrovich, C., & Bumbarger, B. (2001, March 30). The prevention of mental disorders in school-aged children: Current state of the field. *Prevention & Treatment, 4*, Article 1. Retrieved July 16, 2002, from <http://journals.apa.org/prevention/volume4/pre0040001a.html>
- Greenberg, M. T., & Weissberg, R. (2001, June 26). Commentary on "Priorities for prevention research at NIMH." *Prevention & Treatment, 4*, Article 25. Retrieved July 16, 2002, from <http://journals.apa.org/prevention/volume4/pre00400025c.html>
- Greenberg, M. T., Weissberg, R. P., O'Brien, M. U., Zins, J. E., Fredericks, L., Resnik, H., & Elias, M. J. (2003). Enhancing school-based prevention and youth development through coordinated social, emotional, and academic learning. *American Psychologist, 58*, 466-474.
- Greenhouse, J. B., & Iyengar, S. (1994). Sensitivity analysis and diagnostics. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 383-398). New York: Russell Sage Foundation.
- Guralnick, M. J. (Ed.). (1997). *The effectiveness of early intervention*. Baltimore, MD: Paul H. Brookes Publishing.
- Hahn, R. A., Bilukha, O. O., Crosby, A., Fullilove, M. T., Liberman, A., Moscicki, E. K., Snyder, S., Tuma, F., Schofield, A., Corso, P. S., & Briss, P. (2003, October 3). First reports evaluating the effectiveness of strategies for preventing violence: Early childhood home visitation. Findings from the Task Force on Community Preventive Services. *Morbidity and Mortality Weekly Report, 52*, 1-9.
- Hallett, B., Nemeth, M., Stevens, H., & Stewart, D. (2000). *Aboriginal people in Manitoba 2000*. Winnipeg, MB: Human Resources Development Canada, Manitoba Family Services and Housing, and Manitoba Aboriginal Affairs Secretariat.
- Halpern, R. (2000). Early intervention for low-income children and families. In J. P. Shonkoff & S. J. Meisels (Eds.), *Handbook of early childhood intervention* (2nd ed., pp. 361-386). New York: Cambridge University Press.
- Hamilton, C. E. (2000). Continuity and discontinuity of attachment from infancy through adolescence. *Child Development, 71*, 690-694.
- Harlow, L. L., Mulaik, S. A., & Steiger, J. H. (Eds.). (1997). *What if there were no significance tests?* Mahwah, NJ: Lawrence Erlbaum Associates.
- Healthy Child Manitoba. (2000). *BabyFirst support manual*. Winnipeg, MB: Author.
- Heinicke, C. M., Goorsky, M., Moscov, S., Dudley, K., Gordon, J., Schneider, C., & Guthrie, D. (2000). Relationship-based intervention with at-risk mothers: Factors affecting variations in outcome. *Infant Mental Health Journal, 21*, 133-155.

- Henry, W. P., & Strupp, H. H. (1994). The therapeutic alliance as interpersonal process. In A. O. Horvath & L. S. Greenberg (Eds.), *The working alliance: Theory, research, and practice* (pp. 51-84). New York: John Wiley & Sons.
- Henry, W. P., Strupp, H. H., Schacht, T. E., & Gaston, L. (1994). Psychodynamic approaches. In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change* (4th ed., pp. 467-508). New York: Wiley.
- Hiatt, S. W. (2000). Editorial [Special issue: Henry Kempe's Center: Past, present, future]. *Child Abuse and Neglect*, 24, 5-7.
- Hill, N. E., & Herman-Stahl, M. A. (2002). Neighborhood safety and social involvement: Associations with parenting behaviors and depressive symptoms among African American and Euro-American mothers. *Journal of Family Psychology*, 16, 209-219.
- Hilliard, R. B., Henry, W. P., & Strupp, H. H. (2000). An interpersonal model of psychotherapy: Linking patient and therapist developmental history, therapeutic process, and types of outcome. *Journal of Consulting and Clinical Psychology*, 68, 125-133.
- Hoagwood, K., Hibbs, E., Brent, D., & Jensen, P. (1995). Introduction to the special section: Efficacy and effectiveness in studies of child and adolescent psychotherapy. *Journal of Consulting and Clinical Psychology*, 63, 683-687.
- Hogan, B. E., Linden, W., & Najarian, B. (2002). Social support interventions: Do they work? *Clinical Psychology Review*, 22, 381-440.
- Hogben, L. (1957). *Statistical theory*. London: Allen & Unwin.
- Hohmann, A. A., & Shear, M. K. (2002). Community-based intervention research: Coping with the "noise" of real life in study designs. *American Journal of Psychiatry*, 159, 201-207.
- Hops, H., Davis, B., Leve, C., & Sheeber, L. (2003). Cross-generational transmission of aggressive parent behavior: A prospective, mediational examination. *Journal of Abnormal Child Psychology*, 31, 161-169.
- Horvath, A. O. (1981). *An exploratory study of the working alliance: Its measurement and relationship to outcome*. Unpublished doctoral dissertation. University of British Columbia, Vancouver, BC.
- Horvath, A. O. (1994). Empirical validation of Bordin's pantheoretical model of the alliance: The Working Alliance Inventory perspective. In A. O. Horvath & L. S. Greenberg (Eds.), *The working alliance: Theory, research, and practice* (pp. 109-128). New York: John Wiley & Sons.
- Horvath, A. O. (2000). The therapeutic relationship: From transference to alliance. *Journal of Clinical Psychology/In Session: Psychotherapy in Practice*, 56, 163-173.
- Horvath, A. O. (2001). The alliance. *Psychotherapy*, 38, 365-372.

- Horvath, A. O., & Bedi, R. P. (2002). The alliance. In J. C. Norcross (Ed.), *Psychotherapy relationships that work: Therapist contributions and responsiveness to patients* (pp. 37-69). New York: Oxford University Press.
- Horvath, A. O., & Greenberg, L. S. (1986). The development of the Working Alliance Inventory. In L. S. Greenberg & W. M. Pinsof (Eds.), *The psychotherapeutic process: A research handbook* (pp. 529-556). New York: Guilford Press.
- Horvath, A. O., & Greenberg, L. S. (1989). The development and validation of the Working Alliance Inventory. *Journal of Counseling Psychology, 36*, 223-233.
- Horvath, A. O., & Greenberg, L. S. (Eds.). (1994). *The working alliance: Theory, research, and practice*. New York: John Wiley & Sons.
- Horvath, A. O., & Symonds, B. D. (1991). Relation between working alliance and outcome in psychotherapy: A meta-analysis. *Journal of Counseling Psychology, 38*, 139-149.
- House, J. S., Landis, K. R., & Umberson, D. (1988). Social relationships and health. *Science, 241*, 540-545.
- Hubert, N. C., Wachs, T. D., Peters-Martin, P., & Gandour, M. J. (1982). The study of early temperament: Measurement and conceptual issues. *Child Development, 53*, 571-600.
- Human Resources Development Canada & Statistics Canada. (1996). *Growing up in Canada: National Longitudinal Survey of Children and Youth*. Ottawa, ON: Statistics Canada.
- Hunter, J. E., & Schmidt, F. L. (1996). Cumulative research knowledge and social policy formulation: The critical role of meta-analysis. *Psychology, Public Policy, and Law, 2*, 324-347.
- Huston, A. C. (1994). Children in poverty: Designing research to affect policy. *Social Policy Report, 8*, 1-12.
- Huston, A. C. (2002). From research to policy: Choosing questions and interpreting the answers. *New Directions for Child and Adolescent Development, 98*, 29-41.
- Huston, A. C., Duncan, G. J., Granger, R., Bos, J., McLoyd, V., Mistry, R., Crosby, D., Gibson, C., Magnuson, K., Romich, J., & Ventura, A. (2001). Work-based antipoverty programs for parents can enhance the school performance and social behavior of children. *Child Development, 72*, 318-336.
- Huston, A. C., Miller, C., Richburg-Hayes, L., Duncan, G. J., Eldred, C. A., Weisner, T. S., Lowe, E., McLoyd, V. C., Crosby, D. A., Ripke, M. N., & Redcross, C. (2003, June). *New hope for families and children: Five-year results of a program to reduce poverty and reform welfare*. New York: MDRC.
- Infant Health and Development Program. (1990). Enhancing the outcomes of low-birth-weight, premature infants: A multisite, randomized trial. *Journal of the American Medical Association, 253*, 3035-3042.

- Jaffee, S. R., Caspi, A., Moffitt, T. E., Polo-Tomas, M., Price, T. S., & Taylor, A. (2004). The limits of child effects: Evidence for genetically mediated child effects on corporal punishment but not on physical maltreatment. *Developmental Psychology, 40*, 1047-1058.
- Japel, C., Normand, C., Tremblay, R., & Willms, J. D. (2002). Identifying vulnerable children at an early age. In J. D. Willms (Ed.), *Vulnerable children: Findings from Canada's National Longitudinal Survey of Children and Youth* (pp. 105-119). Edmonton, AB: University of Alberta Press and Human Resources Development Canada-Applied Research Branch.
- Judd, C. M., & Kenny, D. A. (1981). *Estimating the effects of social interventions*. Cambridge, MA: Cambridge University Press.
- Karoly, L. A., Greenwood, P. W., Everingham, S. S., Houbé, J., Kilburn, M. R., Rydell, C. P., Sanders, M., & Chiesa, J. (1998). *Investing in our children: What we know and don't know about the costs and benefits of early childhood interventions*. Santa Monica, CA: RAND.
- Kaufman, A. S. (1998). Introduction to the special issue on statistical significance testing. *Research in the Schools, 5*, 1.
- Kempe, H. (1976). *Child abuse and neglect: The family and the community*. Cambridge, MA: Ballinger Publishing Company.
- Keyes, C. L. M., & Haidt, J. (Eds.). (2003). *Flourishing: Positive psychology and the life well-lived*. Washington, DC: American Psychological Association Press.
- Keyes, C. L. M., Shmotkin, D., & Ryff, C.D. (2002). Optimizing well-being: The empirical encounter of two traditions. *Journal of Personality and Social Psychology, 82*, 1007-1022.
- Kiesler, D. J. (1966). Some myths of psychotherapy research and the search for a paradigm. *Psychological Bulletin, 65*, 110-136.
- Kiesler, D. J. (1995). Research classic: "Some myths of psychotherapy research and the search for a paradigm": Revisited. *Psychotherapy Research, 5*, 91-101.
- Kiesler, D. J. (2004). Intrepid pursuits of the essential ingredients of psychotherapy. *Clinical Psychology: Science and Practice, 11*, 391-395.
- Kirk, R. E. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement, 56*, 746-759.
- Kitzman, H. (2004). Effective early childhood development programs for low-income families: Home visiting interventions during pregnancy and early childhood. In R. E. Tremblay, R. G. Barr, & R. DeV. Peters (Eds.), *Encyclopedia on early childhood development* [online]. Montreal, QC: Centre of Excellence for Early Childhood Development. Retrieved November 5, 2004, from <http://www.excellence-earlychildhood.ca/documents/KitzmanANGxp.pdf>

- Kitzman, H., Olds, D. L., Henderson, C. R., Jr., Hanks, C., Cole, R., Tatelbaum, R., McConnochie, K. M., Sidora, K., Luckey, D. W., Shaver, D., Engelhardt, K., James, D., & Barnard, K. (1997). Effect of prenatal and infancy home visitation by nurses on pregnancy outcomes, childhood injuries, and repeated childbearing: A randomized controlled trial. *Journal of the American Medical Association*, 278, 644-652.
- Kitzman, H., Olds, D. L., Sidora, K., Henderson, C. R., Jr., Hanks, C., Cole, R., Luckey, D. W., Bondy, J., Cole, K., & Glazner, J. (2000). Enduring effects of nurse home visitation on maternal life course: A 3-year follow-up of a randomized trial. *Journal of the American Medical Association*, 283, 1983-1989.
- Kivlighan, D. M., Jr., & Shaughnessy, P. (1995). Analysis of the development of the working alliance using hierarchical linear modeling. *Journal of Counseling Psychology*, 42, 338-349.
- Kivlighan, D. M., Jr., & Shaughnessy, P. (2000). Patterns of working alliance development: A typology of client's working alliance ratings. *Journal of Counseling Psychology*, 47, 362-371.
- Kline, R. B. (2004). *Beyond significance testing: Reforming data analysis methods in behavioral research*. Washington, DC: American Psychological Association Press.
- Kohen, D. E., Brooks-Gunn, J., Leventhal, T., & Hertzman, C. (2002). Neighbourhood income and physical and social disorder in Canada: Associations with young children's competencies. *Child Development*, 73, 1844-1860.
- Korfmacher, J. (2000). The Kempe Family Stress Inventory: A review. *Child Abuse and Neglect*, 24, 129-140.
- Korfmacher, J. (2001, July). *Harder than you think: Determining what works, for whom, and why in early childhood interventions*. Occasional paper prepared for the Herr Research Center, Erikson Institute, Chicago, IL.
- Korfmacher, J. (2002). Early childhood interventions: Now what? In H. E. Fitzgerald, K. H. Karraker, & T. Luster (Eds.), *Infant development: Ecological perspectives* (pp. 273-294). New York: RoutledgeFalmer.
- Korfmacher, J., Kitzman, H., & Olds, D. (1998). Intervention process as predictors of outcomes in a preventive home visitation program. *Journal of Community Psychology*, 26, 49-64.
- Korfmacher, J., O'Brien, R., Hiatt, S., & Olds, D. (1999). Differences in program implementation between nurses and paraprofessionals providing home visits during pregnancy and infancy: A randomized trial. *American Journal of Public Health*, 89, 1847-1851.
- Kovach, A. C., Becker, J., & Worley, H. (2004). The impact of community health workers on the self-determination, self-sufficiency, and decision-making ability of low-income women and mothers of young children. *Journal of Community Psychology*, 32, 343-356.
- Kraemer, H. C., Kazdin, A. E., Offord, D. R., Kessler, R. C., Jensen, P. S., & Kupfer, D. J. (1997). Coming to terms with the terms of risk. *Archives of General Psychiatry*, 54, 337-343.

- Kraemer, H. C., Kazdin, A. E., Offord, D. R., Kessler, R. C., Jensen, P. S., Kupfer, D. J. (1999). Measuring the potency of risk factors for clinical or policy significance. *Psychological Methods*, 4, 257-271.
- Kraemer, H. C., Stice, E., Kazdin, A., Offord, D., & Kupfer, D. (2001). How do risk factors work together? Mediators, moderators, and independent, overlapping, and proxy risk factors. *American Journal of Psychiatry*, 158, 848-856.
- Krumboltz, J. D. (1966). *Resolution in counseling: Impacts of behavioral science*. Boston, MA: Houghton-Mifflin.
- Lachman, M. E. (2004). Development in midlife. *Annual Review of Psychology*, 55, 305-331.
- Lamb, M. E. (1998). Nonparental child care: Context, quality, correlates, and consequences. In W. Damon (Series Ed.) & I. E. Siegel & K. A. Renninger (Vol. Eds.), *Handbook of child psychology: Vol. 4. Child psychology in practice* (5th ed., pp. 73-133). New York: John Wiley & Sons.
- Landsverk, J., Carillo, T., Connelly, C. D., Ganger, W. C., Slymen, D. J., Newton, R. R., Leslie, L., & Jones, C. (2002). *Healthy Families San Diego clinical trial: Technical report*. San Diego, CA: San Diego Children's Hospital and Health Center, Child and Adolescent Services Research Center.
- Landy, S., & Tam, K. K. (1998). *Understanding the contribution of multiple risk factors on child development at various ages*. (Working Paper No. W-98-22E). Ottawa: Applied Research Branch, Strategic Policy, Human Resources Development Canada.
- Leventhal, J. M. (2001). The prevention of child abuse and neglect: Successfully out of the blocks. *Child Abuse and Neglect*, 25, 431-439.
- Leventhal, T., & Brooks-Gunn, J. (2000). The neighborhoods they live in: The effects of neighborhood residence on child and adolescent outcomes. *Psychological Bulletin*, 126, 309-337.
- Leventhal, T., & Brooks-Gunn, J. (2004). A randomized study of neighborhood effects on low-income children's educational outcomes. *Developmental Psychology*, 40, 488-507.
- Lewis, M., Feiring, C., & Rosenthal, S. (2000). Attachment over time. *Child Development*, 71, 707-720.
- Lipman, E. L., Boyle, M. H., Dooley, M. D., & Offord, D. R. (2002). Child well-being in single-mother families. *Journal of the American Academy of Child and Adolescent Psychiatry*, 41, 75-82.
- Lipsey, M. W. (1990). *Design sensitivity: Statistical power for experimental research*. Newbury Park, CA: Sage.
- Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *American Psychologist*, 48, 1181-1209.

- Lombard, A. (1981). *Success begins at home*. Lexington, MA: Lexington Books.
- Lovejoy, M. C., Graczyk, P. A., O'Hare, E., & Neuman, G. (2000). Maternal depression and parenting behavior: A meta-analytic review. *Clinical Psychology Review*, 20, 561-592.
- Lykken, D. T. (1968). Statistical significance in psychological research. *Psychological Bulletin*, 70, 151-159.
- Lyons-Ruth, K., & Melnick, S. (2004). Dose-response effect of mother-infant clinical home visiting on aggressive behavior problems in Kindergarten. *Journal of the American Academy of Child and Adolescent Psychiatry*, 43, 699-707.
- Maccoby, E. E. (2000). Parenting and its effects on children: On reading and misreading behavior genetics. *Annual Review of Psychology*, 51, 1-27.
- MacLeod, J., & Nelson, G. (2000). Programs for the promotion of family wellness and the prevention of child maltreatment: A meta-analytic review. *Child Abuse and Neglect*, 24, 1127-1149.
- MacMillan, H. L., MacMillan, A. B., Offord, D. R., & Dingle, J. L. (1996). Aboriginal health. *Canadian Medical Association Journal*, 155, 1569-1578.
- Maddock, J. E., & Rossi, J. S. (2001). Statistical power of articles published in three health psychology-related journals. *Health Psychology*, 20, 76-78.
- Main, M. (1999). Epilogue: Attachment theory: Eighteen points with suggestions for future studies. In J. Cassidy & P. R. Shaver (Eds.), *Handbook of attachment: Theory, research, and clinical applications* (pp. 845-887). New York: Guilford Press.
- Mallinckrodt, B. (1991). Clients' representations of childhood emotional bonds with parents, social support, and formation of the working alliance. *Journal of Counseling Psychology*, 38, 401-409.
- Mallinckrodt, B. (1996). Change in working alliance, social support, and psychological symptoms in brief therapy. *Journal of Counseling Psychology*, 43, 448-455.
- Mallinckrodt, B. (2000). Attachment, social competencies, social support, and interpersonal process in psychotherapy. *Psychotherapy Research*, 10, 239-266.
- Marcantonio, R. J., & Cook, T. D. (1994). Convincing quasi-experiments: The interrupted time series and regression-discontinuity designs. In J. S. Wholey, H. P. Hatry, & K. E. Newcomer (Eds.), *Handbook of practical program evaluation* (pp. 133-154). San Francisco: Jossey-Bass.
- Maris, E. (1998). Covariance adjustment versus gain scores—revisited. *Psychological Methods*, 3, 309-327.

- Martens, P., Bond, R., Jebamani, L., Burchill, C., Roos, N., Derksen, S., Beaulieu, C., MacWilliam, L., Walld, R., Dik, N., Sanderson, D., Health Information and Research Committee, Assembly of Manitoba Chiefs, Tanner-Spence, M., Leader, A., Elias, B., & O'Neil, J. (March 2002). *The health and health care use of registered First Nations people living in Manitoba: A population-based study*. Winnipeg, MB: Manitoba Centre for Health Policy, Department of Community Health Sciences, Faculty of Medicine, University of Manitoba.
- Martin, D. J., Garske, J. P., & Davis, M. K. (2000). Relation of the therapeutic alliance with outcome and other variables: A meta-analytic review. *Journal of Consulting and Clinical Psychology*, 68, 438-450.
- Masse, L. N., & Barnett, W. S. (2002). *A benefit cost analysis of the Abecedarian early childhood intervention*. New Brunswick, NJ: National Institute for Early Education Research.
- Masten, A. S. (1999). Resilience comes of age: Reflections on the past and outlook for the next generation of research. In M. D. Glantz & J. L. Johnson (Eds.) *Resilience and development: Positive life adaptations* (pp. 281-296). New York: Kluwer Academic/Plenum Publishers.
- Masten, A. S. (2001). Ordinary magic: Resilience processes in development. *American Psychologist*, 56, 227-238.
- Masten, A. S., & Coatsworth, J. D. (1998). The development of competence in favorable and unfavorable environments: Lessons from research on successful children. *American Psychologist*, 53, 205-220.
- McCartney, K., & Rosenthal, R. (2000). Effect size, practical importance, and social policy for children. *Child Development*, 71, 173-180.
- McCarton, C. M., Brooks-Gunn, J., Wallace, I. F., & Bauer, C. R. (1997). Results at age 8 years of early intervention for low birthweight premature infants: The Infant Health and Development Program. *Journal of the American Medical Association*, 277, 126-132.
- McCurdy, K. (2001). Can home visitation enhance maternal social support? *American Journal of Community Psychology*, 29, 97-112.
- McCurdy, K., & Daro, D. (2001). Parent involvement in family support programs: An integrated theory. *Family Relations*, 50, 113-121.
- McCurdy, K., Gannon, R. A., & Daro, D. (2003). Participation patterns in home-based family support programs: Ethnic variations. *Family Relations*, 52, 3-11.
- McGuigan, W. M., Katzev, A. R., & Pratt, C. C. (2003a). Multi-level determinants of mothers' engagement in home visitation services. *Family Relations*, 52, 271-278.
- McGuigan, W. M., Katzev, A. R., & Pratt, C. C. (2003b). Multi-level determinants of retention in a home-visiting child abuse prevention program. *Child Abuse and Neglect*, 27, 363-380.
- McLanahan, S., & Adams, J. (1987). Parenthood and psychological well-being. *Annual Review of Sociology*, 13, 237-257.

- Meehl, P. E. (1967). Theory-testing in psychology and physics: A methodological paradox. *Philosophy of Science*, 34, 103-115.
- Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology*, 46, 806-834.
- Meehl, P. E. (1990a). Appraising and amending theories: The strategy of Lakatosian defense and two principles that warrant it. *Psychological Inquiry*, 1, 108-141.
- Meehl, P. E. (1990b). Why summaries of research on psychological theories are often uninterpretable. *Psychological Reports*, 66, 195-244.
- Meehl, P. E. (1997). The problem is epistemology, not statistics: Replace significance test by confidence intervals and quantify accuracy of risky numerical predictions. In L. L. Harlow, S. A. Mulaik, & J. H. Steiger (Eds.), *What if there were no significance tests?* (pp. 393-425). Mahwah, NJ: Lawrence Erlbaum Associates.
- Meehl, P. E. (1998, May 23). *The power of quantitative thinking*. Speech delivered upon receipt of the James McKeen Cattell Fellow award at the meeting of the American Psychological Society, Washington, DC.
- Meyer, B., & Pilkonis, P. A. (2001). Attachment style. *Psychotherapy*, 38, 466-472.
- Meyer, B., & Pilkonis, P. A. (2002). Attachment style. In J. C. Norcross (Ed.), *Psychotherapy relationships that work: Therapist contributions and responsiveness to patients* (pp. 367-382). New York: Oxford University Press.
- Millner, J. (1994). Assessing physical child abuse risk: The Child Abuse Potential Inventory. *Clinical Psychology Review*, 14, 547-583.
- Mistry, R. S., Vandewater, E. A., Huston, A. C., & McCloyd, V. C. (2002). Economic well-being and children's social adjustment: The role of family process in an ethnically diverse low-income sample. *Child Development*, 73, 935-951.
- Mohr, D. C. (1995). Negative outcome in psychotherapy: a critical review. *Clinical Psychology: Science and Practice*, 2, 1-27.
- Mohr, L. B. (1995). *Impact analysis for program evaluation* (2nd ed.). Thousand Oaks, CA: Sage.
- Morrison, D. E., & Henkel, R. E. (Eds.). (1970). *The significance test controversy*. Chicago: Aldine.
- Muñoz, R. F. (2001). On the road to a world without depression. *Journal of Primary Prevention*, 21, 325-338.
- Murphy, S., Orkow, B., & Nicola, R. M. (1985). Prenatal prediction of child abuse and neglect: A prospective study. *Child Abuse and Neglect*, 9, 225-235.

- Musick, J., & Stott, F. (2000). Paraprofessionals revisited and reconsidered. In J. P. Shonkoff & S. J. Meisels (Eds.), *Handbook of early childhood intervention* (2nd ed., pp. 439-453). New York: Cambridge University Press.
- National Assessment of Educational Progress. (1997). *NAEP 1996 trends in academic progress*. Washington, DC: National Center for Education Statistics.
- National Research Council & Institute of Medicine. (1999). *Revisiting home visiting: Summary of a workshop*. Washington, DC: National Academy Press.
- National Research Council & Institute of Medicine. (2000). *From neurons to neighborhoods: The science of early childhood development*. Washington, DC: National Academy Press.
- National Research Council and Institute of Medicine. (2002). *Eager to learn: Educating our preschoolers*. Washington, DC: National Academy Press.
- Navaie-Waliser, M., Martin, S. L., Campbell, M. K., Tessaro, I., Kotelchuk, M., & Cross, A. W. (2000). Factors predicting completion of a home visitation program by high-risk pregnant women: The North Carolina Maternal Outreach Worker Program. *American Journal of Public Health, 90*, 121-124.
- Nelson, G., Laurendeau, M.C., & Chamberland, C. (2001). A review of programs to promote family wellness and prevent the maltreatment of children. *Canadian Journal of Behavioural Science, 33*, 1-13.
- Nelson, G., Westhues, A., & MacLeod, J. (2003, December 18). A meta-analysis of longitudinal research on preschool prevention programs for children. *Prevention & Treatment, 6*, Article 31. Retrieved February 16, 2004, from <http://journals.apa.org/prevention/volume6/pre0060031a.html>
- Newcomb, M. D., & Locke, T. F. (2001). Intergenerational cycle of maltreatment: A popular concept obscured by methodological limitations. *Child Abuse and Neglect, 25*, 1219-1240.
- NICHD Early Child Care Research Network. (1993). *The NICHD Study of Early Child Care: A comprehensive longitudinal study of young children's lives*. ERIC Document Reproduction Service No. ED 353 0870.
- NICHD Early Child Care Research Network. (1996). Characteristics of infant child care: Factors contributing to positive caregiving. *Early Childhood Research Quarterly, 11*, 269-306.
- NICHD Early Child Care Research Network. (1997). The effects of infant child care on infant-mother attachment security: Results of the NICHD Study of Early Child Care. *Child Development, 68*, 860-879.
- NICHD Early Child Care Research Network. (1998a). Early child care and self-control, compliance, and problem behavior at twenty-four and thirty-six months. *Child Development, 68*, 860-879.

- NICHD Early Child Care Research Network. (1998b). Relations between family predictors and child outcomes: Are they weaker for children in child care? *Developmental Psychology, 34*, 1119-1128.
- NICHD Early Child Care Research Network. (1999). Child care and mother-child interaction in the first 3 years of life. *Developmental Psychology, 35*, 1399-1413.
- NICHD Early Child Care Research Network. (2000). The relation of child care to cognitive and language development. *Child Development, 71*, 960-980.
- NICHD Early Child Care Research Network. (2001a). Child care and children's peer interaction at 24- and 36-months: The NICHD Study of Early Child Care. *Child Development, 72*, 1478-1500.
- NICHD Early Child Care Research Network. (2001b). Child-care and family predictors of preschool attachment and stability from infancy. *Developmental Psychology, 37*, 847-862.
- NICHD Early Child Care Research Network. (2002). Child-care structure, process, outcome: Direct and indirect effects of child-care quality on young children's development. *Psychological Science, 13*, 199-206.
- Nickerson, R. S. (2000). Null hypothesis significance testing: A review of an old and continuing controversy. *Psychological Methods, 5*, 241-301.
- Normand, C. L., Zoccolillo, M., Tremblay, R. E., McIntyre, L., Boulerice, B., McDuff, P., Pérusse, D., & Barr, R. G. (1996). In the beginning: Looking for the roots of babies' difficult temperament. In Human Resources Development Canada & Statistics Canada, *Growing up in Canada: National Longitudinal Survey of Children and Youth* (pp. 57-68). Ottawa, ON: Statistics Canada.
- O'Connor, T. G. (2002). The 'effects' of parenting reconsidered: Findings, challenges, and applications [Annotation]. *Journal of Child Psychology and Psychiatry, 43*, 555-572.
- Olds, D. L. (2002). Prenatal and infancy home visiting by nurses: From randomized trials to community replication. *Prevention Science, 3*, 153-172.
- Olds, D. L. (2003). Reducing program attrition in home visiting: What do we need to know? *Child Abuse and Neglect, 27*, 359-361.
- Olds, D. L. (2004, September 23). *Prenatal and infancy home visiting by nurses: From trials to practice*. Paper presented at the Annual Meeting of the European Society for Social Pediatrics, Montréal, Québec.
- Olds, D. L., Eckenrode, J., Henderson, C. R., Jr., Kitzman, H., Powers, J., Cole, R., Sidora, K., Morris, P., Pettitt, L. M., & Luckey, D. (1997). Long-term effects of home visitation on maternal life course and child abuse and neglect: 15-year follow-up of a randomized trial. *Journal of the American Medical Association, 278*, 637-643.

- Olds, D. L., Henderson, C. R., Jr., Cole, R., Eckenrode, J., Kitzman, H., Luckey, D., Pettitt, L., Sidora, K., Morris, P., & Powers, J. (1998). Long-term effects of nurse home visitation on children's criminal and antisocial behavior: 15-year follow-up of a randomized controlled trial. *Journal of the American Medical Association, 280*, 1238-1244.
- Olds, D. L., Henderson, C. R., Jr., Kitzman, H. J., Eckenrode, J. J., Cole, R. E., & Tatelbaum, R. C. (1999). Prenatal and infancy home visitation by nurses: Recent findings. *The Future of Children, 9*, 44-65.
- Olds, D. L., & Kitzman, H. (1993). Review of research on home visiting for pregnant women and parents of young children. *The Future of Children, 3*, 53-92.
- Olds, D. L., O'Brien, R. A., Racine, D., Glazner, J., & Kitzman, H. (1998). Increasing the policy and program relevance of results from randomized trials of home visitation. *Journal of Community Psychology, 26*, 85-100.
- Olds, D. L., Robinson, J. R., O'Brien, R., Luckey, D. W., Pettitt, L. M., Henderson, C. R., Ng, R. K., Sheff, K. L., Korfmacher, J., Hiatt, S., & Talmi, A. (2002). Home visiting by paraprofessionals and by nurses: A randomized, controlled trial. *Pediatrics, 110*, 486-496.
- Olds, D. L., Robinson, J., Song, N., Little, C., Hill, P. (1999, August). *Reducing risks for mental disorders during the first five years of life: A review of preventive interventions*. Paper commissioned by the Center for Mental Health Services. Denver, CO: Prevention Research Center for Family and Child Health, University of Colorado Health Sciences Center.
- Omer, H., & Dar, R. (1992). Changing trends in three decades of psychotherapy research: The flight from theory into pragmatics. *Journal of Consulting and Clinical Psychology, 60*, 88-93.
- Orlinsky, D. E., Grawe, K., & Parks, B. K. (1994). Process and outcome in psychotherapy – noch einmal. In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change* (4th ed., pp. 270-376). New York: Wiley.
- Orkow, B. (1985). Implementation of a family stress checklist. *Child Abuse and Neglect, 9*, 405-410.
- Osburn, H. G. (2000). Coefficient alpha and related internal consistency reliability coefficients. *Psychological Methods, 5*, 343-355.
- Osofsky, J. D., Culp, A. M., & Ware, L. M. (1988). Intervention challenges with adolescent mothers and their infants. *Psychiatry, 51*, 236-241.
- Parke, R. D. (2004). Development in the family. *Annual Review of Psychology, 55*, 365-399.
- Paul, G. L. (1967). Strategy of outcome research in psychotherapy. *Journal of Consulting and Clinical Psychology, 31*, 109-118.

- Peterson, L., Tremblay, G., Ewigman, B., & Saldana, L. (2003). Multilevel selected primary prevention of child maltreatment. *Journal of Consulting and Clinical Psychology*, 71, 601-612.
- Postman, L. (1955). The probability approach and nomothetic theory. *Psychological Review*, 62, 218-225.
- Prentice, D. A., & Miller, D. T. (1992). When small effects are impressive. *Psychological Bulletin*, 112, 160-164.
- Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, 1, 385-401.
- Ramey, C. T., Campbell, F. A., Burchinal, M., Skinner, M. L., Gardner, D. M., & Ramey, S. L. (2000). Persistent effects of early childhood education on high-risk children and their mothers. *Applied Developmental Science*, 4, 2-14.
- Ramey, C. T., & Ramey, S. L. (1998). Early intervention and early experience. *American Psychologist*, 53, 109-120.
- Ramey, C. T., & Ramey, S. L. (2002, June). *Early childhood education: The journey from efficacy research to effective, everyday practice*. Paper presented at the Summit on Early Childhood Cognitive Development, Boise, Idaho.
- Reichardt, C. S., & Bormann, C. A. (1994). Using regression models to estimate program effects. In J. S. Wholey, H. P. Hatry, & K. E. Newcomer (Eds.), *Handbook of practical program evaluation* (pp. 417-455). San Francisco: Jossey-Bass.
- Reis, H. T., Collins, W. A., & Berscheid, E. (2000). The relationship context of human behavior and development. *Psychological Bulletin*, 126, 844-872.
- Repetti, R. L., Taylor, S. E., & Seeman, T. E. (2002). Risky families: Family social environments and the mental and physical health of offspring. *Psychological Bulletin*, 128, 330-366.
- Reppucci, N. D., Woolard, J. L., & Fried, C. S. (1999). Social, community, and preventive interventions. *Annual Review of Psychology*, 50, 387-418.
- Reynolds, A. J. (2000). *Success in early intervention: The Chicago Child-Parent Centers*. Lincoln, NK: University of Nebraska Press.
- Reynolds, A. J., & Robertson, D. L. (2003). School-based early intervention and later child maltreatment in the Chicago Longitudinal Study. *Child Development*, 74, 3-26.
- Reynolds, A. J., Temple, J. A., Robertson, D. L., & Mann, E. A. (2001). Long-term effects of an early childhood intervention on educational achievement and juvenile arrest: A 15-year follow-up of low-income children in public schools. *Journal of the American Medical Association*, 285, 2339-2346.
- Reynolds, A. J., Temple, J. A., Robertson, D. L., & Mann, E. A. (2002, February). *Age 21 cost-benefit analysis of the Title I Chicago Child-Parent Centers*. Institute for Research on Poverty, University of Wisconsin-Madison. Discussion Paper no. 1245-02.

- Rogers, C. R. (1951). *Client centered therapy*. Cambridge, MA: Riverside Press.
- Rogers, C. R. (1957). The necessary and sufficient conditions of therapeutic personality change. *Journal of Consulting Psychology, 21*, 95-103.
- Rogosa, D. (1988). Myths about longitudinal research. In K. W. Schaie, R. T. Campbell, W. Meredith, & S. C. Rawlings (Eds.), *Methodological issues in aging research* (pp. 171-210). New York: Springer.
- Rogosa, D. R. (1995). Myths and methods: "Myths about longitudinal research," plus supplemental questions. In J. M. Gottman (Ed.), *The analysis of change* (pp. 3-65). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Rogosa, D. R., Brandt, D., & Zimowski, M. (1982). A growth curve approach to the measurement of change. *Psychological Bulletin, 90*, 726-748.
- Rogosa, D. R., & Willett, J. B. (1983). Demonstrating the reliability of the difference score in the measurement of change. *Journal of Educational Measurement, 20*, 335-343.
- Rogosa, D. R., & Willett, J. B. (1985). Understanding correlates of change by modeling individual differences in growth. *Psychometrika, 50*, 203-228.
- Rohner, R. P. (2004). The parental "acceptance-rejection syndrome": Universal correlates of perceived rejection. *American Psychologist, 59*, 830-840.
- Rosenthal, R. (1976). *Experimenter effects in behavioral research*. New York: Halstead Press.
- Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin, 85*, 185-193.
- Rosenthal, R. (1995). Writing meta-analytic reviews. *Psychological Bulletin, 118*, 183-192.
- Rosnow, R. L., & Rosenthal, R. (2003). Effect sizes for experimenting psychologists. *Canadian Journal of Experimental Psychology, 57*, 221-237.
- Rothbart, M. K., & Bates, J. E. (1998). Temperament. In W. Damon (Series Ed.) & N. Eisenberg (Vol. Ed.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (5th ed., pp. 105-176). New York: John Wiley & Sons.
- Rozeboom, W. W. (1960). The fallacy of the null hypothesis significance test. *Psychological Bulletin, 57*, 416-428.
- Runyan, D. K., Hunter, W. M., Socolar, R. R. S., Amaya-Jackson, L., English, D., Landsverk, J., Dubowitz, H., Browne, D. H., Bangdiwala, S. I., & Mathew, R. M. (1998). Children who prosper in unfavorable environments: The relationship to social capital. *Pediatrics, 101*, 12-18.
- Rutter, M. (2000). Resilience reconsidered: Conceptual considerations, empirical findings, and policy implications. In J. P. Shonkoff & S. J. Meisels (Eds.), *Handbook of early childhood intervention* (2nd ed., pp. 651-682). New York: Cambridge University Press.

- Rutter, M. (2005). Environmentally mediated risks for psychopathology: Research strategies and findings. *Journal of the American Academy of Child and Adolescent Psychiatry*, 44, 3-18.
- Ryff, C. D. (1989). Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *Journal of Personality and Social Psychology*, 57, 1069-1081.
- Ryff, C. D. (1995). Psychological well-being in adult life. *Current Directions in Psychological Science*, 4, 99-104.
- Ryff, C. D., & Keyes, C. L. M. (1995). The structure of psychological well-being revisited. *Journal of Personality and Social Psychology*, 69, 719-727.
- Ryff, C. D., Singer, B. H., & Palmersheim, K. A. (2004). Social inequalities in health and well-being: The role of relational and religious protective factors. In O. G. Brim, C. D. Ryff, & R. C. Kessler (Eds.), *How healthy are we? A national study of well-being at midlife* (pp. 90-123). Chicago: University of Chicago Press.
- Safran, J. D., & Muran, J. C. (2000). Introduction [Millennium issue: The therapeutic alliance]. *Journal of Clinical Psychology/In Session: Psychotherapy in Practice*, 56, 159-161.
- Safran, J. D., Muran, J. C., Samstag, L. W., & Stevens, C. (2001). Repairing alliance ruptures. *Psychotherapy*, 38, 406-412.
- Safran, J. D., Muran, J. C., Samstag, L. W., & Stevens, C. (2002). Repairing alliance ruptures. In J. C. Norcross (Ed.), *Psychotherapy relationships that work* (pp. 235-254). New York: Oxford University Press.
- Sampson, R. J., Morenoff, J. D., & Gannon-Rowley, T. (2002). Assessing "neighborhood effects": Social processes and new directions in research. *Annual Review of Sociology*, 28, 443-478.
- Sanford, N. (1953). Clinical methods: Psychotherapy. *Annual Review of Psychology*, 4, 317-342.
- Santor, D. A., & Coyne, J. C. (1997). Shortening the CES-D to improve its ability to detect cases of depression. *Psychological Assessment*, 9, 233-243.
- Schmidt, F. L. (1992). What do data really mean? Research findings, meta-analysis, and cumulative knowledge in psychology. *American Psychologist*, 47, 1173-1181.
- Schmidt, F. L. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for the training of researchers. *Psychological Methods*, 1, 115-129.
- Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, 8, 350-353.
- Schmutte, P. S., & Ryff, C. D. (1997). Personality and well-being: Reexamining methods and meanings. *Journal of Personality and Social Psychology*, 73, 549-559.
- Schweinhart, L. J. (2004). *The High/Scope Perry Preschool study through age 40: Summary, conclusions, and frequently asked questions*. Ypsilanti, MI: High/Scope Press.

- Schweinhart, L. J., Barnes, H. V., Weikart, D. P., Barnett, W. S., & Epstein, A. S. (1993). *Significant benefits: The High/Scope Perry Preschool study through age 27* (Monographs of the High/Scope Educational Research Foundation, no. 10). Ypsilanti, MI: High/Scope Press.
- Schweinhart, L. J., Montie, J., Xiang, Z., Barnett, W. S., Belfield, C. R., & Nores, M. (in press). *Lifetime effects: The High/Scope Perry Preschool study through age 40* (Monographs of the High/Scope Educational Research Foundation, no. 14). Ypsilanti, MI: High/Scope Press.
- Schweinhart, L. J., & Weikart, D. P. (1980). *Young children grow up: The effects of the Perry Preschool Program on youths through age 15* (Monographs of the High/Scope Educational Research Foundation, no. 7). Ypsilanti, MI: High/Scope Press.
- Seligman, M. E. P., & Csikszentmihalyi, M. (2000). Positive psychology: An introduction [Special issue]. *American Psychologist, 55*(1).
- Serbin, L. A., & Karp, J. (2004). The intergenerational transfer of psychosocial risk: Mediators of vulnerability and resilience. *Annual Review of Psychology, 55*, 333-363.
- Serlin, R. C., & Lapsley, D. K. (1985). Rationality in psychological research: The good-enough principle. *American Psychologist, 40*, 73-83.
- Sharp, E. A., Ispa, J. M., Thornburg, K. R., & Lane, V. (2003). Relations among mother and home visitor personality, relationship quality, and amount of time spent in home visits. *Journal of Community Psychology, 31*, 591-606.
- Shaw, D. S. (2003). Advancing our understanding of intergenerational continuity in antisocial behavior. *Journal of Abnormal Child Psychology, 31*, 193-199.
- Shinn, M., & Toohey, S. M. (2003). Community contexts of human welfare. *Annual Review of Psychology, 54*, 427-459.
- Shonkoff, J. P. (2000). Science, policy, and practice: Three cultures in search of a shared mission. *Child Development, 71*, 181-187.
- Shonkoff, J. P. (2003). Still waiting for the right questions. *American Journal of Preventive Medicine, 24*(3S), 4-5.
- Shonkoff, J. P. (2004, Summer). Evaluating early childhood services: What's really behind the curtain. *The Evaluation Exchange, 10* (2), 3-4.
- Shrout, P. E. (1997). Should significance tests be banned? Introduction to a special section exploring the pros and cons. *Psychological Science, 8*, 1-2.
- Silk, J. S., Sessa, F. M., Morris, A. S., Steinberg, L., & Avenevoli, S. (2004). Neighborhood cohesion as a buffer against hostile maternal parenting. *Journal of Family Psychology, 18*, 135-146.
- Simcha-Fagan, O., & Schwartz, J. E. (1986). Neighborhood delinquency: An assessment of contextual effects. *Criminology, 24*, 667-704.

- Spitzer, R. L., Williams, J. B., Gibbon, M., & First, M. B. (1989). *Structured Clinical Interview for DSM-III-R*. New York: Biometrics Research Department, New York State Psychiatric Institute.
- St. Pierre, R. G., & Layzer, J. E. (1999). Using home visits for multiple purposes: The Comprehensive Child Development Program. *The Future of Children*, 9, 134-151.
- Statistics Canada. (1998, October 28). National Longitudinal Survey of Children and Youth, Cycle 2, 1996. *The Daily*, 1-5.
- Statistics Canada. (2004, July 9). Aboriginal Peoples Survey 2001: Children who live in non-reserve areas. *The Daily*, 7-9.
- Statistics Canada. (2004, October 25). Parenting style and children's aggressive behaviour: 1994 to 2000. *The Daily*, 8-9.
- Statistics Canada & Human Resources Development Canada. (1995). *NLSC: Overview of survey instruments for 1994-95 data collection cycle 1* (Catalogue no. 89F0078XIE). Ottawa, ON: Minister of Industry.
- Stevens, J., Ammerman, R. T., Putnam, F. G., & Van Ginkel, J. B. (2002). Depression and trauma history in first-time mothers receiving home visitation. *Journal of Community Psychology*, 30, 551-564.
- Stiles, W. B., Glick, M. J., Osatuke, K., Hardy, G. E., Shapiro, D. A., Agnew-Davies, R., Rees, A., & Barkham, M. (2004). Patterns of alliance development and the rupture-repair hypothesis: Are productive relationships U-shaped or V-shaped? *Journal of Counseling Psychology*, 51, 81-92.
- Stiles, W. B., Shapiro, D. A., & Elliot, R. (1986). "Are all psychotherapies equivalent?" *American Psychologist*, 41, 165-180.
- Strayhorn, J. M., & Weidman, C. S. (1988). A parent practices scale and its relation to parent and child mental health. *Journal of the American Academy of Child and Adolescent Psychiatry*, 27, 613-618.
- Streiner, D. L. (2002). The 2 "Es" of research: Efficacy and effectiveness trials. *Canadian Journal of Psychiatry*, 47, 552-556.
- Streiner, D. L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment*, 80, 99-103.
- Strupp H. H., Hadley, S. W., & Gomes-Schwartz, B. (1977). *Psychotherapy for better or worse: The problem of negative effects*. New York: Jason Aronson, Inc.
- Strupp, H. H., Hadley, S. W., & Gomes-Schwartz, B. (1994). *When things get worse: The problem of negative effects in psychotherapy* (Master Works Series). New York: Jason Aronson, Inc.

- Sweet, M. A., & Appelbaum, M. I. (2004). Is home visiting an effective strategy? A meta-analytic review of home visiting programs for families with young children. *Child Development, 75*, 1435-1456.
- Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Boston: Allyn and Bacon.
- Taylor, R. D., Roberts, D., & Jacobson, L. (1997). Stressful life events, psychological well-being, and parenting in African American mothers. *Journal of Family Psychology, 11*, 436-446.
- Thistlethwaite, D. L., & Campbell, D. T. (1960). Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational Psychology, 51*, 309-317.
- Thomas, E. M. (2004). *Aggressive behaviour outcomes for young children: Change in parenting environment predicts change in behaviour* (Children and Youth Research Paper Series, Catalogue no. 89-599-MIE). Ottawa, ON: Statistics Canada.
- Thompson, B. (1998). Statistical significance and effect size reporting: Portrait of a possible future. *Research in the Schools, 5*, 33-38.
- Thompson, B. (2002a). "Statistical," "practical," and "clinical": How many kinds of significance do counselors need to consider? *Journal of Counseling and Development, 80*, 64-71.
- Thompson, B. (2002b). What future quantitative social science research could look like: Confidence intervals for effect sizes. *Educational Researcher, 31*(4), 25-32.
- Thompson, R. A. (2000). The legacy of early attachments. *Child Development, 71*, 145-152.
- Thornberry, T. P., Freeman-Gallant, A., Lizotte, A. J., Krohn, M. D., & Smith, C. A. (2003). Linked lives: The intergenerational transmission of antisocial behavior. *Journal of Abnormal Child Psychology, 31*, 171-184.
- Thornberry, T. P., Hops, H., Conger, R. D., & Capaldi, D. M. (2003). Replicated findings and future directions for intergenerational studies: Closing comments. *Journal of Abnormal Child Psychology, 31*, 201-203.
- Tolan, P. H. (1999). Research methods in community-based treatment and prevention. In P. C. Kendall, J. N. Butcher, & Holmbeck, G. N. (Eds.), *Handbook of research methods in clinical psychology* (2nd ed., pp. 403-418). New York: John Wiley & Sons.
- Tolan, P. H. (2002). Family-focused prevention research: "Tough but tender." In H. A. Liddle, D. A. Santisteban, R. F. Levant, & J. H. Bray (Eds.), *Family psychology: Science-based interventions* (pp. 197-213). Washington, DC: American Psychological Association Press.
- Tolan, P. H., Hanish, L. D., McKay, M. M., & Dickey, M. H. (2002). Evaluating process in child and family interventions: Aggression prevention as an example. *Journal of Family Psychology, 16*, 220-236.

- Tookenay, V. F. (1996). Improving the health status of aboriginal people in Canada: New directions, new responsibilities. *Canadian Medical Association Journal, 155*, 1581-1583.
- Tracey, T. J., & Kokotovic, A. M. (1989). Factor structure of the Working Alliance Inventory. *Psychological Assessment, 1*, 207-210.
- Tremblay, R. E., Nagin, D. S., Séguin, J. R., Zoccolillo, M., Zelazo, P. D., Boivin, M., Pérusse, D., & Japel, C. (2004). Physical aggression during early childhood: Trajectories and predictors. *Pediatrics, 114*, e43-e50.
- Trochim, W. M. K. (1984). *Research design for program evaluation: The regression-discontinuity approach*. Beverly Hills, CA: Sage.
- Trochim, W. M. K. (1990). The regression-discontinuity design. In L. Sechrest, P. Perrin, & J. Bunker (Eds.), *Research methodology: Strengthening causal interpretations of nonexperimental data* (pp. 119-139). Washington, DC: U.S. Department of Health and Human Services.
- Trochim, W. M. K. (2000). *The research methods knowledge base* (2nd ed., Electronic version). Retrieved December 6, 2002, from <http://trochim.human.cornell.edu/kb/> (Note: As of August 16, 2004, reference has moved to <http://www.socialresearchmethods.net/kb/>)
- Turcotte, M., & Zhao, J. (2004). *A portrait of Aboriginal children living in non-reserve areas: Results from the 2001 Aboriginal Peoples Survey* (Catalogue no. 89-597-XIE). Ottawa, ON: Minister of Industry.
- Uchino, B. N., Cacioppo, J. T., & Kiecolt-Glaser, J. K. (1996). The relationship between social support and underlying physiological processes: A review with emphasis on underlying mechanisms and implications for health. *Psychological Bulletin, 119*, 488-531.
- Valentine, J. C., & Cooper, H. (2003). *Effect size substantive interpretation guidelines: Issues in the interpretation of effect sizes*. Washington, DC: What Works Clearinghouse.
- Wagner, M. M., & Clayton, S. L. (1999). The Parents as Teachers program: Results from two demonstrations. *The Future of Children, 9*, 91-115.
- Wasik, B. H. (1993). Staffing issues for home visiting programs. *The Future of Children, 3*, 140-157.
- Wasik, B. H., & Roberts, R. N. (1994). Home visitor characteristics, training, and supervision: Results of a national survey. *Family Relations, 43*, 336-341.
- Waters, E., & Cummings, E. M. (2000). A secure base from which to explore close relationships. *Child Development, 71*, 164-172.
- Waters, E., Hamilton, C. E., Weinfield, N. S. (2000). The stability of attachment security from infancy to adolescence and early adulthood: General introduction. *Child Development, 71*, 678-683.

- Waters, E., Merrick, S., Treboux, D., Crowell, J., & Albersheim, L. (2000). Attachment security in infancy and early adulthood: A twenty-year longitudinal study. *Child Development*, 71, 684-689.
- Waters, E., Weinfield, N. S., & Hamilton, C. E. (2000). The stability of attachment security from infancy to adolescence and early adulthood: General discussion. *Child Development*, 71, 703-706.
- Weikart, D. P., Bond, J. T., & McNeil, J. T. (1978). *The Ypsilanti Perry Preschool Project: Preschool years and longitudinal results through fourth grade* (Monographs of the High/Scope Educational Research Foundation, no. 3). Ypsilanti, MI: High/Scope Press.
- Weinfield, N. S., Sroufe, L. A., & Egeland, B. (2000). Attachment from infancy to early adulthood in a high-risk sample: Continuity, discontinuity, and their correlates. *Child Development*, 71, 695-702.
- Weiss, H. B. (1993). Home visits: Necessary but not sufficient. *The Future of Children*, 3, 113-128.
- Weissberg, R. P., & Greenberg, M. T. (1998). School and community competence-enhancement and prevention programs. In W. Damon (Series Ed.) & I. E. Siegel & K. A. Renninger (Vol. Eds.), *Handbook of child psychology: Vol. 4. Child psychology in practice* (5th ed., pp. 877-954). New York: John Wiley & Sons.
- Weissberg, R. P., Kumpfer, K. L., & Seligman, M. E. P. (2003). Prevention that works for children and youth: An introduction. *American Psychologist*, 58, 425-432.
- Weisz, J. R., Donenberg, G. R., Han, S. S., & Weiss, B. (1995). Bridging the gap between laboratory and clinic in child and adolescent psychotherapy. *Journal of Consulting and Clinical Psychology*, 63, 688-701.
- Weisz, J. R., & Jensen, P. S. (1999). Efficacy and effectiveness of child and adolescent psychotherapy and pharmacotherapy. *Mental Health Services Research*, 1, 122-157.
- Werner, E. E., & Johnson, J. L. (1999). Can we apply resilience? In M. D. Glantz & J. L. Johnson (Eds.) *Resilience and development: Positive life adaptations* (pp. 259-268). New York: Kluwer Academic/Plenum Publishers.
- Westhues, A., Nelson, G., & MacLeod, J. (2003, December 18). The long-term impact of preschool prevention programs: Looking to the future. *Prevention & Treatment*, 6, Article 36. Retrieved February 16, 2004, from <http://journals.apa.org/prevention/volume6/pre0060036r.html>
- Whipple, E. (2004, January 14). *The role of therapeutic alliance in a rural Healthy Families America (HFA) program*. Paper presented at the 9th World Congress of the World Association for Infant Mental Health, Melbourne, Australia.
- Whitehurst, G. J., & Lonigan, C. J. (1998). Child development and emergent literacy. *Child Development*, 69, 848-872.

- Wilkinson, L., & the Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, 54, 594-604.
- Willett, J. B. (1988). Questions and answers in the measurement of change. *Review of Research in Education*, 15, 345-422.
- Willett, J. B. (1997). Measuring change: What individual growth modeling buys you. In E. Amsel & K. A. Reninger (Eds.), *Change and development* (pp. 213-243). Mahwah, NJ: Lawrence Erlbaum Associates.
- Willett, J. B., Singer, J. D., & Martin, N. C. (1998). The design and analysis of longitudinal studies of development and psychopathology in context: Statistical models and methodological recommendations. *Development and Psychopathology*, 10, 395-426.
- Windham, A. M., Rosenberg, L., Fuddy, L., McFarlane, E., Sia, C., & Duggan, A. K. (2004). Risk of mother-reported child abuse in the first 3 years of life. *Child Abuse and Neglect*, 28, 645-667.
- Yoshikawa, H. (2000). Community prevention and intervention: Prevention with young children. In A. E. Kazdin (Ed.), *Encyclopedia of psychology: Vol. 2* (pp. 194-197). American Psychological Association and Oxford University Press.
- Young, T. K. (2003). Review of research on aboriginal populations in Canada: Relevance to their health needs. *British Medical Journal*, 327, 419-422.
- Zigler, E., & Styfco, S. J. (2001). Extended childhood intervention prepares children for school and beyond. *Journal of the American Medical Association*, 285, 2378-2380.