J Abnorm Child Psychol (2017) 45:527–542
DOI 10.1007/s10802-016-0178-0

Two-Year Findings from a National Effectiveness Trial: Effectiveness of Behavioral and Non-Behavioral Parenting Programs

Jens Högström[1] · Viveca Olofsson[2] · Metin Özdemir[2] · Pia Enebrink[1] · Håkan Stattin[2]

Published online: 22 June 2016

Abstract Long-term follow-up studies of selective parent training (PT) programs are scarce, particularly in the case of effectiveness trials conducted within regular care settings. This study evaluated the 2-year effects of 4 programs: Comet, Incredible Years, Cope, and Connect and differences in the rate of change among programs were investigated using Latent Growth Modeling (LGM). Participants were parents who had sought help at 30 local service sector units (e.g., child psychiatric clinics and social services centers) for major problems in managing their children's externalizing behavior. Parents of 749 children (63 % boys) with moderate levels of externalizing behavior, aged 3–12, were randomized to one of the 4 PT programs. Assessments included parent-reported measures of child externalizing, hyperactivity and inattention, as well as parenting practices, sense of competence, and parents' stress and depressive symptoms. At 2-year follow-up, there were no differences in any of the child outcomes among the programs. All programs had reduced externalizing behaviors with large effect sizes (d = 1.21 to d = 1.32), and negative parenting practices with moderate to large effect sizes (d = 0.49 to d = 0.83). LGM analyses showed that the 2 behavioral programs, Comet and Incredible Years, produced more rapid reductions in externalizing behavior during the course of the intervention than the non-behavioral program, Connect. Connect, however, was the only program where children continued to improve after the intervention. Overall, the results indicate that the 4 programs were equally effective in a clinical setting, despite differences in their theoretical origin.

Keywords Parent training programs · Externalizing behavior · Long-term follow-up · Effectiveness · Randomized controlled trial

Parent training (PT) programs targeting children's externalizing behaviors are recommended in clinical practice to prevent problematic child behaviors from escalating into more serious concerns during adolescence, such as drug abuse, risky sexual behaviors, and delinquency (National Institute for Health and Clinical Excellence 2013). There is considerable evidence for the moderate short-term effects of some PT programs, at post measurement and at follow-up 4 months after the intervention (Litschge et al. 2010; Maughan et al. 2005;

Electronic supplementary material The online version of this article (doi:10.1007/s10802-016-0178-0) contains supplementary material, which is available to authorized users.

✉ Jens Högström
jens.hogstrom@ki.se

Viveca Olofsson
viveca.olofsson@oru.se

Metin Özdemir
metin.ozdemir@oru.se

Pia Enebrink
pia.enebrink@ki.se

Håkan Stattin
hakan.stattin@oru.se

[1] Department of Clinical Neuroscience, Karolinska Institute, 171 77 Stockholm, Sweden

[2] School of Law, Psychology and Social Work, Örebro University, 702 81 Örebro, Sweden

Serketich and Dumas 1996) and recently reports also indicate that some of the programs evaluated have persisting effects on child externalizing behaviors for as long as 1 year after program completion (Furlong et al. 2012). However, less is known about their longer-lasting effects (Özdemir 2015). Do they actually vaccinate against continued or future child behavior problems? Are some programs more effective than others in reducing externalizing problems over time? And, are there differences in the rates of change among programs? That is, do they reduce problematic child behaviors at different rates, during the program itself and during the follow-up period? Little is known to answer these questions. To assist families, and to make substantive and informed treatment plans, it is important for clinicians to know whether effects are sustained over longer periods of time. The aim of the present study was to report on a 2-year follow-up of the Swedish National Effectiveness Trial of Parenting Programs (Stattin et al. 2015), a unique study where four PT programs with different theoretical backgrounds (behavioral and non-behavioral) were compared in a randomized controlled trial.

Long-term evaluations of group-based stand-alone PT programs (i.e., programs that do not involve the child or the child's teacher) are underrepresented in the literature. Fairly recent meta-analyses (Furlong et al. 2012; Lundahl et al. 2006) and reviews (Sandler et al. 2011; SBU 2010) have highlighted an overall lack of long-term follow-ups (i.e., 6 months or more) of PT programs, regarding both efficacy (under ideal conditions, in a research setting) and effectiveness (under real-life circumstances, in regular care) (Kazdin 2003). The few existing studies suffer from methodological weaknesses that prevent firm conclusions from being drawn (Özdemir 2015). The lack of methodologically rigorous effectiveness trials is of particular concern, since such trials enable firm and reliable conclusions to be reached about the possible magnitude of effects when programs are implemented in a regular health care system. A meta-analysis by Furlong et al. (2012) summarized the findings of 13 effectiveness studies of group-based PT programs, in total including 1078 parents with children aged 3 to 12 years. However, only one of these studies, evaluating *Incredible Years*, had a follow-up period longer than 1 year (Gardner et al., 2006). A recent large-scale study (N = 6143) evaluated the roll-out of eight PT programs in a UK effort to implement interventions for externalizing problems (Lindsay and Strand 2013). Although this study included a 1-year follow-up that indicated positive effects on child and parent outcome measures, it was hampered by significant attrition (53.5 % and 30.5 % response rates at post-test and 1-year follow-up, respectively), and participants were not randomized between conditions. In sum, the lack of effectiveness trials evaluating the long-term effects of group-based PT programs in regular care is both evident and a cause of concern.

We conducted an overview of group-based PT programs with longer-term (18 months or more) follow-ups, and found nine randomized or quasi-randomized evaluations of stand-alone PT programs (i.e., with no other treatment component, such as a child or teacher directed component, than the parent training part) with a follow-up range between 18 months and 12 years (Bywater et al. 2009; DeGarmo et al. 2004; Eisner et al. 2012; Gardner et al. 2006; Hahlweg et al. 2010; Losel and Stemmler 2012; Malti et al. 2011; Wolchik et al. 2002; Zubrick et al. 2005). These studies differed in several methodological aspects, such as design, participant characteristics, measures, choice of primary reporters, and the definitions of post-test and follow-up time-points. Possibly related to these differences, within-group effect sizes describing changes from baseline to follow-up varied considerably among the studies, from no significant effect (Eisner et al. 2012; Malti et al. 2011), small effect (DeGarmo et al. 2004; Losel and Stemmler 2012) to medium (Gardner et al. 2006; Hahlweg et al. 2010; Zubrick et al. 2005) and large effects (Bywater et al. 2009; Wolchik et al. 2002). Other studies have reported results in a way that baseline to follow-up effect sizes cannot be calculated (e.g., Webster-Stratton et al. 2011). Hence, it is clear that current evaluations of longer-term effects differ between studies, and there is a lack of evidence on effects to guide current and future programs.

Current PT programs have different theoretical paradigms. Therefore, both key components and long-term effects may differ between programs, according, for example, to whether they are based on behavioral or non-behavioral theory. To our knowledge, no long-term effects of non-behavioral programs have been reported with a follow-up equal to or longer than 18 months (with evaluations based on behavioral principles). Hence, there are no comparisons of effects at long-term follow-up, or of rates of change (from baseline to post-test and post-test to follow-up), between behavioral and non-behavioral programs. Interestingly, changes in child externalizing behaviors show different trends between behavioral programs. For example, DeGarmo and Forgatch (2005), in an experimental evaluation of the *Parent Management Training-Oregon Model*, found that a group of children exhibited a mostly linear decline in delinquency, from baseline to 36-month follow-up. An efficacy study of the *Incredible Years BASIC parent program* showed a rapid and large effect on child externalizing behavior between baseline and post-test, and subsequent maintenance of this effect at 18-month follow-up (Bywater et al. 2009). Effects have also been described as having a zig-zag pattern from baseline to post-test, through to 1- and 2-year follow-up, when the *Triple P* parenting program

(Hahlweg et al. 2010) was used as a means of universal prevention. Hence, rates of changes may differ even between programs with a similar theoretical origin. The observed differences between program effects and rates of change may be mere artifacts of study designs and sample characteristics (which differ between studies), but they may also stem from the components of different programs. Studying multiple programs, and programs with different theoretical origins, within the same trial, provides a unique opportunity to examine rates of change across different types of programs.

The Swedish National Effectiveness Trial

The present study used the same study design to evaluate four established PT programs: *Comet* (Kling et al. 2010), *Cope* (Cunningham et al. 1995), *Incredible Years* (Webster-Stratton 1984; Webster-Stratton et al. 2004), and *Connect* (Moretti et al. 2013). Connect is based on attachment theory and encourages parents to reflect on different aspects of the parent-child relationship, and how to interpret child behaviors (Moretti et al. 2013). Cope has a broader theoretical base – in behavioral, family-system and group theory (Cunningham et al. 1995) – and, typically, parents discuss problems, and develop their own solutions under the guidance of group leaders (Cunningham et al. 1993). Comet and Incredible Years are behavioral programs, with positive-reinforcement techniques, communications skills, household rule consistency and effective limit setting as core components (Kling et al. 2010; Webster-Stratton 1984).

A previous report on the short-term effects of all four programs showed significant improvements on child externalizing behaviors compared with a waitlist control (Stattin et al. 2015) The Comet program showed larger effects on externalizing behaviors, as measured by the Eyberg Child Behavior Inventory (ECBI; Eyberg and Pincus 1999), than the other three programs, while children of parents in Connect showed the least improvement. Further, the Comet, Incredible Years and Cope programs, but not Connect, decreased inattention problems significantly compared with a waitlist control, as measured by the Swanson, Nolan, and Pelham Questionnaire (SNAP-IV; Swanson 1992). Additionally, Cope and Incredible Years showed significant reductions in hyperactivity symptoms, compared with the waitlist, in contrast to both Comet and Connect. Hence, all the program conditions were effective compared with the waitlist condition, with Comet as the most potent program, closely followed by Incredible Years and Cope. Connect seemed to be the least potent program in the short term.

The overall aim of the present study was to evaluate the 2-year effects of these four selective PT programs. Specifically, we examined: 1) whether the levels of child externalizing behavior, hyperactivity and inattention problems, and also positive and negative parenting practices, parental sense of competence and parents' mental health, differed among the PT programs 2 years after the interventions, and 2) whether different rates of change in child and parent outcomes could be identified across the programs.

Method

Procedure

The current study was part of a larger randomized controlled trial, The National Effectiveness Trial of Parenting Programs, where four PT programs – Comet, The Incredible Years, Cope and Connect – were evaluated (for a detailed methodological description, see Stattin et al. 2015). In addition to the four PT conditions, a self-help condition (reading a book about parent management techniques) and a waitlist condition were included in the trial. Hence, the trial comprised six arms, four active, one passive and one control. In the current evaluation, only results for the four active conditions (the PT programs) are reported. The waitlist was not included in the 2-year follow-up due to the ethical requirement to offer this group a PT program after post-measurement. The trial was conducted at four sites in Sweden: Stockholm, Örebro, Göteborg, and Lund. Each site provided three out of the six conditions, so that parents at any one particular site were randomized to receive: (a) one of the PT programs offered at the site, (b) another PT program offered at the site, or (c) the book condition (during the first year of the trial) or the waitlist condition (during the second year of the trial). The combinations of PT programs offered differed from site to site (mainly the following combinations; Stockholm: Comet and Connect; Örebro: Comet and Connect; Göteborg: Comet and Incredible years; Lund: Incredible years and Cope). At the four sites, a total of 30 primary care providers, the vast majority of whom were child and adolescent psychiatric units, social service centers or schools, were involved in the PT programs.

The study was conducted as an effectiveness trial of selective prevention programs. This was achieved through collaboration with the providers of the PT programs. Personnel employed at the different service units were group leaders, and study participants were recruited according to normal routines. Parents who had sought help at the units were requested to attend an information meeting about the study, where those who were willing to participate gave their informed consent, and responded to the baseline questionnaires before randomization. The programs started approximately 1 to 2 weeks after the information meeting.

Parent-rated questionnaires were gathered at four timepoints: at baseline, post-test (3 to 4 months after baseline), and at follow-ups conducted 1 year and 2 years after PT program completion. For the current 2-year evaluation we used

the baseline, post-test and 2-year follow-up measurement points. Parents responded to the post-test questionnaires in conjunction with the final program session, and the 2-year follow-up was conducted by regular mail.

Participants

For the current study, parents of 749 children were included and randomized to Comet, Incredible Years, Cope, or Connect. The children's ages ranged between 3 and 12. The only exclusion criterion was the presence of an autism spectrum diagnosis. Children in the age range 3 to 8 were randomized to the programs developed for younger children, Comet, Incredible Years and Cope, whereas children in the age range 9 to 12 were randomized to the ones developed for older children, Comet, Cope and Connect. Hence, participants in the Incredible Years groups were aged 3 to 8, participants in the Connect groups 9 to 12, and participants in the Comet and Cope groups 3 to 12. Table 1 shows demographic data and participant characteristics at baseline. The families who were assigned to one of the four programs differed on only two variables, child and parent age. Connect participants were older than those in the other programs, given that only older children could be randomized to this particular program. Accordingly, child age was controlled for in all analyses.

Of the 749 parents who were randomized to receive a PT program, all (100 %) responded at baseline, 661 (88.3 %) at post-test, and 543 (72.5 %) at 2-year follow-up (see Fig. 1). Complete questionnaire data for baseline, post-test and 2-year follow-up were available for 531 (70.9 %) participants. The parent who attended the most sessions was the primary reporter, and responded to the questionnaires. The mother was chosen as primary reporter if two parents attended an equal number of sessions. In 637 (85.0 %) cases, the primary reporter was female (97.3 % biological mothers), and in 112 (15.0 %) male (95.5 % biological fathers).

Parents of 114 (15.2 %) of the original 749 never started the program to which they were assigned, but the remaining 635 (84.8 %) attended at least one group session. The mean number of attended sessions was 6.8 (SD = 3.7) for the full sample (N = 749), and 8.0 (SD = 2.6) for the parents (N = 635) who actually started a program. Attendance rates differed somewhat among the PT programs, $F(3745) = 4.75$, $p < .01$, but only significantly ($p < .01$) between Cope (m = 6.1, SD = 3.4) and Connect (m = 7.3, SD = 3.2), and with a small effect size (d = 0.38).

The Parent Training Programs

The four evaluated PT programs are all group-based. Connect, originally a program designed for parents of adolescents (aged 12–18), was adapted to target 9–12 year-old children in the run-up to the current study. This was done to achieve age alignment with the other study programs, given the study's focus on children aged 3–12. Each program provides an extensive group-leader manual giving structure and content to the weekly sessions. Role-plays and group discussions are features of all the programs, but they differ in some other key aspects. The behavioral programs, Comet and Incredible Years, focus primarily on teaching parents positive

Table 1 Baseline demographic data and characteristics of the 749 study participants

	Comet (N = 207)	Incredible Years (N = 122)	Cope (N = 202)	Connect (N = 218)	F / χ^2	p
Child age M (SD)	7.4 (2.4)	7.0 (2.2)	7.2 (2.6)	9.7 (1.4)	68.2	< .01
Parent age M (SD)	38.0 (6.8)	37.5 (6.1)	37.5 (6.1)	39.9 (5.4)	6.4	< .01
Girls N (%)	73 (35.3)	47 (38.5)	81 (40.1)	78 (35.8)	1.3	.72
Boys N (%)	134 (64.7)	75 (61.5)	121 (59.9)	140 (64.2)	1.3	.72
Immigrant in Sweden [a] N (%)	25 (12.3)	21 (17.6)	28 (14.1)	31 (14.6)	1.8	.61
Parent w. university education N (%)	98 (47.3)	47 (38.8)	105 (52.2)	102 (47.7)	5.4	.14
Parent w. only primary education [b] N (%)	14 (6.8)	9 (7.4)	23 (11.4)	21 (9.8)	3.3	.35
Family income [c] N (%)						
0–3000	33 (16.2)	25 (21.2)	44 (22.0)	53 (24.9)	4.9	.18
3001–6000	73 (35.8)	46 (39.0)	68 (34.0)	66 (31.0)	2.4	.50
6001 or above	98 (48.0)	47 (39.8)	88 (44.0)	94 (44.1)	2.1	.55
Single parents N (%)	50 (24.8)	33 (27.5)	59 (29.4)	62 (28.8)	1.3	.73
Siblings [d] M (SD)	2.2 (0.9)	2.3 (1.0)	2.1 (0.9)	2.3 (1.0)	1.7	.18

[a] Defined as at least one parent born outside Europe

[b] Meaning nine years of primary education

[c] Monthly household income in USD

[d] Total number of children living in the household

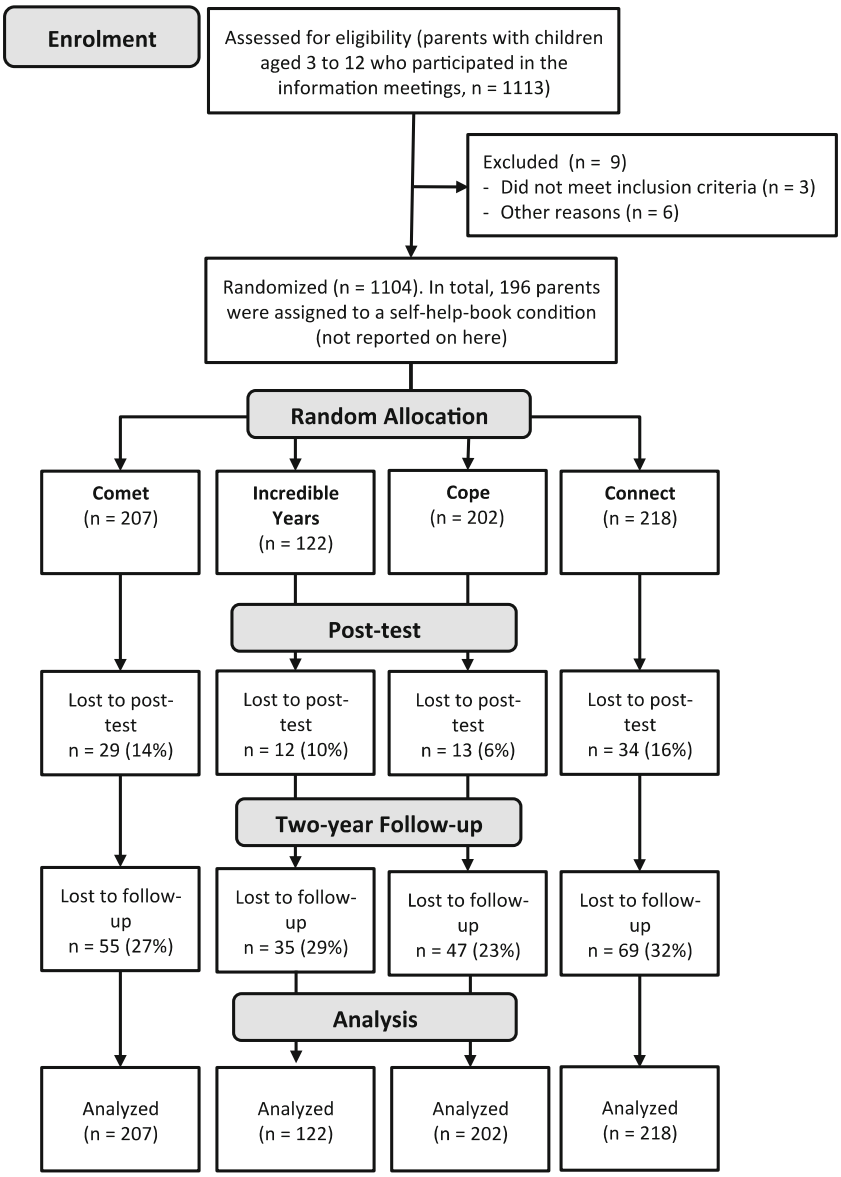

Fig. 1 Flow diagram and response rates

reinforcement through praise and rewards, child-directed play, the importance of consistency in household rules, and non-punitive limit-setting strategies. The programs have different origins. Comet is a Swedish program, whereas Incredible Years is adapted from the USA. Comet derives from, and partly resembles, the *Parent Management Training-Oregon Model* (PMT-O; Forgatch 1994), *Incredible Years* (Webster-Stratton 1984) and *Defiant Children* (Barkley 1997), all programs of North American origin, but Comet has been adjusted to attune with Swedish norms and child rearing traditions (Kling et al. 2006). Cope includes positive reinforcement and rule consistency, just like traditional behavioral PT programs, but it operates through the use of a problem-solving pedagogy where parents collectively reason their way to appropriate strategies under the guidance of group leaders. Also, Cope uses larger groups (up to 30 parents) than Comet and Connect (up to 12 parents), and compared with Incredible Years (up to 16 parents). Connect is based on attachment theory and emphasizes the importance of parents' self-reflection, strengthening the parent-child relationship, and understanding the child's need to develop autonomy. Throughout the program, parents discuss how to improve their attunement, empathic responding, and dyadic emotion regulation, in relation to their child. Examples of the principles that are discussed are that every behavior has a meaning (take a step back) and attachment is life-long (children have different needs in different phases of life). The adaptation from the adolescent version of Connect mostly concerned making the examples and role-plays age appropriate, leaving the core components of the program unaffected.

The number of sessions and the length of sessions vary among the programs. Incredible Years and Cope employ 2-h

sessions and Comet 2.5-h sessions, whereas Connect uses 1-h sessions. Regarding number of sessions, Incredible Years has 12, Comet 11, and both Cope and Connect 10.

Child Outcome Measures

Externalizing Problems The Eyberg Child Behavior Inventory (ECBI; Eyberg and Pincus 1999) was used as the primary outcome measure to investigate parents' perceptions of their child's externalizing behaviors. It is a well-validated instrument that correlates highly with independent measures of conduct problems (Robinson et al. 1980). The ECBI consists of 36 items, and assesses both the frequency of the child's behaviors (*the intensity subscale*) and parents' perceptions of whether the behaviors are problematic for them (*the problem subscale*). Examples of items are "Does not obey house rules," "Whines," and "Constantly seeks attention". The intensity of the child's externalizing problems was rated by parents on a 7-point scale ranging from 1 (*never*) through 7 (*always*). The items in the problem subscale were rated on a 2-point scale: 0 (*not a problem*) and 1 (*a problem*). In the present sample, Cronbach's alphas for the intensity subscale were .93 (baseline), .94 (post-test), and .94 (2-year follow-up). For the problem subscale, Cronbach's alphas were .91 (baseline), .92 (post-test) and .96 (2-year follow-up).

Attention Deficit and Hyperactivity Problems We used the DSM-IV version of the Swanson, Nolan and Pelham Questionnaire (SNAP-IV; Swanson 1992), an instrument that has shown good agreement with clinicians' ratings of ADHD (Alda and Serrano-Troncoso 2013). The SNAP-IV consists of two subscales targeting the DSM-IV criteria for ADHD, and one subscale targeting oppositional defiant disorder (ODD). The ADHD subscales are inattention (nine items, e.g., "Often does not seem to listen when spoken to directly," and "Often is forgetful in daily activities"), and hyperactivity/impulsivity (nine items; e.g., "Often is 'on the go' or often acts as if 'driven by an engine'," and "Often blurts out answers before questions have been completed"). The 8-item subscale for ODD has eight items (e.g., "Often loses temper," and "Often argues with adults"). Parents rated the items on a 4-point scale, ranging from 0 (*not at all*) to 3 (*very much*). Cronbach's alphas for the three subscales at baseline, post-test and 2-year follow-up, respectively, were .89, .91, .93 for the hyperactivity/impulsivity subscale, .91, .92, .91 for the inattention subscale, and .90, .91, .91 for the oppositional defiance subscale.

Parent Outcome Measures

Parents' negative reactions How parents responded to noncompliance was measured in terms of the frequency of angry outbursts and harsh treatment. The angry outbursts instrument (Stattin et al. 2011) uses the question "What do you do when your child does something you really don't like?" with responses like "My first reaction is anger, and I yell at the child," and "I get angry and have an emotional outburst"). Each item is rated on a 3-point scale. Cronbach's alphas for the angry outbursts scale were .79, .78, and .80 at baseline, post-test and 2-year follow-up, respectively.

The *harsh treatment subscale* from the Parents' Practice Interview (PPI; Webster-Stratton 1998) was used to provide a broad measure of harsh reactions to child misbehavior. Parents rate the frequency of negative responses when their child "does something he/she is not supposed to," for example, "Take away privileges (like TV, playing with friends)," or "Slap or hit your child". Items were rated on a 7-point scale. Cronbach's alphas for the harsh treatment scale were .63, .72, and .55 at baseline, post-test and 2-year follow-up, respectively.

Parents' positive reactions The frequency of parents' attempted understanding of child misbehavior (Stattin et al. 2011) was measured by asking them "What do you do when your child does something you really don't like?," with responses like "I try to talk it through without creating new conflicts" and " I try to understand how the child thought and felt". Each item was rated on a three-point scale. Cronbach's alphas were .68, .69, and .70 at baseline, post-test and 2-year follow-up, respectively.

The *rewarding the child* subscale of the Parents' Practice Interview (Webster-Stratton 1998) was used to measure parents' disposition to reward good behavior. The scale includes items such as "Buy something for him/her (such as special food, a small toy) or give him/her money for good behavior," and "Give him/her an extra privilege (such as cake, go to the movies, special activity for good behavior)," which were rated on a 7-point scale. Cronbach's alphas were .79, .77, and .74 at baseline, post-test and 2-year follow-up, respectively.

Parents' sense of competence, stress and depression We used the Parenting Sense of Competence scale (PSOC; Johnston and Mash 1989), which is divided into two subscales measuring parental satisfaction with and the efficacy of parenting. For this study, we used the coding procedure proposed by Gilmore and Cuskelly (2009). Because the subscales were highly correlated, we combined them into a single measure of sense of competence. Cronbach's alphas were .81, .85, and .78 at baseline, post-test and 2-year follow-up, respectively.

The Caregiver Strain Questionnaire (Brannan et al. 1997) was used to measure parental stress. For this study, we used the 10-item *objective strain* subscale to assess parents' levels of stress. Parents were asked to reflect and report on how they had been affected by their children's problems (at pre-test during the past 6 months; and, at post-test and 2-year follow-up during the past month). Cronbach's alphas were .90 at baseline, post-test, and 2-year follow-up.

The Center for Epidemiologic Studies-Depression scale (CES-D; Radloff 1977) was used to measure parents' depressive symptoms. Parents rated the frequency of their symptoms from 0 not at all to 3 often. Cronbach's alphas were .92, .93 and .93 at baseline, post-test and 2-year follow-up, respectively.

Attrition Analyses

Response status at 2-year follow-up (1 = *responded* and 0 = *did not respond*) was entered into a multiple logistic regression on child age and gender, parent education and income, immigrant status, and the baseline measures included in the study. Having immigrated (OR =2.07, p = .01), low parent education (OR = 1.31, p = .01), and being a girl (OR =2.02, p < .01) were significant demographic predictors of non-response status at 2-year follow-up. The participants scoring higher on the ECBI intensity subscale (OR =1.72, p = .02) and those scoring lower on the angry outbursts scale (OR =2.04, p = .02) at baseline were also more likely to fail to respond at 2-year follow-up. Overall, however, the Nagelkerke R^2 value of .14 indicates that responders and non-responders did not differ substantially. Further, there were no differences among the four PT programs regarding the number of participants who responded at 2-year follow-up, $\chi^2(3) = 3.88$, p = .28, and no differences in the numbers of immigrant-status families, education levels or girls among the PT programs (see Table 1). The main source of attrition was the failure of parents to start the PT program to which they were assigned (15.2 % of the 749 participants). Fewer parents started Incredible Years (75.4 % starters) than Comet (83.1 %), Cope (86.6 %), or Connect (89.9 %). This was due to organizational and logistic problems in two of the municipalities where Incredible Years was employed. However, within the Incredible Years group, there were no differences between parents who started a program and those who did not regarding age, gender, parent education and income, immigrant status or any of the outcome measures at baseline. Overall, it was concluded that attrition had given rise to only limited bias.

Statistical Analyses

To compare the levels of all child and parent outcome measures, across the four PT programs at 2-year follow-up, we used ANCOVAs controlling for child age and the baseline level of symptom severity. For these analyses, data from all participants were analyzed, regardless of their actual participation in a PT program or not and the various assessments. Expectation Maximization (EM) imputation was used to estimate missing data (for the ANCOVA analyses only). Partial eta square effect sizes were calculated, with 0.01 regarded as a small effect, 0.06 a medium effect and 0.14 a large effect (Cohen 1988). These analyses were conducted using IBM SPSS Statistics for Macintosh, version 22.0.

To examine how participants in the different PT programs changed between baseline and 2-year follow-up, we fitted a series of multi-group Latent Growth Models (LGMs) (Duncan et al. 2013), using the baseline, post-test (3 months after baseline) and 2-year follow-up measures (24 months after post-test). We expected that most changes in child behavior outcomes would occur between baseline and post-test, while the parents were participating in the programs. Thus, as a first step, we constrained the loadings of baseline to zero, and of post-test to three, to model change over a 3-month period. We freely estimated the loadings of the 2-year follow-up measure to allow the model to estimate the actual amount of change. As a second step, we constrained the loadings of the 2-year follow-up measure to the estimated loadings from the first step. Next, we compared the slopes in the models with each other to see if the rates of change in child outcomes differed among the four PT programs, from baseline to 2-year follow-up. In all models, we included age as a covariate of both the intercept and slope factors to account for the differences in child age among the programs. To aid accurate interpretation of the growth patterns between specific time points, we fitted post-hoc latent change models (Duncan et al. 2013), which provide a direct test of whether any change between two measurement points is statistically significant. The post-hoc analyses were performed to estimate changes from (1) baseline to post-test, and (2) post-test to 2-year follow-up. Latent change models allow measurement error to be included, so as to obtain an unbiased estimate of change. Cohen's d effect sizes were calculated, where 0.20 was considered a small effect, 0.50 a medium effect and 0.80 a large effect (Cohen 1988). All models were fitted using the Robust Maximum Likelihood (MLR) estimator to account for non-normality in the measurements. We used MPlus 7.11 software (Muthén and Muthén 1998-2012). To deal with missing data in the LGM analyses, Full Information Maximum Likelihood (FIML) was used to produce less biased estimates and smaller standard errors (Schafer and Graham 2002) than those produced by alternative methods, such as listwise deletion (Reinecke and Weins 2013).

Results

At the 2-year follow-up, 10.6 % of all participants stated that they had attended another PT program (or part of a program) during the follow-up period, but the proportion of families

who had sought additional help was not significantly different among the programs, $\chi^2(3) = 3.01$, $p = .39$.

Comparing Child Behavior Problems and Parent Outcomes among Programs at the 2-Year Follow-up

The ANOVAs (see Table 2) show that there were no differences among the programs on any of the measures of child externalizing behaviors, parent behaviors, and wellbeing. The largest F-statistic was $F(3, 743) = 2.47$, $p = .06$, $p\eta^2 = .01$, for SNAP-Inattention. Overall, the findings indicate that children and parents in the Comet, Incredible Years, Cope and Connect programs did not differ in the level of child behavior problems, parents' reactions to child behavior, perceived parental competence, or degree of parental stress and depressive symptoms, 2 years after intervention.

Rates of Change in Externalizing Behavior across the Programs

The multi-group LGMs (Table 3) show that the overall change pattern for Comet was significantly different from the change patterns for Incredible Years, Cope, and Connect, as measured by ECBI intensity. Regarding ECBI problems, Comet differed from Connect but not from Incredible Years and Cope. The post-hoc latent change models and effect sizes (Tables 4 and 5) shed some light on these differences. From baseline to post-test, the reductions in ECBI intensity scores were significantly greater for Comet ($d = 1.43$) than for Connect ($d = 1.06$). On the other hand, children of parents in the Connect program decreased significantly more in ECBI intensity ($d = 0.32$) from post-test to 2-year follow-up, compared with Comet, where the child ratings were unchanged during this period. Regarding ECBI problems, post-hoc analyses indicated similar change rates from baseline to post-test for all the programs. During this period, however, the children of parents in Connect were reported to improve significantly more ($d = 0.22$) than the children in Comet, who, again, were unchanged during this phase. In contrast to the indications of different change patterns for ECBI intensity and ECBI problems, in particular between Comet and Connect, SNAP ODD showed no overall differences in change patterns among the programs. Post-hoc analyses confirmed that all the programs produced similar changes, in the small effect-size range, between baseline and post-test, but that Connect was the only program to show continued improvement ($d = 0.23$) during the follow-up period. Despite some differences in the rates of change, all of the programs achieved significant reductions from baseline to 2-year follow-up, with within-group effect sizes (Table 6) for ECBI scores ranging from $d = 1.21$ to

Table 2 Differences in clinical outcomes among the programs at the 2-year follow-up (ANOVAs)

Outcome measures	Comet $N = 207$ M (SD)	Incredible Years $N = 122$ M (SD)	Cope $N = 202$ M (SD)	Connect $N = 218$ M (SD)	F [a]	$p\eta^2$
Child outcomes						
ECBI Intensity	2.74 (0.79)	2.93 (0.79)	2.73 (0.78)	2.69 (0.89)	0.79	0.003
ECBI Problem	0.23 (0.19)	0.27 (0.20)	0.24 (0.20)	0.26 (0.23)	0.32	0.001
SNAP-IV Hyperactivity	0.75 (0.65)	0.82 (0.63)	0.77 (0.61)	0.65 (0.62)	0.80	0.003
SNAP-IV Inattention	0.96 (0.67)	1.01 (0.74)	0.97 (0.70)	1.01 (0.65)	2.47	0.010
SNAP-IV ODD	0.83 (0.59)	0.80 (0.60)	0.76 (0.60)	0.81 (0.61)	1.10	0.004
Negative Parenting						
Angry outbursts	1.76 (0.33)	1.73 (0.33)	1.74 (0.36)	1.69 (0.35)	0.32	0.001
Harsh parenting	1.88 (0.39)	1.89 (0.42)	1.88 (0.48)	1.89 (0.45)	0.17	0.001
Positive Parenting						
Attempted understanding	2.62 (0.30)	2.61 (0.29)	2.64 (0.28)	2.63 (0.28)	1.48	0.006
Rewards	3.95 (0.70)	4.15 (0.63)	4.08 (0.68)	3.80 (0.68)	1.64	0.007
Parenting Competence						
Parent sense of competence	4.37 (0.74)	4.44 (0.81)	4.44 (0.76)	4.21 (0.79)	1.11	0.004
Parental Mental Health						
Stress	1.64 (0.51)	1.72 (0.60)	1.61 (0.55)	1.73 (0.61)	0.65	0.003
Depression	0.75 (0.48)	0.78 (0.58)	0.73 (0.51)	0.84 (0.54)	1.65	0.007

Baseline levels of severity and child age were included as covariates in all the analyses. ECBI = Eyberg Child Behavior Inventory; SNAP-IV = Swanson Nolan and Pelham questionnaire IV; $p\eta^2$ = Partial eta squared

$N = 749$ (with EM imputation for missing data)

[a] None of the F statistics indicated significant differences ($p < .05$) between groups

Table 3 Growth in child and parent outcomes from baseline to 2-year follow-up

		Comet	IY	Cope	Connect	Model Fit Indices			
						Chi-sqr (df)	CFI	RMSEA	SRMR
ECBI Intensity	Intercept	3.58_a	3.75_a	3.58_a	3.54_a	2.82 (11)	1.00	0.00	0.02
	Slope	-0.30_a	-0.27_b	-0.25_b	-0.21_b				
ECBI Problem	Intercept	0.40_a	0.45_a	0.40_a	0.43_a	3.67 (9)	1.00	0.00	0.01
	Slope	-0.06_a	-0.05_{ab}	-0.04_{ab}	-0.04_b				
SNAP-IV Inattention	Intercept	1.11_a	1.12_a	1.15_a	1.28_a	12.52 (12)	0.99	0.02	0.04
	Slope	-0.10_a	-0.11_a	-0.08_a	-0.08_a				
SNAP-IV Hyperactivity	Intercept	1.07_a	1.26_a	1.16_a	1.04_a	13.25 (13)	0.99	0.04	0.05
	Slope	-0.08_a	-0.12_b	-0.10_{ab}	-0.06_a				
SNAP-IV ODD	Intercept	1.18_a	1.16_a	1.13_a	1.24_a	2.69 (12)	1.00	0.00	0.02
	Slope	-0.13_a	-0.13_a	-0.12_a	-0.10_a				
Angry outbursts	Intercept	2.01_a	1.95_a	1.97_a	1.96_a	0.58 (4)	1.00	0.00	0.01
	Slope	-0.10_a	-0.07_b	-0.08_{ab}	-0.07_b				
Harsh parenting	Intercept	2.24_a	2.19_a	2.16_a	2.20_a	9.35 (7)	0.99	0.04	0.08
	Slope	-0.13_a	-0.11_{ab}	-0.11_a	-0.08_b				
Attempted understanding	Intercept	2.53_a	2.60_a	2.56_a	2.61_a	7.84 (7)	0.99	0.03	0.12
	Slope	0.02_a	-0.01_b	0.03_a	0.02_a				
Rewards	Intercept	3.91_a	4.15_b	4.09_b	3.80_a	9.26 (8)	0.99	0.03	0.09
	Slope	0.12_a	0.09_{ab}	0.06_b	0.08_{ab}				
Parents sense of competence	Intercept	3.82_a	3.98_a	3.94_a	3.78_a	0.36 (6)	1.00	0.00	0.02
	Slope	0.19_a	0.10_b	0.12_b	0.14_b				
Stress	Intercept	2.07_a	2.19_a	2.08_a	2.17_a	1.91 (5)	1.00	0.00	0.01
	Slope	-0.14_a	-0.15_a	-0.11_b	-0.14_{ab}				
Depression	Intercept	0.91_a	0.92_a	1.00_a	1.01_a	1.84 (8)	1.00	0.00	0.02
	Slope	-0.07_{ab}	-0.04_b	-0.08_a	-0.06_{ab}				

All mean intercept and slope estimates were statistically significant at $p < .01$ except for the slope of Incredible Years (IY) for Attempted Understanding. Different subscripts indicate significant differences in the intercept or slope means across programs; same subscripts indicate non-significant differences. ECBI = Eyberg Child Behavior Inventory; SNAP-IV = Swanson Nolan and Pelham questionnaire; df = degrees of freedom; CFI = Comparative Fit Index; RMSEA = Root Mean Square Error of Approximation; SRMR = Standardized Root Mean Square Residual

d = 1.32 (for intensity) and from d = 0.72 to d = 0.91 (for problems).

Overall, the analyses indicate that Comet was the most effective program in reducing child externalizing behavior in the short term (baseline to post-test), followed by Incredible Years and Cope, while Connect was the least effective. However, during the 2-year follow-up period, Connect showed a continued reduction in problem behaviors whereas the behaviors of participants in Comet, Incredible Years and Cope were largely unchanged.

Rates of Change in Hyperactivity and Inattention Problems across the Programs

Overall change patterns for SNAP hyperactivity and SNAP inattention are presented in Table 3. Multigroup LGMs showed that participants in Incredible Years changed at a different rate from those in Comet and Connect regarding hyperactivity, but that all the programs had similar change patterns for inattention. The post-hoc latent change analyses indicate that Incredible Years reduced hyperactivity more effectively (d = 0.81) than Connect (d = 0.44) between baseline and post-test. On the other hand, the children of parents in Connect showed a continued decrease (d = 0.41) in hyperactivity from post-test to 2-year follow-up, while participants in Comet, Incredible Years and Cope were unchanged during this period. Notwithstanding the different change rates, particularly between Incredible Years and Connect, all the programs generated significant reductions in hyperactivity and inattention from baseline to 2-year follow-up, with effect sizes ranging from d = 0.68 to d = 0.90 for hyperactivity, and from d = 0.25 to d = 0.54 for inattention (Table 6). All in all, the behavioral program, Incredible Years, produced more rapid change in hyperactivity symptoms during the

Table 4 Estimated rates of change in child and parent outcomes from baseline to post-test

Outcome measures	Comet		Incredible Years		Cope		Connect	
Child outcomes	Slope	*d*	Slope	*d*	Slope	*d*	Slope	*d*
ECBI Intensity	−0.30***$_{a}$	1.43	−0.27***$_{ab}$	1.51	−0.25***$_{ab}$	1.20	−0.22***$_{b}$	1.06
ECBI Problem	−0.06***$_{a}$	0.86	−0.05***$_{a}$	0.80	−0.04***$_{a}$	0.68	−0.04***$_{a}$	0.58
SNAP ODD	−0.13***$_{a}$	0.78	−0.13***$_{a}$	0.69	−0.12***$_{a}$	0.66	−0.10***$_{a}$	0.58
SNAP Hyperactivity	−0.08***$_{ab}$	0.56	−0.11***$_{a}$	0.81	−0.09***$_{ab}$	0.55	−0.06***$_{b}$	0.44
SNAP Inattention	−0.10***$_{a}$	0.59	−0.11***$_{a}$	0.72	−0.08***$_{a}$	0.57	−0.08***$_{a}$	0.57
Negative Parenting								
Angry outbursts	−0.10***$_{a}$	0.83	−0.07***$_{ab}$	0.58	−0.08***$_{ab}$	0.73	−0.07***$_{b}$	0.68
Harsh parenting	−0.13***$_{a}$	0.82	−0.11***$_{ab}$	0.67	−0.11***$_{ab}$	0.68	−0.08***$_{b}$	0.51
Positive Parenting								
Attempted understanding	0.02**$_{a}$	0.22	−0.01$_{a}$	−0.03	0.03***$_{a}$	0.27	0.02**$_{a}$	0.20
Rewards	0.12***$_{a}$	0.45	0.08***$_{a}$	0.30	0.06***$_{a}$	0.28	0.07***$_{a}$	0.33
Parenting Competence								
Parents sense of competence	0.19***$_{a}$	0.76	0.10***$_{b}$	0.42	0.12***$_{ab}$	0.49	0.14***$_{ab}$	0.58
Parental Mental Health								
Stress	−0.14***$_{a}$	0.78	−0.15***$_{a}$	0.66	−0.11***$_{a}$	0.62	−0.13***$_{a}$	0.60
Depression	−0.07***$_{a}$	0.36	−0.04$_{a}$	0.22	−0.08***$_{a}$	0.45	−0.06***$_{a}$	0.33

Unstandardized slope means and within-group effect sizes (Cohen's *d*) are presented. Negative slope values indicate improvement. Child age was included as a covariate in all models. Different subscripts indicate significant differences in the rate of estimated change; same subscripts indicate no significant difference. The Bonferroni-adjusted critical *p*-value ($p < .008$) indicates a significant difference. ECBI = Eyberg Child Behavior Inventory; SNAP-IV = Swanson Nolan and Pelham questionnaire

** $p < .008$ *** $p < .001$

Table 5 Estimated rates of change in child and parent outcomes from post-test to 2-year follow-up

Outcome measures	Comet		Incredible Years		Cope		Connect	
Child outcomes	Slope	*d*	Slope	*d*	Slope	*d*	Slope	*d*
ECBI Intensity	0.03$_{b}$	−0.05	<0.01$_{ab}$	0.02	−0.03$_{ab}$	0.16	−0.08***$_{a}$	0.32
ECBI Problem	<0.01$_{b}$	<0.01	−0.01$_{ab}$	0.10	−0.01$_{ab}$	0.17	−0.02***$_{a}$	0.22
SNAP ODD	0.03$_{b}$	−0.11	0.01$_{ab}$	−0.05	> − 0.01$_{ab}$	0.04	−0.05***$_{a}$	0.23
SNAP Hyperactivity	−0.02$_{a}$	0.17	−0.04$_{a}$	0.20	−0.03$_{a}$	0.22	−0.06***$_{a}$	0.41
SNAP Inattention	0.05***$_{b}$	−0.28	0.06**$_{b}$	−0.35	0.02$_{ab}$	−0.15	−0.01$_{a}$	0.06
Negative Parenting								
Angry Outbursts	0.02**$_{b}$	−0.13	−0.01$_{ab}$	0.06	0.01$_{ab}$	−0.03	−0.02$_{a}$	0.24
Harsh Parenting	0.04***$_{b}$	−0.29	<0.01$_{ab}$	−0.04	0.01$_{ab}$	−0.12	−0.02$_{a}$	0.19
Positive Parenting								
Attempted Understanding	0.01$_{a}$	0.06	0.01$_{a}$	0.09	<0.01$_{a}$	<0.01	−0.02$_{a}$	−0.14
Rewards	−0.11***$_{a}$	−0.46	−0.08***$_{a}$	−0.37	−0.06***$_{a}$	−0.32	−0.08***$_{a}$	−0.33
Parenting Competence								
Parents sense of competence	> − 0.01$_{a}$	<0.01	0.06$_{a}$	0.25	0.05$_{a}$	0.18	<0.01$_{a}$	0.03
Parental Mental Health								
Stress	<0.01$_{b}$	−0.02	> − 0.01$_{ab}$	0.05	−0.06***$_{a}$	0.33	−0.02$_{ab}$	0.14
Depression	0.02$_{a}$	−0.06	<0.01$_{a}$	0.03	−0.01$_{a}$	0.06	0.01$_{a}$	−0.01

Unstandardized slope means and within-group effect sizes (Cohen's *d*) are presented. Negative slope values indicate improvement, and positive values deterioration. Baseline level of severity and child age were included as covariates in all models. Different subscripts indicate significant differences in the rate of estimated change; same subscripts indicate no significant difference. The Bonferroni adjusted critical *p*-value ($p < .008$) indicates a significant difference. ECBI = Eyberg Child Behavior Inventory; SNAP-IV = Swanson Nolan and Pelham questionnaire

** $p < .008$ *** $p < .001$

Table 6 Within-group effect sizes (Cohen's *d*) for changes in child and parent outcomes, between baseline and the 2-year follow-up

	Comet	Incredible years	Cope	Connect
Child Outcomes				
ECBI Intensity	1.30	1.32	1.28	1.21
ECBI Problems	0.90	0.91	0.85	0.72
SNAP-IV Inattention	0.25	0.30	0.34	0.54
SNAP-IV Hyperactivity	0.68	0.90	0.69	0.78
SNAP-IV ODD	0.63	0.62	0.65	0.72
Negative Parenting				
Angry Outbursts	0.68	0.56	0.64	0.83
Harsh Parenting	0.54	0.61	0.49	0.73
Positive Parenting				
Attempted understanding[1]	0.30	0.06	0.26	0.07
Rewards	0.08	−0.02	−0.03	−0.02
Parenting Competence				
Parents sense of competence	0.81	0.61	0.62	0.65
Parental Mental Health				
Stress	0.76	0.76	0.75	0.70
Depression	0.35	0.27	0.53	0.31

N = 749 (EM imputation for missing data). Negative values indicate deterioration. ECBI = Eyberg Child Behavior Inventory; SNAP-IV = Swanson Nolan and Pelham questionnaire

course of intervention, while the non-behavioral program, Connect, was the only program for which parents reported a continued reduction in child hyperactivity during the 2-year follow-up phase.

Rates of Change in Negative and Positive Parenting

The LGM analyses show that change patterns differed among PT programs regarding parents' angry outbursts, where, specifically, Comet differed from Incredible Years and Connect. The post-hoc analyses confirmed that Comet reduced angry outburst ($d = 0.83$) to a greater extent than Connect ($d = 0.68$) in the short term (baseline to post-test). However, Comet participants reported a small deterioration ($d = 0.13$) in angry outbursts at long-term follow-up (post-test to 2-year), while the other three groups reported retention of treatment gains during this period. For harsh parenting, LGM analyses showed that the overall change patterns for Comet and Cope differed significantly from Connect. Again, the post-hoc analyses revealed that Comet produced more change ($d = 0.82$) during the intervention phase than Connect ($d = 0.51$). However, Comet participants had deteriorated by the time of 2-year follow-up ($d = 0.29$), in contrast to all the other groups, which remained stable during the period.

On the measures of positive parenting, the LGM analyses showed that participants in Incredible Years reported an overall change pattern that differed from the other programs regarding parents' attempts to understand their children in situations where the child misbehaved. However, this difference was not reflected in the post-hoc analyses, where all four programs produced changes that were similar between baseline and post-test, and also between post-test and 2-year follow-up. In a similar manner, the LGM analyses showed that there were different rates of change among the programs regarding parents' inclination to reward good behavior. Specifically, parents in Comet reported an overall different change pattern than parents in Cope. Yet, this difference did not appear in the post-hoc analyses, where all four programs showed improved reward tendencies among parents during the intervention, but where parents in all programs also reported deterioration at 2-year follow-up.

All in all, all four programs produced a significant reduction in angry outbursts from baseline to 2-year follow-up, with effect sizes ranging from $d = 0.56$ to $d = 0.83$, and a similarly significant reduction in harsh parenting, with effect sizes ranging from $d = 0.49$ to $d = 0.73$. Parents in Comet experienced a rapid reduction in negative parenting in the short term, but reported greater deterioration at follow-up, than parents in the Connect program. Further, parents in Comet, Cope and Connect generally reported small but significant improvements in attempted understanding from baseline to 2-year follow-up, with effect sizes ranging from $d = 0.07$ to $d = 0.30$, whereas parents in Incredible Years did not report any such improvement. The LGM analyses revealed that all four programs produced an overall improvement of parents' use of rewards in response to good behavior, but the effect sizes were close to zero. Generally, for positive parenting, the PT programs produced small improvements during the intervention period.

However, although improvement was retained at follow-up for attempted understanding, it deteriorated for rewards.

Rates of Change in Parents' Sense of Competence and Mental Health

The LGM analyses show that parents' sense of competence improved more among the Comet participants than those in Incredible Years, Cope and Connect. The post-hoc analyses revealed that Comet improved PSOC significantly more during the intervention period ($d = 0.76$) than did Incredible Years ($d = 0.42$). During the post-test to 2-year follow-up period, all groups retained their improvement, and no between-program differences were found. For parental stress, the LGM change patterns differed significantly when Comet and Incredible Years were compared with Cope. The post-hoc analyses showed that all four programs reduced stress to the same extent, from baseline to post-test, but that Cope was the only program to continue to improve participants from post-test to 2-year follow-up. Regarding symptoms of depression, the LGM analyses show that Incredible Years and Cope had different overall change patterns. However, these differences were only partly reflected in the post-hoc analyses, where all programs except Incredible Years reduced depressive symptoms in the short term (although the slopes did not differ significantly from each other).

Altogether, all four programs significantly strengthened parents' sense of competence from pre-test to 2-year follow-up, with effect sizes ranging from $d = 0.61$ to $d = 0.81$. Similarly, parental stress and depressive symptoms fell among parents in all programs, with effect sizes ranging between $d = 0.70$ and $d = 0.76$, and $d = 0.27$ and $d = 0.53$, respectively.

Discussion

The purpose of the study was to evaluate the long-term effectiveness of four parent training programs implemented in routine care. The PT programs were established group-based interventions targeting child externalizing behavior for treatment-seeking families in 30 municipalities in southern and central Sweden. Overall, the findings show that parent-reported levels of child externalizing behaviors, competent parenting and parental wellbeing were similar for all the programs by the time of the 2-year follow-up, and that all programs had produced significant improvements on these outcomes from baseline to follow-up. The long-term within-group effect sizes for all four programs, $d = 1.21$ to $d = 1.32$, were larger for the main outcome measure (ECBI intensity) in this study than those previously reported in the literature. The latter include $d = 0.87$ (≤ 1 year follow-up; Lundahl et al. 2006), $d = 0.40$ (1-year follow-up; Lindsay and Strand 2013) and $d = 0.47$ (18-month follow-up; Gardner et al. 2006), all with shorter follow-up periods than the current trial. No between-program differences were found for child externalizing problems or for hyperactivity and inattention problems and nor were there any differences among the PT programs regarding parent outcomes at 2-year follow-up. All programs reduced harsh parenting practices, and also parents' angry outbursts, with moderate to large effects sizes, but positive parenting practices were largely unchanged. Parents in all four programs rated their sense of parental competence as moderately to largely improved at 2-year follow-up, and similar but smaller changes were reported for parents' stress and depression.

A related purpose of the present study was to explore whether children of parents in the different programs showed similar rates of change from baseline to post-test, and from post-test to 2-year follow-up. In particular, the study assessed whether programs operating within different theoretical frameworks would produce reductions in externalizing behaviors, and improvements in other child and parent outcomes, at similar or different rates during the two periods. Despite the similar outcomes of all four programs at 2-year follow-up, there were some differences in rates of change among the four programs. The programs mainly derived from behavioral theory (Comet and Incredible Years) were more potent in reducing child externalizing behaviors during the actual course of intervention, whereas the 2-year follow-up period (after completion) for these programs was largely characterized by maintenance of previous treatment gains. The non-behavioral program, Connect, on the other hand, produced changes at a generally lower rate during the course of the intervention, but was the only program where parents reported continued improvements of child problems during the 2-year follow-up period. Children of parents in Cope (with a behavioral as well as a group- and family-system theoretical background) showed change patterns over the various outcome measures that were largely non-significantly different from the other three programs. Hence, changes in child behavior problems among the various PT programs occurred at slightly different rates, from baseline to post-test and from post-test to 2-year follow-up. These differences were more distinct when the behavioral programs, Comet and Incredible Years, were contrasted with the non-behavioral program, Connect. A similar pattern was noted for the reduction of negative parenting over time. The behavioral program, Comet, reduced angry outbursts and harsh parenting significantly more during the intervention phase than did Connect, but parents in Comet reported a small but significant deterioration in these domains during follow-up while Connect parents generally maintained their treatment gains. Overall, the change rates for positive parenting practices, parents' sense of competence and mental health did not differ much among the PT programs.

The observed differences in the rates of change among some of the programs might be related to their different theoretical

underpinnings. The behavioral programs, Comet and Incredible Years, aim to provide parents with a set of manifest techniques, through the use of praise and rewards to promote adaptive child behavior, and the use of limit-setting skills to reduce the frequency of problem behaviors (Kling et al. 2006; Webster-Stratton 1981). Seemingly, this aim reflects a more direct approach than that adopted in the attachment-based program, Connect, whose aim is to enhance parent self-reflection and change perceptions of the child in order to induce a secure parent-child relationship (Moretti et al. 2013). The latter approach may entail a somewhat slower process with delayed effects on actual changes in child and parent behaviors, which is supported by a previous evaluation of the Connect program where there were notable reductions in externalizing behaviors during a 1-year follow-up period (Moretti and Obsuth 2009). Further, behavioral programs rely more on homework assignments, and by the end of the intervention, parents are expected to carry on practicing what they have learned. This requires that parents set aside a certain amount of time during the months, and perhaps years, that follow program participation, doing exercises (e.g., child-directed play) that extend beyond everyday routines. The non-behavioral program Connect, on the other hand, does not rely on homework assignments, and its self-reflecting components may be less time-consuming, and thereby easier for parents to integrate into their everyday life after the end of the intervention. Thus, the continued reduction, during the follow-up phase, in child problem behaviors observed for Connect but not for the other programs may be related to the relative ease with which different program components can be completed.

About 10 % of the participating families reported that they had received additional PT between post measurement and the 2-year follow-up, and although there were no differences in proportions among the programs, there were differences between those who had participated in extra PT and those who had not, with regard the level of reported externalizing problems. Those who tended to seek more PT had more severe problems at post measurement, $t(507) = 2.42$, $p = .02$, and eventually also at the 2-year follow-up, $t(510) = 3.46$, $p < .01$, compared with those who did not obtain any more PT than was provided within the study.

The lack of a control condition limits opportunities to separate program effectiveness from any change associated with spontaneous remission. However, at the 2-year follow-up, the average level of child externalizing behavior across all four programs, as measured by ECBI intensity, was within ½ *SD* of the mean of a normative sample of Swedish children (Axberg et al. 2008). Given that the baseline level was more than 1½ *SD* above the norm mean (i.e., at program start, 27 months prior to the 2-year follow-up), this is an indication that the PT programs were at least reasonably effective in reducing externalizing behaviors.

Limitations and Strengths

The study has some limitations. First, we were not able to directly take the effects of clustered data into account. Due to high model complexity there was not a sufficient number of clusters per group. However, we estimated the intra-class correlations (ICC) to examine how much of the variations in our measures were due to differences across the sites. The ICC values ranged between .001 and .052 with a mean of .013, suggesting that on average only 1 % of the variations in our measures were due to the differences among the sites. Second, the study did not include reports from multiple sources (e.g., teachers or observations), which would have been preferable in terms of validating the findings, particularly in domains where the results rely on parent-reported measures with low internal consistency (parents' harsh treatment and attempted understanding). Third, two of the PT programs were compared despite the fact that children in these programs did not overlap in terms of age. Incredible Years included children aged 3–8, whereas Connect included those aged 9–12. Having child age as a covariate in all analyses may not have adequately compensated for this, and the comparison of program effects, between these particular programs, should be interpreted in light of this limitation. Fourth, we were only able to compare the programs with each other and not with an untreated control group at the 2-year follow-up. Thus, we were only able to draw conclusions about how program effects related to each other. Fifth, several study variables predicted attrition over time, but, despite this, there was no significant difference in attrition across the four programs. Thus, our conclusions regarding the relative effectiveness of the programs should have been minimally influenced by attrition. Lastly, it should be noted that the ECBI intensity subscale had a significant overlap with all three SNAP subscales ($r = .63$ to .70) and these scales can therefore be assumed to measure much the same externalizing behavior construct. It could be argued that these two measures should have been combined, due to this overlap, but as their contents focus on different aspects of externalizing behavior (ECBI items ask about problematic daily situations whereas the SNAP items follow DMS diagnostic criteria), they were reported on separately.

Some strengths also need to be acknowledged. First, this is the only study to compare the long-term effectiveness of four established PT programs with different theoretical underpinnings within the same trial. This makes cross-program comparisons more reliable than any that could be made through meta-analysis. Despite attempts to control for variations in study design, meta-analyses tend to suffer from variability in the design characteristics of their studies. Further, this effectiveness

study retained a large proportion of program participants up to 2-year follow-up.

Implications for Research and Practice

As one of few independent trials (not conducted by program developers) evaluating PT interventions, this 2-year follow-up study contributes to the field by helping clarify what can be expected of the Comet, Incredible Years, Cope and Connect programs when they are offered within a regular health care system. As a consequence, decision-makers and practitioners may be more confident and willing to implement one of these programs as an alternative to the more resource-consuming individual interventions often provided today (Kazdin 2013). The main finding, that effectiveness at long-term follow-up was about the same across programs, directs attention to additional research questions related to cost-effectiveness and program acceptability. In the future, such research may provide support for the use of one of the programs rather than the others.

Conclusions

The present study shows that Comet, Incredible Years, Cope and Connect all reduced child behavior problems, and also improved parenting practices and parental wellbeing. The improvements were generally either retained or continued to grow up to 2-year follow-up. Evidence of differences in change rates from baseline to follow-up, particularly between the attachment-based Connect and the behavioral programs, Comet and Incredible Years, was also found. However, regardless of these differences, levels of child problem behaviors were equal across the programs at 2-year follow-up. The findings support further implementation of these PT programs in routine care.

Acknowledgments We thank the National Board of Health and Welfare in Sweden who funded this trial and we would also like to express our deepest gratitude to the families, care units and group leaders that participated in the study.

Compliance with Ethical Standards

Funding This study was funded by the National Board of Health and Welfare (Socialstyrelsen) in Sweden, grant number 01–12,042/2008.

Conflict of Interest Author Jens Högström declares that he has no conflict of interest. Author Viveca Olofsson declares that she has no conflict of interest. Author Metin Özdemir declares that he has no conflict of interest. Author Pia Enebrink declares that she has no conflict of interest. Author Håkan Stattin declares that he has no conflict of interest.

Ethical Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed Consent Informed consent was obtained from all individual participants included in the study.

References

Alda, J. A., & Serrano-Troncoso, E. (2013). Attention-deficit hyperactivity disorder: agreement between clinical impression and the SNAP-IV screening tool. *Actas Espanolas de Psiquiatria, 41*, 76–83.

Axberg, U., Hanse, J. J., & Broberg, A. G. (2008). Parents' description of conduct problems in their children: a test of the Eyberg child behavior inventory (ECBI) in a Swedish sample aged 3-10. *Scandinavian Journal of Psychology, 49*, 497–505. doi:10.1111/J.1467-9450.2008.00670.X.

Barkley, R. A. (1997). *Defiant children*: *A clinician's manual for assessment and parent training*. New York, NY: Guilford Press.

Brannan, A., Helflinger, A., & Brickman, L. (1997). The caregiver strain questionnaire: measuring the impact on the family of living with a child with severe emotional disturbance. *Journal of Emotional and Behavioral Disorders, 5*, 212–222.

Bywater, T., Hutchings, J., Daley, D., Whitaker, C., Yeo, S. T., Jones, K., et al. (2009). Long-term effectiveness of a parenting intervention for children at risk of developing conduct disorder. *British Journal of Psychiatry, 195*, 318–324. doi:10.1192/bjp.bp.108.056531.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2 ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cunningham, C. E., Davis, J. R., Bremner, R., Dunn, K. W., & Rzasa, T. (1993). Coping modeling problem-solving versus mastery modeling: effects on adherence, in-session process, and skill acquisition in a residential parent-training program. *Journal of Consulting and Clinical Psychology, 61*, 871–877. doi:10.1037//0022-006x.61.5.871.

Cunningham, C. E., Bremner, R., & Boyle, M. (1995). Large group community-based parenting programs for families of preschoolers at risk for disruptive behavior disorders: utilization, cost-effectiveness, and outcome. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 36*, 1141–1159. doi:10.1111/j.1469-7610.1995.tb01362.x.

DeGarmo, D. S., & Forgatch, M. S. (2005). Early development of delinquency within divorced families: evaluating a randomized preventive intervention trial. *Developmental Science, 8*, 229–239. doi:10.1111/j.1467-7687.2005.00412.x.

DeGarmo, D. S., Patterson, G. R., & Forgatch, M. S. (2004). How do outcomes in a specified parent training intervention maintain or wane over time? *Prevention Science, 5*, 73–89. doi:10.1023/B:PREV.0000023078.30191.e0.

Duncan, T. E., Duncan, S. C., & Strycker, L. A. (2013). *An introduction to latent variable growth curve modeling*: *Concepts*, *issues*, *and application*. Mahwah, NJ: Routledge Academic.

Eisner, M., Nagin, D., Ribeaud, D., & Malti, T. (2012). Effects of a universal parenting program for highly adherent parents: a propensity score matching approach. *Prevention Science, 13*, 252–266. doi:10.1007/s11121-011-0266-x.

Eyberg, S., & Pincus, D. (1999). *Eyberg Child Behavior Inventory and Sutter-Eyberg Student Behavior Inventory-Revised*: *Professional manual*. Odessa, FL: Psychological Assessment Resources.

Forgatch, M. S. (1994). *Parenting through change*: *A programmed intervention curriculum for groups of single mothers*. Eugene, OR: Oregon Social Learning Center.

Furlong, M., McGilloway, S., Bywater, T., Hutchings, J., Smith, S. M., & Donnelly, M. (2012). Behavioural and cognitive-behavioural group-based parenting programmes for early-onset conduct problems in children aged 3 to 12 years. *Cochrane Database of Systematic Reviews.* doi:10.1002/14651858.CD008225.pub2.

Gardner, F., Burton, J., & Klimes, I. (2006). Randomised controlled trial of a parenting intervention in the voluntary sector for reducing child conduct problems: outcomes and mechanisms of change. *Journal of Child Psychology and Psychiatry, 47*, 1123–1132. doi:10.1111/j.1469-7610.2006.01668.x.

Gilmore, L., & Cuskelly, M. (2009). Factor structure of the parenting sense of competence scale using a normative sample. *Child: Care, Health and Development, 35*, 48–55. doi:10.1111/j.1365-2214.2008.00867.x.

Hahlweg, K., Heinrichs, N., Kuschel, A., Bertram, H., & Naumann, S. (2010). Long-term outcome of a randomized controlled universal prevention trial through a positive parenting program: is it worth the effort? *Child and Adolescent Psychiatry and Mental Health, 4*, 14. doi:10.1186/1753-2000-4-14.

Johnston, C., & Mash, E. (1989). A measure of parenting satisfaction and efficacy. *Journal of Child & Adolescent Psychology, 18*, 167–175.

Kazdin, A. E. (2003). Research design in clinical psychology (Vol. 3). Boston, MA: Allyn and Bacon Boston.

Kazdin, A. E. (2013). Evidence-based treatment and usual care cautions and qualifications. *JAMA Psychiatry, 70*, 666–667. doi:10.1001/jamapsychiatry.2013.2112.

Kling, Å., Sundell, K., Melin, L., & Forster, M. (2006). Komet för föräldrar: en randomiserad effektutvärdering av ett föräldraprogram för barns beteendeproblem [comet for parents: a randomized evaluation of a parenting program for child behavior problems]: Stockholms Socialtjänstförvaltning: FoU-enheten.

Kling, A., Forster, M., Sundell, K., & Melin, L. (2010). A randomized controlled effectiveness trial of parent management training with varying degrees of therapist support. *Behavior Therapy, 41*, 530–542. doi:10.1016/j.beth.2010.02.004.

Lindsay, G., & Strand, S. (2013). Evaluation of the national roll-out of parenting programmes across England: the parenting early intervention programme (PEIP). *BMC Public Health, 13*. doi:10.1186/1471-2458-13-972.

Litschge, C. M., Vaughn, M. G., & McCrea, C. (2010). The empirical status of treatments for children and youth with conduct problems: an overview of meta-analytic studies. *Research on Social Work Practice, 20*, 21–35. doi:10.1177/1049731508331247.

Losel, F., & Stemmler, M. (2012). Preventing child behavior problems in the Erlangen-Nuremberg development and prevention study: results from preschool to secondary school age. *International Journal of Conflict and Violence, 6*, 214–224.

Lundahl, B., Risser, H., & Lovejoy, M. (2006). A meta-analysis of parent training: moderators and follow-up effects. *Clinical Psychology Review, 26*, 86–104. doi:10.1016/j.cpr.2005.07.004.

Malti, T., Ribeaud, D., & Eisner, M. P. (2011). The effectiveness of two universal preventive interventions in reducing children's externalizing behavior: a cluster randomized controlled trial. *Journal of Clinical Child and Adolescent Psychology, 40*, 677–692. doi:10.1080/15374416.2011.597084.

Maughan, D. R., Christiansen, E., Jenson, W. R., Olympia, D., & Clark, E. (2005). Behavioral parent training as a treatment for externalizing behaviors and disruptive behavior disorders: a meta-analysis. *School Psychology Review, 34*, 267–286.

Moretti, M. M., & Obsuth, I. (2009). Effectiveness of an attachment-focused manualized intervention for parents of teens at risk for aggressive behaviour: the connect program. *Journal of Adolescence, 32*, 1347–1357. doi:10.1016/j.adolescence.2009.07.013.

Moretti, M. M., Braber, K., & Obsuth, I. (2013). *Connect*: *An Attachment-focused treatment group for parents and caregivers – A principle based manual*. Burnaby, BC: Simon Fraser University.

Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus user's guide.* (7th ed.). Los Angeles, CA: Muthén & Muthén.

National Institute for Health and Clinical Excellence. (2013). *Antisocial behaviour and conduct disorders in children and young people*: *recognition*, *intervention and management.* NICE clinical guideline 158. Retrieved from http://www.nice.org.uk/guidance/cg158.

Özdemir, M. (2015). How much do we know about the long-term effectiveness of parenting programmes? Advances, shortcomings, and future directions. *Journal of Children's Services, 10*, 120–132. doi:10.1108/JCS-02-2014-0016.

Radloff, L. S. (1977). The CES-D scale a self-report depression scale for research in the general population. *Applied Psychological Measurement, 1*, 385–401. doi:10.1177/014662167700100306.

Reinecke, J., & Weins, C. (2013). The development of delinquency during adolescence: a comparison of missing data techniques. *Quality & Quantity, 47*, 3319–3334. doi:10.1007/s11135-012-9721-4.

Robinson, E., Eyberg, S., & Ross, A. (1980). The standardization of an inventory of child conduct problem behaviors. *Journal of Clinical Child & Adolescent Psychology, 9*, 22–28.

Sandler, I. N., Schoenfelder, E. N., Wolchik, S. A., & MacKinnon, D. P. (2011). Long-term impact of prevention programs to promote effective parenting: lasting effects but uncertain processes. *Annual Review of Psychology, 62*, 299–329. doi:10.1146/annurev.psych.121208.131619.

SBU. (2010). Program för att förebygga psykisk ohälsa hos barn. En systematisk litteraturöversikt [Methods to Prevent Mental ill-Health in Children. A Systematic Review]. (SBU-report 202). Retrieved january 27th 2014, from Swedish Council on Health Technology Assessment http://www.sbu.se/sv/Publicerat/Gul/Program-for-att-forebygga-psykisk-ohalsa-hos-barn/.

Schafer, J. L., & Graham, J. W. (2002). Missing data: our view of the state of the art. *Psychological Methods, 7*, 147–177. doi:10.1037//1082-989x.7.2.147.

Serketich, W. J., & Dumas, J. E. (1996). The effectiveness of behavioral parent training to modify antisocial behavior in children: a meta-analysis. *Behavior Therapy, 27*, 171–186.

Stattin, H., Persson, S., Burk, W. J., & Kerr, M. (2011). Adolescents' perceptions of the democratic functioning in their families. *European Psychologist, 16*, 32–42. doi:10.1027/1016-9040/a000039.

Stattin, H., Enebrink, H., Özdemir, M., & Giannotta, F. (2015). A national evaluation of parenting programs in Sweden: the short-term effects using an RCT effectiveness design. *Journal of Consulting and Clinical Psychology, 83*, 1069–1084. doi:10.1037/a0039328.

Swanson, J. M. (1992). *School-based assessments and interventions for ADD students*. Irvine, CA: KC Publishing.

Webster-Stratton, C. (1981). Videotape modeling: a method of parent education. *Journal of Clinical Child & Adolescent Psychology, 10*, 93–98.

Webster-Stratton, C. (1984). Randomized trial of two parent-training programs for families with conduct-disordered children. *Journal of Consulting and Clinical Psychology, 52*, 666–678. doi:10.1037/0022-006x.52.4.666.

Webster-Stratton, C. (1998). Preventing conduct problems in head start children: strengthening parenting competencies. *Journal of Consulting and Clinical Psychology, 66*, 715–730. doi:10.1037/0022-006X.66.5.715.

Webster-Stratton, C., Reid, M. J., & Hammond, M. (2004). Treating children with early-onset conduct problems: intervention outcomes for parent, child, and teacher training. *Journal of Clinical Child and Adolescent Psychology, 33*, 105–124. doi:10.1207/s15374424jccp3301_11.

Webster-Stratton, C., Rinaldi, J., & Reid, J. M. (2011). Long-term outcomes of incredible years parenting program: predictors of adolescent adjustment. *Child and Adolescent Mental Health, 16*, 38–46. doi:10.1111/j.1475-3588.2010.00576.x.

Wolchik, S. A., Sandler, I. N., Millsap, R. E., Plummer, B. A., Greene, S. M., Anderson, E. R., et al. (2002). Six-year follow-up of preventive interventions for children of divorce: a randomized controlled trial. *Journal of the American Medical Association, 288*, 1874–1881. doi:10.1001/jama.288.15.1874.

Zubrick, S. R., Ward, K. A., Silburn, S. R., Lawrence, D., Williams, A. A., Blair, E., et al. (2005). Prevention of child behavior problems through universal implementation of a group behavioral family intervention. *Prevention Science, 6*, 287–304. doi:10.1007/s11121-005-0013-2.