The Impact of Behavioural Parent Training on Fathers’ Parenting: A Meta-Analysis of the Triple P-Positive Parenting Program

Behavioural parent training programs have been developed to address child behaviour problems through improvement in parenting practices. Ideally, programs would demonstrate effectiveness with all parents. The Triple P-Positive Parenting Program is widely reported as an effective, evidence-based program for parents. However, in this meta-analysis we demonstrate that there are significant differences in program effectiveness for mothers and fathers. We show that while Triple P has a large positive effect on mothers’ parenting practices, it has a smaller effect on fathers’ parenting practices. Considering that fathers make a significant and unique contribution to child development, we argue that it is important for parenting programs to assess their effectiveness with fathers as well as mothers.

Keywords: fathers, parenting programs, Triple P

It is now well accepted that “poor” parenting practices will have significant impacts on children’s emotional and behavioural wellbeing (Amato & Fowler, 2002; DeKlyen, Speltz, & Greenberg, 1998; Muller & Diamond, 1999; Trautmann-Villalba, Gschwendt, Schmidt, & Laucht, 2006; Verlaan & Schwartzman, 2002). Parents’ harsh inconsistent discipline, inadequate supervision, and a lack of a warm, positive relationship have been shown to increase children’s behaviour problems and later lead to substance abuse, antisocial behaviour, and criminal activity (Fergusson, Horwood, & Ridder, 2005). In order to strengthen parenting competencies and improve parent–child interactions a
wide range of Behavioural Parent Training (BPT) programs have been developed. These programs, which are based on social learning principles, have reported improvements in parental perceptions and parenting skills, improvements in children’s social skills and school adjustment, and reductions in children’s behavioural problems (Barlow & Stewart-Brown, 2000; Taylor & Biglan, 1998). With the increasing recognition of the social and economic costs of children’s behavioural disorders, the possibility of making evidence-based programs such as BPT widely available is being suggested as a practical step toward improving childhood mental health and reducing antisocial and self-harming behaviours (Dodge, 2009; Scott, 2010).

However, an important consideration in extending effective treatments to families in need is to discover which segments of the population are currently accessing and benefiting from the programs and then to identify the adaptations needed to make them acceptable to, and effective with, other groups (Dodge, 2009). In the case of programs addressed to parents of children with behaviour problems, several studies have drawn attention to the unequal participation of mothers and fathers. Budd and O’Brien (1982) examined parent training research reported in behaviour journals over a 12-year period to 1981 and found that among 747 participating families only 97 eligible fathers (13%) were involved. When the exercise was repeated with journals reporting outcomes for parent training addressing behaviour disorders to 1988, the 35 studies reported 20% (248/1253) father involvement (Coplin & Houts, 1991).

More recently Taino and McNeil (2005) reviewed the published reports for three major BPT programs: the individually administered Parent–Child Interaction Program, the five-level Triple P-Positive Parenting Program and Webster-Stratton’s videotape modelling program. They found that few studies collected information from fathers and that fewer again analysed and reported fathers’ data independent of data obtained from mothers. They concluded that “the small body of literature on fathers and BPT that exists from 1970 to 2003 either does not specifically examine direct effects of paternal participation in BPT or suggests that fathers do not enhance the effectiveness of BPT” (p. 19). A similar lack of data for evaluating fathers’ responses to BPT has been reported by subsequent reviews of parent training for families with children Attention-Deficit/Hyperactivity Disorder (Fabiano, 2007) and in two meta-analyses of the Triple P-Positive Parenting Program (de Graaf et al., 2008; Nowak & Heinrichs, 2008).

These findings detailing a lack of father involvement and reporting of outcomes for fathers in previous evaluations of BPT are concerning, especially given that it is now widely accepted that fathers’ parenting has an important influence on children’s development (Lamb, 2004). Furthermore, it has been demonstrated that a father’s influence on a child’s development is separate to that of a mothers (Grossman et al., 2002; Lamb; NICHD Early Child Care Research Network, 2008). On the one hand, fathers’ positive parenting (sensitive to children’s interests, supportive of their autonomy) has been identified as protective against externalizing problems in young children; on the other hand negative parenting by fathers (harsh discipline, unstable presence) has been found to increase the behaviour problems of both boys and girls (DeKlyen et al., 1998; Muller & Diamond, 1999; Trautmann-Villalba et al., 2006; Verlaan & Schwartzman, 2009).
In addition, fathers’ participation in a BPT program has been reported to improve maintenance of behaviour change in children and co-parenting (mother and father supporting each other’s parenting) has been found to reduce the development of behaviour problems in vulnerable children (Bagner & Eyberg, 2003; Schoppe-Sullivan et al., 2009).

Taken together, the findings outlined thus far highlight the importance of directly examining the father’s role in BPT programs separately to that of mothers. While several researchers advocate increasing fathers participation (Bagner & Eyberg, 2003; Fabio, 2007; Tiano & McNeil, 2005), so long as fundamental questions, such as the ability of BPT programs to improve fathers’ parenting, remain unanswered support for increasing fathers’ participation will depend solely on clinicians’ enthusiasm (Bagner & Eyberg). However, as noted above, the reporting of fathers’ program attendance and their engagement with the program content in studies evaluating BPT programs is variable, making it difficult to distinguish between the effects of fathers’ low participation and their response to program content. Explanations offered for fathers’ low participation in BPT programs are uniformly speculative and fathers’ response to the BPT programs is commonly dealt with as something for future investigations (Nowak & Heinrichs, 2008; Leung et al., 2003; de Graaf, 2008). This paper attempts to assess the effects of BPT on fathers by reviewing the existing research on the ability of one of the major BPT programs, the Triple P-Positive Parenting Program, to recruit and engage fathers and to improve fathers’ as well as mothers’ parenting practices.

Triple P offers a number of advantages in identifying the possible effects of a BPT program on fathers’ parenting. As the name suggests, the Triple P-Positive Parenting Programs aim to recruit both mothers and fathers and to produce effects in the parenting practices of both parents. The inclusion of fathers in numerous research reports by the developers of the program, references to “Dads” on the Triple P website (http://www.triplep.net) and specific claims of father inclusion in accreditations of Triple P (http://www.commissioningtoolkit.org) underline that fathers are intended to benefit from their participation in the program. The program has been designed in several formats to be delivered at levels of increasing intensity so that the effect of the program overall as well as the impact of different intensities may be assessed. As well, the large number of studies available may allow the testing of some suggested explanations of fathers’ lack of attendance at parenting programs such as their lack of availability when the target population includes many single mothers (Nowak & Heinrichs, 2008; Sanders, 2003). Importantly, the program has been evaluated in numerous randomized controlled trials over the last decade providing the strongest evidence that any outcome differences between mothers and fathers, such as improved parenting, can be attributed to the program and not to some extraneous factor. The aim of this paper is to answer the following questions in regard to publicised randomised controlled studies of the effectiveness of the Triple P-Positive Parenting Program:

- How many fathers compared to mothers were recruited into these studies?
- What was the level of fathers’ involvement in and attrition from the studies?
- What is the impact of the Triple P-Positive Parenting Program on fathers’ parenting?
THE TRIPLE P-POSITIVE PARENTING PROGRAM

In the Australian-based Triple P-Positive Parenting Program Level 1, Universal Triple P, is the least intensive level of Triple P incorporating a community-wide media strategy to facilitate parents’ access to information on dealing with common child-rearing problems. For example, a 13-episode television program on parenting and family survival skills was aired in New Zealand. Level 2 interventions include the Selected (Individual Support) Triple P, whereby primary care providers offer tip sheets and advice on dealing with specific child rearing issues, and the Triple P Seminar Series, which is designed to reach large audiences of parents. Level 3, Primary Care and Primary Care Teen Triple P utilise four brief (20 to 30 minute) behavioural counselling sessions in order to manage identified problems. Level 4 includes Standard Triple P, Stepping Stones Triple P, Group Triple P, Group Teen Triple P, and Self-Directed Triple P programs. Standard and Stepping Stones (developed for parents of children with a disability) Triple P include 10 sessions on causes of children’s misbehaviour and strategies for encouraging their development and managing their behaviour. The Group Triple P programs combine a total of five group sessions and three follow-up telephone sessions. The Self-Directed Triple P program is a 10-week self-help program that may be supplemented by weekly telephone sessions. Level 5, the most intensive level of the Triple P program is included in addition to Standard or Group Triple P and includes Enhanced and Pathways Triple P. Enhanced Triple P covers such areas as home-based skills training, coping with stress and emotions, and marital communication and support. Pathways Triple P was designed for parents deemed at risk of child maltreatment and focuses on areas such as anger management. (For more detailed descriptions see: Sanders, 1999, 2008; Sanders, Cann, & Markie-Dadds, 2003; Sanders, Markie-Dadds, & Turner, 2003; Sanders, Mazzucchelli, & Studman, 2004.)

Triple P has been shown to effectively reduce common child behaviour problems and increase parenting skills with parents from both clinical and community populations. The program has been used with a wide range of specific groups including: depressed mothers, Japanese parents, parents of children with autism, and Indigenous families (Sanders & McFarland, 2000; Matsumoto, Sofronoff & Sanders 2007; Whittingham, Sofronoff, Sheffield & Sanders, 2009a, 2009b; Turner, Richards & Sanders, 2007). The program materials have been adapted for multiple languages and can be delivered via face-to-face, telephone and electronic modalities (Sanders, 2003).

A previous review by de Graaf et al. (2008) showed that Triple P produces positive changes in parenting skills when parent gender is not analysed. Additionally, a review by Nowak and Heinrichs (2008) showed that fathers had consistently lower overall effect sizes across a range of measures, including parenting and child outcomes, compared to both mothers and teachers. Both reviews concluded that a more extensive analysis on the effects of Triple P for fathers compared with mothers is required. This review extends previous research contributions by examining the efficacy of Triple P on improving parenting practices separately for mothers and fathers. Specifically, this paper addresses the question: Does the Triple P-Positive Parenting Program improve the parenting style of both mothers and fathers?
METHOD

Search Strategy

We searched the following sources: MEDLINE (1950 to June 2011), CINAHL (1982 to June 2011), Embase (1950 to June 2011), PsycINFO (1960 to June 2011) and ERIC (1983 to June 2011) using search terms Triple P and randomized/randomised or clinical trials. The randomized controlled trial papers available through the Triple P website (http://www.triplep.net) were also downloaded.

Selection Criteria

Figure 1 outlines the number of included and excluded studies. We identified all trials of Triple P including group, individual, telephone-delivered and self-directed modes, where individuals or groups of parents were randomly allocated to treatment or control. Although previous meta-analyses have included a range of intervention studies (e.g., de Graaf et al., 2008; Nowak & Heinrichs, 2008), we only selected the Randomized Controlled Trials (RCTs) as they are recognised as the most appropriate means for demonstrating that a program is effective (McLennan & Lavis, 2006; Sanders, 2003). Trials were included if fathers were specifically targeted or where fathers were eligible to participate (programs described as for “parents,” for example). Thus, the following analyses are based on the results reported in 28 separate journal articles.

Sample Characteristics

A total of 37 treatment groups were obtained from the 28 journal articles and are outlined in Table 1. Group Triple P was used in 14 studies, Standard Triple P was used in...
<table>
<thead>
<tr>
<th>Study</th>
<th>Type</th>
<th>% Single Parents</th>
<th>Recruitment (no. fathers/total recruited)</th>
<th>Reports</th>
<th>Level of Father Involvement</th>
<th>Overall</th>
<th>Mothers</th>
<th>Fathers</th>
<th>Overall</th>
<th>Mothers</th>
<th>Fathers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bodenmann et al. (2008) Group</td>
<td>Group</td>
<td>✓</td>
<td>(150/300)</td>
<td>✓</td>
<td>95</td>
<td>77</td>
<td>0.98</td>
<td>48</td>
<td>40</td>
<td>0.76</td>
<td>47</td>
</tr>
<tr>
<td>Bor et al. (2002)</td>
<td>Standard</td>
<td>×</td>
<td>(?/87)</td>
<td>✓</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>21</td>
<td>27</td>
<td>0.79</td>
<td>-</td>
</tr>
<tr>
<td>Connell et al. (1997) Self-Directed</td>
<td>8</td>
<td>✓</td>
<td>(24/48)</td>
<td>✓</td>
<td>24</td>
<td>22</td>
<td>1.67</td>
<td>12</td>
<td>11</td>
<td>1.49</td>
<td>12</td>
</tr>
<tr>
<td>de Graaf et al. (2009) Primary Care</td>
<td>16</td>
<td>✓</td>
<td>(7/129)</td>
<td>✓</td>
<td>75</td>
<td>42</td>
<td>0.4</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Gallart &amp; Matthey (2005) Group</td>
<td>-</td>
<td>✓</td>
<td>(3/49)</td>
<td>×</td>
<td>16</td>
<td>16</td>
<td>0.74</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Hahlweg et al. (2008) Group</td>
<td>Standard</td>
<td>✓</td>
<td>(43/112)</td>
<td>✓</td>
<td>50</td>
<td>47</td>
<td>0.79</td>
<td>32</td>
<td>31</td>
<td>1.02</td>
<td>18</td>
</tr>
<tr>
<td>Hahlweg et al. (2010) Group</td>
<td>22</td>
<td>✓</td>
<td>(201/499)</td>
<td>✓</td>
<td>296</td>
<td>118</td>
<td>0.58</td>
<td>155</td>
<td>61</td>
<td>0.71</td>
<td>141</td>
</tr>
<tr>
<td>Heinrichs et al. (2006) Group</td>
<td>13</td>
<td>✓</td>
<td>(200/479)</td>
<td>✓</td>
<td>247</td>
<td>167</td>
<td>0.8</td>
<td>128</td>
<td>88</td>
<td>0.69</td>
<td>119</td>
</tr>
<tr>
<td>Hoath &amp; Sanders (2002) Enhanced</td>
<td>40</td>
<td>×</td>
<td>(?/21)</td>
<td>×</td>
<td>9</td>
<td>11</td>
<td>0.96</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Ireland et al. (2003)* Enhanced</td>
<td>0</td>
<td>✓</td>
<td>(40/88)</td>
<td>✓</td>
<td>32</td>
<td>-</td>
<td>-</td>
<td>16</td>
<td>-</td>
<td>-</td>
<td>16</td>
</tr>
<tr>
<td>Leung et al. (2003) Group</td>
<td>1</td>
<td>✓</td>
<td>(4/91)</td>
<td>✓</td>
<td>33</td>
<td>36</td>
<td>1.02</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Plant &amp; Sanders (2007) Stepping Stones</td>
<td>22</td>
<td>×</td>
<td>(?/74)</td>
<td>×</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>26</td>
<td>24</td>
<td>0.82</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 1 continued on page 297.
<table>
<thead>
<tr>
<th>Study</th>
<th>Type</th>
<th>% Single Parents</th>
<th>Father Recruitment (no. fathers/total recruited)</th>
<th>Reports Level of Father Involvement</th>
<th>Overall</th>
<th>Mothers</th>
<th>Fathers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N_T</td>
<td>N_C</td>
<td>d_{Overall}</td>
<td>N_T</td>
</tr>
<tr>
<td>Roberts et al. (2006)</td>
<td>Stepping Stones</td>
<td>-</td>
<td>✓ (23/47)</td>
<td>✓ 24</td>
<td>26</td>
<td>1.96</td>
<td>14</td>
</tr>
<tr>
<td>Sanders et al. (2000)</td>
<td>Standard</td>
<td>26</td>
<td>✓ (226/529)</td>
<td>✓ 130</td>
<td>142</td>
<td>1.05</td>
<td>65</td>
</tr>
<tr>
<td></td>
<td>Enhanced</td>
<td></td>
<td></td>
<td>116</td>
<td>142</td>
<td>1.29</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>Self-Directed</td>
<td></td>
<td></td>
<td>122</td>
<td>142</td>
<td>0.32</td>
<td>61</td>
</tr>
<tr>
<td>Sanders et al. (2004)*</td>
<td>Enhanced</td>
<td>30</td>
<td>✓ (7/98)</td>
<td>× 35</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Standard</td>
<td></td>
<td></td>
<td>39</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Sanders et al. (2008)*</td>
<td>Enhanced Self-Directed</td>
<td>20</td>
<td>✓ (14/454)</td>
<td>× 76</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Standard</td>
<td></td>
<td></td>
<td>98</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Turner et al. (2007)</td>
<td>Group</td>
<td>30</td>
<td>✓ (3/51)</td>
<td>× 20</td>
<td>18</td>
<td>0.96</td>
<td>-</td>
</tr>
<tr>
<td>Turner &amp; Sanders (2006)</td>
<td>Primary Care</td>
<td>24</td>
<td>✓ (1/30)</td>
<td>× 12</td>
<td>13</td>
<td>0.88</td>
<td>-</td>
</tr>
<tr>
<td>West et al. (2010)</td>
<td>Group Lifestyle</td>
<td>18</td>
<td>✓ (3/101)</td>
<td>× 52</td>
<td>49</td>
<td>1.06</td>
<td>-</td>
</tr>
<tr>
<td>Whittingham et al. (2009c)</td>
<td>Stepping Stones</td>
<td>14</td>
<td>✓ (4/59)</td>
<td>✓ 29</td>
<td>30</td>
<td>2.02</td>
<td>29</td>
</tr>
<tr>
<td>Wiggins et al. (2009)</td>
<td>Pathways</td>
<td>25</td>
<td>✓ (4/60)</td>
<td>× 27</td>
<td>22</td>
<td>1.61</td>
<td>-</td>
</tr>
<tr>
<td>Zubrick et al. (2005)</td>
<td>Group</td>
<td>12</td>
<td>✓ (16/1610)</td>
<td>✓ 691</td>
<td>774</td>
<td>0.62</td>
<td>-</td>
</tr>
</tbody>
</table>

^ The authors of those studies that did not report separate means for mothers and fathers were contacted, but were unable to provide the necessary data to calculate separate effect sizes.

* Participants were randomly allocated to one of two intervention conditions and so there was no standard Waitlist Control Group in this study.
6 studies, Enhanced Triple P was used in 5 studies, and Self-Directed Triple P was used in 5 studies. Stepping Stones Triple P was used in 4 studies, 2 used Primary Care and 1 used Pathways Triple P. All studies were RCTs, with the majority using a waitlist as the control condition; however, 3 studies used the Standard Triple P intervention as the control and as such a comparable effect size could not be calculated for these studies.

Measurement Outcome

Because the principal aim of our study was to determine the impact of Triple P on mothers’ and fathers’ parenting practices, the Parenting Scale (PS; Arnold, O’Leary, Wolff, & Acker, 1993) was chosen as the dependent variable in our analysis. The PS contains 30 items designed to identify dysfunctional parenting practices: laxness, overreactivity, and verbosity. Laxness refers to ways in which parents give in to their children, allow rules to go unenforced, or provide positive consequences for misbehaviour. Overreactivity covers the times parental mistakes occur, such as displays of anger and irritability. Verbosity reflects lengthy verbal responses and a reliance on talking, even when it is ineffective. The PS has been validated for use with both mothers and fathers (Arnold, O’Leary, & Edwards, 1997; Harvey, Danforth, Ulaszek, & Eberhardt, 2001; Prinzie, Ongena, & Hellinckx, 2007; Rhodes & O’Leary, 2007) in a variety of settings including BPT (e.g., Gardner, Burton, & Klimes, 2006; Gross et al., 2003; Hutchings et al., 2007; Lundahl, Risser, & Lovejoy, 2006). The total score has been shown to have good internal consistency ($\alpha = .84$) and reliability ($r = .84$; Arnold et al., 1993). The PS was used in all of the studies included in this analysis to assess the effectiveness of Triple P in improving parenting behaviours.

Data Analysis

Calculation of effect size. Effect sizes (i.e., the standardized difference between the means of two groups; Cohen’s $d$, see Appendix (p. 312) for formulas used in this analysis) and 95% confidence intervals were calculated for each study comparing the post-intervention treatment group Parenting Scale scores to the post-intervention control group scores. This procedure was repeated for those papers that reported separate means and standard deviations for mothers and fathers.

A comparison of the effect sizes indicates the effectiveness of Triple P in improving mothers’ and fathers’ parenting practices. An effect size greater than $d = 0.8$ may be interpreted as large, whereas effect sizes around $d = 0.5$ are medium, and around $d = 0.2$ are considered to be small (Cohen, 1988). The 95% confidence intervals were used to determine significance. When a confidence interval includes zero, it indicates that an effect size is not significantly different from zero (i.e., the treatment has not had an effect). Furthermore, when two effect sizes are being compared, if their 95% confidence intervals do not overlap, then they are thought to be significantly different from each other.

In addition to calculating the mean effect sizes for the studies included in this meta-analysis overall, and separately for mothers and fathers, we also calculated the number

298
of unpublished studies that would need to exist to reduce these existing effect sizes to a “small” effect using Orwin’s (1983) fail-safe N.

Tests of heterogeneity. Meta-analyses inevitably bring together material with an element of diversity since the included studies will differ in design and conduct as well as in participants, interventions, exposures or outcomes studied (Higgins & Thompson, 2002). Such diversity is commonly referred to as heterogeneity. Statistical heterogeneity exists when the true effects being evaluated differ between studies, and may be detectable if the variation between the results of the studies is above that expected by chance (Higgins & Thompson, 2002). In order to determine whether heterogeneity exists in the sampled studies, we performed Q-tests for the studies overall and separately for mothers and fathers. A significant Q-test, whereby \( p < 0.05 \), is indicative of heterogeneity between studies. However, the Q-test alone does not give a measure of the extent of the heterogeneity between studies (Verhulst & Colliver, 2009). Recently, Higgins and Thompson (2002) proposed the \( I^2 \) Index (with 95% confidence intervals), which can be used to determine the magnitude of heterogeneity (from 0% to 100%) between studies in a meta-analysis (Huedo-Medina, Sánchez-Meca, Botella, & Marín-Martínez, 2006).

Since meta-analyses necessarily combine studies that are diverse in their methodology, heterogeneity in their results is to be expected (Higgins, 2008). An advantage of the \( I^2 \) Index is that it can be used to investigate the cause of the heterogeneity by creating methodologically distinct subgroups (Higgins, Thompson, Deeks, & Altman, 2003). In our analysis, findings of heterogeneity were further investigated by grouping studies by Triple P type, in an attempt to locate the cause of the heterogeneity.

**Results**

The number of fathers recruited, the percentage of single parents in the sample, and the reporting of father involvement as well as the effect sizes for each study can be seen in Table 1. Twenty-one of the 28 studies reported separate recruitment rates for mothers and fathers. Out of the 4959 parents recruited into these studies, 983 (20%) were fathers. Fourteen of the studies (50%) also included information on the fathers’ level of involvement in the program. However the manner of reporting varied widely ranging from noting that families were supplied with separate survey forms for fathers to recording the number of sessions attended by fathers (this was stated explicitly in five papers: Bor et al., 2002; Halweg et al., 2008, 2010; Sanders et al., 2000, 2007). Very few studies reported separate attrition information for mothers and fathers. The total number of participants recruited into the 28 sampled studies was 5342, of which 4,493 completed the pre- and post-intervention (overall attrition rate of 16%). For the 12 studies that reported attrition rates for mothers, the attrition rates ranged from 0–28% (Total number of mothers recruited = 2352). For the 10 studies that reported attrition rates for fathers, the attrition rates ranged from 0% - 100% (Total number of fathers recruited = 700).
Twenty-five of the studies included the number of single parents among the participants however the gender of the single parent was recorded in only 6 of these studies. Assuming that all single parents in the reported studies are mothers, one would expect that fewer fathers would be involved in studies reporting higher numbers of single parents. Although the correlation between percentage of single parents and percentage of involved fathers was in the predicted direction, it was not significant ($r = -.38, p = .07$).

The overall mean effect of Triple P on parenting style was $d_{\text{Total}} = 0.77$, with a 95% confidence interval of 0.71-0.84. This indicates that, overall, Triple P has a large, significant, positive effect on improving parenting practices. The number of studies with a zero effect that would need to be found to reduce this effect size to 0.20 is 69.

When calculated separately, it was found that Triple P has a large, significant and positive impact on improving mothers’ parenting, with $d = 0.77$ and a 95% confidence interval of 0.65–0.87. To reduce this effect size to 0.20, 43 studies with a zero effect would have to be found. Triple P had a significantly smaller, but still moderate, positive effect on father’s parenting compared to mothers’ parenting with $d = 0.51$ and a confidence interval of 0.37–0.63. To reduce this effect size to 0.20, 14 studies with a zero effect would have to be found.¹

We also compared effect sizes for older (studies published prior to 2006, $k = 10$) and newer ($k = 14$) studies. The overall mean effect size of Triple P on parenting style was slightly larger for newer ($d = 0.85, 95\% \text{ CI} = 0.74-0.96$) compared to older studies ($d = 0.72, 95\% \text{ CI} = 0.64-0.80$), but this difference was not significant. When calculated separately, the differences in effect sizes for older and newer studies for mothers (Older: $d = 0.79, 95\% \text{ CI} = 0.60-0.98$; Newer: $d = 0.75, 95\% \text{ CI} = 0.59-0.94$) and fathers (Older: $d = 0.48, 95\% \text{ CI} = 0.28-0.68$; Newer: $d = 0.53, 95\% \text{ CI} = 0.35-0.71$) were not significant.

Given that the studies within this analysis cover a range of Triple P formats, we also examined the effect sizes for each format, both overall, and separately for mothers and fathers. As seen in Figure 2, all Triple P formats have a positive impact on parenting practices when data for mothers and fathers is combined. Primary Care, Self-Directed, and Group Triple P all had significantly smaller effect sizes than Stepping Stones, Enhanced and Pathways Triple P.

Mean effect sizes and 95% confidence intervals can be seen separately for mothers and fathers broken down by Triple P format in Figure 3. There were no significant differences between Triple P formats for mothers and all formats showed a significant positive improvement in mothers’ parenting style. For fathers, with the exception of the Self-Directed format, all formats showed a positive impact on fathers’ parenting style. Stepping Stones Triple P had a significantly greater effect on fathers’ parenting practices than Group and Self-Directed Triple P. Additionally, for all Triple P formats, with the exception of Stepping Stones, fathers had a lower mean effect size than mothers.

¹It is important to note that the 14 studies required to reduce this effect to zero may exist within the studies reported within the meta-analysis that did not provide separate data for mothers and fathers.
THE IMPACT OF BPT ON FATHERING

Figure 2. Mean effect sizes with 95% confidence intervals for each type of Triple P.

Figure 3. Mean effect sizes with 95% confidence for each type of Triple P for mothers (squares) and fathers (circles) separately.
Tests of heterogeneity were performed for the overall data, and separately for mothers and fathers. The results are displayed in Table 2. There was significant heterogeneity for the overall analysis and for mothers. The $I^2$ Index suggested that a moderate to large percentage of variance between studies was due to heterogeneity.

Table 2

<table>
<thead>
<tr>
<th></th>
<th>k</th>
<th>Q</th>
<th>$p$</th>
<th>$I^2$ Index</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Overall</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Sample</td>
<td>24</td>
<td>99.0</td>
<td>&lt;.001</td>
<td>77%</td>
<td>66% - 84%</td>
</tr>
<tr>
<td>Standard Triple P</td>
<td>2</td>
<td>1.1</td>
<td>&gt;.05</td>
<td>0%</td>
<td>-</td>
</tr>
<tr>
<td>Enhanced Triple P</td>
<td>2</td>
<td>0.4</td>
<td>&gt;.05</td>
<td>0%</td>
<td>-</td>
</tr>
<tr>
<td>Group Triple P</td>
<td>13</td>
<td>19.5</td>
<td>&gt;.05</td>
<td>38%</td>
<td>0% - 68%</td>
</tr>
<tr>
<td>Self-Directed Triple P</td>
<td>2</td>
<td>13.7</td>
<td>&lt;.001</td>
<td>93%</td>
<td>76% - 98%</td>
</tr>
<tr>
<td>Stepping Stones Triple P</td>
<td>2</td>
<td>0</td>
<td>&gt;.05</td>
<td>0%</td>
<td>-</td>
</tr>
<tr>
<td>Primary Care Triple P</td>
<td>2</td>
<td>1.1</td>
<td>&gt;.05</td>
<td>0%</td>
<td>-</td>
</tr>
<tr>
<td>Pathways Triple P</td>
<td>1</td>
<td>-</td>
<td></td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Mothers</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Sample</td>
<td>15</td>
<td>31.2</td>
<td>&lt;.01</td>
<td>55%</td>
<td>20% - 75%</td>
</tr>
<tr>
<td>Standard Triple P</td>
<td>3</td>
<td>0.4</td>
<td>&gt;.05</td>
<td>0%</td>
<td>-</td>
</tr>
<tr>
<td>Enhanced Triple P</td>
<td>2</td>
<td>0</td>
<td>&gt;.05</td>
<td>0%</td>
<td>-</td>
</tr>
<tr>
<td>Group Triple P</td>
<td>3</td>
<td>0.1</td>
<td>&gt;.05</td>
<td>0%</td>
<td>-</td>
</tr>
<tr>
<td>Self-Directed Triple P</td>
<td>4</td>
<td>19.0</td>
<td>&lt;.001</td>
<td>84%</td>
<td>60% - 94%</td>
</tr>
<tr>
<td>Stepping Stones Triple P</td>
<td>3</td>
<td>2.3</td>
<td>&gt;.05</td>
<td>0%</td>
<td>-</td>
</tr>
<tr>
<td>Primary Care Triple P</td>
<td>0</td>
<td>-</td>
<td></td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Pathways Triple P</td>
<td>0</td>
<td>-</td>
<td></td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Fathers</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Sample</td>
<td>9</td>
<td>11.3</td>
<td>&gt;.05</td>
<td>29%</td>
<td>0% - 67%</td>
</tr>
<tr>
<td>Standard Triple P</td>
<td>2</td>
<td>0.4</td>
<td>&gt;.05</td>
<td>0%</td>
<td>-</td>
</tr>
<tr>
<td>Enhanced Triple P</td>
<td>1</td>
<td>-</td>
<td></td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Group Triple P</td>
<td>3</td>
<td>0.7</td>
<td>&gt;.05</td>
<td>0%</td>
<td>-</td>
</tr>
<tr>
<td>Self-Directed Triple P</td>
<td>2</td>
<td>1.4</td>
<td>&gt;.05</td>
<td>0%</td>
<td>-</td>
</tr>
<tr>
<td>Stepping Stones Triple P</td>
<td>1</td>
<td>-</td>
<td></td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Primary Care Triple P</td>
<td>0</td>
<td>-</td>
<td></td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Pathways Triple P</td>
<td>0</td>
<td>-</td>
<td></td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Note. The Q-statistic is used to determine whether or not the sample is homogenous. The $I^2$ Index, with its 95% confidence interval, gives an indication of the percentage of total variation across studies due to heterogeneity. Details for calculating each statistic can be found under the heading Data Analysis.
In order to determine the cause of the heterogeneity, $Q$-tests were performed, and the $I^2$ Index was calculated for each Triple P format for the overall data, and separately for mothers and fathers. For mothers and fathers, homogeneity was found in the Standard, Enhanced, Group, and Stepping Stones Triple P formats, while for fathers, homogeneity was also found for the Self-Directed format. The reduction in the $I^2$ Index compared with the total sample supports the use of these subgroups. However, significant heterogeneity was still found for mothers in the Self-Directed subgroup.

**DISCUSSION**

*General Effectiveness of the Triple P-Positive Parenting Program*

A large positive effect of Triple P was found in this meta-analysis, which suggests that the program is effective at improving parenting practices as measured by the self-report PS. This finding confirms previously published reports of decreases in dysfunctional parenting practices upon completion of a Triple P intervention (de Graaf et al., 2008; Nowak & Heinrichs, 2008). However, this result does not take into account the gender of the participating parent. When effect sizes were calculated separately for mothers and fathers it was found that Triple P has a significantly greater effect on improving mothers’ parenting practices than fathers’ parenting practices.3

As numerous reviews have pointed out, the majority of those attending parenting classes and programs are mothers and the research evidence underpinning the program content overwhelmingly derives from studies of mothers and their children (Fabiano, 2007; Ghate, Shaw & Hazel, 2000; Tiano & McNeil, 2005). So it would not be surprising to find that some aspects of program content and delivery were unsuited to fathers. Although no systematic adaptations of behavioural parenting programs to make them more suitable for fathers attending with mothers have been reported, several examples of father-targeted parenting programs have been described and guidelines for program providers have been proposed. Three key recommendations from programs in the UK, Australia and the USA have been identified. Firstly, an appreciation of a father’s concern with his role as provider as well as parent should be recognised and programs for low income fathers should incorporate work-related training. Secondly, programs that recognise fathers’ preference for active learning can boost engagement and learning, especially via sport-related activities with their children. Thirdly, having males as co-facilitators of mixed groups or solo leaders for fathers groups can make programs more acceptable to fathers (Fabiano, 2007; O’Brien & Rich, 2002; Pearson, 2000; Potter, 2008; Raikes, Summers, & Roggman, 2005).

---

2 Note that there were not enough studies to perform tests of heterogeneity for fathers for the Enhanced or Stepping Stones Triple P format.

3 It should be noted that, although the PS has been validated for use with fathers, a study by Rhoades and O’Leary (2007) found that mothers reported significantly higher levels of overactivity than fathers, suggesting that the PS has the potential to be more sensitive to changes in mothers’ than fathers’ parenting practices.
Additionally, it has also been suggested that the poor treatment effects for fathers reported in BPT program studies may be explained by factors to do with the fathers, such as poor motivation and lack of program contact (Fabiano, 2007; Nowak & Heinrichs, 2008; Stemmler, Beelman, Jaursch, & Losel, 2007; Treacy, 2005). Explanations in terms of fathers’ motivation point out that many fathers have relatively limited contact with their children and they may interact with them in settings where behaviour problems are less apparent (e.g., outdoor play). As a result, fathers may not perceive their child’s problems as serious and so may not be motivated to improve their parenting to address their child’s behaviours (Nowak & Heinrichs; Stemmler et al.), rather, they attend to “support the mother” (Treacy, Tripp, & Baird, 2005). Although fathers are reported to score lower than mothers on assessments of problematic behaviours (Connell, Sanders, & Markie-Dads, 1997; Fabiano), several studies have also reported concordance between mothers and fathers’ rating of their child’s behaviour problems (Ireland, Sanders, & Markie-Dads, 2003; Sanders, Bor, & Morawksa, 2007; Sanders, Markie-Dadds, Tully, & Bor, 2000). It may be that fathers recognize the problems as serious and important but see the mother as the person mainly responsible for their child’s behaviours. Whatever their intention in enrolling in the program, fathers seem to attend fewer sessions and complete fewer assessments than mothers. On current evidence, there is no way to determine if the father’s lack of program contact is a consequence of low motivation or is the result of other factors that act as a deterrent, reducing satisfaction and interest in the program.

Effectiveness Across Triple P Formats

Given that Triple P is available in a wide range of formats, and a number of these formats were contained within the sampled studies, we also performed separate analyses for each of the program formats collapsed across parent gender. Overall, the effect sizes ranged from small through to large, with the Primary Care and Self-Directed formats showing the smallest improvement in self-reported parenting practices, and the Stepping Stones format showing the largest improvement.

When mother and father data were examined separately, there was sufficient data available to compare effect sizes across five Triple P formats (Enhanced, Group, Self-Directed, Standard, and Stepping Stones). Importantly, the heterogeneity tests supported our decision to look at effect sizes separately for mothers and fathers across the different Triple P formats. The calculation of the $I^2$ Index assisted us in quantifying the extent of the heterogeneity in our analysis as well as to localise its cause. The usual response to finding heterogeneity in a meta-analysis is to subgroup the studies in the hopes that these subgroups will show homogeneity. However, subgrouping the studies necessarily results in a loss of statistical power, which may contribute to the finding of homogeneity within the subgroups. The $F$ Index gives a measure of the extent of heterogeneity that is not influenced by sample size (Higgins et al., 2003). In this analysis, evidence of both homogeneity and a reduction in the $F$ Index were found when stud-
ies were grouped by parent gender and Triple P format, suggesting true homogeneity, rather than a loss of statistical power.4

Mothers showed significant moderate to very large increases in their positive parenting practices across all Triple P formats. Fathers, on the other hand, showed much smaller improvements in their parenting practices, with the exception of the one study that used the Stepping Stones format. These findings show that all formats of Triple P are very effective for mothers, but are not as effective for fathers.

The finding that the Stepping Stones format (that is, in the one study that reported fathers’ outcomes) produced a large positive change in fathers’ parenting suggests that there are elements within Triple P which are effective with some fathers. Fathers’ higher rates of involvement with the care of children who have a disability have been reported in some studies and might account for the effectiveness of the Stepping Stones format with these fathers (Gray, 2003; Simmerman, Blacher, & Baker, 2001). Equally, the course content and style may be factors. Unfortunately, because so few studies report details of fathers’ attendance or response to BPT programs, explaining the success of just a few studies is difficult.

**Implications for Further Research**

The findings from this meta-analysis raise an important issue in terms of the reporting of research findings, especially because the use of the term “parents” to describe samples with varying numbers of fathers is common (Budd & O’Brien, 1982; Coplin & Houts, 1991; Fabiano, 2007; Tiano & McNeil, 2005). Where different outcomes are found for mothers and fathers, care must be taken not to unintentionally obscure the findings by the use of generic terms such as “parents” and “parenting.” For example, in the study reported by Markie-Dadds and Sanders (2006), where no effect was found for Triple P on fathers’ parenting style, the conclusion that “with respect to parenting skills and parental efficacy … the ESD [Enhanced Self-Directed] condition led to significantly more positive outcomes” (p. 68) may mislead readers into believing that the program was effective for fathers as well as mothers.

Given the relevance of fathers’ parenting to children’s development the underrepresentation of fathers across studies evaluating BPT programs (approximately 21% of intervention group parents in the studies reported here were fathers) should be the subject of investigation and remediation. Studies assessing predictors of attendance at parenting programs (Dumka, Garza, Roosa, & Stoerzinger, 1997; Fox & Gottfredson, 2003; Garvey et al., 2006; Heinrichs, Bertram, Kuschel, & Hahlweg, 2005) and measures of engagement which promise refinement of the participation measures used in

---

4 One notable exception was that the Self-Directed Triple P format showed high levels of significant heterogeneity for mothers. This is not surprising as the lack of supervision in a self-directed intervention makes it difficult to ensure that all participants are interpreting and following the directions in the same manner. This finding suggests that the Self-Directed Triple P format may not be well suited to research studies due to the difficulty in ensuring consistency of practice between participants.
evaluating programs (Nix et al., 1999) are now appearing in the literature. The widespread implementation of Triple P across regions and parent groups can provide a fertile arena for testing the effectiveness of new approaches to recruiting fathers. Recommendations such as providing an operational definition of the fathers based on their rate of involvement with the child, and recruitment strategies specifically addressed to fathers (Bagner & Eyberg, 2003; Matthey & Barnett, 1999; Matthey, Reay, & Fletcher, 2009) could be tested. Tailoring Triple P content to reflect fathers’ perspectives, which has been found to engage fathers’ interest (Fabiano, 2007; Fletcher, Vimpani, & Russell, 2008) could be trialled along with variations to timing of classes, father-only groups and matched trainer-participant groups across divergent geographic and cultural settings. Studies looking at a wider range of parent and child outcome measures are also needed to clarify whether Triple P produces different effects on the basis of gender.

**CONCLUSIONS**

The results from this meta-analysis have a number of implications for the interpretation of previous studies reporting the efficacy of Triple P for parents and for research evaluating parenting programs in other areas. While there is ample evidence that Triple P reliably improves behaviour outcomes for children (e.g., Leung et al., 2003, Matsumoto et al., 2007; Roberts et al., 2006) and we have shown in this meta-analysis that all formats of the Triple P-Positive Parenting Program improve mothers’ parenting skills, the implication that all forms of the program are equally efficacious for fathers’ parenting is not warranted. The findings from this review suggest that evaluations of parenting programs should not assume that mothers and fathers will benefit equally. Further, studies should record fathers’ recruitment, engagement and outcomes separately to those for mothers. The lack of fathers’ engagement with programs intended for parents could then be the subject of specific adaptations and research. Given the widespread acceptance of BPT programs such as Triple P and the potential benefits when both mothers and fathers are involved, BPT programs present an ideal vehicle for researching effective strategies to engage fathers in parenting programs order to benefit all family members.

**REFERENCES**

(* denotes a study included in the meta-analysis)


308


APPENDIX

Calculation of Effect Sizes

Effect sizes (i.e., Cohen’s d) were calculated for each study comparing the post-intervention outcomes for the treatment (T) and control (C) groups using the following formula:

\[ d = (M_T - M_C)/SD_{pooled} \]

where \( d \) is the effect size, \( M_T \) is the mean total PS score for the treatment group, \( M_C \) is the mean total PS score for the control group, and \( SD_{pooled} \) is the pooled standard deviation given by:

\[ SD_{pooled} = \sqrt{(n_T - 1)s_T^2 + (n_C - 1)s_C^2}{n_T+n_C} \]

This procedure was then repeated for those papers that reported separate means and standard deviations for mothers and fathers.

Calculation of Orwin’s Fail-Safe N

\[ N_{fs} = N_e(d_e - d_c) \]

where \( N_e \) is the number of studies used to calculate the mean observed effect size \( d_e \), \( d_c \) is the criterion value of \( d \), and \( d_{0} \) is the mean effect size of the fail-safe studies (often assumed to be zero).

Tests of Heterogeneity: The Q Statistic

The Q statistic follows a Chi-square distribution, and is given by:

\[ Q = \sum w(d_{total} - \bar{d})^2 \]

where \( w \) is the inverse variance weight as given by:

\[ w = \frac{2n_Tn_C(n_T + n_C)}{2(n_T + n_C)^2 + n_Tn_Cd_{total}^2} \]

and \( \bar{d} \) is the weighted mean of \( d_{total} \) as given by:

\[ \bar{d} = \frac{\sum wd_{total}}{\sum w} \]

Tests of Heterogeneity: The I^2 index

\( \hat{I} \) is derived from \( Q \) and is calculated as:

\[ I^2 = \frac{Q - (k - 1)}{Q} \times 100\% \]

where \( Q \) is larger than the number of studies in the meta-analysis \( (k) \) minus 1. Otherwise, \( \hat{I} \) defaults to 0%.

The 95% confidence intervals for \( \hat{I} \) are derived from the 95% confidence intervals for \( H \), an alternative measure of heterogeneity.\(^3\) \( H \), like \( \hat{I} \) is also derived from \( Q \):

\[ H^2 = \frac{Q}{k - 1} \]

The 95% confidence intervals around \( H \) are then calculated using:

\[ e^{\text{ln}(H) \pm 1.96 \times SE[\text{ln}(H)]} \]

where the standard error for the natural log of \( H \) is given by:

\[ SE[\text{ln}(H)] = \frac{1}{2} \times \frac{\ln(Q) - \ln(k - 1)}{\sqrt{2Q} - \sqrt{(2k - 3)}} \]

when \( Q \) is greater than \( k \). Otherwise it is given by:

\[ SE[\text{ln}(H)] = \frac{1}{2(k - 2)} \left( 1 - \frac{1}{3(k - 2)^2} \right) \]

These confidence intervals for \( H \) are then transformed into 95% confidence intervals for \( \hat{I} \), using the following formula:

\[ I^2 = \frac{H^2 - 1}{H^2} \times 100 \]