



UvA-DARE (Digital Academic Repository)

Helping families change: The adoption of the Triple P - Positive Parenting Program in the Netherlands

de Graaf, I.M.

Publication date
2009

[Link to publication](#)

Citation for published version (APA):

de Graaf, I. M. (2009). *Helping families change: The adoption of the Triple P - Positive Parenting Program in the Netherlands*. Trimbos Institute.

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

3 Is Triple P effective on behavior problems in children?*

Abstract

The Triple P Positive Parenting Program is a multilevel parenting program to prevent and offer treatment for severe behavioral, emotional, and developmental problems in children. The aim of this meta-analysis is to assess the effectiveness of Triple P Level 4 interventions in the management of behavioral problems in children by pooling the evidence from relevant literature that included Level 4 Triple P interventions. Level 4 intervention is indicated if the child has multiple behavior problems in a variety of settings and there are clear deficits in parenting skills. Results indicate that Level 4 of Triple P interventions reduced disruptive behaviors in children. These improvements were maintained well over time, with further improvements in long-term follow-up. These effects support the widespread adoption and implementation of Triple P that is taking place in an increasing number of countries in quite diverse cultural contexts around the world.

3.1 The Triple P Positive Parenting Program

The Positive Parenting Program (Triple P) is a multilevel program to prevent and offer treatment for severe behavioral, emotional, and developmental problems in children aged 0 to 16 years through enhancing the knowledge, skills, and confidence of parents. Triple P incorporates five levels of interventions on a tiered continuum of increasing intensity. The rationale for this stepped-care strategy is that there are different levels of dysfunction and behavioral disturbance in children and that parents may have different needs and desires regarding the type, intensity, and mode of assistance they require (Sanders, Markie-Dadds, & Turner, 1999). Triple P is designed as a public health strategy, a population system of interventions that incorporates different delivery modalities (group, individual, and self-directed).

Levels of Intervention

Level 1 is a form of universal prevention, and it delivers psycho educational information on parenting skills to interested parents. Level 2 is a brief intervention of one or two sessions, for parents of children with mild behavioral problems. Level 3 is a four-session intervention, targets children with mild to moderate behavioral difficulties, and includes active skills training for parents. Level 4 is an intensive, 8- to 10-session

* Published as: Graaf, I. de, Speetjens, P., Smit, F., Wolff, M. de, & Tavecchio, L. (2008). Effectiveness of the Triple P Positive Parenting Program on behavioural problems in children. A meta-analysis. *Behaviour Modification*, 32, 714 -35.

parent training program for children with more severe behavioral difficulties or who are at risk of developing such problems, which can be offered either individually or in a group of parents. Parents are taught a variety of child management skills. This intervention is a form of selective or indicated prevention in that the children are at elevated risk levels of developing behavioral problems. Finally, Level 5 is an enhanced behavioral family intervention (BFI) program for families in which parenting difficulties are complicated by other sources of family distress (e.g., marital conflict, parental depression, or high levels of stress; Sanders et al., 1999).

Standard Triple P, Group Triple P, Self-Directed Triple P

This indicated preventive intervention targets high-risk individuals who are identified as having detectable problems but who do not yet meet diagnostic criteria for a behavioral disorder. It should be noted that this level of intervention can target individual children at risk or an entire population to identify individual children at risk. For example, a group version of the program may be offered universally in low-income areas, with the goal of identifying and engaging parents of children with severe disruptive and aggressive behavior. Parents are taught a variety of child management skills including providing brief contingent attention following desirable behavior, how to arrange engaging activities in high-risk situations, and how to use clear, calm instructions, logical consequences for misbehavior, planned ignoring, quiet time (nonexclusionary timeout), and timeout. Parents are trained to apply these skills both at home and in the community. Specific strategies such as planned activities training are used to promote the generalization and maintenance of parenting skills across settings and over time (Sanders & Dadds, 1982). As in Level 3, this level of intervention combines the provision of information with active skills training and support. However, it teaches parents to apply parenting skills to a broad range of target behaviors in both home and community settings with the target child and siblings. Here, it should be noted that there are three delivery formats at Level 4: Standard Triple P, Group Triple P, and Self-Directed Triple P. Standard Triple P is an individual 10-session program for parents. Group Triple P is an 8-session program conducted in groups of 10 to 12 parents with four 15- to 30-min follow-up telephone sessions provided as additional support to the parents. Self-Directed Triple P is a 10-week self-help program for parents and may be augmented by weekly 15- to 30-min telephone consultations.

Level 4 intervention is indicated if the child has multiple behavior problems in a variety of settings and there are clear deficits in parenting skills. If the parent wishes to have individual assistance and can commit to attending a 10-session program, the Standard Triple P program is appropriate. Group Triple P is appropriate as a universal (available to all parents) or selective (available to targeted groups of parents) prevention parenting support strategy; however, it is particularly useful as an early intervention strategy for parents of children with current behavior problems. Self-Directed Triple P is ideal for families who live where access to clinical services is poor (e.g., families in rural or remote areas). It is most likely to be successful with families

who are motivated to work through the program on their own and where literacy or language difficulties are not present.

Theoretical Basis of Triple P

Triple P is based on social learning principles (Patterson, Reid, & Dishion, 1982). This approach to the treatment and prevention of childhood disorders has the strongest empirical support of any intervention with children, particularly those with conduct problems (Kazdin, 1987; Sanders, 1996; Sanders & Dadds, 1993; Taylor & Biglan, 1998; Webster Stratton & Hammond, 1997). Furthermore, the Triple-P program is based on research in child and family behavior therapy, developmental research on parenting every day (Risley, Clark, & Cataldo, 1976; Sanders 1992, 1996), research on social information-processing models (e.g., Bandura, 1977, 1995), research from the field of developmental outcomes in children (e.g., Emery, 1982; Grych & Fincham, 1990; Hart & Risley, 1995; Rutter, 1985), and research on a public health perspective to family intervention (e.g., Biglan, 1995; Mrazek & Haggerty, 1994; National Institute of Mental Health, 1998).

Evaluation

The evaluation of Triple P needs to be viewed in the broader context of evaluations of BFI. There is clear evidence that BFI is beneficial in children with disruptive behavior disorders (Forehand & Long, 1988; Webster Stratton, 1994). Since 1978, the intervention methods of Triple P have been subjected to a series of controlled evaluations (Sanders & Dadds, 1993). Since that time, the intervention methods used in Triple P have been subjected to a series of controlled evaluations using both intrasubject replication designs and traditional randomized control group designs. There is evidence that Triple P is an effective parenting strategy. Several studies have shown that parenting skills training used in Triple P produces predictable decreases in child behavior problems, which have typically been maintained over time. Furthermore, clinically meaningful and statistically reliable outcomes for both children and their parents have been demonstrated for the standard, self-directed, telephone-assisted, group, and enhanced interventions. The population varied in the different studies: parents of children with oppositional behavior, parents of children with oppositional defiant disorder or conduct disorder, or parents reporting concerns about disruptive child behavior. Finally, the program has also been successfully used for several different family types, including two-parent families, single-parent families, stepfamilies, maternally depressed families, maritally discordant families, and families with a child with an intellectual disability (Sanders, Markie-Dadds, & Turner, 2003). In those studies, the following variables were measured: child disruptive behavior, parent-child interaction, parenting style and confidence, parental adjustment (depression, anxiety, stress, self-esteem), parenting conflict, and relationship satisfaction.

3.2 Meta-Analysis

In the current meta-analysis, we examine the effectiveness of Triple P interventions in the management of behavioral problems among children, aged 2 to 12 years old, by pooling the evidence from the pertinent studies. In a meta-analysis, the results of a large and diverse body of studies can be summarized, interpreted, and more readily generalized to an entire population because of the increase in the number of participants (Rosnow & Rosenthal, 2002; Silverman, 2001). Hence, in this meta-analysis, an overall effect size for Level 4 Triple P interventions worldwide is calculated, as is the variability in the set of studies. The systematic coding of study characteristics permits an analytically precise examination of the relationships between study findings and study features such as respondent characteristics, format, design, and nature of intervention (Lipsey & Wilson, 2001). Because most of the relevant Triple P studies that were identified concerned Level 4 of the Triple P system, we decided to restrict the meta-analysis to this level only.

We conducted two meta-analyses. In the first meta-analysis, we assessed the effectiveness of Triple P in behavioral problems of children compared to the control group, as directly measured at the end of the intervention. In the second meta-analysis, we assessed the degree to which post-intervention effects were maintained over time in the intervention group.

In those meta-analyses, we hypothesized that behavior problems of children, aged 2 to 11 years old, decrease after participating in a Level 4 Triple P intervention, both directly after the intervention and after a follow-up of 6 to 12 months. Second, we were also interested in whether the effects of Level 4 of Triple P were moderated by the different delivery formats of the intervention and characteristics of the parents and the children. It was hypothesized that the efficacy of Triple P is independent of whether the intervention was delivered to individual parents, to groups, or in a self-help format. Third, empirical studies have shown that physically aggressive behavior occurs in children 1 year old, increases in the 2nd life year, and then tends to decline from the 3rd birthday onward (Alink et al., 2006; Tremblay et al., 2004). Therefore, we hypothesized that Triple P is more effective when the interventions are given at age 2 to 4 compared with older ages. In addition, it is evident that boys exhibit more externalizing problems than do girls at the age of 2 to 3 years (Alink et al., 2006; Cummings, Davies, & Campbell, 2002; Hudson & Rapee, 2005). Because there is more room for change for boys than for girls, we hypothesized that Triple P is more effective for boys than for girls. Finally, the behavioral problems of the children at the start of the intervention may be of importance. The severity of the problems at the start of the intervention differs across Triple P studies depending on whether they are universal prevention trials, indicated prevention trials, or treatment studies. It was hypothesized that Triple P is more effective for children with higher scores on the Eyberg Child Behaviour Questionnaire (ECBI) because there is more room for change for children with higher ECBI scores.

3.3 Method

We used three different search methods to identify literature for the meta-analysis. First, we searched the literature in two electronic databases, Medline (1975 to February 2006) and PsycINFO (1975 to February 2006). The following keywords were used: *Triple P* and *parent* (and words such as *parenting* or *parental* were also included in the search). Second, we searched all reference lists of studies compiled by the Parenting and Family Support Centre at the University of Queensland in Australia. Third, we asked researchers who had conducted Triple P studies whether they had other relevant unpublished material. We found three Triple P projects in Germany and Switzerland.

Studies had to meet the following inclusion criteria: (a) the study examined the effects of a Level 4 Triple P intervention, an intensive parent training program for children with more severe behavioral difficulties or who are at risk of developing such problems, (b) effectivity of Triple P was assessed using a questionnaire for the parents to evaluate disruptive behavior in their children aged 2 to 11, and (c) sufficient empirical data were reported for the calculation of standardized effect sizes. Because we conducted two meta-analyses, the study had to report posttest data of the intervention and control group (for the first meta-analysis), and preintervention and follow-up data had to be reported separately for the intervention group (for the second meta-analysis). We excluded studies with samples of children older than 11 because Triple P has a separate program for teens.

Measurement

To assess the disruptive behavior of children, the ECBI is often used (Eyberg & Pincus, 1999). The ECBI is a 36-item measure of parental perceptions of disruptive behavior in children aged 2 to 16 years. It provides two measures: frequency of disruptive behaviors (intensity score) rated on 7-point scales and number of disruptive behaviors that parents list as problematic (problem score). The ECBI has a high internal consistency for Intensity Scale ($r = .95$) and for the Problem Scale ($r = .94$) (Robinson, Eyberg, & Ross, 1980). The established cutoff scores (Eyberg & Ross, 1978) of 127 for the Intensity Scale and 11 for the Problem Scale have been validated in clinical studies for both young children (e.g., Webster Stratton, 1984) and adolescents (e.g., Baden & Howe, 1992).

The ECBI is the most frequently used measure in the Triple P interventions to assess behavior problems in children. Other measures used in the studies assessing behavior problems are the Child Behavior Checklist (CBCL) and the Strength and Difficulties Questionnaire (SDQ). The CBCL was used in one study (Heinrichs et al., 2005) and the SDQ in two studies (Leung, Sanders, Leung, Mak, & Lau, 2003; Martin & Sanders, 2003). Because in the last two studies the ECBI was also represented, we decided to include only studies in which behavior problems were assessed with the ECBI.

Selected Studies

We found 48 effect studies in which all levels of Triple P were used and 25 studies that focused on the Level 4 intervention. Of these, 15 studies met the inclusion criteria. In all, 3 studies were excluded from the first meta-analysis because they had no control group, 3 studies were excluded because they were no-effectiveness studies, and 3 studies were eliminated because a questionnaire other than the ECBI was used (CBCL and SDQ).

Selected characteristics of these studies are presented in Table 1. The 15 studies were independently coded by two researchers on design and sample characteristics, delivery format of the Level 4 intervention of Triple P, reliability and validity of the measures, characteristics of the parents and children, initial problems of the children, and duration of follow-up times. Discrepancies between the two researchers were resolved by recalculating and consensus.

Group Triple P was used as the intervention in 9 studies, Standard Triple P in 1 study, and Self-Directed Triple P in 6 studies. In one study (Sanders, Markie-Dadds, Tully, & Bor, 2000), two versions of Triple P, Self-Directed and Standard, were compared with a waitlist control group. In this case, analyses for both interventions were conducted. In one study, the target of the intervention was working parents. In all, 14 studies were randomized clinical trials and 1 study was a nonrandom, two-group, concurrent prospective observation design (Zubrick et al., 2005). In 9 studies, parents reported their child's behavior in the clinically elevated range at preintervention and in 6 studies in the nonclinical range. The percentage of boys was 62.6%, averaged across all studies. We divided the studies into less than 62.6% boys and more than 62.6% boys to have two comparing groups. Boys were overrepresented in all studies, and the studies would have been heterogeneous if we had divided the groups into 50% boys and 50% girls. In 7 studies, more than 62.6% of the children were boys. In 5 studies, the children were younger than 4. A total of 14 studies were selected for the first meta-analysis and also 14 for the second meta-analysis. In 4 studies, follow-up data were presented after 12 months; in 11 studies, the follow-up was conducted at 4 to 6 months; and in 1 study (Bodenman, Cina, Ledenmann, & Sanders, 2007) follow-up measures were taken at both 6 and 12 months.

Meta-Analytic Procedures

For each study, we calculated an effect size: the standardized difference between the means of two groups (Cohen's d). A correction of the standardized mean difference was used (Hedges g) because studies with samples of fewer than 20 tend to show upward bias in their results (Lipsey & Wilson, 2001). Because studies with a larger sample size provide more reliable estimates of the population mean, effect sizes were weighted by sample size.

In the first meta-analysis, we were interested in the differences between mean scores of the experimental and control groups at postmeasurement. The standardized effect size, d , was calculated as $d = (M_E - M_C) / SD_C$, where M_E and M_C are the means of the experimental and control groups, respectively, at postintervention and postmeas-

urement and SDC is the standard deviation of the control group. The standardized effect sizes, d , show by how many standard units (z scores) the experimental group has progressed as compared to the control group at postmeasurement.

In the second meta-analysis, we also calculated the standardized mean difference as $d = (M_p - M_f) / SD$, where M_p and M_f are the means at baseline and follow-up, respectively, and SD_p is the standard deviation at baseline. This within-group effect size thus indicates by how many standard units the recipients of the intervention have improved over time relative to their own baseline score. It can thus be interpreted as a standardized health gain score. An effect size of $d = 0.5$ indicates that the mean of the experimental group is half a standard deviation larger than the mean of the control group. Because the study of Zubrick et al. (2005) was not a randomized clinical trial, we calculated the standardized pre-post change score of the experimental group (d_E) and did the same for the control group (d_C). Then, we calculated their difference using the following formula (d) $\Delta d_E - d_C$.

Basically, meta-analysis amounts to pooling individual d s and obtaining a best overall estimate of the intervention effect, within its 95% confidence interval (95% CI). The meta-analyses were conducted with the computer program Meta-Analysis, Version 5.3 (Schwarzer, 1989). This program is based on the statistical techniques outlined by Hedges and Olkin (1985). We made use of the random effects model. In this model, it is assumed that the variability between the effect sizes is because of sampling error plus variability in the population of effects (Hedges & Olkin, 1985). The meta-analytical outcomes obtained under a random effects model are said to be conservative in that their 95% CIs are usually broad, thus reducing the likelihood of Type II error. The results are shown in Tables 2 and 3 and are corrected for small sample size (transforming Cohen's d into a measure technically known as Hedge's g) and are also corrected for random measurement error in the original scale (i.e., Cronbach's α of the outcome measures as used in the primary studies). From a clinical perspective, an effect size in the range of $d = 0.56$ to 1.2 can be interpreted as a large effect, whereas effect sizes of 0.33 to 0.55 are moderate, and effect sizes of 0.00 to 0.32 are deemed small (Lipsey & Wilson, 1993). In this analysis, it was not possible to calculate the differences between mean scores of the intervention and the control groups because in most cases the control group had no follow-up measurements. We also conducted the homogeneity test, Q , to ascertain whether the various effect sizes that are averaged into the pooled d all estimate the same population effect size (Rosenthal & Rubin, 1982). A rejection of the null hypothesis of homogeneity indicates that there are differences among the effect sizes of the primary studies that cannot be attributed to random sample error and may be related to systematic differences across the original studies—which then need to be further investigated (Schwarzer, 1989). Whenever the Q test for homogeneity was significant, we conducted an outlier analysis. To identify outliers, we conducted cluster analyses with the computer program (Schwarzer, 1989), conducted another meta-analysis without the outlier, and then ascertained whether we had obtained a more homogeneous set of primary studies where the Q test was no longer significant. As the discriminatory power of

Table 1. Selected Characteristics of studies examining the effects of Triple P level 4 on behavioral and emotional problems in children

Study	Conditions	N TP	Target population
Bodenman, Cina, Ledenmann and Sanders (2007)	1. GR 2. no treatment 3. CCET	GR: 51 No treatment: 41	Couples with children aged between 2-12 years recruited by means of public ads in several newspapers in the rural areas of Switzerland.
Connell, Sanders, and Markie-Dadds (1997)	1. SD 2. WL	SD: 12 WL: 11	Families in a rural area, reporting concerns about their child's behavior and rate their child's behavior within the clinical range on the intensity scale of the ECBI
Gallart and Matthey (2005)	1. GR+T 2. GR 3. WL	GR: 33 WL: 16	Parents experiencing difficulties with their children's disruptive behaviors.
Hoath and Sanders (2002)	1. GR 2. WL	GR: 9 WL: 11	Families with a child with a clinical diagnosis of ADHD.
Ireland, Sanders, and Markie-Dadds (2003)	1. GR 2. GR (with partner support module)	GR:16 EGR: 16	Couples with concerns about their child's disruptive behavior, exhibit clinically significant levels of marital conflict over parenting
Leung et al. (2003)	1. GR 2. WL	GR: 33 WL: 36	Parents living in Hong Kong with concerns about their children's behavior
McTaggart and Sanders (2005)	1. GR (Newsletter) 2. GR 3. WL	GR: 79 WL: 244	All families living in a high-risk region (e.g., health, child abuse).
Markie-Dadds and Sanders (2006b)	1. SD 2. WL	SD: 21 WL: 22	Mothers had to rate their children's behavior in the elevated range on the ECBI (IS \geq 127 or PS \geq 11)
Markie-Dadds and Sanders (2006a)	1. SD 2. SD+T 3. WL	SD: 28 WL:12	Mothers had to rate their children's behavior in the elevated range on the ECBI (IS \geq 127 or PS \geq 11)
Martin and Sanders (2003)	1. GR WPTP 2. WL	GR: 16 WL: 11	Academic and general staff at the University of Queensland, Australia. Rating of the child behavioral problems in the clinical range of intensity as measured by the SDQ.
Morawska and Sanders (2006)	1. SD 2. WL	SD: 73 WL: 37	Parents with concerns about their child's behavior.

	Meas.	% DO	Age Child (M)	% gender child is male	Problem and intensity score in E	Problem and intensity score in C	Meta 1/ meta 2
	Pre, post, 6 months, 12 months	12 at FU	6.6 (SD =2.83)	55%	10.52 118.29	- -	Meta 1 and 2
	Pre, 4 months	-	4.27 (SD =1.05)	43%	20.75 155.83	18.55 158.36	Meta 1 and 2
	Pre, post	-	5.4 (SD =1.5)	75%	- 108.3	- 137.1	Meta 1
	Pre, post, 3 months	5 at post 23 at FU	7.7 (SD =1.33)	-	23 164.22	19.55 159.73	Meta 1 and 2
	Pre, 3 months	28 at FU	3.53 (SD =1.12)	58%	11.72 122.16	- -	Meta 2
	Pre, post	24 at post	4.23 (SD =1.06)	64%	13.25 131.38	16.56 137.7	Meta 1
	Pre, post, 6 months	14 at post	-	57%	- 121.6	- 107.1	Meta 1 and 2
	Pre, post, 6 months	-	3.59 (SD = 0.76)	64%	15.71 126.67	15.23 138.5	Meta 1 and 2
	Pre, post, 6 months	0 at post 4 at FU	3.89 (SD =0,96)	64%	19.27 160.2	20.17 145.75	Meta 1 and 2
	Pre, post, 4 months	23 at post	5.8 (SD =0.19)	-	11.89 125.63	17 135.81	Meta 1 and 2
	Pre, post 6 months	11 at post	2.18 (SD = 0.42)	51%	11.16 120.88	10.19 119.9	Meta 1 and 2

Study	Conditions	N TP	Target population
Sanders, Markie-Dadds, Tully, and Bor (2000)a	1. ST 2. SD 3. Enhanced 4. WL	ST: 65 Enhanced: - SD: 61 WL: 71	Parents in a low-income areas, reporting concerns about their child's behavior
Turner, Richards, and Sanders (in press)	1. GR 2. WL	36 GR: 18 WL: 18	Indigenous families with concerns about their child's behavior or their own parenting skills, requesting information or advise at community health sites
Yuki, Matsumoto, Sofronoff, and Sanders (2007)	1. GR 2. WL	TP: 23 WL: 25	Japanese parents in Brisbane (Australia) with reporting behavioral problems in children
Zubrick et al. (2005)	1. GR 2. control region	GR: 691 Control region: 774	Universal population: all families in a higher risk region

GR = Group; EGR = Enhanced Group; GR+T = Group with telephone sessions; SD = Self Directed; SD+T = Self Directed with telephone sessions; S = Standard Triple P; WPTP = Work Place Triple P; CCET = Couples Coping Enhancement Training; WL = Waitlist; Age Child (M):= medium age; DO = percentage of drop-out; FU = Follow Up; meta 1 = meta-analysis 1; meta 2 = meta-analysis 2; E = Experimental Group; C = Control Group. Cut-offs are 127 for ECBI intensity and 11 for ECBI problem score; a. analyses for both Standard Triple P as Self Help Triple P in this study were conducted.

the Q test is not very high, we also looked at the percentage of variance across the primary studies that can be accounted by sample error. The latter can be directly estimated from the random effects model that we used. We assumed homogeneity when at least 80% of the variance across studies could be attributed to random sample error. Outcomes of sufficiently homogeneous sets of primary studies are reported in Table 2.

It should be noted that the use of the Q statistic and indices for sample error allow for a data-driven procedure to identify homogeneous subsets of studies. We also took a content-driven approach and formed subgroups based on the characteristics of the intervention. Again, the Q test was performed to test the idea that these content-driven selections had resulted in homogeneous data sets. In this way, contrasting data sets could be compared. This was done to ascertain, for example, whether a self-help version of Triple P was inferior (or superior) to a therapist-assisted version. When the 95% CIs of the respective effect sizes d were not overlapping, we considered the contrast to be statistically significant. Finally, for each meta-analysis, Orwin's fail-safe number was calculated. This number indicates

	Meas.	% DO	Age Child (M)	% gender child is male	Problem and intensity score in E	Problem and intensity score in C	Meta 1/ meta 2
	Pre, post, 1 yr	14 at post	3.40 (SD = 0.30)	68%	(ST): - 137.4 (SD): - 144.24	(ST): - 138.0 (SD): - 144.24	Meta 1 and 2
	Pre, post 6 months	22 at post 26 at FU	5.72 (SD = 3.19)	67%	17.4 144.23	15.79 130.18	Meta 1 and 2
	Pre, post 3 months	0 at post 4 at FU	4.9 (-)	54%	8.35 108.04	8.0 101.72	Meta 1 and 2
	Pre, post, 12 months 24 months	12 at post 13 at 12 mo 21 at 24 mo	3.73 (SD = 1.17)	57%	- 121.6	- -	Meta 1 and 2

how many (hypothetical) studies with an effect size of zero should be found and included in the meta-analysis to reduce the observed effect size to a smaller value of, say, $d = 0.20$.

3.4 Results

The overall mean effect size for the 14 studies of the child behavior observed by parents at postmeasurement was $d = 0.88$, with a 95% CI of 0.50, 1.27 (Table 2). This effect was statistically significant ($Z = 4.49$, $p < .001$). This was a large effect according to Cohen's criteria. The Q test for the hypothesis of homogeneity across effect sizes had to be rejected, indicating that there was a substantial amount of unexplained variance in the total set of studies that might be attributed to the systematic effects of covariates. Of the variance, 20.4% was caused by random sample error, which left room for a remaining 79.6% that may have systematically covaried with (unknown) covariates. The number of studies with a zero effect that should be found to reduce the effect size to $d = 0.20$ was 47.8. The overall mean effect size concerning the long-term measurement of child behavior was $d = 1.00$, with a 95% CI of 0.55, 1.46. This effect was statistically significant ($z = 4.33$, $p = .001$). This was a large effect. The Q test for the hypothesis of homogeneity across effect sizes had to be rejected. Of the variance, 20.54% was caused by random sample error. The number of studies with a zero effect that should be found to reduce the effect size to

0.20 was 56. At 6 months follow-up, an overall mean effect size of $d = 1.07$ was found ($z = 3.49$, $p = .001$). In the meta-analysis of the four studies with 12-month follow-ups, we found an overall mean effect size of $d = 0.84$ ($z = 2.59$, $p = .001$). However, the results were significantly heterogeneous.

Table 2. Results of meta-analyses examining the effects of the Triple-P level 4 on the ECBI: Eyberg Child Behavior Inventory

	N_{ES}	N	D	95% CI	Q (df)	%SE	F/S-K
Effects at post-measurement							
All studies	14	2537	0.88	0.50-1.27	66.71 (13)***	20.34%	47.8
Only group, outliers excluded (no. 2, 6, 8, 9, 10, 14 in Table 1)	8	2182	0.42	0.33-0.51	13.17 (7)	99.53%	8.8
Long-term effects after 6 and 12 months							
All studies	14	2232	1.00	0.55-1.46	86.5 (13)***	20.54%	56
All studies, outliers excluded (no., 2, 4, 14, in Table 1)	11	2156	0.65	0.44-0.86	40.44 (10)***	57.5%	25
Excluded 3 Self-Directed studies (no. 8, 9, 12, in Table 1)	8	1860	0.37	0.28-0.46	4.60 (7)	100%	6.9
6 months	10	610	1.07	0.47-1.67	52.39 (9)***	18.2%	43
6 months, outliers excluded (2, 8, 14, in Table 1)	7	496	0.49	0.31-0.67	3.87 (6)	100%	10
12 months	4	1622	0.83	0.20-1.46	28.07 (3)***	26.6%	12.7

N_{ES} = Number of effect sizes; N = number of subjects in the studies; d = overall effect size; CI = confidence interval; Q = homogeneity Q; % SE = percentage of the variance accounted for by random sample error; F/S-K = Orwin's Fail/Safe N. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

We conducted an outlier analysis for the set of 14 studies in which a pre-post design was adopted. At a 1% confidence level, four separate clusters of studies were found. We conducted an analysis of one cluster with only group interventions (8 studies). An overall mean effect size of 0.42 was found, which is a moderate effect (95% CI = 0.33, 0.51, $z = 9.46$, $p = .000$). The Q test indicated that this was a homogeneous set of studies.

Furthermore, we conducted an outlier analysis of the follow-up meta-analysis. At a 1% confidence level, three clusters were formed. We examined why 4 studies in two clusters differed from the other 10 studies in the third cluster. In 3 studies, very large

effect sizes were found. For the 4th study of a self-directed intervention, no reasons were found to exclude it from the analysis. We conducted an analysis of 11 studies. These results were also significantly heterogeneous, but the amount of explained variance increased to 57.5%. In a next sub analysis, we excluded 3 studies on self-directed variants of Triple P. These studies were outliers because of their very large effect sizes. An overall mean effect size of $d = .37$ was found, which is a moderate effect (95% CI = 0.28, 0.46, $z = 7.95$, $p = .001$). The Q test indicated that this was a homogeneous set of studies.

Because the results of the 6-month follow-ups are significantly heterogeneous, we again excluded three outlier studies. After exclusion of the outliers, an overall effect size of $d = 0.49$ was found, which is a moderate effect (95% CI = 0.31, 0.67, $z = 5.32$, $p = .001$). The Q test indicated that this was a homogeneous set of studies. We conducted several additional meta-analyses to examine whether effects were moderated by age of children (younger than 4 years vs. older), gender of the children (more than 62.6% boys vs. less than 62.6%), self-directed versus practitioner assisted, individual versus other studies, group versus other studies, and behavior problem scores of the children on the ECBI (problems at pre-test in the clinical range vs. nonclinical range). The cut-off scores of the ECBI are ≥ 127 for the Intensity Scale and ≥ 11 for the Problem Scale (Eyberg & Ross, 1978). We again excluded the outliers by cluster analyses with the computer program (Schwarzer, 1989). The results are summarized in Table 3. Studies with less than 62.6% boys were found to have significantly larger long-term effects on behavior problems than those with more than 62.6% boys ($d = 1.08$, 95% CI = 0.62, 1.54 vs. $d = 0.37$, 95% CI = 0.27, 0.46). Furthermore, studies with an initial behavior problem score in the clinical range (initial intensity score ECBI ≥ 127) were found to have significantly larger long-term effects on behavior problems than those with nonclinical behavior problems ($d = 1.08$, 95% CI = 0.62, 1.54 vs. $d = 0.36$, 95% CI = 0.27, 0.46). None of the other moderator variables were significant.

3.5 Discussion

Main Findings

Level 4 of Triple P has moderate to large effects on behavior problems of children that last in follow-up measurements of 6 to 12 months. A large effect size was found at both postintervention ($d = 0.88$) and long-term follow-up assessment of the child behavior ($d = 1.00$). At 6 and 12 months follow-up, overall mean effect sizes of $d = 1.07$ and $d = 0.84$, respectively, were found. Because the results are significantly heterogeneous, subanalyses were conducted. An overall, homogeneous mean effect size of 0.42 was found at postmeasurement (eight studies). At 6 months, an effect size of 0.49 was found (seven studies). These are moderate effects. Few significant moderators were found, indicating that Triple P can be successfully used with a diverse range of families, types of problems, delivery formats, and ages of children. Studies with a higher proportion of girls have larger

Table 3: Results of meta-analyses of Triple P across modalities

	NES	N	D	95% CI	Q	%SE	F/S-K
Effects directly after the intervention							
age < 4 years	4	1.772	0.54	0.30-0.78	5.52	50.43	6.83
age > 4 years	5	520	0.25	0.06-0.44	3.86	100.0	1.31
<62,6% boys	4	1.990	0.39	0.30-0.48	5.30	100.0	3.83
>62,6% boys	5	302	0.63	0.39-0.97	4.51	10.0	10.81
Initial non-clinical behavior problems	6	2.075	0.40	0.31-0.49	8.32	100.0	5.89
Initial clinical behavior problems	3	217	0.68	0.40-0.97	3.01	100.0	7.22
Self Directed	5	348	1.14	0.40-1.89	21.24***	18.71	23.55
Guided	7	2.121	0.44	0.29-0.59	16.10*	86.84	8.38
Long-term effects after 6 and 12 months							
age <4 years	5	1.664	0.65	0.31-0.99	23.45***	32.12	11.24
age >4 years	5	384	0.66	0.15-1.74	6.19	36.72	11.59
<62,6% boys	5	230	1.08	0.62-1.54	4.18	61.19	17.61
>62,6% boys	5	1.786	0.37	0.27-0.46	3.20	100.0	4.13
Initial non-clinical behavior problems	6	1.818	0.36	0.27-0.46	3.22	100.0	4.92
Initial clinical behavior problems	4	230	1.08	0.62-1.54	4.18	61.19	17.61
Self Directed	5	335	1.09	0.55-1.63	13.07*	42.29	22.34
Guided	8	1.831	0.65	0.28-1.01	27.10***	34.98	17.96

N_{ES} = Number of effect sizes; N = number of subjects in the studies; D = overall effect size; CI = confidence interval; Q = Homogeneity Q ; % SE = percentage of the variance accounted for by random sample error; $F/S-K$ = Orwin's Fail/Safe N . Excluded outliers studies are Nos. 2, 6, 8, 9, 10, 14 in Table 1); * $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$.

long-term effect sizes than do studies with fewer girls ($d = 1.08$ vs. $d = 0.37$). In the long term, the effects in the seven studies with scores in the clinical range on behavior problems at the start of the intervention were larger than in the nine studies with lower scores ($d = 0.36$ vs. $d = 1.08$).

Limitations

The present meta-analysis has several limitations. First, the number of participants in several studies was small (in 73% of the randomized studies, 10 to 50 respondents were included). Second, in the long-term analysis, sometimes other studies were used, as in the postintervention analysis. Consequently, a longitudinal comparison of those effect sizes must be conducted with caution. Third, in this meta-analysis, we took the child as the "unit of analysis" because mothers and fathers reported about

the same child. But it would be interesting to analyze both parents separately to see if they report differently. Fourth, because strict methodological criteria for inclusion were conducted, 10 effect studies were not included in this meta-analysis. This meta-analysis has assurance that the synthesis is based on only the best evidence, but its results may summarize only a narrow research domain.

Directions for Future Research

Despite these limitations, this meta-analysis suggests that the Level 4 system of the Triple P intervention is a worthwhile intervention to both prevent and treat behavior problems in children. At the same time, because of the above-cited limitations, further research is necessary. First, it may be useful to conduct more meta-analyses with all other instruments in the studies on Level 4 Triple P, giving us more insight into the effects of Triple P on Parental competences (De Graaf, Speetjens, Smit, & De Wolff, 2008), giving us more insight into differences between mothers and fathers, and enabling the impact of Triple P on parental mental health to be examined. We are also interested in the differences in effect sizes for the different delivery formats, especially in the Self-Help Triple P, because of the promising effects in this meta-analysis. Furthermore, it would be worthwhile to conduct meta-analyses on some other levels of Triple P. A second direction for future research is to conduct more in-depth analyses on the influences of the age and gender of the child on the effects of the Triple P intervention. Because the long-term effects of studies with fewer boys were found to have significantly larger long-term effects on behavior problems than studies with more boys, it would be interesting to conduct more research on the influence of this moderator. Third, it would be interesting to examine whether the observed maintenance effects up to 3 years postintervention occur over a longer period into children adolescence. A fourth suggestion is to tentatively add one or two more randomized trials on Self-Directed Triple P to this meta-analysis. In a cumulative meta-analysis, it can be established whether the Self-Directed Triple P is more effective than the therapist-assisted Triple P interventions.

3.6 Conclusion

This meta-analysis was conducted to assess the effectiveness of Level 4 of the Triple P multilevel intervention system on behavioral and emotional problems of children across different target groups and intervention modalities. This level of intervention is part of a multilevel suite of interventions designed as a public health strategy to promote better parenting. It contains different delivery formats. We were interested in the pooled effect sizes of the measures of disruptive behavior in children directly after the intervention and after 6 and 12 months. The results indicate that the interventions using Level 4 of Triple P improve the behavior of the child, as observed by the parents. Improvements in children's behavior are sustained over time and seem to even improve somewhat in the long term. Because the analyses involved both prevention universal

samples and high-risk samples, the effect sizes are very large for a universally offered public health intervention. The positive effects of Triple P shown in this study seem to support the widespread adoption and implementation of the program in an increasing number of countries in quite diverse cultural contexts around the world.

References

- Alink, L. R. A., Mesman, J., Zeijl, J. van, Stolk, M. N., Juffer, F., Koot, H. M., et al. (2006). The early childhood aggression curve: Development of physical aggression in 10-to-50 month-old children. *Child Development, 77*(4), 945-966.
- Baden, A. D., & Howe, G. W. (1992). Mothers' attributions and expectancies regarding their conduct-disordered children. *Journal of Abnormal Child Psychology, 20*, 467-485.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioural change. *Psychological Review, 84*, 191-215.
- Bandura, A. (1995). *Self-efficacy in changing societies*. New York: Cambridge University Press.
- Biglan, A. (1995). Translating what we know about the context of antisocial behaviour into a lower prevalence of such behaviour. *Journal of Applied Behaviour Analysis, 28*, 479-492.
- Bodenman, G., Cina, A., Ledenmann, T., & Sanders, M. R. (2007). *The efficacy of Positive Parenting Program (Triple P) in improving parenting and child behavior: A comparison with two other treatment conditions*. Manuscript in preparation.
- Connell, S., Sanders, M. R., & Markie-Dadds, C. (1997). Self-directed behavioural family intervention for parents of oppositional children in rural and remote areas. *Behavior Modification, 21*, 379-408.
- Cummings, M. E., Davies, P. T., & Campbell, S. B. (2002). *Developmental psychopathology and family process*. New York: Guilford.
- Emery, R. E. (1982). Interparental conflict and the children of discord and divorce. *Psychological Bulletin, 92*, 310-330.
- Eyberg, S. M., & Pincus, D. (1999). *Eyberg Child Behaviour Inventory and Sutter Eyberg Student Behaviour Inventory-Revised: Professional manual*. Odessa, FL: Psychological Assessment Resources.
- Eyberg, S., & Ross, A. W. (1978). Assessment of child behavior problems: The validation of a new inventory. *Journal of Clinical Child Psychology, 7*, 113-116.
- Forehand, R. L., & Long, N. (1988). Outpatient treatment of the acting out child: Procedures, long term follow-up data, and clinical problems. *Advances in Behaviour Research and Therapy, 10*, 129-177.
- Gallart, S. C., & Matthey, S. (2005). The effectiveness of Group Triple P and the impact of the four telephone contacts. *Behaviour Change, 22*, 71-80.
- Graaf, I. de, Speetjens, P., Smit, F., & Wolff, M. de (2008). *Effectiveness of the*

Positive Parenting Program on parenting styles and parents competences: A meta-analysis. Family Relations, 57, 566

- Grych, J. H., & Fincham, F. D. (1990). Marital conflict and children's adjustment: A cognitive-contextual framework. *Psychological Bulletin, 108*, 267-290.
- Hart, B., & Risley, T. R. (1995). Incidental teaching of language in the preschool. *Journal of Applied Behaviour Analysis, 8*, 411-420.
- Hedges, L., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic press.
- Heinrichs, N., Hahlweg, K., Bertram, H., Kuschel, A., Naumann, S., & Harstick, S. (2005). Die langfristige wirksamkeit eines elterntrainings zur universellen prävention kindlicher verhaltensstörungen: Ergebnisse aus sicht der mütter und väter [The long-term effectiveness of a parenting program for universal prevention of child behavior problems: Results from mothers and fathers]. *Zeitschrift für Klinische Psychologie und Psychotherapie, 35*, 82-91.
- Hoath, F. E., & Sanders, M. R. (2002). A feasibility study of Enhanced Group Triple P-Positive Parenting Program for parents of children with attention deficit hyperactivity disorder. *Behaviour Change, 19*(4), 191-206.
- Hudson, J. L., & Rapee, R. M. (2005). *Current thinking on psychopathology and the family*. Amsterdam: Elsevier.
- Ireland, J. L., Sanders, M. R., & Markie-Dadds, C. (2003). The impact of parent training on marital functioning: A comparison of two group versions of the Triple P-Positive Parenting Program for parents of children with early-onset conduct problems. *Behavioural and Cognitive Psychotherapy, 31*, 127-142.
- Kazdin, A. E. (1987). Treatment of antisocial behaviour in children: Current status and future directions. *Psychological Bulletin, 102*, 187-203.
- Leung, C., Sanders, M. R., Leung, S., Mak, R., & Lau, J. (2003). An outcome evaluation of the implementation of the Triple P-Positive Parenting Program in Hong Kong. *Family Process, 4*(4), 95-108.
- Lipsey, M. W. (2001). *Practical meta-analysis*. London: Sage.
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. London: Sage.
- Markie-Dadds, C., & Sanders, M. R. (2006a). A controlled evaluation of an enhanced self-directed behavioural family intervention for parents of children with conduct problems in rural and remote areas. *Behaviour Change, 23*(1), 55-72.
- Markie-Dadds, C., & Sanders, M. R. (2006b). Self-directed Triple P (Positive Parenting Program) for mothers with children at-risk of developing conduct problems. *Behavioural and Cognitive Psychotherapy, 34*(3), 259-275.
- Martin, A. J., & Sanders, M. R. (2003). Balancing work and family: A controlled evaluation of the Triple P-Positive Parenting Program as a work-site intervention. *Child and Adolescent Mental Health, 8*(4), 161-169.
- McTaggart, P. G., & Sanders, M. R. (2003). The transition to school project: Results from the classroom. *Australian e-Journal for Advancement of Mental Health, 2*(3). Available from the Auseinet Web site, <http://www.auseinet.com/journal/vol2iss3/mctaggart.pdf>.

- Morawska, A., & Sanders, M. R. (2006). Self-administered behavioural family intervention for parents of toddlers: Part I—Efficacy. *Journal of Clinical and Consulting Psychology*, 74(1), 10-19.
- Mrazek, P., & Haggerty, R. J. (1994). *Reducing the risks for mental disorders*. Washington, DC: National Academy Press.
- National Institute of Mental Health. (1998). *Priorities for prevention research at NIMH: A report by the National Advisory Mental Health Council Workgroup on Mental Disorders Prevention Research* (NIH Pub. No. 98-4321). Washington, DC: Government Printing Office.
- Patterson, G. R., Reid, J. B., & Dishion, T. J. (1982). *Antisocial boys*. Eugene, OR: Castalia.
- Risley, T. R., Clark, H. B., & Cataldo, M. F. (1976). Behavioural technology for the normal middle class family. In E. J. Mash, L. A. Hamerlynck, & L. C. Handy. (Eds.), *Behaviour modification and families* (pp. 34-60). New York: Brunner/Mazel.
- Robinson, E. A., Eyberg, S. M., & Ross, A. W. (1980). The standardization of an inventory of child conduct problem behaviours. *Journal of Clinical Child Psychology*, 9, 22-29.
- Rosenthal, R., & Rubin, D. B. (1982). Comparing effect sizes of independent studies. *Psychological Bulletin*, 92, 500-504.
- Rosnow, R. L., & Rosenthal, R. (2002). *Beginning behavioral research: A conceptual primer*. Upper Saddle River, NJ: Prentice Hall.
- Rutter, M. (1985). Family and school influences on behavioural development. *Journal of Child Psychology and Psychiatry*, 26, 349-368.
- Sanders, M. R. (1992). *Every parent: A positive approach to children's behaviour*. Sydney, Australia: Addison Wesley.
- Sanders, M. R. (1996). New directions in behavioural family intervention with children. In T. H. Ollendick & R. J. Prinz (Eds.), *Advances in clinical child psychology* (pp. 283-330). New York: Plenum.
- Sanders, M. R., & Dadds, M. R. (1982). The effects of planned activities and child management training: An analysis of setting generality. *Behaviour Therapy*, 13, 1-11.
- Sanders, M. R., & Dadds, M. R. (1993). *Behavioral family intervention*. Needham Heights, MA: Allyn & Bacon.
- Sanders, M. R., Markie-Dadds, C., Tully, L., & Bor, W. (2000). The Triple P-Positive Parenting program: A comparison of enhanced, standard, and self directed behavioural family intervention for parents of children with early conduct problems. *Journal of Consulting and Clinical Psychology*, 69, 624-640.
- Sanders, M. R., Markie-Dadds, C., & Turner, K. M. T. (1999). *Practitioner's manual for Enhanced Triple P*. Brisbane, Australia: Families International.
- Sanders, M. R., Markie-Dadds, C., & Turner, K. M. T. (2003). Theoretical, scientific and clinical foundations of the Triple P-Positive Parenting Program: A population approach to the promotion of parenting competence. *Parenting Research and Practice Monograph*, 1, 1-21.

- Schwarzer, R. (1989). *Meta-analysis*. Berlin, Germany: Freie Universitat.
- Silverman, D. (2001). *Interpreting qualitative data: Methods for analyzing talk, text and interaction*. London: Sage.
- Taylor, T. K., & Biglan, A. (1998). Behavioural family interventions for improving child-rearing: A review of the literature for clinicians and policy makers. *Clinical Child and Family Psychology, 1*, 41-60.
- Tremblay, R. E., Nagin, D. S., Jean, R. S., Zoccolillo, M., Zelazo, P. D., Boivin, M., et al. (2004). Physical aggression during early childhood: Trajectories and predictors. *Pediatrics, 114*(1), 43-50.
- Turner, K. M. T., Richards, M., & Sanders, M. R. (in press). A randomised clinical trial of a group parent education program for Australian indigenous families. *Journal of Pediatrics and Child Health*.
- 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.
- Webster Stratton, C. (1994). Advancing videotape parent training: A comparison study. *Journal of Consulting and Clinical Psychology, 62*(3), 583-593.
- Webster Stratton, C., & Hammond, M. (1997). Treating children with early-onset conduct problems: A comparison of child and parent training interventions. *Journal of Consulting and Clinical Psychology, 65*, 93-109.
- Yuki Matsumoto, Y., Sofronoff, K., & Sanders, M. R. (2007). *The acceptability and effectiveness of the Triple P program with Japanese parents in a cross cultural context: Results of an efficacy trial*. Manuscript submitted for publication.
- 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, 3*, 1-18.