

## **Effectiveness of the *Triple P* Positive Parenting Program on behavioural problems in children: meta-analysis**

Ireen de Graaf<sup>1</sup>, Paula Speetjens<sup>2</sup>, Filip Smit<sup>3</sup>, Marianne de Wolff<sup>4</sup>, Louis Tavecchio<sup>5</sup>

1. Ireen de Graaf, MSc, is Research Fellow at the Trimbos Institute (the Netherlands Institute of Mental Health and Addiction), Utrecht. She is engaged in several research, development and implementation projects involving interventions for children and adolescents. She coordinates Dutch research in Triple P interventions. She studied Social Sciences and worked in the Youth Care. Her main interests are research in behavior and emotional problems in children and parenting/ child-rearing.

2. Paula Speetjens, MSc, is Research Fellow at the Trimbos-institute, (the Netherlands Institute of Mental Health and Addiction), Utrecht. She studied Child and Family Studies at Leiden University, the Netherlands. For her master thesis, she conducted a meta-analysis on the effectiveness of the parenting program Triple P. This thesis has been awarded the 'Triple P Student Research Award' in March 2007 at the annual 'Helping Families Change' conference in Charleston, USA.

3. Filip Smit, PhD, is Senior Research Fellow at the Trimbos Institute (the Netherlands Institute of Mental Health and Addiction), Utrecht, and at the Department of Clinical Psychology of the VU University, Amsterdam. His key-interests are evidence-based public mental health, psychiatric epidemiology and health economics. He has published over 60 articles.

4. Marianne de Wolff, PhD, is lecturer in the Center for Child and Family studies of Leiden University. She graduated upon a meta-analysis on parental sensitivity and infant attachment. She received the ISED International Article Award (1996), a yearly award for graduate students publishing in major international journals. From 1997- 2000 she worked as a Postdoctoral Fellow on a study on moral development in adolescents. Main interests are meta-analysis, moral development and attachment.

5. Louis Tavecchio studied psychology at the University of Amsterdam, where he received his PhD in 1977. Currently he is professor of Research in Child Care at the University of Amsterdam, where he leads the *National Day Care Study*, an eight year research project funded by the government. His main research interests focus on quality of center daycare and after-school care, parenting/child-rearing, and male gender role development. He has published many dozens of articles in (inter)national scientific journals.

## **Abstract**

The Triple P-Positive Parenting Program is a multi-level parenting program to prevent and offer treatment for severe behavioural, emotional and developmental problems in children. The aim of this meta-analysis was to assess the effectiveness of Triple P level 4 interventions in the management of behavioural 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 behaviour problems in a variety of settings and there are clear deficits in parenting skills. The results indicated that level 4 of Triple P interventions reduced disruptive behaviours in children. These improvements maintained well over time with further improvements in long term follow up. These effects seem to support the widespread adoption and implementation of Triple P that is taking place in an increasing number countries in quite diverse cultural contexts around the world.

**Keywords:** Triple P, parenting program, behavioural problems in children, meta-analysis

**Abbreviations:** Triple P: Positive Parenting Program; ECBI: Eyberg Child Behaviour Questionnaire.

### *The Triple P – Positive Parenting Program*

The Triple P-Positive Parenting Program (Triple P) is a multi-level program to prevent and offer treatment for severe behavioural, 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 behavioural disturbance in children, and parents may have different needs and desires regarding the type, intensity and mode of assistance they require (Sanders, 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 behavioural problems. Level 3 is a four-session intervention, targets children with mild to moderate behavioural difficulties and includes active skills training for parents. Level 4 is an intensive eight to ten session parent training program for children with more severe behavioural 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 behavioural problems. Finally, level 5 is an enhanced behavioural family intervention program for families where parenting difficulties are complicated by other sources of family distress, e.g., marital conflict, parental depression, or high levels of stress (Sanders, 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 behavioural 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 behaviour. Parents are taught a variety of child management skills including providing brief contingent attention following desirable behaviour, how to arrange engaging activities in high-risk situations, and how to use clear calm instructions, logical consequences for misbehaviour, planned ignoring, quiet time (non exclusionary time-out), and time out. 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 generalisation 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 behaviours 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 ten-session program for parents. Group Triple P is an eight-session program conducted in groups of 10-12 parents with four 15 to 30 minute 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 minute telephone consultations.

Level 4 intervention is indicated if the child has multiple behaviour 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 behaviour problems. Self-Directed Triple P is ideal for families 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, 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 en Dadds, 1993; Taylor & Biglan, 1998; Webster-Stratton, & Hammond, 1997). Furthermore the Triple-P program is based on research in child and family behaviour therapy, developmental research on parenting in every day (Risley et al, 1976; Sanders 1992, 1996); 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 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 behavioural family intervention (BFI). There is clear evidence that BFI is beneficial in children with disruptive behaviour 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 intra-subject replication designs and traditional randomised control group designs. There is evidence that Triple P is an effective parenting strategy. Several studies showed that parenting skills training used in Triple P produces predictable decreases in child behaviour 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 behaviour, with oppositional defiant disorder or conduct disorder or reporting concerns about disruptive child behaviour. Finally, the program has also been successfully used for several different family types including two-parent families, single parents, stepfamilies, maternally depressed families, martially discordant families, and families with a child with an intellectual disability (Sanders et al, 2003). In those studies the following variables were measured: child disruptive behaviour, parent-child interaction, parenting style and confidence, parental adjustment (depression, anxiety, stress, self-esteem), parenting conflict and relationship satisfaction.

### *Meta-analysis*

In the current meta-analysis we examine the effectiveness of Triple P interventions in the management of behavioural 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 of 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 well as 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 as respondent characteristics, format, design, 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 behavioural problems of children compared to the control group as measured directly at the end of the intervention. In the second meta-analysis we assessed the degree to which post intervention effects are maintained over time in the intervention group.

In those meta-analyses we hypothesised that behaviour problems of children, aged 2 to 11 years old, decrease after participating in level 4 Triple P intervention, directly after the intervention and after follow-up of six to twelve 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 hypothesised that the efficacy of Triple P is independent on whether the intervention was delivered to individual parents, groups, or in a self-help format. Third, empirical studies have shown that physically aggressive behaviour occurs in children of 1 year old, increases in the second life year, and then tends to decline from the third birthday onward (Alink et al., 2006; Tremblay, 2004). Therefore, we hypothesized that Triple P is more effective when the interventions are given at age two to four compared other older ages. In addition, it is evident that boys exhibit more externalising problems than girls at the age of two and three 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 behavioural 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 hypothesised that Triple P is more effective for children with higher scores on the ECBI, because there is more room for change for the children with higher ECBI-scores.

## Methods

We used three different search methods to identify literature for the meta-analysis. First we searched literature in two electronic databases Medline, 1975- February 2006 and Psychinfo 1975-February 2006. The following keywords were used: 'Triple P' and 'parent' indicating that also words as parenting or parental are 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 on level 4 of Triple P intervention: an intensive parent training program for children with more severe behavioural 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 behaviour in their children aged 2-11 year, 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 post-test data of the intervention and control group (for the first meta-analysis), and pre- 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 with the age above 11 years old, because Triple P has a separate program for teens.

### *Measurement*

To assess the disruptive behaviour of the children The Eyberg Child Behaviour Questionnaire is used (ECBI) (Eyberg & Pincus, 1999). The ECBI is a 36-item measure of parental perceptions of disruptive behaviour in children aged 2-16 years. It provides two measures: frequency of disruptive behaviours (intensity score) rated on 7-point scales; and the number of disruptive behaviours that parents list as problematic (problem score). The ECBI has a high internal consistency for intensity ( $r = 0.95$ ) as for problem score ( $r = 0.94$ ) (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 behaviour problems in children. Other measures used in the studies assessing behaviour problems are the Child Behaviour 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 et al, 2003; Martin & Sanders, 2003). Because in the last two studies the ECBI was also represented, we decided to include only studies in which behaviour 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, fifteen studies met the inclusion criteria. Three studies were excluded of the first meta-analysis because they had no control group, three studies were excluded since these studies were no effectiveness studies, three 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.

Insert Table 1 here

Group Triple P was used as the intervention in nine studies, Standard Triple P in one study and Self-Directed Triple P in six studies. In one study (Sanders et al, 2000) two versions of Triple P, the Self-Directed intervention and the Standard Triple 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. Fourteen studies were randomised clinical trials, one study was a non-random two group concurrent prospective observation design (Zubrick et al, 2005). In nine studies parents reported their child's behaviour in the clinically elevated range at pre-intervention, and in six studies in the non-clinical range. The percentage of boys was 62.6%, averaged across all studies. We divided the studies in less than 62,6% boys and more than 62,6% boys in order to have two comparing groups. Boys were overrepresented in all studies, and the studies would be heterogeneous if we had divided the groups in 50% boys and 50% girls. In seven studies more than 62.6% of the children were boys. In five studies the children were younger than four. Fourteen studies were selected for the first meta-analysis and also fourteen for the second meta-analysis. In four studies follow-up data were presented after twelve months; in eleven studies the follow-up was conducted at four to six months, in one study (Bodenman et al., in press) follow-up measures were taken at both six and twelve 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 standardised mean difference was used (Hedges  $g$ ), because studies with samples of less than 20 tend to show upward bias in their results (Lipsey, 2001). Since studies with a larger sample size provide more reliable estimates of the population mean, effect sizes were weighted by  $1/S_E$ .

In the *first* meta-analysis we were interested in the differences between mean scores of the experimental and the control group at post-measurement. The standardised 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 group respectively at post intervention and post measurement and  $SD_C$  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 with the control group at post measurement.

In the *second* meta-analysis we also calculated the standardised mean difference as  $d = (M_P - M_F) / SD_P$ , 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 standardised health gain score. An effect size of  $g = 0.5$  indicates 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 randomised clinical trial we calculated the standardised 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:  $\Delta(d) = d_E - d_C$ .

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 due to sampling error plus variability in the population of studies (Hedges & Olkin, 1985). The meta-analytical outcomes obtained under a random effects model are said to be conservative in that their 95% CI 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  $\alpha$  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, while 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 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 estimated directly 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 datasets 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% confidence intervals of the respective effect sizes  $d$  were not overlapping, then we considered the contrast to be statistically significant.

Finally, for each meta-analysis the Orwin's Fail/Safe Number was calculated. This number indicates how many (hypothetical) studies with an effect size of zero should be found and included in the meta-analysis in order to reduce the observed effect size to a smaller value of, say,  $d = 0.20$ .

## Results

The overall mean effect size for the fourteen studies of the child behaviour observed by parents at post-measurement was  $d = 0.88$ , with a 95% confidence interval of 0.50 – 1.27 (Table 2). This effect is statistically significant ( $Z = 4.49, p < .001$ ). This is 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 is a substantial amount of unexplained variance in the total set of studies that might be attributed to the systematic effects of covariates. Twenty percent (of the variance is caused by random sample error, which leaves room for a remaining 79.6% which may systematically co-vary with (unknown) covariates. The number of studies with a zero-effect that should be found in order to reduce the effect size to  $d = 0.20$  is 47.8.

The overall mean effect size concerning the long-term measurement of child behaviour was  $d = 1.00$ , with a 95% confidence interval of 0.55-1.46 This effect is statistically significant ( $Z = 4.33, p < .001$ ). This is a large effect. The  $Q$ -test for the hypothesis of homogeneity across effect sizes had to be rejected. 20.54% of the variance is caused by random sample error. The number of studies with a zero-effect that should be found in order to reduce the effect size to 0.20 is 56. At six months follow-up an overall mean effect size of  $d = 1.07$  was found ( $Z = 3.49, p < 0.001$ ). In the meta-analysis of the four studies on the twelve months follow-up studies, we found an overall mean effect size of  $d = 0.84$  ( $Z=2.59, p < 0.001$ ). However, the results are significantly heterogeneous.

Table 2

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 (eight studies). An overall mean effect size of 0.42 was found, which is a moderate effect (95% CI: 0.33-0.51,  $Z=9.46, p=0.000$ ). 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 four studies in two clusters differ from the other ten studies in the third cluster. In three studies very large effect sizes were found. For the fourth study of a self-directed intervention no reasons were found to exclude it from the analysis. We conducted an analysis of eleven studies. These results are also significantly heterogeneous, but the amount of explained variance increased to 57.5%. In a next sub-analysis we excluded three 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  $d = 0.37$  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 six months follow-up 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 the age of children (younger than four years versus older); gender of the child (more than 62.6% boys versus less than 62.6%); self-directed versus practitioner assisted; individual versus other studies; group versus other studies; behaviour problems scores of the children on the ECBI (problems at pre-test in clinical range versus non-clinical range). The cut-off scores of the ECBI are  $\geq 127$  for the intensity scale and  $\geq 11$  for the problem scale (Eyberg & Ross, 1978). We excluded again 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 behaviour problems than those with more than 62.6% boys ( $d = 1.08$ : 95%CI 0.62-1.54 versus  $d=0.37$ ; CI 0.27-0.46). Furthermore studies with an initial

behaviour problems score in the clinical range (initial intensity score ECBI  $\geq 127$ ); were found to have significantly larger long term effects on behaviour problems than non-clinical behaviour problems ( $d = 0.36$ ; 95% CI 0.27-0.46 versus  $d = 1.08$ ; 95% CI 0.62-1.54). None of the other moderator variables were significant.

Table 3

## Discussion

### *Main findings*

Level 4 of Triple P has moderate to large effects on behaviour problems of children which lasted in the follow-up measurements of six to twelve months. A large effect size was found at both the post-intervention ( $d = 0.88$ ) and at long-term follow up assessment of the child behaviour ( $d = 1.00$ ). At six and twelve months follow-up an overall mean effect size of  $d = 1.07$  and  $d = 0.84$  respectively was found. Because the results are significantly heterogeneous, sub-analyses were conducted. An overall, homogeneous, mean effect size of 0.42 was found at post-measurement (eight studies). At six 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 successfully used with a diverse range of families, (e.g., types of problems, delivery formats, age of the children). Studies with a higher proportion of girls have larger long-term effect sizes than studies with fewer girls ( $d = 1.08$  versus  $d = 0.37$ ). In the long term, the effects in the seven studies with scores in the clinical range on behaviour problems at the start of the intervention, were larger than in the nine studies with lower scores ( $d = 0.36$  versus  $d = 1.08$ ).

### *Limitations*

The present meta-analysis has several limitations. First, the number of participants in several studies was small (in 73% of the randomised studies 10 to 50 respondents were included). Second, in the long-term analysis sometimes other studies were used as in the post-intervention 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 report about the same child. But it would be interesting to analyse both parents separately to see if they report differently. Fourth, because strict methodological criteria for inclusion were conducted ten 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.

### *Direction for future research*

Despite these limitations this meta-analysis suggests that the level four system of Triple P intervention is a worthwhile intervention to both prevent and treat for behaviour problems in children. At the same time, due to the above-cited limitations, further research is necessary. First, it may be useful to conduct more meta-analyses on all other instruments and data in the studies on level four of Triple P. It will give us more insight into the effects of Triple P on parental competencies (De Graaf et al., in prep), in differences between mothers and fathers and enable 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. Further, 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 behaviour problems than studies with more boys, it would be interesting to conduct more research on the influence of this moderator.

Thirdly, it would be interesting to examine whether the observed maintenance effects up to three years post intervention occur over a longer period into children adolescence. A fourth suggestion is to add tentatively one or two more randomised trials on Self-Directed Triple P to this meta-analysis. In a cumulative meta-analyse it can be established whether the Self-Directed Triple P is more effective than the therapist assisted Triple P interventions.

## **Conclusion**

This meta-analysis was conducted to assess the effectiveness of level four of the Triple P multilevel intervention system on behavioural 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 size of the measures of disruptive behaviour in children directly after the intervention and after 6 and 12 months.

The results indicate that the interventions on level four of Triple P improve the behaviour of the child, as observed by the parents. Improvements in children's behaviour are sustained over time, and seem even to 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 as shown in this study seem to support the widespread adoption and implementation of the program in an increasing number countries in quite diverse cultural contexts around the world.

## **Acknowledgments**

We would like to thank all researchers in Triple P worldwide for providing us with all the necessary information for conducting the meta-analysis, especially Prof. Dr. M. Sanders.

## **Correspondence to**

I. de Graaf  
Dept. of Prevention  
Trimbos-institute, The Netherlands Institute of Mental Health and Addiction  
PO Box 725, 3500 AS, Utrecht, The Netherlands  
Phone: +31 30 2959243  
Fax: +31 30 2971111  
email: [igraaf@trimbos.nl](mailto:igraaf@trimbos.nl)

## Reference list

- Alink, L. R. A., Mesman, J., Zeijl, J. van, Stolk, M. N., Juffer, F., Koot, H. M., Bakermans-Kranenburg, M. J., & IJzendoorn, M. H., van, (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.
- Chalmers, I., & Altman, D. G. (1995). *Systematic reviews*. London: BMJ Publishing Group.
- 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* (4), 379-408.
- Cummings, M. E., Davies, P. T., and 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 follos-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.
- Glass, G., McGaw, B. & Smith, M.L. (1981). *Meta-analysis in social research*. Beverly Hills: Sage University Paper.

Graaf, de, I., Speetjens, P., Smit, F., Wolff, de, M. Effectiveness of the Positive Parenting Program on parenting styles and parents competences: meta-analysis. Manuscript in preparation.

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: 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*. Manuscript in preparation. *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 CognitivePsychotherapy*, 31, 127-142.

Kazdin, A. E. (1987). Treatment of antisocial behaviour in children: Current status and future directions. *Psychological Bulletin*, 102, 187–203.

Kazdin, A. E. (1998), John R. Weisz. Identifying and Developing Empirically Supported Child and Adolescent Treatments. *Journal of Consulting and Clinical Psychology*, Vol. 66 (1), 19-36.

Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioural treatment. Confirmation from meta-analysis. *American Psychologist*, 48 (12), 1181-1209.

**Lipsey, M. W. (2001).** *Practical meta-analysis. Applied social research methods series*, 49. London: Sage.

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.

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). [www.auseinet.com/journal/vol2iss3/mctaggart.pdf](http://www.auseinet.com/journal/vol2iss3/mctaggart.pdf)

Markie-Dadds, C., & Sanders, M. R. (2006) 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.

Markie-Dadds, C., & Sanders, M. R. (2006) 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

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.

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: National Academy Press.

Mullen, B. (1989). *Advanced BASIC Meta-Analysis*. Hillsdale: Lawrence Erlbaum Associates.

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 Publication No. 98-4321). Washington, DC: U. S. 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.

Roberts, C., Mazzucchelli, T., Studman, L., & Sanders, M. R. (2006). Behavioural family Intervention for children with developmental disabilities and behavioural problems. *Journal of Clinical Child and Adolescent Psychology*, 35 (2), 180-193.

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*. New Jersey: 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, NSW, 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*, 283-330. New York: Plenum Press.

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., & Turner, K. M. T. (1999). *Practitioner's manual for Enhanced Triple P*. Brisbane, Australia: Families International Publishing.

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.

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., & McFarland, M. L. (2000). Treatment of depressed mothers with disruptive children: A controlled evaluation of cognitive behavioral family intervention. *Behavior Therapy* 31 (1), 89-112.

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: Freie Universitat.uitgever

Silverman, D. (2001). *Interpreting qualitative data : methods for analyzing talk, text and interaction*. London: Sage.

Stallman, H. M., Ralph, A., & Sanders, M. R. (2007). *Evaluation of self-directed Teen Triple P: A behavioural family intervention to reduce risk factors for adolescent behavioural and emotional problems*. Manuscript submitted for publication.

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., Pérusse D., Japel C. (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.

Wolf, F. M. (1981). *Meta-analyses, Quantitative methods for research synthesis*. London: Sage University Papers.

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., Robertson, D., & Sanders, M. (2005). Prevention of child behavior problems through universal implementation of a group behavioral family intervention. *Prevention Science, 3*, 1-18.