

## A LONG-TERM FOLLOW-UP OF A RANDOMIZED CONTROLLED TRIAL OF MOTHER–INFANT PSYCHOANALYTIC TREATMENT: OUTCOMES ON THE CHILDREN

MAJLIS WINBERG SALOMONSSON, KIMMO SORJONEN, AND BJÖRN SALOMONSSON

*Karolinska Institutet*

**ABSTRACT:** A randomized controlled trial (RCT) compared two groups of mother–infant dyads in a Stockholm sample. One had received mother–infant psychoanalytic treatment (MIP group) and the other Child Health Center care (CHCC group). Effects were found on mother-reported depression and expert-rated mother–infant relationship qualities and maternal sensitivity. When the children were 4½ years old, they were followed up with assessments of attachment representations, socioemotional development, and global functioning. They also were divided into two types according to individual characteristics and psychological well-being: the “OK” and the “Troubled” children. Of 80 dyads in the mother–infant RCT, data were gathered from 66 cases approximately 3½ years after treatment. The children in the MIP group had better results on global functioning. There were more OK children in the MIP group and more Troubled children in the CHCC group. No other between-group differences were found. A relatively brief mother–infant psychotherapy seemed to help the children function better and be less troubled 3½ years after therapy.

**RESUMEN:** Trasfondo: un ensayo controlado al azar (RCT) ha comparado dos grupos de díadas madre-infante en un grupo muestra en Estocolmo. Un grupo había recibido tratamiento psicoanalítico para madre-infante (grupo “MIP”) y el otro, cuidado del Centro de Salud Infantil (grupo “CHCC”). Se encontraron los efectos en la depresión reportada por las madres y las características de la relación madre-infante y la sensibilidad maternal según evaluación de expertos. Método: cuando los niños tenían 4 años y medio de edad, se les dio seguimiento con evaluaciones de representaciones de la afectividad, desarrollo social y emocional, y funcionamiento global. También se les dividió en dos tipos según las características individuales y el bienestar psicológico; los niños “OK” y los niños “Con Problemas.” De las 80 díadas en este RCT madre-infante, se recogió información de 66 casos aproximadamente 3 años y medio después del tratamiento. Resultados: los niños en el grupo “MIP” tuvieron mejores resultados en cuanto a funcionamiento global. Hubo más niños “OK” en el grupo “MIP” y más niños “Con Problemas” en el grupo “CHCC.” No se encontraron otras diferencias entre grupos. Conclusiones: una relativamente breve psicoterapia madre-infante pareció ayudar a los niños a funcionar mejor y tener menos problemas tres años y medio después de la terapia.

**RÉSUMÉ:** Le contexte: Une étude randomisée contrôlée a comparé deux groupes de dyades mère-b’ebé chez un échantillon à Stockholm en Suède. L’un des deux groupes a reçu un traitement psychanalytique mère-bébé (le groupe “MIP”) et l’autre groupe a été placé sous les soins du Centre de Soins de l’Enfant (le groupe “CHCC”). Les effets ont été trouvés sur la dépression auto-rapportée de la mère ainsi que sur la sensibilité maternelle et les qualités de la relation mère-bébé évaluées par un expert. Méthode: Lorsque les enfants ont eu 4 ans et demi ils furent suivis avec des évaluations de représentations d’attachement, de développement social et émotionnel, de fonctionnement global. Ils ont aussi été divisés en deux types selon les caractéristiques individuelles et le bien-être psychologique: le groupe “OK” et le groupe des enfants “Perturbés”. Sur les 80 dyades de l’étude randomisée contrôlée, des données ont été recueillies à partir de 66 cas à peu près 3 ans et demi après le traitement. Les Résultats: Les enfants du groupe MIP ont eu de meilleurs résultats pour ce qui concerne le fonctionnement global. Il y avait plus d’enfants OK dans le groupe MIP et plus d’enfants perturbés dans le groupe CHCC. Aucun autre différence entre les groupes n’a été trouvée. Conclusions: Un psychothérapie mère-bébé relativement brève a semblé aider les enfants à mieux fonctionner et à être moins perturbés 3ans et demi après la thérapie.

**ZUSAMMENFASSUNG:** Hintergrund: Eine randomisierte kontrollierte Studie (RCT) hat zwei Gruppen von Mutter-Kind-Dyaden anhand einer Stockholmer-Stichprobe verglichen. Eine Gruppe erhielt eine psychoanalytische Behandlung für Mutter und Kind (die "MIP"-Gruppe) und die andere Gruppe erhielt Dienstleistungen aus einem Kindergesundheitszentrum (die "CHCC" Gruppe). Effekte wurden hinsichtlich der von der Mutter

We thank the mothers and children who participated in the study as well as Professor Per-Anders Rydelius, Associate Professor Marie-Louise Ögren, Ph.D. Agneta Thorén, Professor Rolf Sandell, and Professor Andrzej Werbart for their valuable suggestions. We also thank the raters, psychologists Britta Blomberg, Elisabeth Cleve, Anders Schiöler, and Anna Lundh, M.D., Ph.D. The study was supported by the Ax:son Johnson, Children’s Welfare Foundation Sweden, Olle Engkvist Byggmästare, Clas Groschinsky, Signe and Ane Gyllenberg, Karolinska Institutet, Kempe-Carlgrén, Mayflower Charity, Solstickan and Wennborg Foundations, and the Research Advisory Board of the International Psychoanalytical Association.

Direct correspondence to: Majlis Winberg Salomonsson, Child and Adolescent Psychiatric Unit, Department of Women’s and Children’s Health, Karolinska Institutet, Astrid Lindgren Children’s Hospital, SE-171 76 Stockholm, Sweden; e-mail: majlis.winberg-salomonsson@ki.se.

berichteten Depression und der von Experten bewerteten Qualität der Mutter-Kind-Beziehung sowie dem mütterlichen Einfühlungsvermögen gefunden. Methode: Im Alter von 4,5 Jahren wurden die Kinder zu einer Follow-up-Untersuchung zur Beurteilung der Bindungsrepräsentationen, der sozialen und emotionalen Entwicklung und dem globalen Funktionsniveau wiedergesehen. Sie wurden anhand individueller Eigenschaften und psychischem Wohlbefinden in zwei Kategorien aufgeteilt: die "OK" und die "troubled" Kinder. Von 80 Dyaden aus der Mutter-Kind-RCT wurden Daten von 66 Fällen rund 3,5 Jahre nach der Behandlung gesammelt. Ergebnisse: Die Kinder der MIP-Gruppe hatten bessere Ergebnisse hinsichtlich des globalen Funktionsniveaus. Es gab mehr "OK" Kinder in der MIP-Gruppe und mehr "troubled" Kinder in der CHCC Gruppe. Es wurden keine anderen Unterschiede zwischen den Gruppen gefunden. Schlussfolgerungen: Eine relativ kurze Mutter-Kind-Psychotherapie schien den Kindern zu helfen, ein besseres Funktionsniveau zu entwickeln und 3,5 Jahre nach der Therapie weniger von Problemen belastet zu sein.

Abstract: 抄録:背景:ランダム化対照試験(RCT)で、ストックホルムの2群の母子ペアを比較した。一つのグループは母子精神分析治療を受けた(「MIP」群)。そしてもう一つのグループは児童保健センターケアを受けた(「CHCC」群)。母親が報告する抑うつと、専門家が評価する母親と乳児の関係性の質と母親の感受性に、効果が認められた。方法:子どもが4.5歳の時、子どもたちはフォローアップで、愛着表象、社会性と情緒の発達、および全般的な機能を評価された。子どもたちはまた個人の特徴と心理学的な健康によって、二つのタイプに分けられた。それらは、「OK」と「問題のある」子どもたちである。母親乳児RCTの80組のうち、データは66組から治療後約3.5年で、集められた。結果:MIP群の子どもは、全般的な機能においてよりよい結果を示した。OKの子どもがMIP群により多く、問題のある子どもがCHCC群により多かった。それ以外の群間差は見られなかった。結論:比較的短い母親と乳児の精神療法が、治療3.5年後に、子どもがよりよく機能し、問題が少ないことを助けているようだった。

摘要: 背景:一項隨機對照試驗比較了兩組在斯德哥爾摩樣本中的母嬰二人組合。一組曾接受母嬰心理治療(下稱“MIP”組),另一組接受兒童健康中心的護理(下稱“CHCC”組)。我們發現這兩種不同的處置對母親報告抑鬱症和專家鑒定的母嬰關係質量和母親的敏感性有影響。方法:當孩子為4.5歲時,隨訪會評估他們的依附表徵,社會和情感發展,以及整體功能。根據兒童的個體特徵和心理健康,他們會被分為兩種類型:“OK型”和“困擾型”。在這母嬰隨機對照試驗中的80二人組合,有66個案在療程約3.5年後再被收集數據。結果:MIP組的孩子們在整體功能上有較好的結果。在MIP組中有較多OK型兒童,CHCC組中則有較多困擾型兒童。沒有發現其他的組間差異。結論:比較上簡單的母嬰心理療法似乎可幫助孩子們更好地運作,並在療程約3.5年後會有少些困擾。

الملخص: خلفية: تم مقارنة محاكمة عشوائية تسيطر على (RCT) بمجموعتين من مزدوجي الصبغة للأم والرضيع في عينة في ستوكهولم. تلقى واحد منها العلاج النفسي للأم والرضيع (مجموعة "MIP") وبينما الآخر لمركز رعاية صحة الطفل (مجموعة "CHCC"). وجدت آثار في تقرير الاكتئاب للأم وتصنيف الخبراء لصفات العلاقة بين الأم والرضيع وحساسية الأمومة. الطريقة: عندما كان الأطفال في عمر 4.5 سنة، كانت تجري متابعتها مع تقييمات تمثيل التعلق، والتنمية الاجتماعية والعاطفية، والسير العالمي. تم تقسيمها أيضاً إلى نوعين وفقاً للخصائص الفردية والنفسية، الـ"موافق" والأطفال "المتعثر". من بين 80 من مزدوجي الصبغة في RCT للأم والرضيع، تم جمع بيانات من 66 حالة بعد حوالي 3.5 سنوات العلاج. النتائج: كانت نتائج الأطفال في الفئة MIP أفضل على السير العالمي. وكان هناك أكثر أطفال بإجابة موافق في مجموعة MIP وأكثر الأطفال بالمتعثر في المجموعة CHCC، لذا لا توجد اختلافات أخرى بين المجموعة. الاستنتاجات: ظهر موجز نسبي للعلاج النفسي للأم والرضيع لمساعدة الأطفال على نحو أفضل ولكي تكون أقل اضطراباً بعد 3.5 سنوات بعد العلاج.

\* \* \*

Several studies (Grossmann, Grossmann, & Waters, 2005; Massie & Szajnberg, 2002; Sroufe, Egeland, Carlson, & Collins, 2005) have shown that the quality of the infant's relationships to his or her caregivers is of crucial importance for the child's future development (Iluoma et al., 2001; Laucht, Schmidt, & Esser, 2004). This has been confirmed in longitudinal population studies on infants from birth onward (Skovgaard et al., 2005, 2007) as well as in clinical follow-up studies of psychological treatment or social support (Murray, Cooper, Wilson, & Romaniuk, 2003). So far, studies have more thoroughly investigated the mother-infant relationship than the father-infant relationship.

Infants with insecure attachment and socioemotional problems may later develop behavioral problems and psychopathology (Laucht et al., 2004; Lyons-Ruth, Dutra, Schuder, & Bianchi, 2006; Ogawa, Sroufe, Weinfeld, Carlson, & Egeland, 1997). Studies in this field seldom have described clinical samples that have received treatment, but they have indicated the correlations between early problems and later pathology. Infant disorders can have different sources; hereditary or environmental. To the extent

that they emerge from emotional problems in the relationship, we assumed that a psychotherapy aiming at ameliorating the mother-infant interaction might promote a benign development in toddlerhood.

Randomized controlled trials (RCTs) of parent-infant psychotherapies are not numerous, and follow-up studies are even rarer. In a meta-analysis of psychotherapies for postpartum depression, Dennis (2004) found it difficult to reach definitive conclusions regarding their efficacy due to the lack of well-designed investigations. In Singleton's (2005) meta-analysis of parent-infant interventions, only 60% of the studies were of high or fair methodological quality. One recurrent methodological problem with studies on parent-infant psychotherapies is the brief follow-up periods. Some RCTs targeting therapies of depressed mothers with infants have used follow-up periods of only 3 months (Clark, Tluczek, & Wenzel, 2003; Mulcahy, Reay, Wilkinson, & Owen, 2010; O'Hara, Stuart, Gorman, & Wenzel, 2000). Other studies (Cohen, Lojkasek, Muir, Muir, & Parker, 2002; Cohen et al., 1999; Hayes & Matthews, 2008; Letourneau & Stewart,

2011; Lieberman, Weston, & Pawl, 1991; Ravn et al., 2012; Robert-Tissot et al., 1996; Santelices & Guzmán, 2011) have used follow-up periods from zero to 12 months' posttreatment.

We found one study with a lengthy follow-up (Cooper, Murray, Wilson, & Romaniuk, 2003; Murray et al., 2003); it investigated if therapy for depressed mothers might improve dyadic relationships and child development. Participants were randomized to one of three active 10-week therapies (cognitive-behavioral therapy, psychodynamic mother–infant therapy, and nondirective counseling) or to routine primary care. The active treatment groups lowered their depression scores immediately posttreatment while only the mother–infant therapy group improved on the clinical interviews. The follow-ups at a child age 9 months and 1½ and 5 years showed limited effects on mother-reported relationship problems, and the psychodynamic group even reported more behavioral problems posttreatment. Sensitivity only improved among mothers at social risk who received counseling (Holden, Sagovsky, & Cox, 1989). Thus, most outcomes failed to show any long-term benefit of the active treatments.

A Swedish RCT (Salomonsson & Sandell, 2011a, 2011b) compared mother–infant psychoanalytic therapies (MIP; Norman, 2001) with Swedish standard Child Health Center Care (CHCC). Outcome assessments were made 6 months after intake. We will describe this study and then present the results of our follow-up investigation. For a summary of the design of these two studies, please refer to Figure 1.

### THE INFANT STUDY

The study was performed from 2005 to 2008 on mothers and infants below 1½ years of age recruited from Child Health Centers, a delivery ward, and advertisements on parenting Internet sites. Inclusion criteria were that the mothers had expressed “baby worries” (Salomonsson, 2010); that is, concerns about their role as mothers, their infants' well-being, or the mother–baby relationship. The babies displayed problems with sleep, breast-feeding, weaning, mood, anxiety, and attachment. The mothers were depressed, anxious, or ambivalent about motherhood. The cases were thus included if they reflected disturbances within each of the dyad's participants as well as between the two. The problems should have lasted for more than 2 weeks. The mothers should live in Stockholm and speak Swedish to participate in treatments. Cases of psychosis or substance abuse were excluded if precluding cooperation. Eighty dyads were seen in a video-recorded interview and randomized to MIP or CHCC. Assessments were made at intake and 6 months later of mother-reported depression, stress, general psychological distress and infant functional problems, interviewer-assessed relationship qualities, and externally rated interactions. Intent-to-treat analyses could be made on 75 cases; 38 in the MIP group and 37 in the CHCC group.

After a pediatric examination, treatment started. MIP cases were treated by analysts in the Mother-Infant Psychoanalysis Project in Stockholm (MIPPS). The method (Norman, 2001, 2004; Salomonsson, 2007, 2011, 2014) implies that both mother and

baby are regarded as active participants who relate to the analyst. Concerning the baby, it is assumed

(1) that a relationship can be established between the infant and the analyst, (2) that the infant has a primordial subjectivity and self as a base for intersubjectivity and the search for containment, (3) that the infant has an unique flexibility in changing representations of itself and others that comes to an end as the ego develops, and (4) that the infant is able to process aspects of language. (Norman, 2001, p. 83)

The method thus aims at bringing the infant's disturbance into the session for containment. The mother is given ample opportunities of venting her distress and working with her “ghosts in the nursery” (Fraiberg, Adelson, & Shapiro, 1975), “negative attributions” (Silverman & Lieberman, 1999), and “projective distortions” (Cramer & Palacio Espasa, 1993). Session frequency and treatment duration are adapted to the disorder and the mother's motivation and possibilities of continuing therapy. A high frequency is preferred to facilitate working through the distress of mother and baby and to improve the baby's attachment.

Swedish CHCC implies that nurse calls follow a schedule; weekly the first month, monthly up to 4 months, and every second month during the rest of the first year. Checkups at 1½, 3, 4, and 5 years comprise weighing and measuring the baby and providing inoculations, nutritional advice, and guidance (Lojkasek, Cohen, & Muir, 1994). The nurse also seeks to promote a secure attachment and to detect depression.

### Instruments

*Maternal distress.* Three self-report questionnaires were used: the Edinburgh Postnatal Depression Scale (EPDS; Cox, Holden, & Sagovsky, 1987), the Swedish Parental Stress Questionnaire (SPSQ; Östberg, Hagekull, & Wettergren, 1997), and the Symptom Check List-90 on general psychological distress (SCL-90; Derogatis, 1994; Fridell, Cesarec, Johansson, & Malling Thorsen, 2002). The EPDS has been validated on Swedish samples (Wickberg & Hwang, 1997). It has 10 items, each rated on a scale of 0 to 2. The SPSQ is a Swedish modified version of the Parenting Stress Index (Abidin, 1990), with 34 items rated on a scale of 0 to 4. In the SCL-90, 90 items on a scale of 0 to 4 are subsumed into a mean score, the General Severity Index (GSI). Internal consistencies ranged from .82 to .96.

*Infant distress.* We used a Swedish translation of the mother-report Ages & Stages Questionnaire: Social-Emotional (ASQ:SE; Squires, Bricker, Heo, & Twombly, 2002). Each item is rated on a scale of 0 to 2. The mother can add if the symptom causes concern. Since there are three versions for our age range, we used mean scores for each questionnaire to enable comparisons. Internal consistencies ranged from .66 to .79.

*Mother–infant relationship.* The interviewer used the Parent-Infant Relationship Global Assessment Scale (PIR-GAS; ZERO

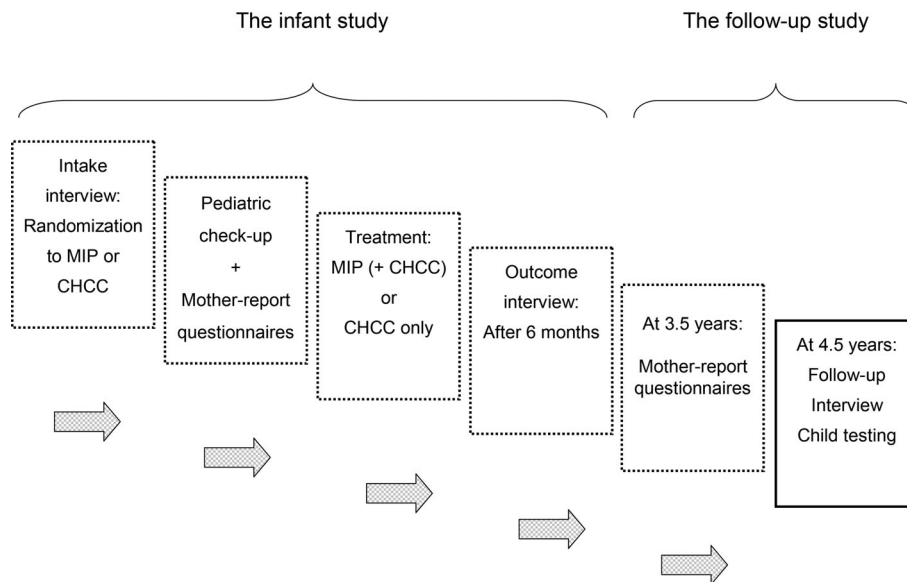


FIGURE 1. Design of the infant and the follow-up studies.

TO THREE, 2005). To check interrater reliability, an independent psychologist rated 20 pre- and posttreatment interviews. Intraclass correlation coefficients (ICCs) were .90 and .86, respectively.

*Mother–infant interaction.* Ten-minute video recordings were assessed by two independent raters using the Emotional Availability Scales (EAS; Biringen, Robinson, & Emde, 1998). The dimensions are maternal sensitivity, structuring, nonintrusiveness, and nonhostility, plus infant responsiveness and involvement. Each dimension score was divided by its range. Nonhostility was omitted due to low interrater reliability. ICCs for the remaining subscales ranged between .68 and .84.

## RESULTS

The sample's educational level was slightly higher than the Stockholm average. Pretreatment mean scores were at clinical levels for depression, stress, psychological distress, and infant behavioral and interaction problems. The mean score on the depression measure, the EPDS, was 11.9 ( $SD = 4.7$ ), which is about twice as high as scores among nonclinical Swedish mothers of babies (Seimyr, Edhborg, Lundh, & Sjögren, 2004; Wickberg & Hwang, 1997). The mean GSI scores on general psychological distress were 0.97 ( $SD = 0.56$ ), which is two (Fridell et al., 2002) or even three times (Börjesson, Ruppert, & Bågedahl-Strindlund, 2005) as high as that for nonclinical Swedish samples. The mean score on the stress measure, the SPSQ, was 2.97 ( $SD = 0.54$ ) to compare with a population-based sample with a mean score of 2.52 ( $SD = 0.56$ ; Östberg, 1998). The mothers' ratings of their babies' dysfunction on the ASQ:SE also were about twice as high ( $M = 2.0$ ,  $SD = 1.15$ ), as compared with a nonclinical American sample (Squires, Bricker, & Twombly, 2004). Maternal sensitivity as measured by the EAS was at 60% of the optimal level. The quality of mother–infant relationship as measured by

the PIR-GAS amounted to a mean score of 68.1 ( $SD = 12.1$ ). All in all, these figures reflected the sample's psychiatric medium-risk level. Furthermore, somatic illnesses such as multiple sclerosis, diabetes, epilepsy, and ulcerative colitis affected more than 10%.

The median of the MIP therapies was 23 sessions, with 2 to 3 hr per week. Adherence was rated posttreatment in interviews with the analyst and the mother, respectively. A nine-item list of salient features covered the analyst's contact with the mother and child and the interventions she or he used. Each item was rated on a scale of 1 to 4, where 4 indicated optimal adherence. The optimal total score was 36 and the mean score of 29 indicated an acceptable adherence. All mothers were asked about the scheduled CHC visits. There were no signs of any between-group differences concerning the mothers' compliance or the nurses' adherence to the CHC routines. One third of the CHCC dyads received brief additional psychological support. The results favored MIP on maternal depression, mother–infant relationships, maternal sensitivity, and on a marginally significant level, maternal stress. Effect sizes (Cohen's  $d$ ) were small to moderate. When measured by a standardized mean-change score that accounts for pretreatment differences between groups (Becker, 1988), they were higher.

It was concluded that MIP holds promise for more widespread use. A follow-up study was launched, and we will now report on its child outcomes. A second article will report on maternal outcomes and mother–child interactions, and a third article will be devoted to a qualitative study of the sample.

## THE PRESENT STUDY

### *Aims and Hypotheses*

The aim was to compare the long-term efficacy of two treatments. No infant effects had been found at the 6-months' follow-up. This could reflect a true null hypothesis or that the effects emerged



as “sleeper-effects,” but it also might result from measurement problems for such a young population. At 4½ years, assessments of children’s behavior are more reliable. The children have more means of expressing themselves and a vocabulary to participate in tests for which language is needed. Moreover, they can be alone with an interviewer. This enables the use of a comprehensive test battery unrelated to the mother’s assessments.

We hypothesized that the infants should develop better from MIP than from CHCC on measures of social, emotional, and global functioning including attachment representations. Such domains are likely to be influenced by the quality of the mother–infant interactions, which ought to be improved more by psychotherapy than by routine care. In contrast, we did not believe that cognitive functioning would be improved by psychotherapy in such a young population. We primarily included a measurement of intelligence to see if it confounded any between-group differences on the other measures. Secondarily, it was used to get a picture of the child’s power of perseverance and concentration.

One exception to our general hypothesis was mother-reported child functioning since we knew that the validity of such a measure might be problematic (Salomonsson & Sled, 2010); the mothers’ reports could reflect more closely their own psychological well-being than the children’s functioning as assessed by external raters. We hypothesized that such questionnaires would yield no between-group differences.

The interviewer also sought to categorize her impressions of the children’s characteristics and psychological well-being in a systematic and clinically meaningful way. In case a ranking order could be obtained, we hypothesized that there would be more children with optimal characteristics in the MIP group.

## METHOD

*Participants.* Of the 75 cases analyzed in the infant study, 71 appeared for the outcome interview 6 months after intake. They were asked if they were willing to post questionnaires at a child age of 3½ years and bring the child for an interview at 4½ years; everyone consented. At the time of the 3½-year questionnaires, 3 mothers declined further participation; 1 in the MIP group, 2 in the CHCC group. Of these 68 women, 1 mother in each group declined participation in the 4½-year interviews. Among the remaining 66 mothers, 6 provided incomplete data; 3 had just delivered a baby and declined the interview, but completed the questionnaires. Two mothers, 1 from each group, arrived with her child, but without the questionnaires. Despite several reminders, they did not return them. Finally, 1 mother did not want her child to be interviewed but consented to being interviewed herself. In summary, at the 4½-year follow-up, we gathered complete or partial data from 66 cases, 33 from each group, of the 80 cases originally randomized in the infant study, for a response rate of 82.5%. Figure 2 illustrates the participants, including the dropout cases.

*Interview procedure.* All interviews were video recorded. One author (B.S.), an experienced psychiatrist who had interviewed all

participants in the infant study, now interviewed the mother. Another author (M.W.S.), an experienced child psychologist, interviewed the child. M.W.S. was blind concerning the children’s previous history and treatment assignment. The interviews started with a joint welcoming, whereupon mother and child were separated into two adjacent rooms with video cameras. The mother was asked by B.S. to give a picture of her internal representations of her child and their relationship. Meanwhile, M.W.S. began by testing the child’s cognitive functioning since these tests caught the child’s interest and made him or her acquainted with the interviewer. About one third of the children had difficulties in leaving their mothers, but they were supported by the safe and well-structured test situation. Toward the end of this part of the interview, the child received a Lego toy with the instruction to assemble it with mother. Thereupon, mother and child reunited. After 12 min, they were served cookies and lemonade. In total, they were together for 20 min and then separated once again.

In the final part of the interview, the mother was asked by B.S. about major life events since the time of the infant study, such as divorce, the birth of new siblings, unemployment, and psychiatric or somatic disorders and treatments of mother or child. He also asked standardized questions about her child’s daily functioning such as behavior at home and at preschool, the relationships with friends and others, and so on. Finally, he handed her the questionnaires. He asked if she consented to requesting the preschool teacher to fill in a questionnaire on the child’s functioning. Meanwhile, the child was tested by M.W.S. on social and emotional functioning and was asked standardized questions about daily life, such as “What do you do when you are at home, at the preschool, what do you do together with your pals?” Thereafter, mother and child reunited, and there was a joint farewell.

## Child Measurements

Information was collected from four sources; reports from the mother and the preschool teacher, impressions of the interviewer, and assessments by the external raters.

## Instruments.

*Cognitive functioning.* The Wechsler Preschool and Primary Scale of Intelligence-III (WPPSI-III; Wechsler, 2002, 2005) was used to investigate if any between-group differences in cognitive functioning might confound treatment effects. These scores were thus assumed to reflect an independent variable. The test also gave impressions of the child’s perseverance and ability to concentrate, which were used as material for the categorization into *Ideal* types and the Children’s Global Assessment Scale (CGAS; Shaffer et al., 1983) ratings, as described later. We used the seven core subtests Block Design, Information, Matrix Reasoning, Vocabulary, Picture Concepts, Coding, and Comprehension, and then calculated three composites: Full-Scale IQ (WPPSI-III IQ), Verbal (WPPSI-III V), and Performance IQ (WPPSI-III P).

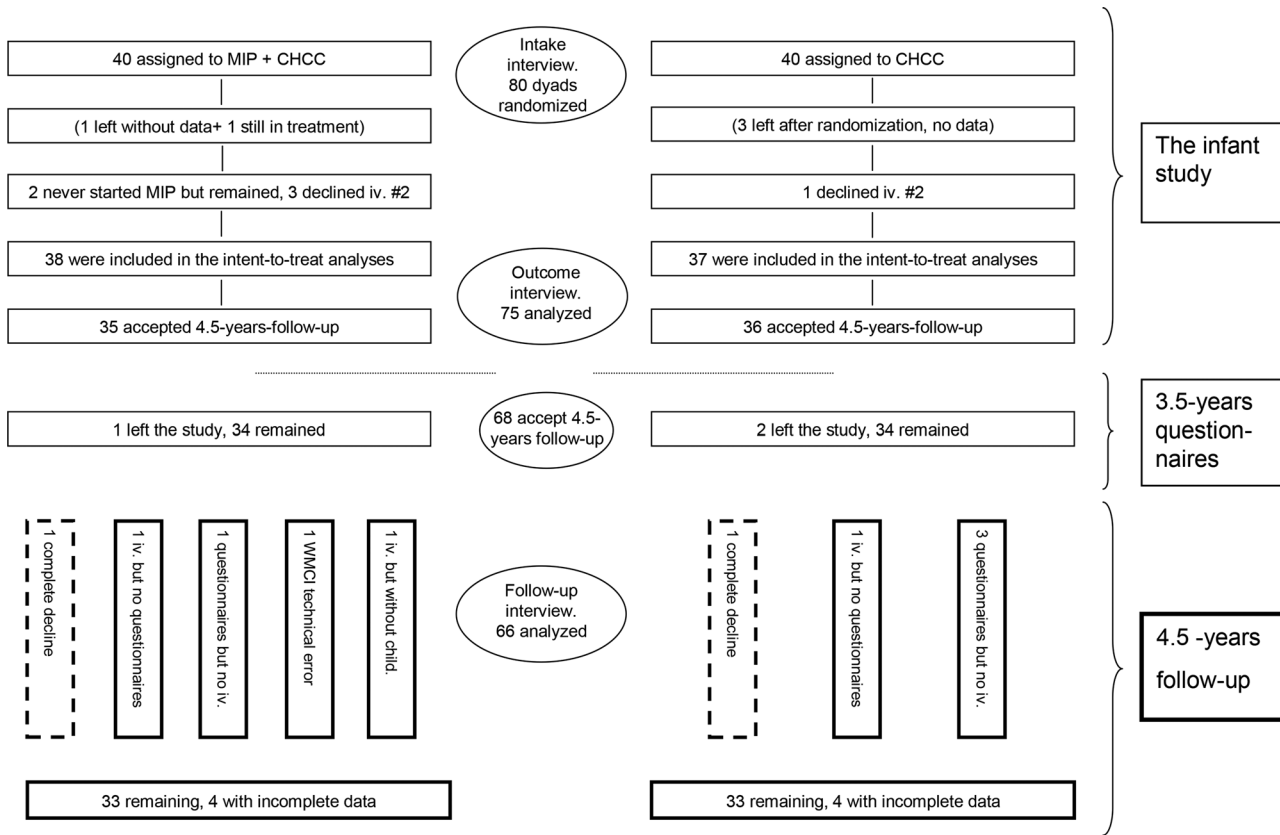


FIGURE 2. Flowchart of the infant and follow-up studies.

Joint welcoming of mother and child



FIGURE 3. The interview procedure.

The WPPSI-III is widely used in clinical settings and research. In its original version, reliability coefficients for the three composites ranged from .83 to .95. Concurrent validity was measured by correlating scores with the Mental scale of the Bayley Scales of Infant Development (BSID-II; Bayley, 1969), with correlations ranging from .61 to .80 (Wechsler, 2002). The reliability coefficients for the three composites of the Swedish version of the WPPSI-III (Wechsler, 2005) ranged from .96 to .97. Concerning

the choice of subtests for this age group, consultations were made with the researcher who has validated the Swedish version (E. Tideman, personal communication, September 10, 2009).

In the Machover Draw a Person test (Machover, 1949, 1951), the child is asked to draw two persons. Standardized questions are posed, such as “Is s/he happy/sad/angry? Does s/he have friends? What are his/her three major wishes?” It has been suggested (Blomberg & Cleve, 1997; Machover, 1949) that the drawings

express the child's impulses, emotional defenses, and object relations. This test thus measures both cognitive and emotional functioning (Daud, 2008).

The Machover test is still in clinical use. In contrast, researchers have not used it much due to the difficulties in assessing the drawings in a reliable and valid way (Smith & Dumont, 1995). We decided to create and quantify two dimensions: (a) the age adequacy of the first drawing's formal aspects (*Machover Formal*) and (b) the regulation of emotions expressed in the drawing and in the child's accompanying comments (*Machover Emotional*). The Machover formal targeted the completeness and wealth of details of the human body, and the coherence of the drawing. The scoring was made on an ordinal scale (1 = *below age*, 2 = *age adequate*, and 3 = *above age*). Level 1 implied a drawing where one could not see the human figure; only a circle with lines without facial details. Alternatively, the different body parts were unconnected. Level 2 implied a human figure with a head having eyes, a mouth, or both. The legs should be fixed to the head or the trunk. The drawing also might contain two arms and some fingers and toes. On Level 3, one also could identify details such as hands, feet, hair, eyelashes, teeth, nose, ears, clothes, facial expressions, or additional objects that clarified the situation.

All ratings were made by two external and uninformed raters with long experience of assessing and publishing on children's drawings (Britta Blomberg and Elisabeth Cleve). Each rater coded half of the material. Intercoder agreement was calculated by comparing 15 assessments; Spearman's  $\rho = .81$ ,  $p < .001$ .

*Social and emotional functioning.* The Machover Emotional scale measured the child's capacity to handle emotional material linked with the drawing. The scoring was made on an ordinal scale (1 = *emotions handled nonadequately*, 2 = *emotions handled adequately*, and 3 = *emotions handled creatively*). Level 1 implied that the child verbalized no emotions at all or acted them out in a dysregulated way, either in words ("this is just poo-poo, fart, snot," etc.), or in behavior (jumping, crawling under the table, running around). On Level 2, emotions were expressed more coherently, as when a child pointed to a stern face saying "He is angry." On Level 3, the child also elaborated on the characters' emotions in a way that indicated a more developed mentalizing capacity, for example, by reflecting on their relationships. Intercoder agreement was assessed similarly to that for the Machover Formal scale; Spearman's  $\rho = .81$ ,  $p < .001$ .

The Ages and Stages Questionnaire: Social-Emotional (ASQ:SE; Squires, Bricker, Heo, & Twombly, 2001) is a parent-completed questionnaire designed to identify developmental problems in children up to 5 years of age. It focuses on the child's social and emotional behavior in the areas of self-regulation, compliance, communication, adaptive behavior, autonomy, affect, and interactions with people. In the constructors' study, internal consistency was .91 for 5-year-olds. Test-retest reliability was .94. Our internal consistency was .80 for 3½-year-olds and .81 for 4½-year-olds.

The Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997, 2001) is a screening instrument for prosocial

behavior and psychopathology in children aged 3 to 16 years. The Swedish version (Smedje, Broman, Hetta, & von Knorring, 1999) was given to the mother (SDQ-M) and the preschool teacher (SDQ-T). It was instituted only at 4½ years when the mother was interviewed and could be asked for her consent to the teacher-report version. The SDQ contains 25 propositions covering the child's prosociality, hyperactivity, emotional symptoms, conduct problems, and peer problems. The score range is 0 to 50, with 0 indicating the most optimal score. In an earlier study (Goodman, 2001), retest stability after 4 to 6 months was .62, and internal consistency was .73. The validity was gauged by comparing with the *Diagnostic and Statistical Manual for Mental Disorders, fourth edition* (American Psychiatric Association, 1994) diagnoses assessed through interviews. Among the 10% of the children with the least optimal scores, the odds ratio of having a diagnosis was 15. Our internal consistency was .82 for the SDQ-M and .84 for the SDQ-T.

The general level of functioning was rated through the CGAS (Shaffer et al., 1983). Assessments do not rely on diagnostic terms but on descriptions of functioning and symptoms subsumed in one score ranging from 0 to 100, with 0 being the least optimal. The study by Shaffer et al. (1983) found a test-retest reliability after 6 months of .85. Concerning validity, Weissman, Warner, and Fendrich (1990) investigated the correspondence between CGAS ratings based on interviews with mother and child, respectively, and external information. The agreement between the three sources of information was satisfactory. Validity studies with other measures have been performed (Bird, Canino, Rubio-Stipec, & Ribera, 1987; Mathiassen et al., 2012; Schorre & Vandvik, 2004; Sourander & Piha, 1997).

In one study (Lundh, Kowalski, Sundberg, Gumpert, & Landén, 2010), the ICC was .73 among untrained raters and .92 among trained raters. Our assessments were made by the most experienced rater in Sweden (Anna Lundh) in consensus with the interviewer of the children (M.W.S.). They were comprehensive in that they were based on the child's answers and behavior in the interview, the mother's interview transcripts, and scores on the ASQ:SE, the SDQ-M, and the SDQ-T.

*Attachment representations.* We used the Story Stem Assessment Profile (SSAP; Hodges, Steele, Hillman, Henderson, & Kaniuk, 2003). The interviewer told the child the beginning of a story highlighting everyday family scenarios containing an inherent dilemma. The child was then asked to show and tell what happened next. This allowed an assessment of how the child perceived family roles, attachments, and relationships. Seven stories were used.

The coding has been manualized (Hodges, Hillman, Steele, & Henderson, 2002). Each story completion is rated for different themes on a scale of 0 to 2. These points are then subsumed into one score for each attachment dimension: secure, avoidant, ambivalent, and disorganized. High scores are optimal for the secure dimension and nonoptimal for the other dimensions. The constructors regard the scores as reflecting continuous variables.

A study by Oppenheim, Emde, and Warren (1997) assessed intercoder agreement on the MacArthur Story-Stem Narratives (Emde, Wolf, & Oppenheim, 2003), a test similar to the SSAP. Cohen's  $\kappa$  was .92. Validity was measured by comparing with the CBCL. At age 4½ years, externalizing behavior problems correlated with narrative coherence and aggressive themes in the stories,  $r = -.27$ ,  $p < .05$ , and  $r = .64$ ,  $p < .001$ , respectively. In our study, an external and trained rater scored all transcripts from video-recorded sessions. Interrater reliability was assessed on 15 children by comparing with the interviewer's scores. The ICCs exceeded .88. The outcome analyses used the external rater's scores.

*Characteristics and psychological well-being.* The child interviewer used Ideal types (Wachholz & Stuhr, 1999). This is a method in which the observer extracts a common denominator out of a series of single cases. It is an inductive procedure of analyzing data (Philips, Wennberg, & Werbart, 2007; Philips, Werbart, Wennberg, & Schubert, 2007) in which one subsumes impressions into a few succinct descriptors. Ideal types have been used to categorize psychoanalytic processes and results (Leuzinger-Bohleber & Target, 2002; Werbart et al., 2011) and types of suicidal behavior (Lindner, Fiedler, Altenhofer, Gotze, & Happach, 2006).

The Ideal type method also had been used in the infant project. The interviewer subsumed his clinical impression of the mothers into five types, which were later distilled into two overarching maternal Ideal types. They reflected the mother's sense of agency in the baby's disorder and their interest in taking part in psychotherapeutic work. The infants were divided according to their clinical status; those who were affected and unaffected, respectively, by the relationship disorder. The validity of especially the infant types was corroborated by their correlations with several quantitative measures of infant functioning and interactive behavior. Both infant and maternal types proved to moderate outcomes on maternal sensitivity and mother–infant relationships (Salomonsson & Sandell, 2011b).

In the present study, the child interviewer used the same method to describe the child's essential characteristics. We hesitate to apply the term *personality* to a 4½-year-old child since this term refers to an operationalized concept. On the other hand, it is customary to speak of children as being cheerful, gloomy, shy, quarrelsome, and so on—adjectives which subsume the child's *characteristics and psychological well-being*. During her meeting with the child, the interviewer assembled her impressions of the child and labeled them with personally invented words. To enlarge the basis of her categorizations, she also took into account the material upon which assessments of the CGAS, the Machover, the SSAP, and the WPPSI-III were based. The Ideal types did not aim to ascribe *numbers* to the levels of functioning or psychopathology; rather, they described the child's individuality in idiosyncratic *words*. The interviewer first sketched her impressions of each child into formulations such as “curious,” “don't you get inside me,” “good girl,” “a fighter,” “full of ideas,” “a troublemaker,” “scared,” “precocious,” and “a teaser.” These simple words subsumed her observations of and her emotional reactions to the child. In a

second step, she compared these catchwords across all children, with the aim of reaching a limited number of descriptors. She opted for four types:

- The *Open* child seemed lively, confident, and open with the interviewer. The child was often relaxed, constructive, and creative in his or her play, and took initiatives. The mood was mostly happy or cheerful.
- The *Orderly* child was competent, kind, and often a bit moderate. Sometimes she or he was strict or inhibited. She or he seemed expectant, cautious, or task-oriented with the interviewer. The mood was often neutral.
- The *Anxious* child showed overt anxiety, inhibition, or marked shyness. She or he often had pronounced difficulties in separating from mother and demanded her to stay with the interviewer in a craving or clinging way. Sometimes the child was restless. The capacity for attention seemed to vary with the level of anxiety.
- The *Provocative* child was annoying, sometimes rejecting the interviewer's instructions. She or he also could be overtly aggressive or spiteful. The mood was often irritable, anxious, or cocky.

For a reliability check, an external child psychotherapist rated video-recorded interviews of 15 cases. Since the child types were qualitative categories, intercoder agreement was used to measure reliability. Cohen's  $\kappa = .71$ ,  $p < .001$ . To arrive at a more parsimonious division, one that also reflected the interviewer's clinical concern, she then collapsed the four groups into two overarching types; the *OK* children and the *Troubled* children. These two terms may seem scientifically imprecise. We emphasize that they denote the interviewer's clinical impression of each child rather than a strictly defined diagnosis. The *OK* type consisted of the *Open* and *Orderly* children, and the *Troubled* type covered the *Anxious* and *Provocative* children. Cohen's  $\kappa = .62$ ,  $p = .010$ .

### Statistics

SPSS Version 21.0 was used. Scores were considered to be univariate outliers if  $z$ -transformed scores exceeded 3.29,  $p < .001$ , two-tailed test (Tabachnik & Fidell, 2007); three such scores were removed. Multivariate outliers were identified by calculating Mahalanobis distance through a multiple regression analysis; no such cases were found. Missing data were very rare, except for the SDQ-T; one fifth of the mothers declined handing the questionnaire to the preschool teacher. Little's MCAR test on all outcome measures yielded  $\chi^2 = 38.421$ ,  $df = 30$ ,  $p = .139$ . Data were thus missing at random, and no scores were imputed.

Outcomes on the ASQ:SE, the only variable with several measurement points, were analyzed with multilevel modeling (MLM). The intent-to-treat analysis utilized all the cases from the infant study, including all dropouts up to 4½ years of age ( $n = 75$ ). For the remaining variables with only one measurement



**TABLE 1.** Principal Components Analysis With Oblimin Rotation. Factor Loadings > .40. Questionnaire Scores, External Ratings of Child Functioning, Drawings, and Attachment Representations (Correlation coefficients between the components are presented at the bottom of the table.)

	Components			
	Drawings	Functioning	Amb.-Disorg. Att.	Sec - Avoidant Att
Machover Formal	.877			
Machover Emotional	.824			
SDQ-T	-.558			
SDQ-M		.918		
ASQ:SE		.932		
CGAS		-.409		-.456
SSAP Secure				-.912
SSAP Avoidant				.889
SSAP Ambivalent			.926	
SSAP Disorganized			.958	
%Variance Explained	26.8	21.1	17.7	11.6
Drawings	–	-.193	-.136	-.237
Functioning		–	-.070	-.035
Amb-Disorg. Att			–	.053

Note. Amb.-Disorg. Att = Ambivalent-Disorganized Attachment; ASQ:SE = the Ages & Stages Questionnaire: Social-Emotional; SDQ = the Strengths and Difficulties Questionnaire; Sec.-Avoidant Att. = Secure-Avoidant Attachment; CGAS = the Children's Global Assessment Scale; M = mother. Machover Formal = the Machover test, Formal scale; Machover Emotional = the Machover test, Emotional scale; SSAP Secure = the Story Stem Assessment Profile, secure attachment representation; SSAP Avoidant = the avoidant representation of SSAP; SSAP Ambivalent = the ambivalent representation of SSAP; SSAP Disorganized = the disorganized representation of SSAP; T = teacher.

point at 4½ years, we used Mann-Whitney *U* tests,  $\chi^2$  tests, and *t* tests. Data from 60 to 66 participants could be used, except for the SDQ-T where 51 teachers provided data.

### Ethical Approval

The project was approved by the Regional Ethical Review Board in Stockholm, Dnr 2009/1334-32. Mothers were informed that they could leave the project at any time. Video recordings were made only with their consent. For further details, we refer to the infant study (Salomonsson & Sandell, 2011a).

## RESULTS

### Child Functioning: Distribution of Scores

The Full-Scale IQ (WPPSI-III IQ) in the MIP group was 105.1, and 106.4 in the CHCC group,  $t = 0.427$ ,  $p = .671$ , and the results were similar on the WPPSI-III V and the WPPSI-III P. The randomization had thus yielded two groups with equal intelligence. To investigate patterns and intercorrelations among the dependent variables at 4½ years, a principal components analysis was performed with Oblimin rotation. We used the scores on all questionnaires, the CGAS, the Machover ratings, and the SSAP ratings. The analysis yielded four factors with eigenvalues > 1. The solution explained 77% of the variance. The first component, the "Functioning" component, covered the mother-report questionnaires on child functioning and the CGAS. The "Ambivalent-Disorganized Attachment" component subsumed the two corresponding SSAP scales. The "Secure-Avoidant Attachment" component comprised

the corresponding SSAP scales and the CGAS. The "Drawings" component covered the two Machover scales, but also included loadings on the SDQ-T. The correlations between the components ranged in strength between .035 and .237. Analysis is presented in Table 1.

Mean values on ASQ:SE at 4½ years were at nonclinical levels ( $M = 0.93$ ,  $SD = 0.79$ ) whereas the average CGAS score ( $M = 73.70$ ,  $SD = 14.49$ ) was just above the clinical cutoff point of  $\leq 70$  as cases and  $> 70$  as noncases (Bird et al., 1990). Among the SSAP measures, only the Secure Attachment could be considered nonoptimal. The girls had more optimal scores on the WPPSI-III IQ, the CGAS, the Machover Formal and Emotional, and the SSAP Secure.

As for the two overarching Child types, there were more girls in the OK group and more boys in the Troubled group,  $\chi^2 = 3.921$ ,  $p = .048$ . We also investigated their associations with the other measurements. As seen in Table 2, the OK children often had more favorable scores than did the Troubled children.

### Child Functioning: Between-Group Differences

There were no differences on major life events, except for a tendency,  $p < .10$ , among the CHCC children to be more afflicted by somatic illness: One case of benign heart disease, one of pyelonephritis, and one of recurrent ear, nose, and/or throat infections with hospitalization. There also was a tendency for MIP mothers to have conceived another child after the infant study. No further changes (e.g., unemployment, social upheavals, or death in the family) were reported.

**TABLE 2.** Differential Scores Concerning the Two Overarching Child Types. Only Instruments With Any Significant Differences Are Reported. Positive Values Indicate a More Optimal Rating for the First Type, Except for the SSAP Avoidant

	OK–Troubled
CGAS	20.54. $F(1, 59) = 61.36, p = .000.$
WPPSI-III IQ	8.09. $F(1, 58) = 8.27, p = .006.$
WPPSI-III P	8.61. $F(1, 59) = 8.89, p = .004.$
Machover Formal	0.29. $F(1, 58) = 4.09, p = .048.$
Machover Emotional	0.38. $F(1, 58) = 5.50, p = .022.$
SSAP Secure	0.62. $F(1, 59) = 4.34, p = .042.$
SSAP Avoidant	-0.35. $F(1, 59) = 8.28, p = .006.$

Note. WPPSI-III IQ = the Wechsler Preschool and Primary Scale of Intelligence, full scale; WPPSI-III P = Performance subscale of the WPPSI-III; Machover Formal = the Machover test, Formal scale; Machover Emotional = the Machover test, Emotional scale; SSAP Secure = the Story Stem Assessment Profile, secure attachment representation; SSAP Avoidant = the avoidant representation of SSAP.

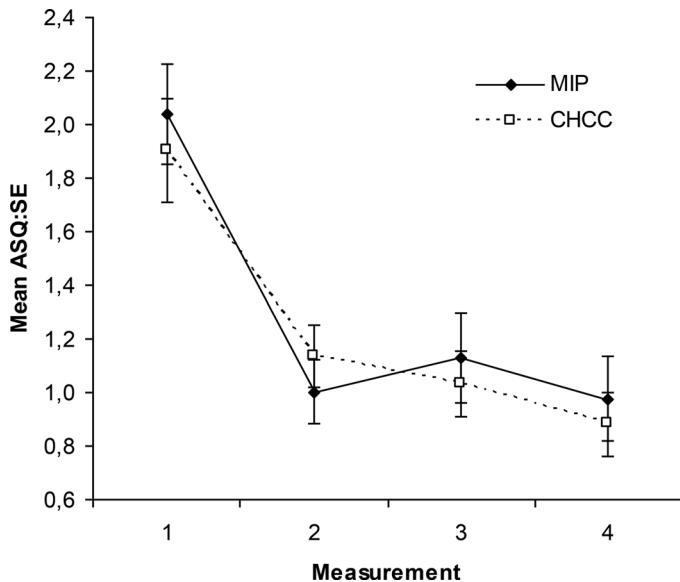


FIGURE 4. ASQ:SE scores at intake, 6 onths later, and at 3½ and 4½ years.

Machover ratings were regarded as reflecting a ranking scale and the Child types were analyzed as nominal variables whereas the remaining instruments were treated as continuous variables. For the ASQ:SE, we had data from four measurement points: at intake in infancy, 6 months later, and at child ages of 3½ and 4½ years. To calculate the development of the between-group differences, we performed an MLM analysis. Treatment group, measurement points, and their interaction were used as predictors. As seen in Figure 4, the ASQ:SE decreased over time,  $p < .001$ . No differences were found between the treatment groups or the interactions between measurement points and treatment groups. We then performed separate analyses for each interval (i.e., Measurements 1 and 2, 2 and 3, 3 and 4). The scores decreased between Points 1

and 2,  $p < .001$ , and Points 3 and 4,  $p = .020$ . Again, between-group differences and interactions were not significant. Since the score distributions at 3½ and 4½ years were positively skewed, we also performed an MLM analysis on scores obtained through logarithmic transformations. The results were essentially the same.

Three more variables were positively skewed: the SDQ-M, the SDQ-T, and the SSAP Disorganized. We also performed Mann–Whitney  $U$  tests, and the nonsignificant effects did not substantially differ from the  $t$  tests.

As seen in Table 3, the CGAS showed significant between-group differences in favor of the MIP group, with an effect size (Cohen’s  $d$ ) of .69. Using a cutoff point of 70/71 (Bird et. al, 1990), we found significantly more noncases in the MIP group. We then investigated how the two Child types were distributed in the two treatment groups. As Table 3 indicates, there were more than twice as many OK children and less than half as many Troubled children in the MIP group, as compared with the CHCC group. To measure the effect size, we calculated the odds ratio (5.78; 95% CI = 1.91–17.44) and the relative risk (2.54; 95% CI = 1.38–4.82). The Gender  $\times$  Treatment interaction effect, as analyzed with binary logistic regression, was not significant. Treatment had thus affected boys and girls on an equal level on this outcome.

We then analyzed the Gender  $\times$  Treatment interaction effects on the quantitative variables, and the only significant effect was found on the SDQ-T; boys in CHCC treatment had 5.0 less optimal points than did MIP boys,  $t = 2.326, p = .031$ , while the effect of treatment was not significant for girls. We also analyzed if the number of MIP sessions affected outcomes and found no significant correlations. In the CHCC group, almost one third of the mothers had received additional psychological support during the infant study. We analyzed if this had affected outcomes, but no significant correlations were found.

## DISCUSSION

A sample of 4½-year-old children in Stockholm had been randomized during infancy to MIP or to CHCC. About 4 years later, this follow-up study showed effects in favor of MIP on global functioning. In addition, the MIP group now contained more “OK” children, as rated by an experienced and external clinician. In contrast, the CHCC group contained more “Troubled” children. No effects were shown on other measurements on socioemotional development, including attachment representations as rated by mothers, teachers, and external raters.

These findings raise several topics of discussion: (a) Why would a psychotherapy provided in infancy yield effects 4 years later, and how could one explain the mechanisms of change? (b) On which domains would psychotherapy be likely to yield effects, and were the assessment methods relevant in tapping these domains? (c) To what extent were our findings confounded by extra-treatment factors such as gender and intelligence? (d) To what extent did pretreatment data and posttreatment results among our children differ from other investigated samples? (e) How do we interpret, on a general level, our findings?

TABLE 3. Between-Group Comparisons at 4½ Years of Age With Statistical Tests

<i>n</i>	MIP			CHCC			Comparison		
	<i>M</i>	<i>SD</i>		<i>n</i>	<i>M</i>	<i>SD</i>	Statistics	Value	<i>p</i>
Marital Discord or Divorce	9			10			$\chi^2$	0.198	.657
Psychotherapy, M	15			18			$\chi^2$	1.071	.301
Psychotropic Drugs, M	3			7			$\chi^2$	2.230	.135
New Somatic Illness, M	2			1			$\chi^2$	0.286	.593
New Somatic Illness, Child	0			3			$\chi^2$	3.363	.067
More Children in Family	20			12			$\chi^2$	2.721	.090
ASQ:SE	32	.98	.90	32	.88	.68	<i>t</i> test	0.480	.633
SDQ-M	32	8.17	5.54	31	7.39	5.19	<i>t</i> test	0.580	.564
SDQ-T	24	5.71	4.32	27	6.59	5.31	<i>t</i> test	-0.647	.520
CGAS	31	78.39	12.80	30	68.87	14.74	<i>t</i> test	2.696	.009
CGAS Cases/Noncases	8/23			16/14			$\chi^2$	4.841	.028
Machover Formal	31	1.98	.58	29	1.85	.55	M-W	404.5	.431
Machover Emotional	31	1.87	.68	29	1.79	.62	M-W	423.5	.671
SSAP Secure	31	2.22	1.05	30	2.32	1.33	<i>t</i> test	-0.304	.762
SSAP Avoidant	31	1.05	.48	30	1.16	.52	<i>t</i> test	-0.915	.364
SSAP Ambivalent	31	.96	.73	30	.84	.61	<i>t</i> test	0.719	.475
SSAP Disorganized	31	.80	.84	30	.63	.58	<i>t</i> test	0.883	.381
OK/Troubled Children	21/10			8/22			$\chi^2$	10.314	.001

Note. CHCC = Child Health Centre Care; M = mother; MIP = Mother-Infant Psychoanalysis; ASQ:SE = the Ages & Stages Questionnaire: Social-Emotional; SDQ-M = the Strengths and Difficulties Questionnaire-Mother; SDQ-T = the Strengths and Difficulties Questionnaire-Teacher; SDQ = the Strengths and Difficulties Questionnaire; Sec.-Avoidant Att. = Secure-Avoidant Attachment; CGAS = the Children's Global Assessment Scale; M = mother. Machover Formal = the Machover test, Formal scale; SSAP Secure = the Story Stem Assessment Profile, secure attachment representation; SSAP Avoidant = the avoidant representation of SSAP; SSAP Ambivalent = the ambivalent representation of SSAP; SSAP Disorganized = the disorganized representation of SSAP; M-W = Mann-Whitney *U* Test; OK/Troubled Children = the Overarching Child types; Value = value of the statistical parameter.

### Long-Term Effects

It might seem unfounded to hypothesize that a relatively brief mother-infant psychotherapy would help children 3½ years later. A British study (Cooper et al., 2003; Murray et al., 2003) had shown no such effects, but the therapeutic setting and methods were not the same, and therefore the mechanisms of change might differ. The MIP treatments were longer and with a higher frequency and included a double focus on mother and infant, which probably gave better possibilities of handling distress in both participants. Similarly to the techniques in the British study, the MIP mothers could vent their distress with the therapists. In addition, and specific to the MIP mode, the therapist addressed the infant about his or her distress, which may have diminished the child's feelings of helplessness and loneliness. This was thought to promote the child's attachment to the mother. Furthermore, the Swedish mothers had searched actively for help with their babies rather than being recruited on the basis of depression questionnaires. Thus, they were plausibly more motivated to enter psychotherapy. Finally, the length of therapy was adapted to each mother's personal needs. Mothers with a more profound distress could thus receive more extensive help.

### The Investigated Domains

When assessing psychotherapy outcomes, changes in symptoms and behaviors must be measured. Psychodynamic therapy also

aims at a positive development of the patient's personality, attachments, and interactions (Target, 2002); accordingly, research studies must find suitable assessment methods. A specific challenge in child psychotherapy is to find appropriate measurements on these domains. Our children's symptoms and behaviors were assessed by mothers, teachers, and a clinician via questionnaires and an interview. These assessments were thought to cover various effects of psychotherapy (Lambert, 2004). The principal component analysis showed that the instrument scores were not substantially intercorrelated. It was possible to identify four factors which covered several differential measurement areas. We also included qualitative assessments to achieve triangulation (Elliott, Fischer, & Rennie, 1999; Jick, 1979; Malterud, 2001). This is regarded as a more fruitful way of measuring psychotherapy outcomes than using either qualitative or quantitative methods (Hill, Chui, & Baumann, 2013).

### Confounding Factors

The children's intelligence was thought to be uninfluenced by the treatments. Since there were no between-group differences on the WPPSI-III scores, we concluded that intelligence had not confounded any outcomes. Among other possible confounding factors, we identified family life events and child gender. For life events, there were no significant between-group differences, and they thus could not explain the differential effects. For gender, in the entire

sample, the girls had more optimal scores on the WPPSI-III IQ, the CGAS, the Machover Formal and Emotional, and the SSAP Secure. Concerning treatment effects, however, MIP seemed to have affected boys and girls on an equal level. One exception was the teachers' ratings of child functioning where the MIP boys scored better than did the CHCC boys.

### *Comparison With Other Samples*

According to the mother and the teacher reports, our children functioned at a normal level. In contrast, many expert-rated measures were close to clinical levels. The mean CGAS score was just above the recommended clinical cutoff point (Bird et al., 1990); 68.9 for the CHCC children and 78.4 for the MIP children. Other studies have shown mean pretreatment scores between 53 (Target & Fonagy, 1994a) and 60 (Muratori, Picchi, Bruni, Patarnello, & Romagnoli, 2003) and posttreatment scores between 64 (Lundh, 2012) and 78 (Muratori et al., 2003).

Our sample had certain difficulties concerning attachment representations in comparison with a British community sample of 4- to 6-year-olds (Hillman & Hodges, 2011); our sample scored lower on secure attachment representations. For the Machover ratings and the Ideal types, no community data exist for comparison.

### *Interpretation of Findings*

The CGAS showed a mean effect of 10 points in favor of the MIP children. The effect size ( $d = 0.69$ ) was moderate (Cohen, 1988). The infant study had not used this instrument since it was not devised for such young children. At intake, there had been no between-group differences on any infant assessments. We therefore interpret that the CGAS effects at 4½ years cannot be explained by differences in functioning at the start. Thus, they may mirror a true treatment effect. We speculate that the scores may be influenced by psychotherapy in the following way: The ratings were based on how the child was functioning at home and in preschool, and how she or he dealt with strong emotions and conflicts. To the extent that these capacities were affected by conflicts and difficulties in the mother–infant relationship, we assumed that an early intervention might have a positive, long-term effect.

Regarding clinical significance of the CGAS effects, this instrument has been used as an outcome variable in a number of child psychotherapy studies. In a study of child psychoanalysis (Target & Fonagy, 1994a, 1994b), the 2- to 6-year-old children improved by 9 points after nonintensive treatments and 13 points after intensive treatments; these effects were obtained directly after therapy. Odhammar, Sundin, Jonson, and Carlberg (2011) investigated a group of 5- to 10-year-old children in lengthy therapies and found an improvement of 17 points directly after therapy. Ratings were made by the therapists, who may have been biased in their evaluations. Muratori et al. (2002) investigated effects of short-term psychotherapy on a group of 6- to 10-year-old children. After 18 months, the CGAS had improved 16 points. In a larger sample from this study (Muratori, Picchi, Bruni, Patarnello, & Romagnoli,

2003), scores had improved 14 points after 6 months and 17 points after 2 years. Our follow-up was made around 3½ years after treatment, and we conclude that the range of score improvements are comparable to these studies and are clinically significant.

The qualitative assessments of the children's psychological well-being (Child types) were set up with no implicit ranking order. Nevertheless, a ranking order between the OK and Troubled children was supported by their differential scores on several quantitative variables. There was a preponderance of OK children in the MIP group and of Troubled children in the CHCC group. The method of Ideal types also had been used in the infant study to classify mothers and infants (Salomonsson & Sandell, 2011b). Two types of babies were established at intake: those who seemed clinically affected and nonaffected, respectively, by the relationship disorder. Since these baby types were evenly distributed in the two treatment groups, the between-group differences in our present study should reflect a true treatment effect. To interpret the magnitude of the odds ratio (5.78), we refer to two conventions. One (Chinn, 2000) has suggested an ln-transformation of the odds ratio and then dividing it by 1.81. This yields a value comparable to Cohen's  $d$  of 0.97; that is, a large effect size. Another convention (Haddock, Rindskopf, & Shadish, 1998, p. 342) has suggested that an odds ratio above 3.0 indicates "strong relationships."

Regarding the questionnaires on child functioning, our null hypothesis could not be rejected for the mothers' reports. One might argue that this reflected a reporter bias; however, such an argument could not be applied to the preschool teachers since they were less emotionally involved with the children. They reported no between-group differences; however, note that they considered the boys in the MIP group to be functioning better than their peers in the CHCC group.

The SSAP scores did not show any between-group effects. The relevance of this finding is not clear. It has been argued that there is no direct connection between attachment categories and psychopathology (Minnis et al., 2009; Sroufe et al., 2005). On the other hand, instrument scores have been shown to be associated with maternal psychopathology (Pass, Arteché, Cooper, Creswell, & Murray, 2012), children's attention and achievements at school (Jacobsen, Edelstein, & Hofmann, 1994), and preschoolers' verbal IQ (Stievenart, Roskam, Meunier, & van de Moortele, 2011). Another study (Stadelmann, Perren, von Wyl, & von Klitzing, 2007) has shown that children's representations on a story stem task were the only predictor of changes in symptoms/strengths between 5 and 6 years of age. Most of these studies thus confirm the validity of the story stem test. For this instrument, our hypothesis was disconfirmed.

The Machover test has been used in clinical settings, but quantitative scoring methods for research purposes are not well-established. Especially for young children, there are difficulties in assessing the drawings and getting valid answers to the questions (Blomberg & Cleve, 1997). In an effort at quantification, we developed two scales based on the raters' assessments of the drawings and the child's comments; the Machover Formal and Emotional scales. No between-group differences were found. Our



hypothesis was thus disconfirmed. We speculate that this test might be a valuable method of approaching the child's internal world in a psychotherapeutic setting. In contrast, it seems less valid for assessing his or her emotional and cognitive functioning in a research study.

To interpret the variety of our findings, we will first discuss, on a more general level, the validity of our outcome measures. Three factors need to be pointed out; they were amassed from informants with different biases, they were both explicit and implicit (Josephs & Bornstein, 2011), and they were collected in situations with different stress levels. Informants are known to differ when they assess childhood psychopathology (De Los Reyes & Kazdin, 2005); they have "different motivations for providing ratings of children and have different thresholds or perceptions of what constitutes abnormal behavior in a given child" (p. 483). Our mothers might have been biased when assessing their children since they were emotionally close to them. As for the teachers' ratings, they were generally more optimal than were those of the mothers. This could reflect their different thresholds of rating a disturbed behavior or, alternately, that the children showed less behavior problems in preschool. We must leave this question unanswered.

Josephs and Bornstein (2011) recommended combining explicit measures that "best predict conscious attitudes and goal-directed behaviors" with implicit measures that "predict spontaneous behaviors that are exhibited 'mindlessly,' without conscious awareness" (p. 423). Some of our measures were explicit (the questionnaires), and some were implicit (the Child types). The CGAS, the SSAP, and the Machover scales quantified implicit phenomena, but only the CGAS was based on numerous informants and also comprised explicit measures. The CGAS is thus a more comprehensive measure of the child's functioning and well-being. Another measure based on an overall evaluation was the Child types. Children's ability to report verbally on their well-being is restricted. Furthermore, they cannot fill in a questionnaire, and the mother might be biased in her reports. We therefore speculate that comprehensive assessments by a clinician should be considered a valuable tool in evaluating young children's functioning and emotional well-being.

Another issue concerns to what extent the assessments were influenced by the children's stress during data collection. In the interview, they were separated from their mothers to be with a stranger, they were asked questions that challenged their performance, and they were requested to comment on anxiety-laden stories. This stressful test situation enabled the interviewer to discern flaws in the children's psychological functioning, which their parents or teachers perhaps did not consider in their reports. It is known that a subject's test responses may be influenced not only by his or her stress tolerance but also by other character traits which one can distinguish only after "considering test data in the light of the patient's history, observed behavior, and circumstances surrounding the evaluation" (Meyer, 1999, p. 6). We suggest that the interviewer was in a particularly suitable position when she made the CGAS and the Child types assessments. She could

weigh together various "hard" data such as the questionnaires scores with "soft" data such as her impressions from the interview.

Arguing in an opposite direction, one might claim that the interviewer's assessments represented unfounded generalizations from a stressful test situation (De Los Reyes & Kazdin, 2005). However, the substantial intercorrelations between several quantitative instruments and the Child types and the CGAS indicate that the latter two reflected something more substantial than the interviewer's idiosyncratic impressions. The interview stress helped reveal dormant child psychopathology whereas the questionnaires did not always differentiate between "genuine" and "illusory" mental health (Cousineau & Shedler, 2006; Shedler, Mayman, & Manis, 1993). This leaves us with the task of explaining why no differences emerged on the Machover and the SSAP, which were collected under stress. It might reflect a true absence of effects. Alternatively, as argued earlier, these instruments covered the child's functioning in a less comprehensive way than did the CGAS and the Child types.

One also needs to consider the interviewers' allegiance (Luborsky et al., 1999; Markin & Kivlighan, 2007; Munder, Brüttsch, Leonhart, Gerger, & Barth, 2013), which might have affected the evaluations. Such influence is seldom clearly acknowledged in mother-infant psychotherapy studies. Luborsky et al. (1999) stated that the researchers' allegiances are significantly associated with the effect sizes of the treatment outcomes. In our study, both interviewers were psychoanalysts. Everything was done to minimize the effect of allegiance, such as guaranteeing that the child interviewer, similarly to all other raters, was blind to treatment assignments and case histories.

One also must consider the influence of the mothers' demand characteristics (Orne, 1962). Every mother knew about her treatment assignment and that both interviewers were psychoanalysts. The MIP mothers might have had more positive expectations of belonging to "the researchers' team." In addition, in the infant study, they had shown more favorable outcomes and might be more positive to their treatments than the CHCC mothers. If so, this could be transmitted to the children, making them more at ease with the interviewer. It also might have influenced the mothers' interview answers about their children's functioning. If such factors were at work, they would influence the two ratings that were based on the interviewer's impressions: the CGAS and the Child types. On the other hand, such bias, if existent, also should have affected the mothers' questionnaire ratings. This was not the case. In the end, we acknowledge the possible influence of the interviewers' allegiance and the mothers' demand characteristics, although their impact cannot be calculated.

Concerning the dose effect, outcomes did not differ between the two thirds in the CHCC group who merely received nurse care and the one third who received additional, brief psychological help. In the MIP group, outcomes were uncorrelated with the number of sessions. This was in contrast to Singleton's (2005) findings that treatments of 11 to 20 sessions were better than either shorter or longer treatments. Longer treatments were perhaps less effective because, as she speculated, "the families that receive long term

therapy could have more difficulties, or the effects of [parent–infant interaction interventions] do not last” (Singleton, 2005, p. 95). In the infant study, the interviews had confirmed that mothers with a more substantial psychopathology received longer treatments. To better understand the mechanisms of change, between-group analyses such as the one in this article must be complemented by longitudinal analyses of each case. Our aim is to do this in an upcoming study.

## CONCLUSIONS AND CLINICAL IMPLICATIONS

In a long-term follow-up study of an RCT comparing two different ways of helping mothers and infants with baby worries, one of them established and contained mostly supportive elements (CHCC) and the other new and contained psychodynamically based interventions (MIP), we found significant between-group differences on two measures. The MIP children showed better global functioning and were more often categorized as being OK in the interview situation. In contrast, two other measures failed to yield the hypothesized effects: the SSAP and the Machover. They represented efforts at quantifying implicit phenomena, they were collected under stress, and they were assessed by experts. These characteristics also applied to the CGAS and the Child types, which, however, were based on a personal meeting with the child where the interviewer could weigh together several sources of information. These two features should be taken into account when one seeks to understand why the CGAS and the Child types, but not the SSAP and the Machover, yielded effects. The two most comprehensive measures seemed to capture information which the other measures were not able to cover.

Several studies have pointed to the short-term effects of mother–infant psychotherapies (e.g., Cohen et al., 2002; Lieberman et al., 1991; Salomonsson & Sandell, 2011a). Their long-term effects are difficult to evaluate and interpret; thus, we do not know much about their persistent value. An earlier study found that the long-term effects were meager (Murray et al., 2003). Our study supports the notion that babies with relationship disorders may benefit from early interventions and thus should be offered psychotherapeutic help when problems emerge. Treatment was carried out at a crucial point in their lives and in statu nascendi of the pathogenesis. This was thought to help the dyad turn a destructive interactive helix into a benign development. The long-term effects on the mothers and the mother–child interaction will be reported in a separate article.

## LIMITATIONS

Caution must be exercised in attempts to generalize the findings of this study. We do not know to what extent MIP might help samples with other social, psychological, and psychiatric characteristics. Another limitation is that CHCC was a more heterogeneous treatment modality than was MIP. The MIP group received more intense treatment than did the CHCC group. Furthermore, the MIP therapists were specially trained in mother–infant psychotherapy. As

for the nurses in the CHCC group, some had EPDS training while others had no such education. This should be taken into account when comparing between-group effects. Further studies should compare MIP with other well-defined mother–infant psychotherapies, preferably with the same intensity of treatment.

As argued earlier, the influence of the interviewers’ allegiance needs to be considered in an evaluation of the results, although its possible quantitative impact cannot be calculated. We also discussed the possible impact on outcomes that might emanate from the mothers’ demand characteristics.

Another question of validity emerges from the fact that the child interviewer was involved in two outcome assessments, the CGAS and the Child types. One could argue that they were idiosyncratic and nongeneralizable. However, the CGAS ratings were made in consensus with an expert rater. The Child types showed an acceptable level of agreement with an external rater and correlated substantially with several other measures.

## REFERENCES

- Abidin, R.R. (1990). *Parenting Stress Index (PSI)—Manual*. Odessa, FL: Psychological Assessment Resources.
- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
- Bayley, N. (1969). *Manual for the Bayley Scales of Infant Development*. Berkeley, CA: Institute of Human Development.
- Becker, B.J. (1988). Synthesizing standardized mean-change measures. *British Journal of Mathematical and Statistical Psychology*, 41, 257–278.
- Bird, H., Canino, G., Rubio-Stipec, M., & Ribera, J. (1987). Further measures of the psychometric properties of the Children’s Global Assessment Scale. *Archives of General Psychiatry*, 29(5), 821–824.
- Bird, H., Yager, T., Staghezza, B., Gould, M., Canino, G., & Rubio-Stipec, M. (1990). Impairment in the epidemiological measurement of childhood psychopathology in the community. *Journal of the American Academy of Child & Adolescent Psychiatry*, 29, 796–803.
- Biringen, Z., Robinson, J.L., & Emde, R.N. (1998). *Emotional Availability Scales* (3rd ed.). Unpublished manual, Colorado State University.
- Blomberg, B., & Cleve, E. (1997). *Machovertestet. Handbok med utvecklingsteoretiska perspektiv* (The Machover test. Handbook with perspectives of developmental theory). Stockholm: Psykologiförlaget.
- Börjesson, K., Ruppert, S., & Bågedahl-Strindlund, M. (2005). A longitudinal study of psychiatric symptoms in primiparous women: Relation to personality disorders and sociodemographic factors. *Archives of Women’s Mental Health*, 8(4), 232–242.
- Chinn, S. (2000). A simple method for converting an odds ratio to effect size for use in meta-analysis. *Statistics in Medicine*, 19(22), 3127–3131.
- Clark, R., Tluczek, A., & Wenzel, A. (2003). Psychotherapy for postpartum depression: A preliminary report. *American Journal of Orthopsychiatry*, 73(4), 441–454.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: Erlbaum.

- Cohen, N.J., Lojkasek, M., Muir, E., Muir, R., & Parker, C.J. (2002). Six-month follow-up of two mother–infant psychotherapies: Convergence of therapeutic outcomes. *Infant Mental Health Journal*, 23(4), 361–380.
- Cohen, N.J., Muir, E., Parker, C.J., Brown, M., Lojkasek, M., Muir, R., & Barwick, M. (1999). Watch, wait and wonder: Testing the effectiveness of a new approach to mother–infant psychotherapy. *Infant Mental Health Journal*, 20(4), 429–451.
- Cooper, P.J., Murray, L., Wilson, A., & Romaniuk, H. (2003). Controlled trial of the short- and long-term effect of psychological treatment of post-partum depression. 1. Impact on maternal mood. *British Journal of Psychiatry*, 182(5), 412–419.
- Cousineau, T.M., & Shedler, J. (2006). Predicting physical health: Implicit mental health measures versus self-report scales. *Journal of Nervous and Mental Disease*, 194(6), 427–432.
- Cox, J., Holden, J., & Sagovsky, R. (1987). Detection of postnatal depression: Development of the 10-item Edinburgh Postnatal Depression Scale. *British Journal of Psychiatry*, 150, 782–786.
- Cramer, B., & Palacio Espasa, F. (1993). *La pratique des psychothérapies mères-bébés. Études cliniques et techniques (The practice of mother-infant psychotherapies. Clinical and technical studies)*. Paris: Presses Universitaires de France.
- Daud, A. (2008). Post-traumatic stress disorder and resilience in children of traumatised parents: A transgenerational perspective. Stockholm: Karolinska Institutet.
- De Los Reyes, A., & Kazdin, A.E. (2005). Informant discrepancies in the assessment of childhood psychopathology: A critical review, theoretical framework, and recommendations for further study. *Psychological Bulletin*, 131(4), 483–509.
- Dennis, C.-L.E. (2004). Treatment of postpartum depression, Part 2: A critical review of nonbiological interventions. *Journal of Clinical Psychiatry*, 65(9), 1252–1265.
- Derogatis, L.R. (1994). *Symptom Checklist-90-R: Administration, scoring and procedures manual (3rd ed., Rev.)*. Minneapolis, MN: National Computer Systems.
- Elliott, R., Fischer, C.T., & Rennie, D.L. (1999). Evolving guidelines for publication of qualitative research studies in psychology and related fields. *British Journal of Clinical Psychology*, 38(3), 215–229.
- Emde, R., Wolf, D.P., & Oppenheim, D. (2003). *Revealing the inner worlds of young children*. New York: Oxford University Press.
- Fraiberg, S., Adelson, E., & Shapiro, V. (1975). Ghosts in the nursery. A psychoanalytic approach to the problems of impaired infant-mother relationships. *Journal of the American Academy of Child Psychiatry*, 14(3), 387–421.
- Fridell, M., Cesarec, Z., Johansson, M., & Malling Thorsen, S. (2002). *Svensk normering, standardisering och validering av symptomskalan SCL-90 (A Swedish standardization and validation of the SCL-90)*. Stockholm: Statens Institutionsstyrelse.
- Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 38(5), 581–586.
- Goodman, R. (2001). Psychometric properties of the Strengths and Difficulties Questionnaire. *Journal of the American Academy of Child & Adolescent Psychiatry*, 40(11), 1337–1345.
- Grossmann, K.E., Grossmann, K., & Waters, E. (Eds.). (2005). *Attachment from infancy to adulthood: The major longitudinal studies*. New York: Guilford Press.
- Haddock, C., Rindskopf, D., & Shadish, W. (1998). Using odds ratios as effect sizes for meta-analysis for dichotomous data: A primer on methods and issues. *Psychological Methods*, 3, 339–353.
- Hayes, L., & Matthews, J. (2008). A randomized controlled trial of a mother–infant or toddler parenting program: Demonstrating effectiveness in practice. *Journal of Pediatric Psychology*, 33(5), 473–486.
- Hill, C.E., Chui, H., & Baumann, E. (2013). Revisiting and reenvisioning the outcome problem in psychotherapy: An argument to include individualized and qualitative measurement. *Psychotherapy*, 50(1), 68–76.
- Hillman, S., & Hodges, J. (2011). *SSAP Standard Comparison Tables*. Unpublished manuscript.
- Hodges, J., Hillman, S., Steele, M., & Henderson, K. (2002). “Little Pig” narrative story stem coding manual. Anna Freud Centre, Great Ormond Street, Coram Family Research Project on Attachment Adoption, London. Unpublished manuscript.
- Hodges, J., Steele, M., Hillman, S., Henderson, K., & Kaniuk, J. (2003). Changes in attachment representations over the first year of adoptive placement: Narratives of maltreated children. *Clinical Child Psychology and Psychiatry*, 8(3), 351–367.
- Holden, J.M., Sagovsky, R., & Cox, J.L. (1989). Counselling in a general practice setting: Controlled study of health visitor intervention in treatment of postnatal depression. *British Medical Journal*, 298(6668), 223–226.
- Iluoma, I., Tamminen, T., Kaukonen, P., Laippala, P., Puura, K., Salmelin, R., & Almqvist, F. (2001). Longitudinal study of maternal depressive symptoms and child well-being. *Journal of the American Academy of Child & Adolescent Psychiatry*, 40(12), 1367–1374.
- Jacobsen, T., Edelstein, W., & Hofmann, V. (1994). A longitudinal study of the relation between representations of attachment in childhood and cognitive functioning in childhood and adolescence. *Developmental Psychology*, 30, 112–124.
- Jick, T. (1979). Mixing qualitative and quantitative methods: Triangulation in action. *Qualitative Methodology*, 24(4), 602–611.
- Josephs, L., & Bornstein, R. (2011). Beyond the illusion of structural change: A process priming approach to psychotherapy outcome research. *Psychoanalytic Psychology*, 28(3), 420–434.
- Lambert, M.J. (Ed.). (2004). *Bergin and Garfield’s handbook of psychotherapy and behavior change (5th ed.)*. New York: Wiley.
- Laucht, M., Schmidt, M.H., & Esser, G. (2004). The development of at-risk children in early life. *Educational and Child Psychology*, 21(1), 20–31.
- Letourneau, N., & Stewart, M. (2011). Effect of home-based peer support on maternal-infant interactions among women with postpartum depression: A randomized, controlled trial. *International Journal of Mental Health Nursing*, 20(5), 345–357.

- Leuzinger-Bohleber, M., & Target, M. (2002). A follow-up study—Critical inspiration for our clinical practice? In M. Leuzinger-Bohleber & M. Target (Eds.), *Outcomes of psychoanalytic treatment: Perspectives for therapists and researchers* (pp. 143–173). London: Whurr.
- Lieberman, A.F., Weston, D.R., & Pawl, J.H. (1991). Preventive intervention and outcome with anxiously attached dyads. *Child Development*, 62(1), 199–209.
- Lindner, R., Fiedler, G., Altenhofer, A., Gotze, P., & Happach, C. (2006). Psychodynamic ideal types of elderly suicidal persons based on counter transference. *Journal of Social Work Practice*, 20(3), 347–365.
- Lojkasek, M., Cohen, N.J., & Muir, E. (1994). Where is the infant in infant intervention? A review of the literature on changing troubled mother–infant relationships. *Psychotherapy: Theory, Research, Practice, Training*, 31(1), 208–220.
- Luborsky, L., Diguier, L., Seligman, D.A., Rosenthal, R., Krause, E.D., Johnson, S. et al. (1999). The researcher’s own therapy allegiances: A “wild card” in comparisons of treatment efficacy. *Clinical Psychology: Science and Practice*, 6(1), 95–106.
- Lundh, A. (2012). *On the Children’s Global Assessment Scale (CGAS)*. Stockholm: Karolinska Institutet.
- Lundh, A., Kowalski, J., Sundberg, C., Gumpert, C., & Landén, M. (2010). Children’s Global Assessment Scale (CGAS) in a naturalistic clinical setting: Inter-rater reliability and comparison with expert ratings. *Psychiatry Research*, 177, 206–210.
- Lyons-Ruth, K., Dutra, L., Schuder, M., & Bianchi, I. (2006). From infant attachment organization to adult dissociation: Relational adaptations or traumatic experiences? *Psychiatric Clinics of North America*, 29, 63–86.
- Machover, K. (1949). *Personality projection in the drawing of the human figure*. Springfield, IL: Thomas.
- Machover, K. (1951). Drawings of the human figure: A method of personality investigation. In H. Anderson & G. Anderson (Eds.), *An introduction to projective techniques* (pp. 341–369). Engelwood Cliffs, NJ: Prentice Hall.
- Malterud, K. (2001). Qualitative research: Standards, challenges, and guidelines. *The Lancet*, 358(9280), 483–488.
- Markin, R.D., & Kivlighan, D.M., Jr. (2007). Bias in psychotherapist ratings of client transference and insight. *Psychotherapy: Theory, Research, Practice, Training*, 44(3), 300–315.
- Massie, H., & Szajnberg, N. (2002). The relationship between mothering in infancy, childhood experience and adult mental health: Results of the Brody prospective longitudinal study from birth to age 30. *International Journal of Psychoanalysis*, 83(1), 35–55.
- Mathiassen, B., Brondbo, P.H., Waterloo, K., Martinussen, M., Eriksen, M., Hanssen-Bauer, K., & Kvernmo, S. (2012). IQ as a predictor of clinician-rated mental health problems in children and adolescents. *British Journal of Clinical Psychology*, 51(2), 185–196.
- Meyer, G.J. (1999). The convergent validity of MMPI and Rorschach scales: An extension using profile scores to define response and character styles on both methods and a reexamination of simple Rorschach response frequency. *Journal of Personality Assessment*, 2(1), 1–35.
- Minnis, H., Green, J., O’Connor, T.G., Liew, A., Glaser, D., Taylor, E. et al. (2009). An exploratory study of the association between reactive attachment disorder and attachment narratives in early school-age children. *Journal of Child Psychology and Psychiatry*, 50, 931–942.
- Mulcahy, R., Reay, R.E., Wilkinson, R.B., & Owen, C. (2010). A randomised control trial for the effectiveness of group interpersonal psychotherapy for postnatal depression. *Archives of Women’s Mental Health*, 13, 125–139.
- Munder, T., Brüttsch, O., Leonhart, R., Gerger, H., & Barth, J. (2013). Researcher allegiance in psychotherapy outcome research: An overview of reviews. *Clinical Psychology Review*, 33, 501–511.
- Muratori, F., Picchi, L., Bruni, G., Patarnello, M., & Romagnoli, G. (2003). A two-year follow-up of psychodynamic psychotherapy for internalizing disorders in children. *Journal of the American Academy of Child & Adolescent Psychiatry*, 42(3), 331–339.
- Muratori, F., Picchi, L., Casella, C., Tancredi, R., Milone, A., & Patarnello, M.G. (2002). Efficacy of brief dynamic psychotherapy for children with emotional disorders. *Psychotherapy and Psychosomatics*, 71(1), 28–38.
- Murray, L., Cooper, P.J., Wilson, A., & Romaniuk, H. (2003). Controlled trial of the short- and long-term effect of psychological treatment of post-partum depression: 2. Impact on the mother–child relationship and child outcome. *British Journal of Psychiatry*, 182(5), 420–427.
- Norman, J. (2001). The psychoanalyst and the baby: A new look at work with infants. *International Journal of Psychoanalysis*, 82(1), 83–100.
- Norman, J. (2004). Transformations of early infantile experiences: A 6-month-old in psychoanalysis. *International Journal of Psychoanalysis*, 85(5), 1103–1122.
- Odhammar, F., Sundin, E.C., Jonson, M., & Carlberg, G. (2011). Children in psychodynamic psychotherapy: Changes in global functioning. *Journal of Child Psychotherapy*, 37(3), 261–279.
- Ogawa, J.R., Sroufe, L., Weinfeld, N.S., Carlson, E., & Egeland, B. (1997). Development and the fragmented self: A longitudinal study of dissociative symptomatology in a nonclinical sample. *Development and Psychopathology*, 9, 855–879.
- O’Hara, M.W., Stuart, S., Gorman, L.L., & Wenzel, A. (2000). Efficacy of interpersonal psychotherapy for postpartum depression. *Archives of General Psychiatry*, 57(11), 1039–1045.
- Oppenheim, D., Emde, R.N., & Warren, S. (1997). Children’s narrative representations of mothers: Their development and associations with child and mother adaptation. *Child Development*, 68(1), 127–138.
- Orne, M.T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, 17(11), 776–783.
- Östberg, M. (1998). Parental stress, psychosocial problems and responsiveness in help-seeking parents with small (2–45 months old) children. *Acta Paediatrica*, 87(1), 69–76.
- Östberg, M., Hagekull, B., & Wettergren, S. (1997). A measure of parental stress in mothers with small children: Dimensionality, stability and validity. *Scandinavian Journal of Psychology*, 38(3), 199–208.



- Pass, L., Arceche, A., Cooper, P., Creswell, C., & Murray, L. (2012). Doll play narratives about starting school in children of socially anxious mothers, and their relation to subsequent child school-based anxiety. *Journal of Abnormal Child Psychology*, 40, 1375–1384.
- Philips, B., Wennberg, P., & Werbart, A. (2007). Ideas of cure as a predictor of premature termination, early alliance and outcome in psychoanalytic psychotherapy. *Psychology and Psychotherapy: Theory, Research and Practice*, 80(Pt. 2), 229–245.
- Philips, B., Werbart, A., Wennberg, P., & Schubert, J. (2007). Young adults' ideas of cure prior to psychoanalytic psychotherapy. *Journal of Clinical Psychology*, 63(3), 213–232.
- Ravn, I.H., Smith, L., Smeby, N.A., Kynoe, N.M., Sandvik, L., Bunch, E.H., & Lindemann, R. (2012). Effects of early mother–infant intervention on outcomes in mothers and moderately and late preterm infants at age 1 year: A randomized controlled trial. *Infant Behavior & Development*, 35(1), 36–47.
- Robert-Tissot, C., Cramer, B., Stern, D.N., Serpa, S.R., Bachmann, J.P., Palacio-Espasa, F. et al. (1996). Outcome evaluation in brief mother–infant psychotherapies: Report on 75 cases. *Infant Mental Health Journal*, 17(2), 97–114.
- Salomonsson, B. (2007). “Talk to me baby, tell me what’s the matter now.” Semiotic and developmental perspectives on communication in psychoanalytic infant treatment. *International Journal of Psychoanalysis*, 88(1), 127–146.
- Salomonsson, B. (2010). “Baby worries”—A randomized controlled trial of mother–infant psychoanalytic treatment. Stockholm: Karolinska Institutet.
- Salomonsson, B. (2011). The music of containment. Addressing the participants in mother–infant psychoanalytic treatment. *Infant Mental Health Journal*, 32(6), 599–612.
- Salomonsson, B. (2014). *Psychoanalytic therapy with infants and parents: Practice, theory and results*. London: Routledge.
- Salomonsson, B., & Sandell, R. (2011a). A randomized controlled trial of mother–infant psychoanalytic treatment: I. Outcomes on self-report questionnaires and external ratings. *Infant Mental Health Journal*, 32(2), 207–231.
- Salomonsson, B., & Sandell, R. (2011b). A randomized controlled trial of mother–infant psychoanalytic treatment: II. Predictive and moderating influences of qualitative patient factors. *Infant Mental Health Journal*, 32(3), 377–404.
- Salomonsson, B., & Sleed, M. (2010). The Ages & Stages Questionnaire: Social–Emotional: A validation study of a mother-report questionnaire on a clinical mother–infant sample. *Infant Mental Health Journal*, 31(4), 412–431.
- Santelices, M.P., & Guzmán, G.M. (2011). Promoting secure attachment: Evaluation of the effectiveness of an early intervention pilot programme with mother–infant dyads in Santiago, Chile. *Child: Care, Health and Development*, 37(2), 203–210.
- Schorre, B.E.H., & Vandvik, I.H. (2004). Global assessment of psychosocial functioning in child and adolescent psychiatry: A review of three unidimensional scales (CGAS, GAF, GAPD). *European Child & Adolescent Psychiatry*, 13, 273–286.
- Seimyr, L., Edhborg, M., Lundh, W., & Sjögren, B. (2004). In the shadow of maternal depressed mood: Experiences of parenthood during the first year after childbirth. *Journal of Psychosomatic Obstetrics & Gynecology*, 25(1), 23–34.
- Shaffer, D., Gould, M.S., Brasic, J., Ambrosini, P., Prudence, F., Bird, H., & Aluwahlia, S. (1983). A Children’s Global Assessment Scale (CGAS). *Archives of General Psychiatry*, 40(11), 1228–1231.
- Shedler, J., Mayman, M., & Manis, M. (1993). The illusion of mental health. *American Psychologist*, 48(11), 1117–1131.
- Silverman, R., & Lieberman, A. (1999). Negative maternal attributions, projective identification, and the intergenerational transmission of violent relational patterns. *Psychoanalytic Dialogues*, 9(2), 161–186.
- Singleton, J.L. (2005). *Parent–infant interaction interventions: A meta-analysis*. Dissertation Abstracts International: Section B: Sciences and Engineering (Doctoral dissertation). University of Northern Colorado.
- Skovgaard, A.M., Houmann, T., Christiansen, E., Landorph, S., Jorgensen, T., C.C.C.S. Study Team et al. (2007). The prevalence of mental health problems in children 1½ years of age—The Copenhagen Child Cohort 2000. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 48(1), 62–70.
- Skovgaard, A.M., Olsen, E.M., Houmann, T., Christiansen, E., Samberg, V., Lichtenberg, A., & Jorgensen, T. (2005). The Copenhagen County child cohort: Design of a longitudinal study of child mental health. *Scandinavian Journal of Public Health*, 33(3), 197–202.
- Smedje, H., Broman, J.E., Hetta, J., & von Knorring, A.L. (1999). Psychometric properties of a Swedish version of the “Strengths and Difficulties Questionnaire.” *European Child & Adolescent Psychiatry*, 8(2), 63–70.
- Smith, D., & Dumont, F. (1995). A cautionary study: Unwarranted interpretations of the Draw-A-Person Test. *Professional Psychology: Research and Practice*, 26(3), 298–303.
- Sourander, A., & Piha, J. (1997). Parent, teacher and clinical ratings on admission to child psychiatric inpatient treatment: A study of cross-informant correlations using the CBCL, Rutter scales and the CGAS. *Nordic Journal of Psychiatry*, 51(5), 365–370.
- Squires, J., Bricker, D., Heo, K., & Twombly, E. (2001). Identification of social-emotional problems in young children using a parent-completed screening measure. *Early Childhood Research Quarterly*, 16(4), 405–419.
- Squires, J., Bricker, D., Heo, K., & Twombly, E. (2002). *Ages & Stages Questionnaires: Social-Emotional*. A parent-completed, child-monitoring system for social-emotional behaviors. Baltimore: Brookes.
- Squires, J., Bricker, D., & Twombly, E. (2004). Parent-completed screening for social emotional problems in young children: The effects of risk/disability status and gender on performance. *Infant Mental Health Journal*, 25(1), 62–73.
- Sroufe, L.A., Egeland, B., Carlson, E., & Collins, W.A. (2005). *The development of the person. The Minnesota Study of Risk and Adaptation from Birth to Adulthood*. New York: Guilford Press.

- Stadelmann, S., Perren, S., von Wyl, A., & von Klitzing, K. (2007). Associations between family relationships and symptoms/strengths at kindergarten age: What is the role of children's parental representations? *Journal of Child Psychology and Psychiatry*, 48(10), 996–1004.
- Stievenart, M., Roskam, I., Meunier, J.C., & van de Moortele, G. (2011). The reciprocal relation between children's attachment representations and their cognitive ability. *International Journal of Behavioral Development*, 35(1), 58–66.
- Tabachnik, B.G., & Fidell, L.S. (2007). *Using multivariate statistics* (5th ed.). Boston: Allyn & Bacon.
- Target, M. (2002). The problem of outcome in child psychoanalysis: Contributions from the Anna Freud Centre. In M. Leuzinger-Bohleber & M. Target (Eds.), *Outcomes of psychoanalytic treatment: Perspectives for therapists and researchers* (pp. 240–251). London: Whurr.
- Target, M., & Fonagy, P. (1994a). Efficacy of psychoanalysis for children with emotional disorders. *Journal of the American Academy of Child & Adolescent Psychiatry*, 33(3), 361–371.
- Target, M., & Fonagy, P. (1994b). The efficacy of psychoanalysis for children: Prediction of outcome in a developmental context. *Journal of the American Academy of Child & Adolescent Psychiatry*, 33(8), 1134–1144.
- Wachholz, S., & Stuhr, U. (1999). The concept of ideal types in psychoanalytic follow-up research. *Psychotherapy Research*, 9(3), 327–341.
- Wechsler, D. (2002). *WPPSI-III Technical and Interpretive Manual*. San Antonio: Psychological Corporation.
- Wechsler, D. (2005). *WPPSI-III. Manual, del 1. Svensk version (WPPSI-III. Swedish version)*. Stockholm: Psykologiförlaget.
- Weissman, M.M., Warner, V., & Fendrich, M. (1990). Applying impairment criteria to children's psychiatric diagnosis. *Journal of the American Academy of Child & Adolescent Psychiatry*, 29(5), 789–795.
- Werbart, A., Grunbaum, C., Jonasson, B., Kempe, H., Kusz, M., Linde, S. et al. (2011). Changes in the representations of mother and father among young adults in psychoanalytic psychotherapy. *Psychoanalytic Psychology*, 28(1), 95–116.
- Wickberg, B., & Hwang, C.P. (1997). Screening for postnatal depression in a population-based Swedish sample. *Acta Psychiatrica Scandinavica*, 95(1), 62–66.
- ZERO TO THREE. (2005). *Diagnostic classification of mental health and developmental disorders of infancy and early childhood (DC 0–3:R)*. Washington: DC: ZERO TO THREE Press.