The development of children in foster care

Goemans, A.

2

DEVELOPMENTAL OUTCOMES OF FOSTER CHILDREN: A META-ANALYTIC COMPARISON WITH CHILDREN FROM THE GENERAL POPULATION AND CHILDREN AT-RISK WHO REMAINED AT HOME

Published
Goemans, A., Van Geel, M., Van Beem, M., & Vedder, P. (2016). Developmental outcomes of foster children: A meta-analytic comparison with children from the general population and children at-risk who remained at home. *Child Maltreatment, 21,* 198-217.

Chapter 2

ABSTRACT

Foster care is often preferred to other placement options for children in the child welfare system. However, it is not clear how the developmental outcomes of foster children relate to children in other living arrangements. In this study, a series of meta-analyses are performed to compare the cognitive, adaptive and behavioral functioning of children placed in foster care (N = 2,305) with children at risk who remained with their biological parents (N = 4,335), and children from the general population (N = 4,971). A systematic literature search in PsycINFO, MEDLINE, ERIC, and ProQuest identified 31 studies suitable for inclusion (N = 11,611). Results showed that foster children had generally lower levels of functioning than children from the general population. No clear differences were found between foster children and children at risk who remained at home, but both groups experienced developmental problems. Improving the quality of foster care and future research to identify which children are best served by either foster care or in-home services are recommended.

INTRODUCTION

Foster care is a way of providing a traditional family life for children in the child welfare system who cannot live with their own parents. This form of child protective services provides stability and continuity of caregivers, the opportunity to build close relationships with substitute parent figures, and ideally opportunities for positive development (Dozier, Kaufman, Kobak, O'Connor, et al., 2014; Roy et al., 2000; Tizard & Hodges, 1978). However, the high percentage of breakdowns, defined as the placement not lasting as long as planned, is in stark contrast with these optimistic claims. Between 20 and 50% of foster care placements result in breakdown (Minty, 1999) and the negative consequences of breakdown may compromise the positive effects of foster care (Newton, Litrownik, & Landsverk, 2000). Moreover, although almost no systematic reviews on the comparison of foster children with children from a general population have been performed (however, for an example on foster care and education see Scherr, 2007), individual studies suggest that foster children perform worse on developmental outcomes than children from the general population (Frank, 1980; Lawrence et al., 2006; Lehmann et al., 2013; Leslie, Gordon, Meneken, et al., 2005; Lloyd & Barth, 2011). Given these results and the high risk of breakdown, the effects of out-of-home placement into foster care remain the subject of controversy (Lawrence et al., 2006).

An alternative to foster care, which has gained popularity over the last decade (Janssens & Deboutte, 2010), is the option of remaining at home while receiving some form of in-home support services. At the moment, it is not clear how the developmental outcomes of this group of children relate to that of foster children (Mennen et al., 2010). Information on the comparison of the developmental outcomes for children in different living arrangements (i.e., foster care, at home care with support from the child welfare system, and regular care at home) can provide evidence on which living arrangement is best for which children and hence have consequences for service delivery for different groups of children and their caregivers (Janssens & Deboutte, 2010; Wald et al., 1988). The current study therefore aims to systematically compare the developmental outcomes (cognitive, adaptive and behavioral functioning) of children in foster care with children from the general population and with children at risk who remained at home by using meta-analyses.

Placement Settings for Children in the Child Welfare System

Children cared for by the child welfare system either can be placed in out-of-home care or remain with their families of origin (Leslie, Gordon, Ganger, & Gist, 2002). Although many children reside in foster families, growing up within the family of origin is in principle preferred over growing up in a foster family (United National General Assembly, 1989). The majority of children in the child welfare system remain with their biological parents while receiving some form of in-home support services aimed at preventing out-of-home placement (Mennen et al., 2010). Remaining at home prevents the potential lack of permanence in case of out-of-home placement and the negative effects associated with the separation of the child from the biological parents. On the other hand, children remaining at home are suggested to be at risk for repeated maltreatment (Campbell, Thomas, Cook, & Keenan, 2012; Murphy,

Bishop, Jellinek, Quinn, & Poitrast, 1992; Runyan & Gould, 1985) and for developmental problems due to less optimal home environments such as poverty, parental psychopathology, substance abuse, and family violence (Campbell et al., 2012; Lloyd & Barth, 2011). However, for the evaluation of placement settings the possible negative consequences of staying in a dysfunctional home, which may be moderated by in home services, have to be balanced against foster care and its accompanying effects of separation from parents, the child's removal from home and the challenge of getting used to a new home and family in terms of the child's well-being and developmental outcomes (Britner & Mossler, 2002; Kinard, 1982). Some studies have found better outcomes for children who were placed in foster care (Beatty, 1995a; Colton, Heath, & Aldgate, 1995), while others found better outcomes for children who remained home (Bada et al., 2008; Heflinger, Simpkins, & Combs-Orme, 2000). In short, these separate studies have not convincingly established how the developmental outcomes of children in foster care relate to those of children at risk who remained at home (Janssens & Deboutte, 2010; Mennen et al., 2010). However, an important issue in the individual studies comparing children in different placement settings, is whether children in foster care can be meaningfully compared to children from the general population and to children at risk who remained at home. The evidently unavoidable pre-existing differences between children in particular living arrangements complicate simple comparison; differences on many dimensions and aspects of child development, in caregivers' competences and personalities, as well as in contextual affordances make that 'common' ceteris paribus reasoning (i.e., all other things being equal), which is basic to systematic comparisons, is challenged.

However, this does not diminish the importance of finding support for the difficult process of decision making when it comes to matching children with the best possible developmental future. Such decisions in child welfare, on whether children should be placed in foster care or can remain with their family of origin, are among the most difficult decisions that child welfare professionals have to make. Apart from the pre-existing differences between families, children, and the professionals involved, placement decisions are also influenced by historical and political trends (Gilbert, Parton, & Skivenes, 2011; Wulczyn, 2004), for example the Adoption and Safe Families Act of 1997. This Act reflects the change in focus from family preservation to the importance of achieving permanency and stability for children. Other determinants, such as the availability of foster homes and professionals' judgments in the process of decision-making (Britner & Mossler, 2002; Jones, 1993; Lindsey, 1992; Osmo & Benbenishty, 2004), also play a role in placement decisions. Together, these factors and the processes involved, make placement decisions complex and difficult, which can partly be resolved by developing or adapting empirical models supporting evidence-based decision making (Zuravin, Orme, & Hegar, 1995). Researchers need to provide input for this challenge by using the best available instruments to weigh available knowledge and derive at the most valid integration possible.

Meta-Analysis on Comparison Studies

The existing studies comparing foster children to children from a general population (e.g., Bruce et al., 2013; Jacobsen, Moe, Ivarsson, Wentzel-Larsen, & Smith, 2013), or comparing foster children to at risk children remaining at home (e.g., Bada et al., 2008; Salo et al., 2009) have been very valuable in gaining knowledge about the correspondence

between developmental outcomes and care provided, as well as in developing notions about the correspondence between child and care characteristics on the one hand and the care needed to achieve particular developmental outcomes on the other hand. In short, they provide information needed for deciding what placement decision or what decision on care provision is desirable (Janssens & Deboutte, 2010). Unfortunately, many of the available studies (e.g., Bruce et al., 2013; Roy et al., 2000; Salo et al., 2009) have relatively small sample sizes, and effect sizes between studies have varied widely (cf., Min, Minnes, Yoon, Short, & Singer, 2014; Victor, Wozniak, & Chang, 2008), which makes direct comparisons and overall conclusions difficult.

These challenges can be addressed in a meta-analysis. Meta-analysis allows for the combination of results from individual studies into a statistical summary. An important incremental value of meta-analysis is that a more convincing and reliable overall effect size is obtained than in the individual studies themselves (Borenstein et al., 2009). In the current study, we provide meta-analyses on cognitive, adaptive, and behavioral functioning. This provides both foster care professionals and researchers with a statistical summary of some of the most important developmental outcomes of foster children in comparison to the general population and children at risk who remained at home. Furthermore, it may demonstrate for what type of developmental outcomes foster children most likely need extra help. In addition to providing a statistical summary, meta-analysis also allows to assess the heterogeneity between the effect sizes of included studies, and it allows the analysis of potential publication bias. Publication bias may occur because studies with high effect sizes are more likely to be published than studies with lower effect sizes. This may erroneously lead professionals and researchers to believe that effects are larger than they actually are. In a meta-analysis, the extent to which publication bias affects an overall effect size can be analyzed, and, if needed, an effect size adjusted for publication bias can be estimated (Rothstein et al., 2006).

Current Study

It seems that both the option of remaining at home and the option of placement in foster care have their own benefits and risks, and as a result, placement decisions cannot always be made unequivocally. Information on the comparison of developmental outcomes for children in different placement settings can guide toward a placement setting that is in the best developmental interest of children and more important, it can guide service delivery for different groups (Wald et al., 1988). Based on the results of individual comparison studies, it is difficult to make such decisions and guidelines. The current article therefore aims to systematically compare the developmental outcomes of children in foster care and children at risk who remained at home by using meta-analysis. Furthermore, a systematic comparison between foster children and children from the general population is made. The focus is on cognitive, adaptive, and behavioral functioning. To also give insight into the characteristics of the individual studies and samples, a qualitative review is given for each group comparison. For both quantitative and qualitative analysis, attention is paid to the inclusion criteria and the matching of samples within individual studies. This way we hope to add in a systematic and transparent way to the evaluation of comparability of samples with pre-existing differences. Because earlier studies reported inconclusive findings, no specific hypotheses are formulated

regarding the developmental outcomes of foster children compared with children at risk who remained at home. Regarding the comparison of foster children with children from the general population, we expect children from the general population to function better than foster children (Farmer et al., 2001; Garwood & Close, 2001; Lanier, Kohl, Raghavan, & Auslander, 2014; Meltzer, Gatward, Corbin, Goodman, & Ford, 2003).

METHOD

Search Strategy

We used four online databases to systematically search for relevant studies on a comparison of the development of children in foster care, children at risk who remained at home, and children from the general population. PsycINFO, MEDLINE, ERIC and ProQuest Dissertations & Theses were searched for publications until April 2014 with the search terms *foster child** or *foster care* combined with *internalizing, externalizing, behavio*, SDQ, VABS, CBC*, development*, disorder*, cognitive, IQ, intelligence, intellect*, mental problem*, mental health, psychological problem*, compar*, psychopatho*, Vineland,* or *adaptive*. The reference lists of the collected studies were searched for relevant earlier studies. This search resulted in 3,371 studies which included articles, reports, dissertations and book chapters. A flow diagram of our search is presented in Figure 2.1.

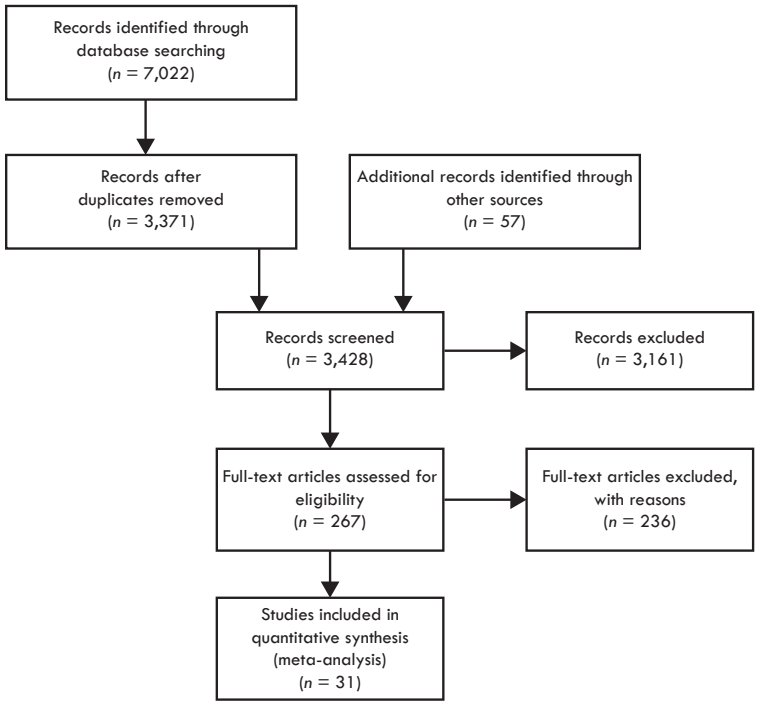

Figure 2.1. Flow diagram of all stages of the literature search.

Inclusion and Exclusion Criteria

A study had to meet the following four criteria to be included in the meta-analysis. First, studies had to include children (0-18 years) from both foster care (kinship and / or non-kinship) and at least one comparison group; either children at risk who remained with their biological parents, or children from the general population (community comparisons). Studies using norm scores as representative of children from the general population, for instance, Child Behavior Checklist (CBCL) norm scores, were excluded. Also excluded were studies that compared groups before children had been actually placed in different caregiving settings. Only studies on children from regular foster care were included, because children who receive more intensive forms of foster care might develop in a way that is different from foster children receiving 'care-as-usual'; studies on children from therapeutic foster care or on foster children receiving an intensive intervention were excluded .

Second, studies were included if they reported on cognitive, adaptive, or behavioral outcomes. Although we included studies which reported proportions of foster children with CBCL clinical scores (e.g., Heflinger et al., 2000), we excluded studies which reported on the percentage of psychiatric diagnoses (Cederna-Meko, Koch, & Wall, 2013; Harman, Childs, & Kelleher, 2000). This distinction is made for the reason that psychiatric diagnoses cannot be fairly compared to clinical CBCL scores, because it takes more than only high CBCL scores to receive a psychiatric disorder classification (Hudziak, Copeland, Stanger, & Wadsworth, 2004).

Third, studies were only included if they reported enough data to compute an effect size. Authors of studies wherein statistics relevant to be included in the meta-analyses were lacking were emailed with a request for additional information. This resulted in the inclusion of four additional studies (Bruce et al., 2013; Carbone, Sawyer, Searle, & Robinson, 2007; Mennen et al., 2010; Min et al., 2014).

Fourth, if two or more articles were based on the same sample, we chose to include the study with the larger sample size; if sample sizes were equal we coded the study with the most detailed information (Carbone et al., 2007; Colton et al., 1995). The 31 studies summarized in Table 2.1 met the inclusion criteria for the meta-analyses. The third column in this table gives an overview of the groups compared. The last two columns describe which inclusion criteria were used by the individual studies and whether and how the samples compared within studies were matched.

For a study to be included in our meta-analysis, language was not used as an exclusion criterion. Thirty studies were written in English, and one was written in Turkish (Üstüner, Erol, & Simsek, 2005). We included six dissertations (Beatty, 1995; Davis, 2000; Maroufi, 2003; Rork, 2007; Schiefer, 1994; Shepherd, 2009), and two books (McAuley, 1996; Wald et al., 1988). Each study eligible for inclusion in the meta-analysis was also included in the qualitative review.

Table 2.1. *Studies included in the meta-analyses.*

Study (Year of Publication)	Country	Groups[A]	*N*	Age Range in years (Mean)	% Female	Measure[B]
Bada et al. (2008)[1,4,8]	USA	FC RH GP	152 317 514	Tested at 3 years Tested at 3 years Tested at 3 years	52.4% 46.8% 46.8%	CBCL, VABS
Beatty (1995)[2]	USA	FC RH	43 42	4-16 (8.0) 4-16 (9.9)	46.5% 50.0%	CBCL
Bernedo et al. (2012)[1,2]	Spain	FC GP	97 97	5-18 (10.9) 5-18 (n/a)	44.3% 44.3%	TRF
Bruce et al. (2013)[1]	USA	FC GP	11 11	9-12 (11.0) 9-12 (10.9)	45.0% 45.0%	CBCL
Burns et al. (2004)[1,5,8]	USA	FC RH	323 3411	2-14 (n/a) 2-14 (n/a)	50.3% 50.3%	CBCL
Carbone et al. (2007)[1,4,6,8]	Australia	FC GP	91-326 1273-3255	6-17 (11.4) 6-17 (11.4)	46.0% 50.2%	CBCL, YSR
Colton et al. (1995)[2,6]	UK	FC RH	24-49 29-58	8-14 (n/a) 8-14 (n/a)	46.9% 41.4%	Rutter Behaviour Scales
Damjanoviç et al. (2012)[1,4,8]	Serbia	FC GP	104 238	8-18 (13.2) 8-18 (12.8)	61.0% 55.0%	PedsQL
Davis (2000)[2]	USA	FC GP	55 30	6-16 (n/a) 6-16 (n/a)	41.2% 50.0%	WISC-III
Fernandez (2008)[1,4,7]	Australia	FC GP	43 42	4-15 (8.8) n/a (n/a)	50.9% 46.5%	Adaptive functioning scale, TRF
Heflinger et al. (2000)[1,3,4,5]	USA	FC RH	105 72	2-18 (n/a) 2-18 (n/a)	54.3% 54.3%	CBCL

Inclusion Criteria	Matching Samples[c]
FC, RH, and GP: age. FC and RH: prenatally drug exposed.	Yes. Matching on gender, race, birth weight, head circumference, and maternal age (see p.177, Table 3). Quantitative data used in meta-analysis is based on subgroups of groups reported on in Table 3.
FC and RH: from foster or custodial parents, age 4-16 years, Tarrant and Dallas county area, backgrounds of parental neglect/abuse.	No (see Table 1, p.30 and Table 2, p.31).
FC and GP: 5-18 years. FC: not severely disabled, from Malaga, Jaen and Granada (Spain).	Yes. Matching on gender, age, and social class background (see p.617, paragraph 2.1.4.).
FC and GP: 9-12 years. GP: low income, parental education less than a 4-year college degree, no involvement child welfare. Several other eligibility criteria (see p.933, paragraph 'participants').	Yes. Matching on gender, age, ethnicity, and parental education (see p.933, Table 1 and paragraph 'participants').
FC and RH: National Survey of Child and Adolescent Well-being (NSCAW) consists of two cohorts of randomly selected children. The NSCAW sample design involves a stratified two-stage sample with the primary sampling units being county child welfare agencies and the secondary sampling units were selected from lists of closed investigations or assessments from the sampled agencies. Sampling within primary sampling units was stratified by age, sexual abuse and child welfare services.	No. This article does not report sample characteristics or the matching of the FC and RH group.
FC and GP: 6-17 years. FC: Adelaide metropolitan region, under guardianship, case managed by 'Families SA', within foster family for at least 1 month. GP: 6-17 years, random sample from Australian National Survey of Child and Adolescent Mental Health and Wellbeing.	Yes. Matching on gender and age (see p.1160, Table 1).
FC and RH: 8-14 years, same school. FC: long-term foster care, stable placements (no changes). RH: similar backgrounds, families receiving preventive social work support.	Yes. Matching on age band and FC and RH children attended the same schools (see p.68).
FC and GP: 8-18 years, literacy. FC: Children from three major regional childcare centers in Serbia. GP: from four elementary schools in Serbia.	Yes. Matching on gender (see p.471, Table 1).
FC and GP: 6-16 years, same geographical areas. GP: no history of involvement with legal system and never been abused.	Yes. Matching on geographical areas (assuming: same SES and cultural background) (see p.16). Quantitative data included in our meta-analyses is controlled for age (see p.23).
FC and GP: 4-16 years. FC: new foster placement, long-term foster care, majority non-kinship. GP: classmate from FC, matched for age and sex.	Yes. Matching on gender and age (see p. 1289).
FC and RH: 2-18 years, random and representative sample of children and youth in state custody, State of Tennessee.	No. This article does not report sample characteristics or the matching of the FC and RH group.

Table 2.1. *(continued)*

Study (Year of Publication)	Country	Groups[A]	N	Age Range in years (Mean)	% Female	Measure[B]
Hulsey & White (1989)[1,8]	USA	FC GP	65 65	4-8 (5.8) 4-8 (5.5)	52.3% 44.6%	CBCL
Jacobsen et al. (2013)[1,7,8]	Norway	FC GP	60 42	22-25 months (23.3 months) 22-24 months (23.2 months)	40.0% 50.0%	ITSEA, MSEL
Janssens & Deboutte (2010)[1,4,6,8]	Belgium	FC RH	7-55 29-59	3-17 (10.2) 3-17 (12.6)	54.1% 47.5%	CBCL, SDQ, TRF, YSR
Johnson et al. (2010)[1,3,5,7,8]	USA / Romania	FC GP	63 72	5-32 months (21.0 months) 5-32 months (19.3 months)	50.0% 57.0%	BSID, WPPSI-II
Leifer & Shapiro (1993)[1,3]	USA	FC RH	32 28	5-16 (8.9) 5-16 (8.9)	100% 100%	CBCL, CDI, WISC-R
Maroufi (2003)[2]	USA	FC RH	39 39	2-36 months (12.7 months) 2-36 months (12.5 months)	38.5% 38.5%	MSEL, VABS
McAuley (1996)[2,7]	Northern Ireland	FC GP	17 17	4-12 (8.4) n/a	36.8% n/a	TRF
Mennen et al. (2010)[1,4,8]	USA	FC RH GP	138 164 151	9-12 (10.8) 9-12 (10.8) 9-12 (11.1)	50.0% 50.0% 40.7%	CBCL, CDI, MASC, SPPA, YSR
Min et al. (2014)[1,4,6,7]	USA	FC RH	45-49 95-100	12-15 (12.1) 12-15 (12.1)	60.0% 52.6%	YSR
Pears & Fisher (2005)[1,8]	USA	FC GP	99 54	3-6 (4.4) 3-6 (4.3)	48.5% 48.2%	WPPSI-R

Inclusion Criteria	Matching Samples[c]
FC and GP: 4-8 years. FC: placed for physical abuse or neglect, at least 3 months with foster family, first time in foster care, eligible for medical assistance (assure a match for SES). GP: children attending a Baltimore City Health Department well-child clinic, eligible for Medicaid.	Yes. Matching on children's gender, age, race, and FC and GP attended the same school. Matching on mother education, natural mother age, sibling number (see p. 505 and 506, Table 1 and 2, and paragraph 'Results').
FC and GP: 2-3 years.	Yes. Matching on gender, marital status, ethnic origin of caregivers and mean family income (see p.669, Table 1).
FC and RH: 3-17 year, children from child welfare services in the region of Antwerp (representative for all Flemish services regarding capacity of the different types of placement, sex, and age distributions), being in care for at least 4 weeks.	Yes. Matching on gender and education level of mother (see p.355, Table 1 and paragraph 'background characteristics').
FC and GP: born at same hospitals. FC: <32 months, without serious handicapping conditions. GP: no history of institutional care.	Yes. Matching on gender and age (see p. 508, paragraph 'never institutionalized group').
FC and RH: 5-16 years, African-American female, history of sexual abuse.	Yes. Matching on SES (64% supported by Public Aid, remaining participants dispersed among Hollingshead classes 3-7, see p.31, paragraph 'subjects'). Quantitative data included in our meta-analyses is controlled for age, SES, and number of abuse incidents (see p.37).
FC and RH: 2-36 months, all prenatally drug exposed. FC: placed in foster care after birth.	Yes. Matching on gender and age (see p.61, Table 1).
FC and GP: 4-11 years, Northern Ireland. FC: long-term foster care, new foster placement, non-kinship foster care. GP: classmate from FC.	Yes. Matching on gender and age (see p.67, paragraph 4.4.6.ii).
FC, GP and RH: 9-12 years. FC and RH: newly opened cases in 10 zip codes in Los Angeles, cases of maltreatment, either African-American, Latino, or white. GP: same zip codes.	Yes. Matching on age, ethnicity, and zip codes (see p.1677, Table 1). Quantitative data included in our meta-analyses is adjusted for gender, ethnicity and age (see p.1679, Table 4, footnote).
FC and RH: 12-15 years, recruited from urban county hospital with high-risk maternal population screened for drug use. No children with Down syndrome, FASD, medical illness. Several exclusion criteria for mothers (see p.2, paragraph 'sample').	No. Quantitative data included in our meta-analysis is provided by the authors and based on subgroups of the groups presented in article. Therefore, we cannot be sure whether the information on matching as provided (see p.7, Table 4) is true for the FC and RH groups included in our meta-analysis.
FC and GP: 3-6 years. FC: new foster placements. GP: child had lived consistently with at least on biological parent, maximum for household income and parental education level, no previous involvement with CWS.	Yes. Matching on gender, age, and ethnicity (see p.114, paragraph 'participants').

Table 2.1. *(continued)*

Study (Year of Publication)	Country	Groups[A]	*N*	Age Range in years (Mean)	% Female	Measure[B]
Rork (2008)[2]	Australia	FC GP	30 30	2-11 (7.6) 2-11 (5.9)	43.3% 50.0%	CBCL
Roy et al. (2000)[1,9]	UK	FC GP	19 19	Primary school age (6.9) Primary school age (6.8)	36.8% 36.8%	Rutter Behaviour Scales, Classroom observation, WISC
Salo et al. (2009)[1]	Finland	FC RH	14 7	n/a (3 years ±3 months) n/a (3 years ±3 months)	n/a n/a	BSID
Schiefer (1994)[2,4,6]	USA	FC GP	56-80 91-114	4-18 (n/a) 4-18 (n/a)	n/a n/a	CBCL
Shepherd (2009)[2]	USA	FC GP	42 37	12-18 (15.6) 12-18 (15.0)	47.6% 48.7%	YSR
Tininenko et al. (2010)[1,4]	USA	FC GP	31 44	3-7 (5.4) 3-7 (5.1)	52.0% 55.0%	PDR
Tizard & Hodges (1978)[1,5,9]	England	FC GP	3 29	tested at 8 years tested at 8 years	35.3% 35.3%	WISC
Üstuner et al. (2005)[2,3,4,5,6]	Turkey	FC GP	15-31 30-62	6-17 (10.7) 6-17 (10.7)	66.7% 74.2%	CBCL, TRF, YSR
Victor et al. (2008)[1,8]	USA	FC RH	117 19	6-12 (9.4) 6-12 (9.9)	45.3% 42.1%	CBCL, WISC-III

Inclusion Criteria	Matching Samples[c]
FC and GP: 2-11 years. FC: regular foster care. GP: no previous involvement with child welfare, no adoptive or step-parents.	Yes. Matching on children's gender, ethnicity, and medical problems. Matching on parent gender, marital status, parent ethnicity, and number of hours spend with their child on the weekend (see p.116, Table 1). Quantitative data included in our meta-analyses is adjusted for parent age (see p.122, Table 4, footnote).
FC and GP: primary school age, attending same school. FC: long-term foster care, placed in that foster family under the age of 1. GP: received no substitute care during their lives.	Yes. Matching on gender, age and FC and GP attended the same school (see p.140, paragraph 'sample selection').
FC and RH: age, born at the same hospital, prenatally drug-exposed, child protective services involved. FC: mostly in non-kinship foster care.	Yes. Matching on infant characteristics (gestational age, height, weight), maternal education level (see p.249, Table 2 and p.251, paragraph 'demographic and perinatal data').
FC and GP: 4-18 years, Niagara county.	No. A description of demographic data is given (see p.47, paragraph 'demographic data'), but there were no reported efforts or statistics for matching of groups.
FC and GP: 12-18 years, from same family (GP consist of biological children of foster parents), no developmental disabilities. FC: at least 6 months in foster family.	Yes. Matching on gender, age, ethnicity, and SES (see p. 50, paragraph 'participants'). Descriptives of both samples are also provided (see p.62, Table 2). Quantitative data included in our meta-analyses is controlled for infant birth weight and height, gestational age, maternal SES, and number of placements (see p.251, paragraph 'results').
FC and GP: 3-7 years. FC: new foster placement. GP: consistently lived with at least 1 biological parents, no previous involvement with child welfare.	Yes. Matching on gender, age, ethnicity (see p.670, Table 1 and paragraph 'participants'). Quantitative data included in our meta-analyses is controlled for age and gender (see p.671, paragraph 'preliminary analyses').
FC and GP: age. FC: healthy full-term babies, admitted before the age of 4 months and continuously institutionalized until at least the age of 2. GP: working-class, London.	Yes. Matching on age and SES (see p.101, paragraph 'parents of later adopted and restored children').
FC and GP: 6-17 years, Ankara, no mental or physical disabilities. FC: child welfare service Ankara.	Yes. Matching on age (see p.131 and p.132, paragraph 'yöntem').
FC and RH: 6-12 years, FASD.	Yes. Matching on gender, age, ethnicity, and FASD (see p.293, Table 1).

Table 2.1. (continued)

Study (Year of Publication)	Country	Groups[A]	*N*	Age Range in years (Mean)	% Female	Measure[B]
Wald et al. (1988) [2,4,6,7]	USA	FC RH GP	8-13 18-19 42-48	5-10 (n/a) 5-10 (n/a) 5-10 (n/a)	38% 31% n/a	WISC-R, WPPSI, social behavior (teacher assessment), CBS

Note. n/a = not applicable

[A] Groups: FC = Foster care, RH = Remained Home, GP = General Population.

[B] Measures: BSID = Bayley Scales for Infant Development, CBCL = Child Behavior Checklist, CBS = Child Behavior Scales, CDI = Children's Depression Inventory, ITSEA = Infant-Toddler Social and Emotional Assessment, MASC = Multidimensional Anxiety Scale for Children, MSEL = Mullen Scales of Early Learning, PDR = Parent Daily Report Checklist, PedsQL = Pediatric Quality of Life Inventory, SDQ = Strengths and Difficulties Questionnaire, SPPA = Self Perception Profile for Adolescents, TRF = Teacher's Report Form, VABS = Vineland Adaptive Behavior Scales, WISC = Wechsler Intelligence Scale for Children, WPPSI = Wechsler Preschool and Primary Scale of Intelligence, YSR = Youth Self Report.

[C] Matching samples: coded as yes or no (for sensitivity analyses). Studies are coded with a 'yes' if matching occurred on individual or a combination of relevant characteristics (e.g., gender, age, SES). See for more information the 'Method' section.

[1] Peer-reviewed journal. [2] Non-peer-reviewed text. [3] Age not specified. [4] Sample size as used in the analyses. [5] Percentage females not specified. [6] Sample size varies by measure/scale. [7] Sample size, age range and mean age at first measurement. [8] Included in sensitivity analysis (criteria: peer-reviewed, sample of at least 50 foster children, validated instruments, matched samples). [9] Edition for measures is not specified.

Inclusion Criteria	Matching Samples[c]
FC, RH and GP: 5-10 years. FC and RH: abused and neglected children. GP: from schools from which a large number of reports of abuse and neglect were perceived each year.	Yes. Matching on gender, age (see p.52, Table 4), and ethnicity (quantitative data for non-Black and non-Hispanic children).

Coding of Studies

Quantitative data were extracted from text and tables and used to compute Hedges g. Like Cohen's *d*, Hedges *g* is an effect size measure but it is computed differently. It incorporates an adjustment which removes the bias of Cohen's *d*. Hedges *g* is defined as the difference between the two means, divided by the pooled standard deviation (Borenstein et al., 2009). If an article provided longitudinal outcome data (e.g., Jacobsen et al., 2013) or multiple dependent samples (e.g., multiple informants for the same study population) (e.g., Janssens & Deboutte, 2010), the findings were averaged in the meta-analyses (Borenstein et al., 2009). If articles included multiple independent samples (e.g., boys and girls) (Bernedo, Salas, García-Martín, & Fuentes, 2011), these were entered in the meta-analyses separately. Because samples included in a meta-analysis must be independent, multiple independent samples were averaged into the meta-analyses for articles which included multiple independent samples for one group, but not for the comparison group (e.g., kinship and non-kinship foster children compared to one group of children from the general population). One article reported both CBCL t-scores and raw scores (Janssens & Deboutte, 2010). We chose to code the t-scores because this was most in line with the other included articles.

For the qualitative analyses and quantitative sensitivity analyses, studies were also coded on several other study or sample characteristics. Studies were coded on country, presence or absence of peer review, groups, sample size, age range, mean age, percentage of female and outcome measures. In addition, each study was coded on type of informant, sampling design, non-response, inclusion criteria, whether or not the effect sizes were adjusted for confounders, and whether or not the compared groups within a study were matched on sample characteristics (e.g., gender, age, ethnicity, parental education, neighborhood). In order to include matching as a criterion for the sensitivity analysis, we sought for scholarly precedence for ascribing and evaluating matching of groups. We did not find any example. Existing measurement tools or scoring systems for the quality of reviews (Detsky, Naylor, O'Rourke, McGeer, & L'Abbé, 1992; Moher, Liberati, Tetzlaff, Altman, & Group, 2009; Sanderson, Tatt, & Higgins, 2007; Shea et al., 2009; Wells et al., 2011; Zeng et al., 2015) either were not suitable for our meta-analysis or did not include matching. In addition, available systems have been criticized (e.g., Stang, 2010). One of the main criticisms is that no empirical evidence of their validity has been presented (cf., Moayyedi, 2004). Our own evaluations and the criticisms of other scholars made us decide to design our own matching protocol wherein

studies were labeled with a 'yes' or a 'no', indicating whether the study was matched or not (see Table 2.1). To classify for a 'yes' on matching, studies had to have matched their samples on relevant variables such as socioeconomic status (SES; e.g., family income, parental education, neighborhood), gender and age. These variables are shown to be related to children's developmental functioning (Bongers, Koot, Van der Ende, & Verhulst, 2003; Bradley & Corwyn, 2002) and are therefore likely to be related to decisions as regards children's living arrangement (Curtis, Alexander, & Lunghofer, 2001; Scholte, 1997). We chose to code matching as a categorical variable, in order to facilitate the objectivity of evaluations and accommodate the replicability of the process and outcomes. A simple matching protocol allows for clear and easy adaptation when needed in future studies. The coding of matching in a quantitative manner (i.e., count on how many variables samples were matched) would not have given a fair or adequate representation of the quality of the matching (i.e., some studies did match on a substantial number of variables, but these variables were neither necessarily relevant nor was the matching of these studies always of better quality than those of studies that matched on fewer, but more relevant variables). Moderator analyses based on study and sample characteristics were not performed because there were too few studies to perform such analyses in a reliable manner.

Two authors coded the quantitative data as well as the characteristics of the individual studies which were used for the qualitative review and sensitivity analyses by using a standardized abstraction form. Differences between coders were resolved by discussion. Prior to discussion, the authors coded identically 95% of the time.

Analyses

The meta-analyses were performed using the program Comprehensive Meta-Analysis version 2.2 (Borenstein, Hedges, Higgins, & Rothstein, 2005). For most studies, means and standard deviations were used to compute effect sizes. For some studies, we coded proportions (Burns et al., 2004; Colton et al., 1995; Heflinger et al., 2000; Wald et al., 1988), or sample sizes and *t* or *F* test statistics (Davis, 2000; Tininenko, Fisher, Bruce, & Pears, 2010). One study only reported that there were no significant differences between groups (Leifer & Shapiro, 1995). The effect size for this study was fixed at $r = .00$ (Mullen, 2013). For the study of Mennen et al. (2010), standard errors were converted into standard deviations (Moore & McCabe, 2006).

The current article reports on ten meta-analyses: two different group comparisons (foster care vs general population and foster care vs children at risk who remained at home) for five developmental outcome domains (internalizing problems, externalizing problems, total problem behavior, adaptive behavior and cognitive functioning). A random effects model was used to analyze the data, because a random effects model does not assume a common underlying effect size for all included studies and is commonly more appropriate for meta-analyses based on a literature search than a fixed effects model (Borenstein et al., 2009). The homogeneity of studies was tested using the Q test. If the Q test is significant, there is true heterogeneity between effect sizes of the studies included in the meta-analysis. To quantify this heterogeneity, the I^2 was used. The I^2 is the percentage of total variability in a set of effect sizes due to true heterogeneity (Huedo-Medina et al., 2006). To identify studies with

large influence on the overall effect size estimate, the jackknife procedure was used. This 'one-study-removed' procedure indicates whether the overall effect size of the meta-analysis is biased due to the influence of any one study (Borenstein et al., 2009). Sensitivity analyses were performed to assess the robustness of the results to different assumptions and inclusion criteria (Egger et al., 1997). These analyses only included studies which were published in peer-reviewed journals, had a sample of at least 50 foster children, made use of validated instruments, and compared matched groups (see Table 2.1). We examined whether results from these studies differed from the overall findings of the meta-analyses wherein all studies were included.

Publication bias was assessed with Duvall and Tweedie's trim-and-fill procedure, the Kendall's τ method and the Failsafe *N*. The Duvall and Tweedie's (2000) trim-and-fill procedure provides a more unbiased estimate of the effect size than the observed effect size, by repeatedly imputing effect sizes until the error distribution closely approximates normality. The Kendall's τ method represents the association between the standardized effect sizes and the variance of these effect sizes. A non-significant Kendall's τ coefficient suggests the absence of publication bias, and a significant Kendall's τ indicates that small studies with non-significant results tend not to be published (Rothstein et al., 2006). Finally, if the overall effect of the meta-analysis turned out to be significant, Rosenthal's Failsafe *N* and Orwin's Failsafe *N* were computed. Rosenthal's Failsafe *N* indicates how many studies are required to nullify the effect, that is to a statistically non-significant effect. The Orwin's Failsafe *N* indicates how many studies are required to bring the overall effect to a specified level other than zero (Cooper et al., 2009; Rothstein et al., 2006).

RESULTS

Children in foster care (N = 2,305) were compared with children who remained home (N = 4,335) and children from the general population (N = 4,971). For each of these group comparisons, a qualitative review is given of characteristics of the individual studies and samples (e.g., publication outlet, type of informant, age and gender distribution, range and mean of sample sizes, matching of groups, inclusion criteria, sampling design, non-response, adjustment for confounders). This should facilitate the understanding of differences between studies that affect their comparability and quality. We furthermore present the findings of the meta-analysis on the basis of a table and forest plot, which give an overview of findings with respect to cognitive functioning, adaptive functioning, internalizing behavior problems, externalizing behavior problems, and total problems. In these tables the number of studies eligible for inclusion in the meta-analysis is characterized by *k*. The *N* indicates how many foster children and children in the group comparison were included in each meta-analysis. The results are represented by Hedges g, its accompanying confidence intervals (CI), and the *p*-value. In addition, the Q-statistic, I^2 and Kendall's τ are reported. The last two columns of each table display the results of the sensitivity analysis for high quality studies.

In the forest plots, a graphic representation of the effect sizes of the included studies is given, as well as the overall effect size. For the meta-analyses on cognitive and adaptive functioning, a positive effect size means that foster children perform better than the comparison group. For the meta-analyses on behavioral problems, a positive effect size indicates higher levels of behavior problems for foster children.

Qualitative Review of the Studies on Foster Children and Children Remaining at Home

We found 13 studies on the comparison of children in foster care (N = 1,129) with children at risk who remained home (N = 4,335), reporting 13 independent effect sizes. Nine studies were published in peer-reviewed journals. Ten studies were performed in the United States, and three in Europe. Most studies (k = 7) made use of only one informant, while six studies made use of a combination of reports by different informants. Ten studies made use of caregiver reports (parents or foster parents), three studies of teacher reports, and eight studies of child reports or developmental assessment with children (e.g., Wechsler intelligence scales). Regarding the sampling design and inclusion criteria, most studies tried to sample the entire target population in a certain area (e.g., Beatty, 1995; Janssens & Deboutte, 2010; Mennen et al., 2010), from certain hospitals (Min et al., 2014; Salo et al., 2009), the same school (Colton et al., 1995), or tried to give a representative random sample of children in state care (Burns et al., 2004; Heflinger et al., 2000). Non-response was not reported for almost half of the studies (k = 6), and for the other seven studies the reported non-response varied widely from 17% (Heflinger et al., 2000) to 87% (Beatty, 1995). Two of the thirteen studies reported effect sizes adjusted for confounders (Leifer & Shapiro, 1995; Mennen et al., 2010); effect sizes were adjusted for varying demographic characteristics, such as age, gender, SES, and ethnicity (Leifer & Shapiro, 1995; Mennen et al., 2010). Leifer and Shapiro (1995) also controlled for the number of abuse incidents or the number of placements. The sample sizes between studies ranged from 13 (Wald et al., 1988) to 323 (Burns et al., 2004) for foster children, and from seven (Salo et al., 2009) to 3,411 (Burns et al., 2004) for children at risk who remained at home. The average number of respondents per sample was 87 for foster children and 333 for children at risk who remained at home. Within the majority of the studies, the group of foster children and the group of children who remained at home were matched on gender and age. Except three studies (Leifer & Shapiro, 1995; Maroufi, 2003; Wald et al., 1988), almost all studies included an approximately equal distribution of boys and girls. Some studies also matched on other characteristics, such as SES-related variables (Janssens & Deboutte, 2010; Leifer & Shapiro, 1995; Salo et al., 2009) or on ethnicity (Mennen et al., 2010; Victor et al., 2008). The age of the children ranged from 2 to 18 years. Between studies, the group of children at risk who remained with their parents differed as regards the extent to which they received additional home support services (cf., Beatty, 1995; Burns et al., 2004). But also within studies the amount of support services could vary widely (cf., Leifer & Shapiro, 1995).

Meta-Analyses on the Comparison of Foster Children with Children Remaining at Home

The characteristics and results for the meta-analyses comparing foster children with children at risk who remained at home are displayed in Table 2.2; all comparisons were non-significant. This means that foster children do not differ from children who remained at home with respect to cognitive functioning, adaptive functioning and behavior problems. The forest plots are provided in Figures 2.2, 2.3, 2.4, 2.5, and 2.6. Sensitivity analyses for studies with high-quality study characteristics did not show different outcomes.

For the meta-analysis on cognitive functioning, adaptive functioning and the two meta-analyses on behavioral problems (internalizing and externalizing), the jackknife procedure showed that the overall effect remained the same when one study at a time was removed from the meta-analyses. When the study of Min, Minnes, Yoon, Short and Singer (2014) was removed from the meta-analysis on total behavior problems, the overall effect became significant with children in foster care showing fewer total behavioral problems than children at risk who remained at home.

The Kendall's τ suggested absence of publication bias for each meta-analysis (see Table 2.2). For the meta-analysis on adaptive functioning, the Duvall and Tweedie's trim-and-fill procedure suggested that one study to the left of the mean needed to be imputed to shift the point estimate from 0.13 (95% CI: [-0.34, 0.08]) to -0.18 (95% CI: [-0.39, 0.05]), which still indicates a non-significant overall effect. For the meta-analysis on cognitive functioning, one study to the right of the mean needed to be imputed to shift the point estimate from 0.11 (95% CI: [-0.13, 0.36]) to 0.17 (95% CI: [-0.07, 0.42]), still a non-significant overall effect. For internalizing behavior problems, one study to the left of the mean would shift the observed point estimate from 0.13 (95% CI: [-0.31, 0.05]) to 0.21 (95% CI: [-0.40, -0.01]), a significant overall effect. Imputation of two studies to the left of the mean would be required in the meta-analysis on externalizing behavior problems to shift the observed point estimate from 0.05 (95% CI: [-0.12, 0.21]) to 0.15 (95% CI: [-0.02, 0.33]), which still indicates a non-significant overall effect. For total behavior problems, imputation of one study to the right of the mean would shift the observed point estimate from -0.12 (95% CI: [-0.29, 0.05]) to -0.11 (95% CI: [-0.28, 0.06]), which again still indicates a non-significant overall effect.

Table 2.2. *Meta-analyses on the comparison of foster care (fc) with children who remained at home (rh).*

Outcome domain	***k* (samples)**	***N* fc / *N* rh**	***g* (95% CI)**	***p***	***Q (p)***
Cognitive functioning	6 (6)	263 / 211	.12 (-.13, .36)	.36	11.74 (p = .04)
Adaptive functioning	6 (6)	368 / 573	-.13 (-.34, .08)	.22	16.10 (p = .01)
Internalizing behavior problems	5 (5)	529 / 600	-.13 (-.31, .05)	.15	9.27 (p = .06)
Externalizing behavior problems	5 (5)	542 / 668	.05 (-.12, .21)	.58	10.77 (p = .03)
Total behavior problems	11 (11)	1,032 / 4,377	-.12 (-.29, .05)	.18	72.45 (p < .001)

Note. n/a = not applicable

[A] Failsafe *N* analyses were only performed when the overall effect of the meta-analyses turned out to be significant.

[B] Sensitivity analyses were performed for selected studies (criteria: peer-reviewed, sample of at least 50 foster children, validated instruments, matched samples). Sensitivity analyses were performed if at least two studies qualified for inclusion.

Table 2.3. *Meta-analyses on the comparison of foster care (fc) vs. general population (gp).*

Outcome domain	***k* (samples)**	***N* fc / *N* rh**	***g* (95% CI)**	***p***	***Q (p)***
Cognitive functioning	5 (5)	280 / 227	-2.41 (-3.33,-1.48)	< .001	153.54 (p < .001)
Adaptive functioning	7 (7)	424 / 1,096	-.36 (-.68, -.03)	.03	58.82 (p < .001)
Internalizing behavior problems	10 (11)	818 / 4,306	.21 (-.04, .47)	.10	107.85 (p < .001)
Externalizing behavior problems	10 (11)	818 / 4,306	.59 (.28, .90)	< .001	158.58 (p < .001)
Total behavior problems	16 (17)	1,073 / 4,543	.54 (.25, .84)	< .001	330.90 (p < .001)

Note. n/a = not applicable

[A] Failsafe N analyses were only performed when the overall effect of the meta-analyses turned out to be significant.

[B] Sensitivity analyses were performed for selected studies (criteria: peer-reviewed, sample of at least 50 foster children, validated instruments, matched samples). Sensitivity analyses were performed if at least two studies qualified for inclusion.

I^2	Kendall's *(z, p)*	Rosenthal's Failsafe N^A	Orwin's Failsafe N^A	*k* (samples) Sensitivity analysis[B]	*g* (CI), *p* Sensitivity analysis[B]
57.41	.27 (*z* = .75, *p* = .23)	n/a	n/a	n/a	n/a
68.94	-.00 (*z* = .00, *p* = .50)	n/a	n/a	2 (2)	-.19 (-.55, .16), *p* = .29
56.86	-.10 (*z* = .24, *p* = .40)	n/a	n/a	3 (3)	-.16 (-.40, .09), *p* = .20
62.86	-.00 (*z* =.00, *p* = .50)	n/a	n/a	3 (3)	-.02 (-.11, .10), *p* = .90
86.20	-.11 (*z* = .47, *p* = .32)	n/a	n/a	4 (4)	-.20 (-.45, .05), *p* = .16

I^2	Kendall's *(z, p)*	Rosenthal's Failsafe N^A	Orwin's Failsafe N^A	*k* (samples) Sensitivity analysis[B]	*g* (CI), *p* Sensitivity analysis[B]
97.40	-.50 (*z* = 1.22, *p* = .11)	525	53	3 (3)	-1.09 (-1.55, -.64), *p* <.001
89.80	-.29 (*z* = .90, *p* = .18)	52	4	4 (4)	-.33 (-.74, .06), *p* = .10
90.73	.11 (*z* = . 47, *p* = .32)	n/a	n/a	5 (6)	.25 (-.08, .58), *p* = .14
93.69	.25 (*z* = 1.09, *p* = .14)	839	87	5 (6)	.65 (.25, 1.05), *p* = .00
95.17	.14 (*z* = .78, *p* = .21)	1,062	78	6 (7)	.52 (.01, 1.0), *p* = .05

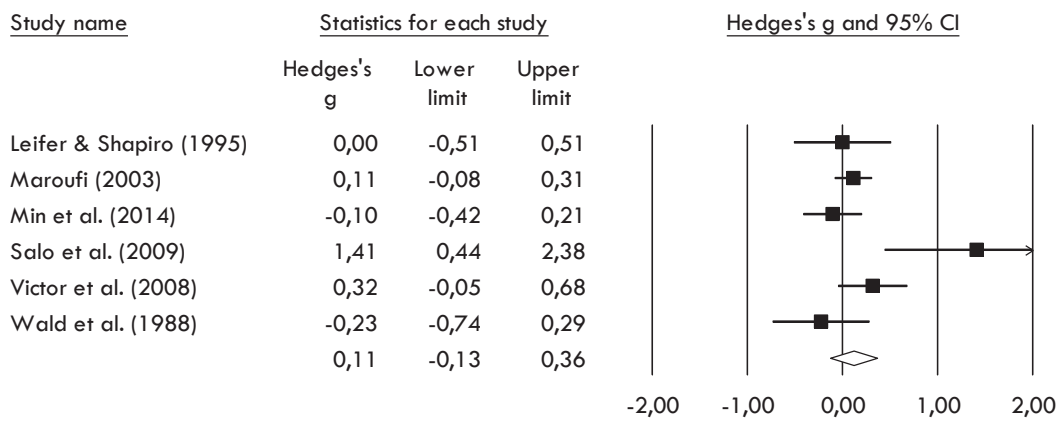

Figure 2.2. Forest plot for the meta-analysis on the comparison of foster care with children remaining at home on cognitive functioning.

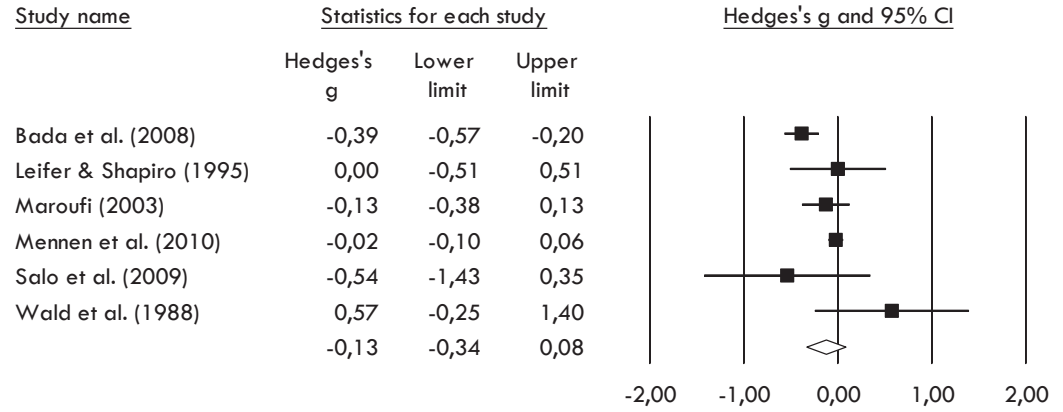

Figure 2.3. Forest plot for the meta-analysis on the comparison of foster care with children remaining at home on adaptive functioning.

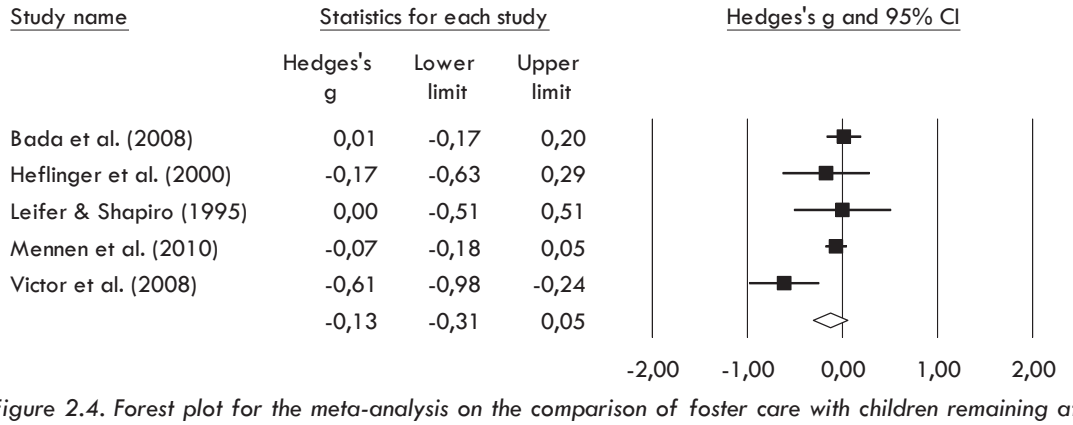

Figure 2.4. Forest plot for the meta-analysis on the comparison of foster care with children remaining at home on internalizing behavior problem.

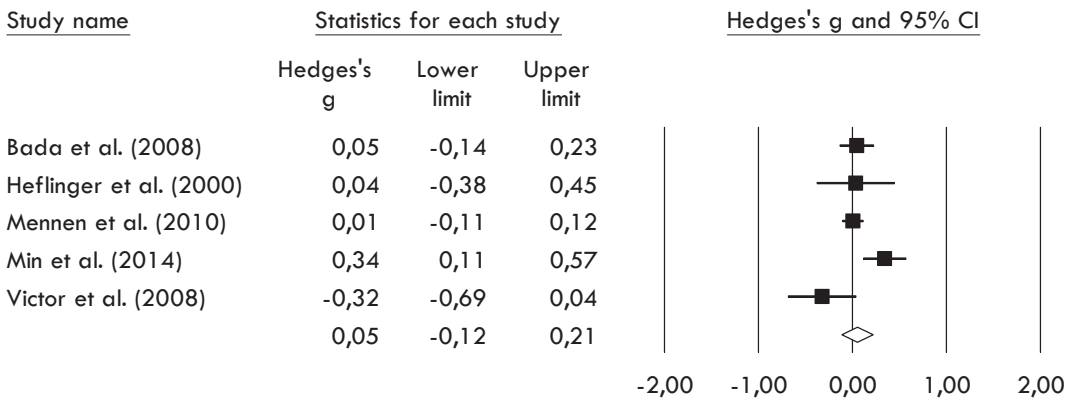

Figure 2.5. Forest plot for the meta-analysis on the comparison of foster care with children remaining at home on externalizing behavior problems.

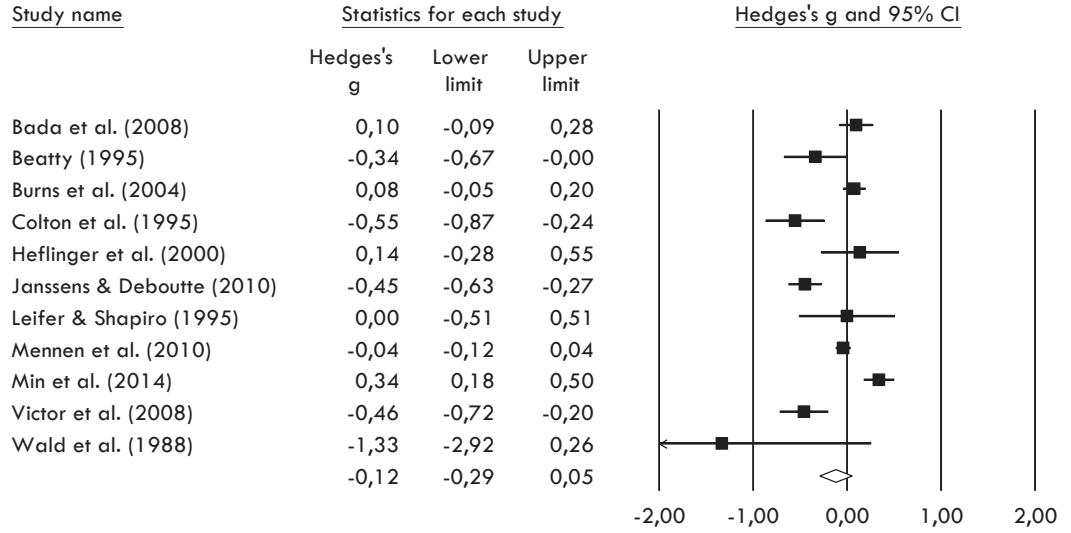

Figure 2.6. Forest plot for the meta-analysis on the comparison of foster care with children remaining at home on total behavior problems.

Qualitative Review of the Studies on Foster Care and the General Population

We found 21 studies, with 22 independent samples, that compared the developmental outcomes of children in foster care (N = 1,479) with that of children from the general population (N = 4,971). The majority of the studies (k = 14) was published in peer-reviewed journals. Approximately half of the studies were performed in the United States (k = 11), the other studies came from European countries and Australia. Approximately three-quarters of the studies (k = 15) included the reports of one informant only. Studies used caregiver reports (k = 12), teacher reports (k = 6), and/or children's self-report or assessment tests (k = 11). Foster children were often sampled from several participating foster care institutions, or from all foster children in a certain area (e.g., Carbone et al., 2007; Fernandez, 2008), although it was not always reported which inclusion criteria were used for selection of foster children within these institutions or areas (e.g., Damnjanoviç et al., 2012). Sampling designs for children from the general population varied widely; children were randomly picked from the general population (Carbone et al., 2007), recruited with flyers or advertisements (e.g., Bruce et al., 2013), chosen because they were a classmate of the included foster child (e.g., Fernandez, 2008), or included because they came from Medicaid families (Hulsey & White, 1989). As a consequence of different sampling designs, matching of foster children and children from the general population also varied widely. Most often, but not always (e.g., Damnjanoviç et al., 2012), the groups were matched on gender and age and sometimes also on social background (e.g., Bernedo et al., 2011; Hulsey & White, 1989; Pears & Fisher, 2005). Six studies matched their compared samples on SES-related variables (Bernedo et al., 2011; Bruce et al., 2013; Hulsey & White, 1989; Jacobsen et al., 2013; Shepherd, 2009; Tizard & Hodges, 1978) and a few others mentioned that the samples were matched on characteristics like SES and cultural background, a matching that was not very precise, but argued from the fact that the samples were all sampled from the same geographical areas or schools (Davis, 2000; Mennen et al., 2010; Roy et al., 2000). Non-response was not reported for the majority of the studies (k = 14). For the studies for which non-response was known, rates varied from 7% (Roy et al., 2000) to 83% (Üstüner et al., 2005). Effect sizes were adjusted for confounders in only a few studies (e.g., Damnjanoviç et al., 2012; Mennen et al., 2010). Confounders controlled for were gender, ethnicity, and/or age. Sample sizes differed between studies, ranging from 7 (Tizard & Hodges, 1978) to 326 (Carbone et al., 2007) for foster care samples, and from 11 (Bruce et al., 2013) to 3,255 (Carbone et al., 2007) for samples of children from the general population. For foster care samples, the average number of respondents was 70, compared to 237 of the general population. The age of the children ranged from 2 to 18 years. With only a few exceptions (McAuley, 1996; Roy et al., 2000; Tizard & Hodges, 1978; Wald et al., 1988), the studies reported an equal number of boys and girls in the analyses.

Meta-Analyses on the Comparison of Foster Care versus the General Population

Table 2.3 displays the characteristics and results of the meta-analyses on the comparison of foster children with children from the general population. All analyses, except the analysis on internalizing behavior problems, revealed an overall significant difference between the effect sizes of the two groups. Foster children showed lower levels of cognitive and adaptive functioning, and had significantly more externalizing and total behavior problems than children

in community samples. Figures 2.7, 2.8, 2.9, 2.10, and 2.11 provide a graphic representation of the meta-analyses. Sensitivity analyses showed that except for adaptive functioning, the results remained the same for all outcomes. For adaptive functioning, the result changed from significant to non-significant, meaning that in studies with high quality study characteristics no differences in adaptive functioning were found between foster children and children from the general population.

For the meta-analysis on adaptive functioning, the jackknife procedure showed that the overall effect became non-significant ($p > .05$) when the study of Bada et al. (2008), the study of Jacobsen, Moe, Ivarsson, Wentzel-Larsen and Smith (2013), the study of Mennen et al. (2010), or the study of Schiefer (1994) were removed from the meta-analysis. The meta-analysis on internalizing behavior problems became significant when the study of Fernandez (2008) and Üstüner, Erol and Simsek (2005) were removed. For the meta-analyses with respect to the other three domains, the overall effect remained the same when one study at a time was removed.

The Kendall's τ suggested absence of publication bias for either meta-analysis. For the meta-analysis on cognitive functioning, the Duvall and Tweedie's trim-and-fill procedure suggested that an imputation of one study to the left of the mean would shift the observed point estimate from 2.41(95% CI: [-3.33, -1.48]) to -3.31 (95% CI: [-4.83, -1.79]), which is still significant. For the meta-analyses on adaptive and behavioral functioning, the Duvall and Tweedie's trim-and-fill procedure suggested that the analyzed studies yielded an unbiased estimate that was the same as the observed effect size. Rosenthal's Failsafe N varied from 49 to 1,062 and Orwin's Failsafe *N* varied from 4 to 87 (see Table 2.3).

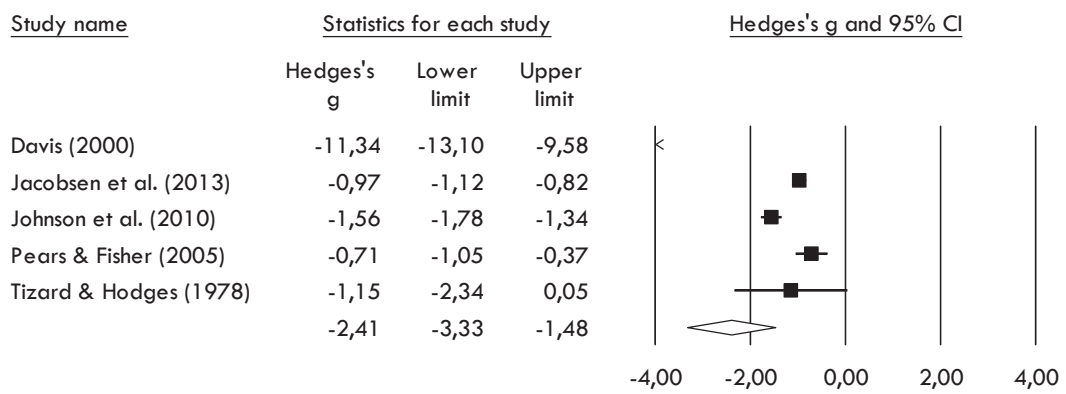

Figure 2.7. Forest plot for the meta-analysis on the comparison of foster care with the general population on cognitive functioning.

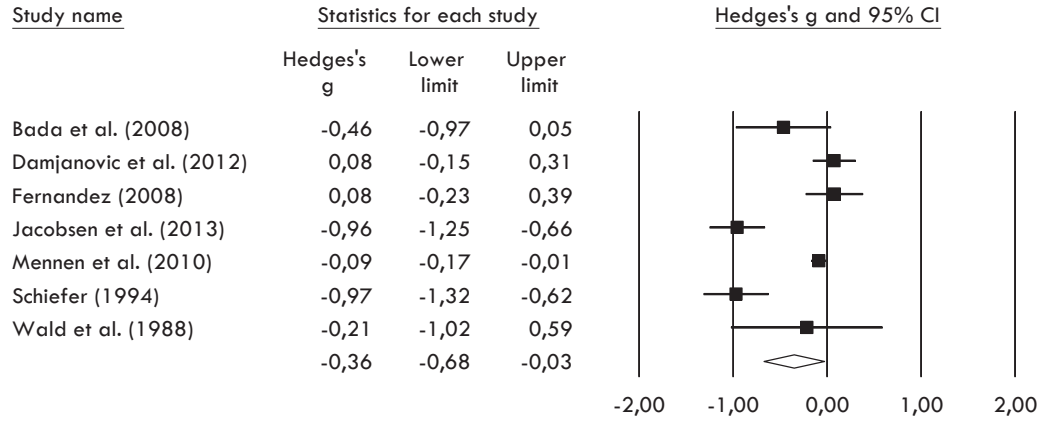

Figure 2.8. Forest plot for the meta-analysis on the comparison of foster care with the general population on adaptive functioning.

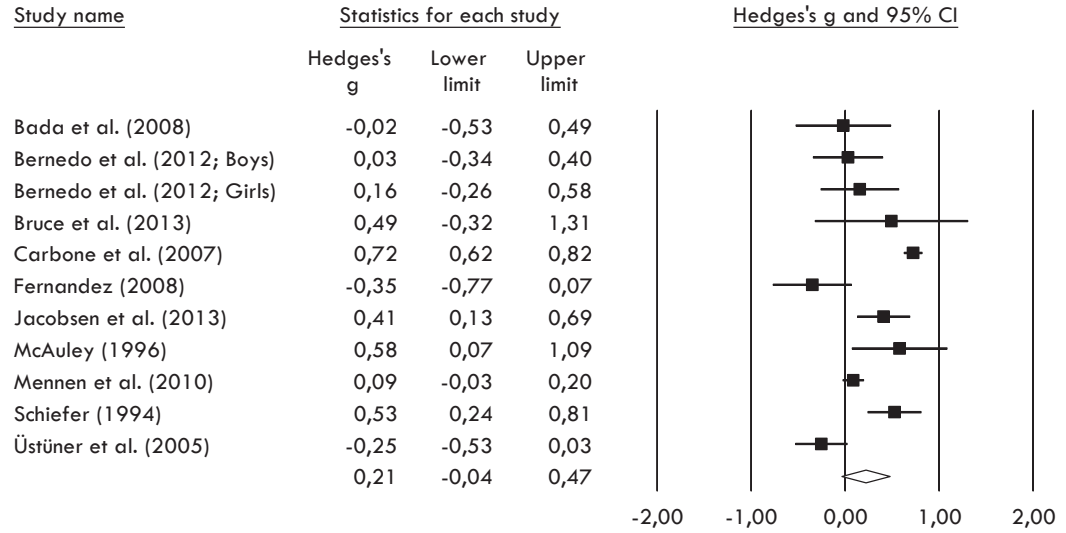

Figure 2.9. Forest plot for the meta-analysis on the comparison of foster care with the general population on internalizing behavior problems.

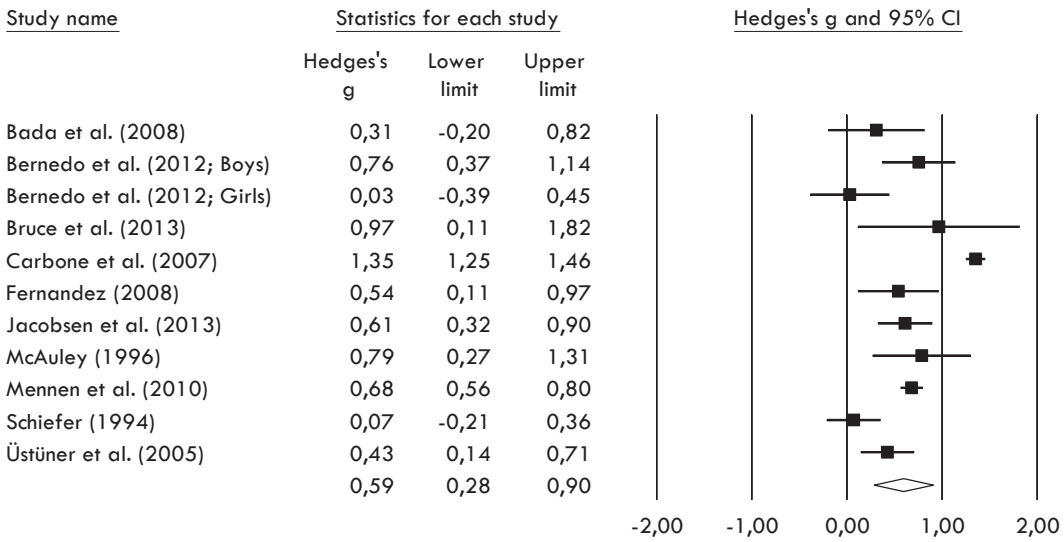

Figure 2.10. Forest plot for the meta-analysis on the comparison of foster care with the general population on externalizing behavior problems.

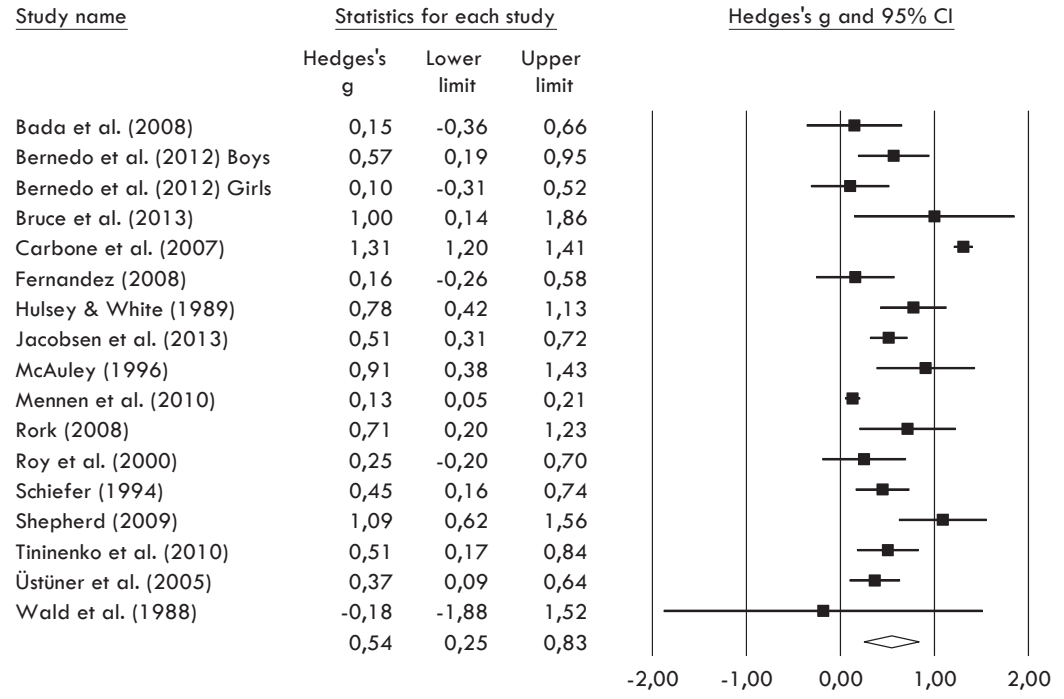

Figure 2.11. Forest plot for the meta-analysis on the comparison of foster care with the general population on total behavior problems.

Chapter 2

DISCUSSION

This study shed light on how the developmental outcomes of foster children relate to those of children from the general population and children at risk who remained at home. Our meta-analyses provided a systematic overview of comparison studies for cognitive, adaptive, and behavioral functioning simultaneously, and suggested that children in foster care perform worse on cognitive and adaptive functioning and on behavioral functioning as regards externalizing problems and overall behavioral problems than children from the general population. Furthermore, in comparison with children at risk who remained at home, foster children showed similar levels of functioning on all developmental domains. Publication bias did not seem to have a large influence on the effects, except for the meta-analysis on adaptive functioning in the comparison of foster care and children from the general population. Contrary to our hypothesis, foster children did not significantly differ from children in the general population on internalizing behavior problems. A possible explanation is that externalizing behaviors are more prominent and readily observable than internalizing behaviors (Kolko & Kazdin, 1993). Moreover, the type of informant may also have played a role, because previous studies have reported that there were significant differences between foster parents' and foster children's reports on internalizing behaviors (Strijker, Van Oijen, & Knot-Dickscheit, 2011).

While this meta-analysis showed that foster children had generally lower levels of functioning than children from the general population, a different trend is found for the comparison of foster children with children at risk who remained at home. This study showed that foster children and children at risk do equally well on all included domains of developmental functioning. The fact that foster children are similar to children at risk who remained in their homes of origin could be considered a positive sign. Both groups experience several childhood adversities and come from families characterized by at risk home environments. Although, as suggested in the introduction, several historical and political trends might affect placement decisions, it could be argued that the pre-existing differences between children in different placement settings, their histories and caregiving environments are to the disadvantage of foster children because, eventually, it is not without reason that the decision for out-of-home placement is made (Biehal, 2007; Van Santen, 2010). As a consequence, children placed in foster care might have experienced additional trauma compared with children at risk who remained at home. In addition, subsequent broken attachment due to the separation from their biological parents confronts foster children with another risk (Kinard, 1982). Despite these potentially challenging experiences, foster care apparently offers children a safe and stimulating home environment which partly helps to recover and to reach similar levels of functioning compared with children who initially might have been better off, like high risk children staying at home (Joseph, O'Connor, Briskman, Maughan, & Scott, 2014; Victor et al., 2008). A possible explanation is that the benefits of the more supportive home environment which foster care can offer, are eventually balanced with the higher initial levels of developmental problems of foster children compared to children who remained at home (Lloyd & Barth, 2011). At the same time, the impact of out-of-home placement for children

should not be underestimated. Trying to prevent separation from the biological parents is an important goal. Therefore, it would be of interest to examine whether options like part-time fostering as a family support service (Roberts, 2016) and shared family care (Barth & Price, 1999) combine the benefits of both foster care and remaining at home into one type of living arrangement.

The similarity between foster children and children at risk who remained at home should not trigger complacency. After all, similarity does not say anything about the level on which these children function. It is known that children enter foster care with developmental problems, and a recent meta-analysis on the adaptive and behavioral development of foster children showed that these children did not improve their functioning during their stay in foster care (Goemans, Van Geel, & Vedder, 2015). This suggests that both foster children and children who remained at home experience developmental problems. This is also supported by the current study, which demonstrated that foster children stay behind on cognitive, adaptive and most aspects of behavioral functioning when compared with children from the general population. We agree with scholars who earlier suggested that improvements in foster care are necessary (Dozier, Albus, Fisher, & Sepulveda, 2002; Fisher, Burraston, & Pears, 2005; Fisher, Chamberlain, & Leve, 2009; Jee et al., 2011; Leslie, Gordon, Lambros, et al., 2005; Price et al., 2008). An often suggested way to improve the functioning of children in foster care is to carefully screen and monitor their safety and development (Chambers, Saunders, New, Williams, & Stachurska, 2010; Garwood & Close, 2001; Jee et al., 2010; Tarren-Sweeney, 2007). This standardized record keeping should represent a complete picture of the child, focusing not only on negative outcomes but also on several well-being indicators such as child education, health and behavioral functioning (Harden, 2004; Jonson-Reid & Drake, 2016). This information can be used by child welfare professionals for monitoring and clinical casework decisions, and by researchers to perform their studies and seek and find support for evidence-based practice.

Limitations and Implications for Future Research

Although this study yielded clear findings, several limitations should be noted. In general, meta-analysis relies on the confidence regarding the comparability between individual studies and samples. In the current study, not only the caregiving settings differed between individual studies, but likely also relevant characteristics of children and caregivers engaged in these settings (Beatty, 1995; Pears & Fisher, 2005; Pottick, Warner, & Yoder, 2005). Nevertheless, one of the advantages of meta-analysis is to take this variety and variance in studies into account (Q-test statistics, I^2, sensitivity analyses). This is especially relevant here, because, differences did exist not only between placement options but also within specific placement forms (Winokur, Holtan, & Batchelder, 2014). The current study did not differentiate between possibly relevant types of foster care (e.g., kinship/non-kinship foster care) because the distinctions were often not made by the individual studies included in the meta-analyses. However, both kinship and non-kinship foster care are community placements that contrast to in-home services. Still, many other potential moderating effects on foster children's functioning were not accounted for (e.g., children's age, placement duration, placement history, provision of additional services). In addition, it could be that different developmental outcomes are

present for different groups of children at risk remaining at home while receiving different types and intensities of in-home support services (Bada et al., 2008; Beatty, 1995). In the current study, it was shown that the heterogeneity between the effect sizes of included studies was large, which was possibly due to the differences between the children, caregivers, and environments involved in the settings studied. Unfortunately, it was not possible to perform moderator analysis using this heterogeneity as a moderator variable because there were too few studies to perform a moderator analysis in a reliable manner.

Another issue, which is especially relevant in our meta-analysis, is the matching of samples within studies. We examined the degree to which the individual studies included in the meta-analysis took pre-existing differences of children, caregivers, and caregiving settings, into account by giving an overview of the inclusion criteria and matching of the samples for each study. The studies included in our meta-analyses varied widely on the strictness of the inclusion criteria and the quality of the matching. Groups were often matched on characteristics like age or gender, but only less than a third of the studies matched their samples on SES-related variables. Unfortunately, samples were almost never matched on experienced adversities, such as type or severity of maltreatment. As a consequence, we could not include these variables as matching variables in our meta-analysis. On the one hand, there are several studies in which the matching is not described at all, or where the information provided was not equally clear (samples were approximately matched), which makes it difficult to draw conclusions. Sometimes this is because the comparison of the samples on developmental outcomes is not the primary goal of the studies (e.g., Burns et al., 2004). This complicates analyzing baseline differences between non-matched samples. On the other hand, it seems that several researchers are well aware of the importance of matching and comparability, but nevertheless did not or not fully accomplished to match their samples on predefined characteristics (e.g., Jacobsen et al., 2013; Wald et al., 1988). For these studies, it is possible to indicate baseline differences (such as differences in parental education, ethnicity, or type of abuse). These differences should be taken into account when interpreting the results of the individual studies. Furthermore, if matching is difficult to realize, controlling for relevant variables might be a solution. Few studies adjusted their findings for confounders, and sometimes the results were controlled for variables on which the samples were matched (e.g., Mennen et al., 2010; Tininenko et al., 2010). It can be concluded that matching on relevant variables is challenging (e.g., Wald et al., 1988). Our sensitivity analyses do not suggest that studies in which matching was rigorous provided very different results from studies wherein there was less of an effort to match the samples. Nonetheless, pre-existing differences will remain a problem in the interpretation of the results of any non-experimental study wherein foster children are compared to another population, and therefore the pre-existing differences are also a challenge in our meta-analyses.

Still, our meta-analyses establish that children from foster care *in general* fare worse than children from the general population and equally well compared to children at risk who remained at home. By using sensitivity analyses, in which one of the criteria to be included was the matching of samples within studies, we have excluded the explanation that differences in the methodological quality of the studies explain our results. Only for the meta-analysis comparing the adaptive functioning of foster children and children from the general population, we found that the sensitivity analyses showed a different result compared to the meta-analysis for all

studies. It should be noted that we cannot draw conclusions about whether the difference between the original meta-analysis and the sensitivity analysis is statistically significant, because in sensitivity analyses, informal comparisons are made between different ways of estimating the same phenomenon instead of formal statistical comparisons. In the high quality studies, there appeared no difference in adaptive functioning between children in foster care and children in the general population, whereas in the meta-analysis with all studies included, foster children showed worse adaptive functioning than children in the general population. At the same time, we must be careful with the interpretation of the meta-analysis and sensitivity analysis, because it was this same meta-analysis which showed indications for publications bias.

Although we have attempted to exclude the explanation that differences in methodological quality of the studies explain our results, our meta-analyses do not provide full insight in the factors that explain why children in foster care generally fare worse than children in the general population and just as good as children at risk who remained at home. Pre-existing differences might partly be responsible for differences in children's developmental outcomes, but future studies should examine this in more detail. Moreover, research exploring whether foster care or remaining at home with extra support services is in the best interest of children would definitely benefit the quality of placement decisions. A suggestion toward this end is careful matching of groups within studies. Future comparisons should preferably include samples that are as similar as possible or provide better controls for possible confounders (Whetten et al., 2014). Ideally, matching of groups is done not only on factors such as age, gender, and SES but also on experienced adversities. Moreover, more efforts should be made to systematically register and compare the development and functioning of children in the child welfare system (Wald et al., 1988). Although random assignment at agency level to either foster care or the option of remaining at-home while receiving additional support is not possible, countries wherein child welfare practices can differ because of differences in policies on state levels (Russell & Macgill, 2015; Wald et al., 1988) might allow for a more 'natural experiment' on this matter. It may well be that children with certain characteristics are better served by either foster care or in-home services; in the current meta-analysis we could not include child characteristics. Future studies should establish which children are best served by either foster care or in-home services (Washington, 2004).

Conclusions

For children in foster care and for children remaining at home while receiving intensive home support services we found comparable developmental outcomes. Although child welfare policy favors family preservation, it has been suggested that foster children have better access to mental health services than children at risk who remained at home (Burns et al., 2004; Hurlburt et al., 2004; Ringeisen, Casanueva, Cross, & Urato, 2009). In addition, the risk of repeated maltreatment is lower for children in foster care (Campbell et al., 2012; Runyan & Gould, 1985). This does not mean that foster care agencies should not be wary of maltreatment (Benedict, Zuravin, Somerfield, & Brandt, 1996). Continued attention, standardized record keeping of (repeated) maltreatment by child welfare and foster care professionals, and further research is recommended to assess the risks of both remaining at home and living in a foster family, and to identify whether child safety is ensured (Jellinek

et al., 1995). Furthermore, the current study showed that compared to normative standards both children in foster care and children who remained at home face serious developmental challenges as regards cognitive, adaptive, and behavioral functioning. It is important that, irrespective of whether the child is in foster care or remained at home, policy makers and care professionals are aware of the developmental challenges for these children. Careful developmental monitoring and additional support services are necessary to assure that foster children and children in the child welfare system develop in the best possible manner.

2