

# Prevention of Child Behavior Problems Through Universal Implementation of a Group Behavioral Family Intervention

Stephen R. Zubrick,<sup>1,2,4</sup> Kristine A. Ward,<sup>2</sup> Sven R. Silburn,<sup>1,2</sup> David Lawrence,<sup>1,2</sup>  
Anwen A. Williams,<sup>2</sup> Eve Blair,<sup>2</sup> Deborah Robertson,<sup>2</sup> and Matthew R. Sanders<sup>3</sup>

Published online: 14 September 2005

The aim of this mental health promotion initiative was to evaluate the effectiveness of a universally delivered group behavioral family intervention (BFI) in preventing behavior problems in children. This study investigates the transferability of an efficacious clinical program to a universal prevention intervention delivered through child and community health services targeting parents of preschoolers within a metropolitan health region. A quasi-experimental two-group (BFI,  $n = 804$  vs. Comparison group,  $n = 806$ ) longitudinal design followed preschool aged children and their parents over a 2-year period. BFI was associated with significant reductions in parent-reported levels of dysfunctional parenting and parent-reported levels of child behavior problems. Effect sizes on child behavior problems ranged from large (.83) to moderate (.47). Positive and significant effects were also observed in parent mental health, marital adjustment, and levels of child rearing conflict. Findings are discussed with respect to their implication for significant population reductions in child behavior problems as well as the pragmatic challenges for prevention science in encouraging both the evaluation and uptake of preventive initiatives in real world settings.

**KEY WORDS:** prevention; conduct problems; parent training.

Approximately 11–15% of children under 13 years of age, and 13–17% of young people aged 14–18 years experience significant mental health problems (Sawyer *et al.*, 2000; Silburn *et al.*, 1996; Zubrick *et al.*, 1995, 1997). These findings and others which highlight the global burden of mental health problems (Murray & Lopez, 1996), have moved Australian health authorities to promote an evidence-based action plan for a comprehensive population approach to promotion, prevention, and early intervention in mental health (Commonwealth Department of Health and Aged Care, 2000). A central focus of this action plan has been the preven-

tion of serious childhood conduct problems and the implementation of evidence-based early intervention programs (Marshall & Watt, 1999; National Crime Prevention, 1999).

Attempts to prevent serious conduct problems need to address the quality of parenting children receive and their family relationships. There is clear evidence linking parenting and family risk factors to the development of conduct problems. Specifically, the lack of a warm, positive relationship with parents; insecure attachment; harsh, inflexible, rigid, or inconsistent discipline practices; inadequate supervision of and involvement with children; marital conflict and breakdown; and parental psychopathology (particularly maternal depression and high levels of parenting stress) increase the risk that children develop major behavioral and emotional problems, including conduct problems, substance abuse, antisocial behavior, and participation in delinquent activities (Coie, 1996; Loeber & Farrington, 1998; Patterson, 1982).

Of the interventions targeting parenting variables, behavioral family interventions (BFI) based

<sup>1</sup>Centre for Developmental Health, Curtin University of Technology, Perth, Australia.

<sup>2</sup>Telethon Institute for Child Health Research, Perth, Australia.

<sup>3</sup>School of Psychology, The University of Queensland, Brisbane, Australia.

<sup>4</sup>Correspondence should be directed to Stephen R. Zubrick Telethon Institute for Child Health Research, 855, West Perth, 6872, Western Australia, Australia; e-mail: S.Zubrick@curtin.edu.au.

on social learning models (Patterson, 1982) have sufficient empirical support to warrant consideration for broader population level application. BFI's are the most thoroughly evaluated interventions available to assist children with conduct problems (Brestan & Eyberg, 1998; Lochman, 1990; Sanders *et al.*, 1996a; Taylor & Biglan, 1998). Typically, parents are taught to increase positive interactions with children and to reduce coercive and inconsistent parenting practices. These interventions often produce positive changes in parental perceptions and parenting behaviors as well as changes in child behaviors (Barlow & Stewart-Brown, 2000; Webster-Stratton, 1998). These programs are associated with large effect sizes (Serketich & Dumas, 1996), often generalize to a variety of home and community settings (McNeil *et al.*, 1991; Sanders & Dadds, 1982), and are maintained over time (Long *et al.*, 1994). They are also associated with high levels of consumer satisfaction (Webster-Stratton, 1989).

For an intervention to be considered suitable for a population-level implementation several conditions need to be met (Sanders *et al.*, 2000). These include having (1) evidence available about the prevalence of the target problem (child behavior) and the targeted risk factors (parenting variables); (2) evidence that shows that modifying the risk factors are associated with improvements in the targeted problem; (3) an available, effective, and culturally appropriate intervention that can be readily disseminated, and (4) a delivery mechanism to enable the program to be widely implemented in the community. Importantly, the implementation of a universal parenting strategy is based on the rationale that while the association of many risk factors for problems such as child conduct problems may be quite "weak" (i.e., small odds ratios) a large proportion of the child population is exposed to these risks. Because dysfunctional parenting falls along a continuum much of which can be addressed by preventive interventions, interventions that modify the population exposure may result in a significant reduction in poor outcome despite modest effects at the individual level (Doll, 1996; Rose, 1995).

While there have been a small number of effectiveness trials of BFI within regular clinical services (Dishion *et al.*, 2002; Spoth & Redmond, 2002; Stormshak *et al.*, 2002) there have been no large-scale evaluations of a universal parenting intervention within a population health framework delivered through primary health care. The Triple-P Program is one example of a BFI that has extensive empirical

evidence supporting its efficacy (see Sanders, 1999 for a review of this evidence). It was designed to reduce the use of dysfunctional parenting methods; increase the use of positive parenting behaviors; reduce parental depression, anxiety, and stress; and reduce the general level of marital conflict associated with raising children. These exposures are hypothesized to be part of the causal pathway leading to serious behavior problems (Commonwealth Department of Health and Aged Care, 2000).

The present study extends the literature on the prevention of serious conduct problems by evaluating the effects of a large-scale, universally accessible, population-level application in a primary health care setting of the Triple P-Positive Parenting Program (Sanders, 1999; Sanders *et al.*, 1996a, 2000; Turner *et al.*, 1998). We report here the immediate, 1-year and 2-year outcomes for 804 children whose parent(s) participated in a regionally based universal program of group Triple-P. This is the largest evaluation of the effectiveness of Triple-P to date and the first to assess the population-based delivery of BFI within a general health services context.

## METHOD

As this research was conducted as an effectiveness trial in the context of regular health service delivery there were certain constraints on the type of evaluation that was possible. The program evaluation was funded and commissioned by the Western Australian Department of Health. The intention of the evaluation was to inform the funder about the effectiveness of universal delivery of a behavioral family intervention through community and child health services. The research methodology was constrained by the requirements of contract stipulated by the funder. The Department of Health was interested in assessing program outcomes, process, and transferability. This included assessing the effectiveness of participant recruitment and retention, mode of delivery and uptake through primary care, and the potential for program sustainability using existing services. The research design and methodology was selected on the basis of the funder's requirement for a large-scale population level intervention using core health services in a community setting.

The requirement of universal opportunity of access to BFI within a socioeconomically deprived region for a programme conducted through existing services precluded evaluation by randomized

controlled trial with subject randomization at the individual level within the region. A quasi-experimental design was therefore adopted, obtaining a control group by similar means from another region with access to similar health care services and serially evaluating key targeted child behavior problems, parenting practices, parental adjustment, and consumer satisfaction, using parent report measures that could be used routinely within existing health services in all regions. Therefore, this evaluation of Triple-P regionwide was principally aimed at assessing the *effectiveness* of the program under “real life” conditions of delivery.

We hypothesized that immediately postintervention and again at 1- and 2-year follow-up, compared with parents of the comparison group, parents participating in Triple-P would report: (1) lower levels of dysfunctional parenting; (2) lower levels of child behavior problems in their children; (3) lower levels of depression, anxiety and stress, parent conflict over child rearing, and marital dissatisfaction.

#### Participants and Recruitment

Parents in the intervention group were recruited from the Eastern Metropolitan Health Region of Western Australia. This region was selected because it has a higher proportion of families in receipt of Family Crisis Program benefits and higher rates of child abuse notifications relative to other metropolitan health regions within the state. This region also had a high proportion of preschool aged children, and a population growth rate above 2%.

Program promotion and program recruitment occurred through local media, professional referral, and later through participant recommendation and advocacy. Program promotion employed a variety of strategies including: posters, letters, and registration brochures to schools, preschools, kindergartens, daycare and family centers, doctors surgeries, health clinics, recreation centers, and clubs. Other strategies included recruiting at preschool registration days, radio advertisements and features in community newspapers, posters and displays in shopping centres, direct referral and personalized invitations from health, education, and welfare professionals. Program registration, follow-up, and reminder telephone calls confirmed parent commitment and childcare needs. Both the program and childcare were available at no cost and the program was offered at times during the day that suited most parents. All parents

living within the region were encouraged to participate if they had a child within the age range 3–4 years.

The comparison group comprised parents from the South Metropolitan Health Region. While the population growth rate, the proportion of families receiving Family Crisis Program benefits, and the rate of child abuse notifications within this region were higher than the state average, they were nonetheless lower than those of the intervention region which had the highest proportions and rates of any of the metropolitan regions. Parents were invited to participate in a health services survey of child behavior and were chiefly recruited from daycare centers, by direct invitation of health professionals and at enrolment days for preschool. Research staff were in regular telephone and postal contact with these families over the 2-year period and developed a good knowledge of the participating families. Considerable attention was paid to monitoring changes of address, seeking returns of questionnaires, and requesting additional information where returns were incomplete.

#### Measures

Data were gathered principally from one parent—usually the mother—or the father if he was the sole parent.

#### *Family Background and Demographic Details*

Parents completed a standardized questionnaire about their level of education, family structure, employment status and occupation, income, family support payments, and use of health and mental health services (Zubrick *et al.*, 1995). Parents also answered questions about their child’s vision and hearing, any physical disabilities or chronic conditions, speech development, use of medication, and use of mental health care services. These details were requested at each assessment point and updated as they changed.

#### *Eyberg Child Behavior Inventory (ECBI)*

The ECBI measures parental perceptions of disruptive behavior in children aged 2–16 years (Eyberg & Pincus, 1999). It incorporates a measure of frequency of disruptive behaviors (Intensity score) and a measure of the number of disruptive behaviors that are a problem for parents (Problem score).

The ECBI has high internal consistency for both the Intensity ( $r = .95$ ) and Problem ( $r = .94$ ) scores and good test-retest reliability ( $r = .86$ ). The ECBI allows categorization of clinical status based on the Intensity score ( $\geq 127$ ) or Problem score ( $\geq 11$ ) (Eyberg & Ross, 1978). For the purposes of this report we have used the continuous ECBI Intensity score. There are extensive data attesting to its utility in clinical and population samples (Burns *et al.*, 1991; Burns & Patterson, 1990).

#### *Parenting Scale (PS)*

This 30-item questionnaire measures three dysfunctional discipline styles: Laxness (permissive discipline); Overreactivity (authoritarian discipline, displays of anger, meanness, and irritability); and Verbosity (overly long reprimands or reliance on talking). The PS Total score (range = 1–7) increases with increasingly dysfunctional parenting, has good internal consistency ( $\alpha = .84$ ), good test-retest reliability ( $r = .84$ ), and reliably discriminates between parents of clinical and nonclinical children where scores in excess of 3.1 denote “clinical” levels of dysfunctional parenting (Arnold *et al.*, 1993).

#### *Parent Problem Checklist (PPC)*

The PPC measures conflict between partners over child rearing, rating parents’ ability to cooperate and work together in family management, and was therefore not administered to sole parents. Items explore the extent to which parents disagree over rules and discipline for child misbehavior. Items rate the occurrence of open conflict over childrearing issues and the extent to which parents undermine each other’s relationship with their children. High scores signify greater interparental conflict in areas of child rearing with scores  $\geq 5$  being in the clinical range. The PPC has a moderately high internal consistency ( $\alpha = .70$ ) and high test-retest reliability ( $r = .90$ ) (Dadds & Powell, 1991).

#### *Abbreviated Dyadic Adjustment Scale (ADAS)*

The ADAS (Sharpley & Rogers, 1984) is an abbreviated, 7-item version of the 32-item Spanier Dyadic Adjustment Scale (Spanier, 1976). The total score reliably distinguishes between distressed

and nondistressed couples on relationship satisfaction drawing upon aspects of communication, intimacy, cohesion, and disagreement. Higher scores represent better relationship adjustment. The measure is moderately reliable ( $\alpha = .76$ ), has an item total correlation of .57, and interitem correlations of .34–.71 (Sharpley & Rogers, 1984).

#### *Depression Anxiety Stress Scales (DASS)*

The DASS assesses symptoms of depression, anxiety, and stress in adults. The scale has high reliability for the Depression ( $\alpha = .91$ ), Anxiety ( $\alpha = .81$ ), and Stress ( $\alpha = .89$ ) scales, and good discriminant and concurrent validity (Lovibond & Lovibond, 1995a,b).

#### *Client Satisfaction*

Client satisfaction was administered to the intervention group immediately after the 8-week exposure to Triple-P. Fifteen questions were used to assess satisfaction with the intervention content, format, and materials and one question required an overall rating to the statement: “Overall, I would rate this program...” where a rating of 1 corresponded to “Poor” and a rating of 5 corresponded to “Excellent.”

#### *Intervention*

Group Triple-P is described extensively elsewhere (Sanders *et al.*, 2000; Turner *et al.*, 1998). Briefly, enrolled parents participated in a 2-hr training workshop in groups of about 10 parents (representing on average about 8 children), once a week for 4 weeks, followed by a 15-min telephone support session once a week for 4 weeks. Each family received a copy of the text, “Every Parent” (Sanders, 1992), “Every Parent’s Workbook for Groups” (Markie-Dadds *et al.*, 1997), and a video to support their participation in the program (Sanders *et al.*, 1996b). These educational resources also served as a self-directed package for use at home with partners who did not attend.

The program involved teaching parents core child management strategies, designed to promote children’s competence and development, and to help parents manage misbehavior. These strategies covered three key areas: (1) strategies to promote

children's development, (2) strategies to manage misbehavior, and (3) strategies that involved planned activities and routines. Parents were taught to apply parenting skills to a broad range of target behaviors in both home and community settings with the target child and all relevant siblings. By working through the exercises in their workbook, parents learned to set and monitor their own goals for behavior change and enhance their skills in observing their child's and their own behavior.

The program facilitators were 16 community and child health nurses, social workers, health promotion officers, and psychologists that were recruited from community and child health services within the health region. All facilitators were required to attend a 3-day intensive training program in behavioral family intervention, and commit to cofacilitate a minimum of three programs with an experienced facilitator, within 12 months. In all, 101 Triple-P groups were conducted over 18 months.

Several factors ensured the program was delivered to meet the standardized Group Level 4 training and the Triple-P curriculum. Program fidelity was supported by a detailed manual of the 8-session curriculum, highly structured training, and the use of performance criteria to assess integrity of learning to ensure consistency of program delivery. A clinical psychologist was employed part-time as case manager for the project. A checklist of the contents covered and key learnings was monitored and forwarded to the evaluation team at the conclusion of each session. Trainers were paired together with experienced facilitators at random to deliver each program. This supported transfer of learning and expertise. Regular debriefing sessions were conducted following each program cycle to ensure program implementation issues and concerns were addressed.

### Comparison Group

Participants in the comparison region were able to access health care and family support services as usual, but did not participate in Group Triple-P.

### Design

A quasi-experimental two-group longitudinal design was employed. Participants in the intervention group were asked to complete questionnaires on four occasions: prior to the delivery of the program (preintervention), approximately 9 weeks later immediately postprogram, and then at 12, and

24 months following the postprogram assessment. Participants in the comparison group were asked to complete identical questionnaires on enrolment in the study and approximately 9 weeks later and subsequently at 12 and 24 months following enrolment. All participation in the evaluation was voluntary and written informed consent was obtained from the participating parents.

### Statistical Analysis

Initial differences in distributions of categorized demographic characteristics between the two groups were assessed using chi-squared analyses, as were the effects of attrition on the composition of each group. Statistical significance was accepted at .05 or less.

To assess changes in the Parenting Scale (PS), the Eyberg Intensity Score (ECBI), and carer variables (PPC, ADAD, and DAS) over time and assess the differences within and between the intervention and comparison groups we have used linear mixed modelling (SAS Institute Inc, 2000). This method was selected because repeated measurements relating to the same subject will be correlated over time.

Mixed linear models, or hierarchical linear models as they are often called, relax the standard assumption of ordinary modelling techniques that all data points are independent. This is achieved by fitting a correlation structure to the data. For this analysis we fitted separate correlation matrices for the intervention and control groups to allow for the fact that the groups come from different health regions and for the possibility that any intervention effects could effect the correlation structure within the intervention group. Apart from handling the correlated data structure, SAS PROC MIXED has the advantage that it allows a variable pattern of missing data and can simultaneously adjust for the effects of several covariates. We have included variables that were distributed differently between the two groups at baseline as covariates in the model. This allows the effect of the intervention to be separated from the effects of time and of the initial differences between the intervention and comparison groups.

## RESULTS

### Sample Characteristics

Table 1 shows the family characteristics of the 804 parents that were enrolled in Triple-P. The

Table 1. Characteristics of Participating Parents<sup>a,b</sup>

Variable	Intervention group ( <i>n</i> = 804)		Comparison group ( <i>n</i> = 806)		$\chi^2$ or ( <i>t</i> ) <sup>c</sup>	Prob
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
<i>Demographic variables</i>						
<i>Child's age in months</i>	43.9	7.5	45.6	6.5	(4.6)	<0.0001
	%		%			
<i>Gender of child (male)</i>	58.7		54.7		2.6	0.106
<i>Type of household</i>						
Original family	82.2		83.7		7.99	0.018
Step/blended	4.2		6.3			
Sole parent	13.6		9.9			
<i>Parent's education</i>						
<Year 10	5.4		5.3		5.42	0.144
Year 10	32.3		36.4			
Year 11	17.9		19.5			
Year 12	44.0		38.8			
<i>Marital status</i>						
First marriage	72.2		73.6		3.23	0.358
De Facto	8.7		9.5			
Remarried	5.6		6.2			
Single	13.5		10.7			
<i>Parent's qualifications</i>						
No postschool	37.9		45.2		10.5 <sup>a</sup>	0.015
Trade/certificate	26.7		26.5			
Diploma/degree or higher	8.7		19.5			
Other	25.0		8.8			
Not stated	1.6		0			
<i>Family income</i>						
<\$20,000	13.6		13.3		0.925 <sup>a</sup>	0.921
\$20,001– 30,000	16.9		18.5			
\$30,001–40,000	20.4		23.2			
\$40,001–60,000	29.4		32.0			
\$60,000 and over	11.4		11.7			
Not stated	8.3		1.4			
<i>Carer response variables</i>						
<i>Parenting scale (total)</i>						
In clinical range	61.0		43.1		52.0	<0.0001
<i>Parental depression</i>						
In clinical range	16.2		7.1		32.5	<0.0001
<i>Parental anxiety</i>						
In clinical range	7.6		5.6		2.63	0.105
<i>Parental stress</i>						
In clinical range	18.0		8.2		34.3	<0.0001
<i>Interparental conflict</i>						
In clinical range	44.9		32.4		24.0	<0.0001
<i>Marital satisfaction</i>						
In clinical range	10.0		8.8		0.65	0.420
<i>Child response variable (ECBI)</i>						
Intensity score in clinical range	41.5		21.5		74.7	<0.0001
Laxness in clinical range	39.6		56.7		47.0	<0.0001
Verbosity in clinical range	69.5		69.4		0.00	0.96
Overreactivity in clinical range	40.6		58.9		54.1	<0.0001
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
ECBI intensity score	121.6	27.7	107.1	26.5	(–10.7)	<0.0001

<sup>a</sup>“not stated” category not included in  $\chi^2$  test.

<sup>b</sup>Eyberg Child Behavior Inventory.

<sup>c</sup>*df* = 1608.

children in these families were on average 43.9 months ( $SD = 7.5$ ,  $R = 27.0\text{--}66.1$ ) and more were male (58.7%). The majority of families were original (i.e., biological or adoptive) two-parent families (82.2%) with the remaining families being step/blended (4.2%) or sole parent (13.6%). Seventy-two percent of the participating parents were in their first marriage. Five percent (5.4%) of primary caregivers reported having less than the current mandated minimum of 10 years of education and 13.6% of parents reported a family income below \$A20,000 per year.

Not all of the children whose parents enrolled for the program came from the original target postcode areas. Available census data estimated that there were approximately 706 eligible children (aged 36–48 months inclusive) in the postcode areas targeted for recruitment (Australian Bureau of Statistics, 1997). Of the 804 children whose parents enrolled for the program, 464 (66%) children came from these postcode areas. The remaining 346 children had parents who were recruited from adjacent postcode areas in the same region.

With respect to child behavior, the mean parent-reported Eyberg Intensity score for children of participating parents was 121.6 ( $SD = 27.7$ ) with 41.5% of these children being in the Eyberg clinical range ( $\geq 127$ ). Sixty-one percent of caregivers reported Parenting Scale scores in the clinical range (i.e.,  $\geq 3.1$ ). A high rate of interparental conflict (PPC  $\geq 5$ ) about child rearing was reported (44.9%) as well as a high level of parental stress (18.0%).

We compared the 804 parents who enrolled in the Triple-P sessions with the 806 parents in the comparison group to assess the differences in these groups at the outset of the study and to determine potential confounding factors (i.e., covariates) that would need to be modelled in interpreting any changes in the ECBI intensity score over time (Table 1).

Children in the comparison group were significantly older ( $M_{\text{compar}} = 45.6$  months,  $R = 31.5\text{--}67.4$ ,  $SD = 6.5$  vs.  $M_{\text{inter}} = 43.9$  months,  $R = 27.0\text{--}66.1$ ,  $SD = 7.5$ ;  $t_{1608} = 4.6$ ;  $p < .0001$ ) and more likely to be from step/blended families (6.3 vs. 4.2%,  $\chi^2 = 7.99$ ,  $df = 2$ ,  $p = .018$ ) and have mothers with no postschool qualifications (45.2% vs. 37.9%,  $\chi^2 = 10.5$ ,  $df = 3$ ,  $p = .015$ ) than children in the intervention group. Children in the comparison group were also more likely to have entered the trial with much lower levels of child behavior problems. The comparison group had a significantly lower mean Eyberg intensity score on entry to the trial ( $M_{\text{compar}} = 107.1$ ,

$SD = 26.5$  vs.  $M_{\text{inter}} = 121.6$ ,  $SD = 27.7$ ,  $t_{1608} = -10.7$ ;  $p < .0001$ ). Expressed as a measure of clinical severity, the comparison group was significantly less likely to be in the clinical range on entry to the trial (21.5 vs. 41.5%,  $\chi^2 = 74.7$ ,  $df = 1$ ,  $p < .0001$ ). These initial differences between the comparison and intervention groups required that these variables be included as covariates in the data analysis.

Finally, in keeping with these differences, comparison group caregivers were less likely to score in the clinical range of the Parenting Scale (43.1 vs. 61.0%,  $\chi^2 = 52.0$ ,  $df = 1$ ,  $p < .0001$ ), and had lower levels of depression (7.1 vs. 16.2%,  $\chi^2 = 32.5$ ,  $df = 1$ ,  $p < .0001$ ), lower levels of stress (8.2 vs. 18.0%,  $\chi^2 = 24.0$ ,  $df = 1$ ,  $p < .0001$ ), and lower levels of parental conflict about child rearing (32.4 vs. 44.9%,  $\chi^2 = 24.0$ ,  $df = 1$ ,  $p < .0001$ ).

### Study Retention and Program Attendance

Study retention rates for both the intervention and comparison groups were measured at each follow-up period. Of the 804 parents in the intervention group 691 (86.0%) provided immediate posttest follow-up data, with 650 (80.8%) and 587 (73.0%) providing 12- and 24-month follow-up data, respectively. Of the 806 parents in the comparison group 774 (96.0%) provided immediate posttest follow-up data, with 758 (94.0%) and 691 (85.7%) providing 12- and 24-month follow-up data, respectively.

The intervention entailed a total dose of 9hr, which involved four 2-hr workshops plus another hour in the form of four 15-min telephone follow-up contacts. Measures of program attendance showed that, of the 804 participating parents, 803 (99.8%) completed the first workshop, 718 (89.4%) the second, 692 (86.1%) the third, and 658 (81.8%) the fourth workshop. Success in contacting participants with the four follow-up phone-calls ranged from 601 (74.8%) to 640 (79.7%) parents. In summary, parents received on average 7.8 hr (1.9  $SD$ ) of total program exposure.

Study retention for both the comparison and intervention groups as well as the program attendance rates for the intervention group were very high and merit some comment. Study retention was maintained through regular contact with the comparison group and intervention group families. Follow-up details included residential and e-mail addresses, as well as residential and work telephone numbers. Careful recording of parent names and name changes allowed searches of the electoral roll in the event

contact was lost. Contact details were also recorded for a nonresident friend or relative who was knowledgeable of the enrolled family's whereabouts. In addition, the name and address of the family's general practitioner was noted. These details permitted contact tracing of families where unnotified moves occurred during the study.

Program participation was very high. This reflects the considerable effort that was expended on initial social marketing and focus group research by the funder in order to determine community attitudes toward parenting, tolerance of and interest in parenting programs, and assessing those program features that parents thought would be essential. These features were reflected in promotional materials that were widely disseminated. These materials were nonstigmatizing, promoted child development in the context of opportunities to meet with other parents and obtain structured information, and attend at times suitable to family life and routine. The provision of free crèche facilities was a critical component allowing parents to have their child cared for while they attended the program. Finally, staff were enthusiastic and encouraged parents to attend and, once enrolled, promoted the need to complete the program.

We compared the characteristics of the participating parents who received less than 7 hr of Triple-P ( $N = 157$ ) with those who received 7 or more hours ( $N = 647$ ). Relative to those parents with high levels of participation, those parents with lower levels of participation were significantly more likely to be in step and blended families (8.3 vs. 3.2%) or sole parent families (25.5 vs. 10.7%) ( $\chi^2 = 34.1$ ,  $df = 2$ ,  $p < .0001$ ), have less than Year 10 schooling (12.7 vs. 3.6%,  $\chi^2 = 23.9$ ,  $df = 3$ ,  $p < .0001$ ), earn less than \$20,000 per year in family income (17.8 vs. 12.5%,  $\chi^2 = 13.6$ ,  $df = 5$ ,  $p < .009$ ), and were more likely to be single (25.6 vs. 10.5%) or in a defacto relationship ( $\chi^2 = 34.2$ ,  $df = 3$ ,  $p < .0001$ ) rather than in a first marriage. Parents with lower levels of Triple-P participation were also more likely to report significantly higher levels of depression (25.5 vs. 13.9%,  $\chi^2 = 12.5$ ,  $df = 1$ ,  $p < .001$ ), anxiety (13.4 vs. 6.2%,  $\chi^2 = 9.32$ ,  $df = 1$ ,  $p < .002$ ), and stress (25.5 vs. 16.2%,  $\chi^2 = 7.31$ ,  $df = 1$ ,  $p < .007$ ). There were no statistically significant differences in the proportion of children with clinical elevation in their ECBI scores (46.2 vs. 40.3%) or in mean ECBI scores ( $M_{low} = 124.1$ ,  $SD = 31.0$  vs.  $M_{high} = 121.0$ ,  $SD = 26.8$ ,  $t_{802} = -1.26$ ;  $p < .208$ ) for parents with lower versus higher levels of Triple-P participation.

In 683 of the families (85%) only mothers attended, for another 11 families (1.4%) only the fathers attended, for another 86 families (10.7%) the male partner (either the biological father of the child, or a stepfather) also attended either three or four group sessions, while in the remaining 24 families (3%) the male partner attended one or two sessions. Where both partners attended, information was collected from both mothers and fathers, however in the data analysis, only information from the mothers was used.

## Parenting and Child Behavior Outcomes

### Parenting Behavior

The total score on the Parenting Scale (PS) (Arnold *et al.*, 1993) was used to measure parenting behavior, and fitted in a linear mixed model. Table 2 contains the model estimates ( $B$ 's), standard errors, associated 95% confidence intervals, and estimated effects of the intervention effects on both the Parenting Total Score and the scores for verbosity, laxness, and overreactivity components.

The immediate effect of the intervention was an improvement in the adjusted mean Parenting Scale total score by an estimated 0.62 points (95%  $CI = 0.57, 0.67$ ). This effect is calculated by taking the pretest mean PS Total Score of the intervention group adjusted for child's age, parent's education, and family type and income, subtracting the immediate posttest adjusted mean PS score of the intervention group, and adding the adjusted mean group  $\times$  time effect for the comparison group shown in Table 2 [i.e.,  $(3.24 - 2.57) + (-0.046) = 0.624$ ]. At 12 and 24 months postintervention, this improvement in parenting style, while not as large, was still significant with decreases in the adjusted mean PS score of 0.34 (95%  $CI = 0.29, 0.39$ ) and 0.32 (95%  $CI = 0.27, 0.38$ ), respectively.

We were able to independently estimate a mean (2.84) and standard deviation (0.58) on the PS from a local random sample of 4-year-old children ( $N = 1,431$ ) (D. M. Lawrence, personal communication, May 6, 2004). This permitted calculating effect sizes of 1.08, 0.59, and 0.56 for each of the immediate, 12-, and 24-month postintervention periods. This corresponds to large and moderate effect sizes (Cohen, 1988).

Finally, Table 2 also shows that the changes in the PS total score could be attributed equally to



Table 2. Summary of Linear Mixed Model Estimates for the Parenting Scale (PS) Total Score ( $N = 1,610$ )

	<i>B</i>	<i>SE</i>	95% <i>CI</i>	
Intercept	2.920	0.106	2.174, 3.128	
<b>Covariates</b>				
<i>Child's age</i> at preintervention	0.002	0.001	-0.002, 0.005	
<b>Family type</b>				
Original <sup>a</sup>	0	—	—	
Step/blended family	-0.062	0.059	-0.178, 0.054	
Sole parent family	-0.049	-0.049	-0.148, 0.049	
<b>Education</b>				
Grade 12 or less	0.003	0.049	-0.092, 0.099	
Trade qualification	-0.013	0.051	-0.113, 0.088	
Diploma/Degree <sup>a</sup>	0	—	—	
Other qualification	-0.176	0.053	-0.279, -0.072	
Not stated	-0.064	0.171	-0.400, 0.272	
<b>Family income</b>				
Not stated	-0.041	0.076	-0.189, 0.109	
<\$20,001	<b>0.174</b>	<b>0.059</b>	<b>0.057, 0.290</b>	
20,001–30,000	<b>0.108</b>	<b>0.050</b>	<b>0.009, 0.207</b>	
30,001–40,000	0.008	0.048	-0.086, 0.101	
40,001–60,000	-0.031	0.045	-0.120, 0.058	
+60,000	0	—	—	
<b>Group</b>				
Comparison pretest <sup>a</sup>	0	—	—	
Comparison posttest	<b>-0.046</b>	<b>0.014</b>	<b>-0.734, -0.019</b>	
Comparison 12 months	<b>-0.097</b>	<b>0.015</b>	<b>-0.126, -0.068</b>	
Comparison 24 months	<b>-0.108</b>	<b>0.017</b>	<b>-0.143, -0.074</b>	
Intervention pretest	<b>0.273</b>	<b>0.029</b>	<b>0.214, 0.331</b>	
Intervention posttest	<b>-0.395</b>	<b>0.032</b>	<b>-0.458, -0.332</b>	
Intervention 12 months	<b>-0.169</b>	<b>0.032</b>	<b>-0.224, -0.097</b>	
Intervention 24 months	<b>-0.161</b>	<b>0.032</b>	<b>-0.224, -0.097</b>	
<b>Estimated intervention effects<sup>b</sup></b>				
	Reduction in PS total score (95% <i>CI</i> )	Reduction in laxness total score (95% <i>CI</i> )	Reduction in verbosity total score (95% <i>CI</i> )	Reduction in overreactivity total score (95% <i>CI</i> )
$T_1$ Posttest	<b>.624 (0.57, 0.67)</b>	<b>0.60 (0.53–0.67)</b>	<b>0.70 (0.62–0.78)</b>	<b>0.69 (0.62–0.76)</b>
$T_2$ 12 months	<b>.344 (0.29, 0.39)</b>	<b>0.40 (0.33–0.47)</b>	<b>0.35 (0.26–0.43)</b>	<b>0.34 (0.29–0.41)</b>
$T_3$ 24 months	<b>.325 (0.27, 0.38)</b>	<b>0.38 (0.30–0.46)</b>	<b>0.29 (0.20–0.38)</b>	<b>0.33 (0.25–0.40)</b>

Note. Bolded entries are significant.

<sup>a</sup>Reference category.

<sup>b</sup>In addition to the PS total score, the effects for each of the subscores on laxness, verbosity, and overreactivity are also provided.

changes in all three of the subscales (i.e. Laxness, Verbosity, and Over-reactivity).

In summary, parent-reported levels of dysfunctional parenting behavior declined in the comparison group as their children grew older. Adjusting for this effect in the intervention group, and simultaneously adjusting for the pre-test differences between the groups, parent-reported dysfunctional parenting behavior in the intervention group showed a significant decline immediately post-intervention. This effect attenuated over time, but remained significantly below the pre-test level and below comparison group levels at 12- and 24-month post-intervention.

### Child Behavior

The response variable was the continuous ECBI intensity score and was fitted in a linear mixed model using the approach described above. Model estimates, their standard errors, 95% confidence intervals, and estimated intervention effects are presented in Table 3.

The immediate effect of the intervention was an improvement in parent-reported child behavior as measured by the decrease in adjusted mean ECBI by an estimated 22.4 points (95% *CI* = 20.38, 24.48). At 12- and 24-month

Table 3. Summary of Linear Mixed Model Estimates for the Eyberg Child Behavior Inventory Intensity (ECBI) Score ( $N = 1,610$ )

	<i>B</i>	<i>SE</i>	95% <i>CI</i>
Intercept	112.90	4.617	103.85, 121.96
<i>Covariates</i>			
<i>Child's age</i>			
at preintervention	-0.148	0.084	-0.313, 0.016
<i>Family type</i>			
Original <sup>a</sup>	0	—	—
Step/blended family	<b>8.055</b>	<b>2.620</b>	<b>2.915, 13.195</b>
Sole parent family	<b>5.765</b>	<b>2.156</b>	<b>1.537, 9.994</b>
<i>Family income (\$A)</i>			
Not stated	-3.429	3.275	-9.853, 2.993
<20,001	3.649	2.571	-1.395, 8.693
20,001–30,000	<b>5.431</b>	<b>2.179</b>	<b>1.157, 9.705</b>
30,001–40,000	-1.081	2.080	-5.162, 2.999
40,001–60,000	-1.467	1.967	-5.325, 2.390
+60,000 <sup>a</sup>	0	—	—
<i>Education</i>			
Grade 12 or less	-0.171	2.123	-4.335, 3.992
Trade qualification	0.0178	2.220	-4.337, 4.373
Diploma/Degree <sup>a</sup>	0	—	—
Other qualification	-3.824	2.277	-8.291, 0.642
Not stated	11.173	7.459	-3.458, 25.803
<i>Group</i>			
Comparison pretest <sup>a</sup>	0	—	—
Comparison post-test	0.649	0.612	-0.550, 1.849
Comparison 12 months	<b>-5.413</b>	<b>0.740</b>	<b>-6.863, -3.962</b>
Comparison 24 months	<b>-7.664</b>	<b>0.821</b>	<b>-9.275, -6.054</b>
Intervention pretest	<b>14.488</b>	<b>1.366</b>	<b>11.810, 17.166</b>
Intervention posttest	<b>-7.296</b>	<b>1.338</b>	<b>-9.919, -4.672</b>
Intervention 12 months	-2.214	1.361	-4.882, 0.453
Intervention 24 months	<b>-6.097</b>	<b>1.356</b>	<b>-8.756, -3.439</b>
<i>Estimated intervention effect</i>			
	Reduction in ECBI total score (95% <i>CI</i> )		
- $T_1$ Posttest	<b>22.43 (20.38, 24.48)</b>		
- $T_2$ 12 months	<b>11.29 (9.08, 13.49)</b>		
- $T_3$ 24 months	<b>12.92 (10.44, 15.39)</b>		

Note. Bolded entries are significant at  $p < .05$ .

<sup>a</sup>Reference category.

postintervention, this improvement attenuated but was still statistically significant with decreases in the adjusted mean ECBI score of 11.3 (95%  $CI = 9.1, 13.5$ ) and 12.9 (95%  $CI = 10.4, 15.4$ ), respectively.

Using these mean changes in the ECBI score and noting that the ECBI has a standard deviation of 27.0 points (Burns & Patterson, 1990), then the immediate impact of the intervention was to improve child behavior by .83 of a standard deviation which corresponds to a large effect size (Cohen, 1988, pp. 25–26). At 12 months this improvement had diminished to .41 and at 24 months it was .47 of a standard deviation corresponding to a medium effect size.

### Effects on Other Parental Outcomes

For each of the immediate postprogram, 12- and 24-month self-report results, linear mixed models were fitted for the parental outcome variables in the same way as for our analysis of the PS and ECBI intensity scores. For economy of space only the estimated intervention effects with their 95% confidence intervals are presented in Table 4.

### Effects on Caregiver Depression, Anxiety, and Stress (DAS)

Adjusted mean DAS scores declined by 7.2 points (95%  $CI = 5.7, 8.7$ ) immediately

Table 4. Estimated Intervention Effects (95% Confidence Interval)<sup>a</sup>

	Total score (95-CI)		
	Depression, Anxiety and Stress (DASS)	Parent Problem Checklist (PPC)	Abbreviated Dyadic Adjustment Scale (ADAS)
<i>T</i> <sub>2</sub> Post-test	7.20 (5.72–8.68)	3.49 (2.43–4.55)	–1.01 (–1.36–0.66)
<i>T</i> <sub>3</sub> 12 months	5.49 (3.91–7.06)	2.32 (1.13–3.53)	–0.74 (–1.16–0.32)
<i>T</i> <sub>4</sub> 24 months	4.41 (2.78–6.04)	3.34 (2.01–4.68)	–0.73 (–1.21–0.25)

<sup>a</sup>Adjusted for child-s age at preintervention, family type, education, family income, and age trend from comparison group.

postintervention. At 12- and 24-month postintervention this improvement attenuated but remained significant with decreases in the adjusted mean DAS score of 5.5 (95% *CI* = 3.9, 7.1) and 4.4 (95% *CI* = 2.8, 6.0), respectively.

Using norms available from a large ( $N = 1,771$ ) general adult British population (Crawford & Henry, 2003) where the total DAS score had a mean of 18.38 and standard deviation of 18.82, results in small effect sizes of 0.38, 0.29 and 0.23 in the immediate, 12- and 24-month post-intervention periods. Overall, there was a small but significant improvement in parent-rated mental health, as measured by the DAS. While this effect declines over time, it is still significant at 24 months post-intervention.

#### *Effects on Conflict Between Partners Over Child-rearing (PPC)*

The adjusted mean PPC score decreased by 3.5 points (95% *CI* = 2.4, 4.5). This improvement diminished non-significantly at 12 months to 2.3 points (95% *CI* = 1.1, 3.5) and returned to 3.3 points (95% *CI* = 2.0, 4.7) at 24 months (Table 4). Good quality population norms for the PPC have not been published. Dadds and Powell (1991) report a mean of 2.59 and a standard deviation of 2.40 among a nonclinic sample of mothers of boys, and a mean of 2.86 and a standard deviation of 2.69 among a nonclinical sample of mothers of girls. However, we chose to independently calculate a PPC mean of 5.3 and standard deviation of 3.7 based on a random sample ( $N = 1,694$ ) of 2–3-year-old children (D. M. Lawrence, personal communication, May 6, 2004). Based on this, we estimated effect sizes of 0.95, 0.62, and 0.89 corresponding to moderate to large effect sizes. In general these findings suggest that the intervention significantly decreased the level

of parent-reported conflict over childrearing in the immediate, 12-, and 24-month time periods.

#### *Effects on Quality of Marital Dyadic Relationship Adjustment (ADAS)*

Mean ADAS scores improved (decreased) immediately (–1.01, 95% *CI* = –1.4, –0.6), and at 12 months (–0.74, 95% *CI* = –1.2, –0.3) and 24 months (–0.73, 95% *CI* = –1.2, –0.3) after the intervention. We again used data from a local random sample of 3-year-old children ( $N = 1,568$ ) to estimate a population mean and standard deviation for the ADAS (D. M. Lawrence, personal communication, May 6, 2004). Using a mean ADAS score of 25.3 and a standard deviation of 5.2 resulted in small effect sizes of 0.19, 0.14, and 0.14 in each of the immediate, 12- and 24-month postintervention periods.

#### *Program Acceptability and Satisfaction*

Of the 666 (93%) parents offered the client satisfaction questionnaire, 355 (53.3%) rated the intervention program as “Excellent,” 240 (36%) as “Very good,” 64 parents (9.6%) as “Average,” and 7 parents (1.05%) as “Below average.”

#### *Biases That May Affect Interpretation of the Results*

Selection bias is demonstrated by the initial differences between the groups (Table 1). Given the higher prevalence and greater intensity of behavioral disturbance and higher levels of dysfunctional parenting, regression to the mean could be anticipated to be greater in the intervention group.

In order to investigate this possibility, the data were reanalyzed following stratification

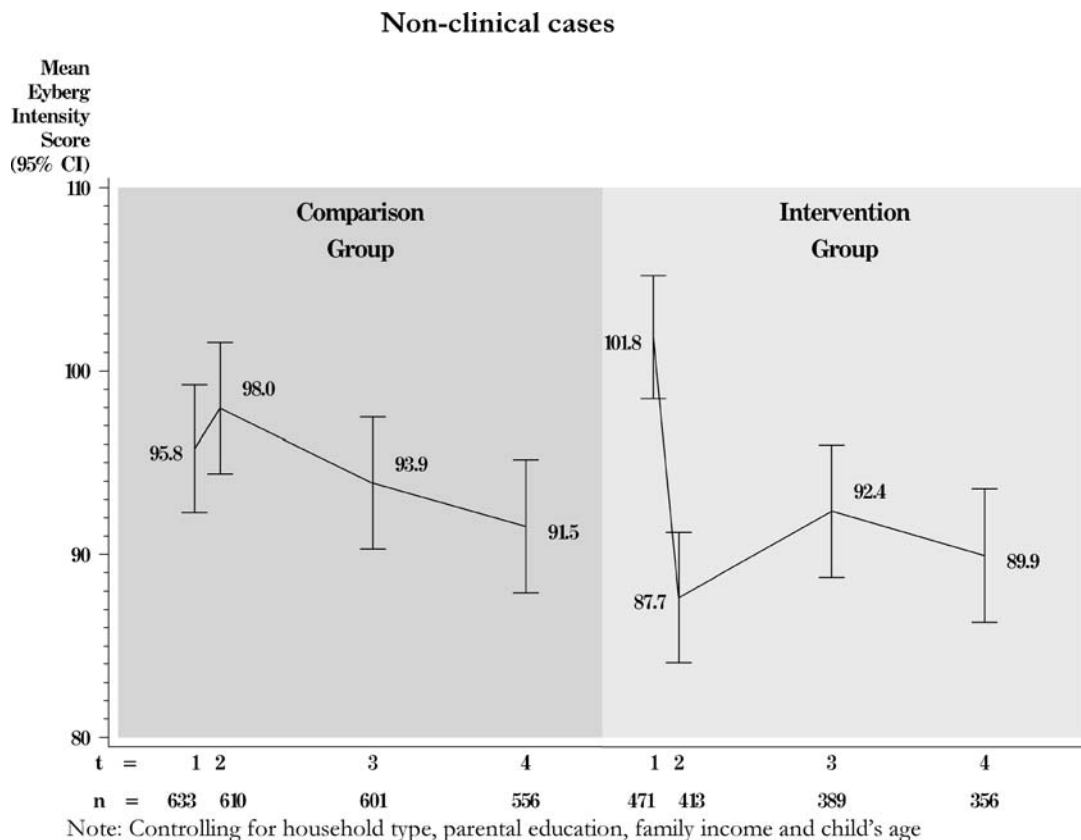


Fig. 1. Estimated Eyberg intensity score (ECBI) at pre-test (1), immediate post-test (2), twelve (3) and twenty-four months (4).

of both groups on ECBI intensity score, using  $\geq 127$ , the level defining clinical status, as the cut point.

For children with initial ECBI in the nonclinical range (Fig. 1) the intervention reduced the ECBI score from an initial value of 101.8 by 14.4 (95%  $CI = 14.2, 18.6$ ), 7.6 (95%  $CI = 5.3, 9.9$ ), and 7.7 (95%  $CI = 5.2, 10.1$ ) points at posttest, 12- and 24 months, respectively, over and above the effects associated with advancing developmental age as measured in the comparison group with ECBI in the nonclinical range. Clearly there is a measurable preventive effect in the children who are in the normal range at the outset of the program. Their levels of reported behavior problems declined in response to the intervention and remain at or lower than those in the comparison group 24 months later.

For children with initial ECBI in the clinical range (Fig. 2) the intervention reduced the ECBI score from an initial score of 151.0 by 27.3 (95%  $CI = 23.4, 31.3$ ), 8.4 (95%  $CI = 3.8, 13.1$ ), and 12.7

(95%  $CI = 7.1, 18.3$ ) points at posttest, 12- and 24 months, respectively, over and above the effects of advancing developmental age as measured in the comparison group with initial ECBI in the clinical range. Thus, the tendency for parents to report few behavior problems in children as they grow older, combined with the effects of the intervention is sufficient to keep the children below the clinical threshold. For these children, a treatment effect is evident in a reduction of the high level of parent-reported child behavior problems immediately posttest and at 12- and 24 months.

Overall, these findings are not merely the result of regression to the mean owing to bias in the recruitment of more seriously behaviorally disturbed children in the treatment group. Indeed, the results disentangle a complex interplay of prevention effects in the lowering of behavior problem levels in otherwise clinically unaffected children, treatment effects in the lowering of behavior problem levels in clinically affected children, and developmental changes

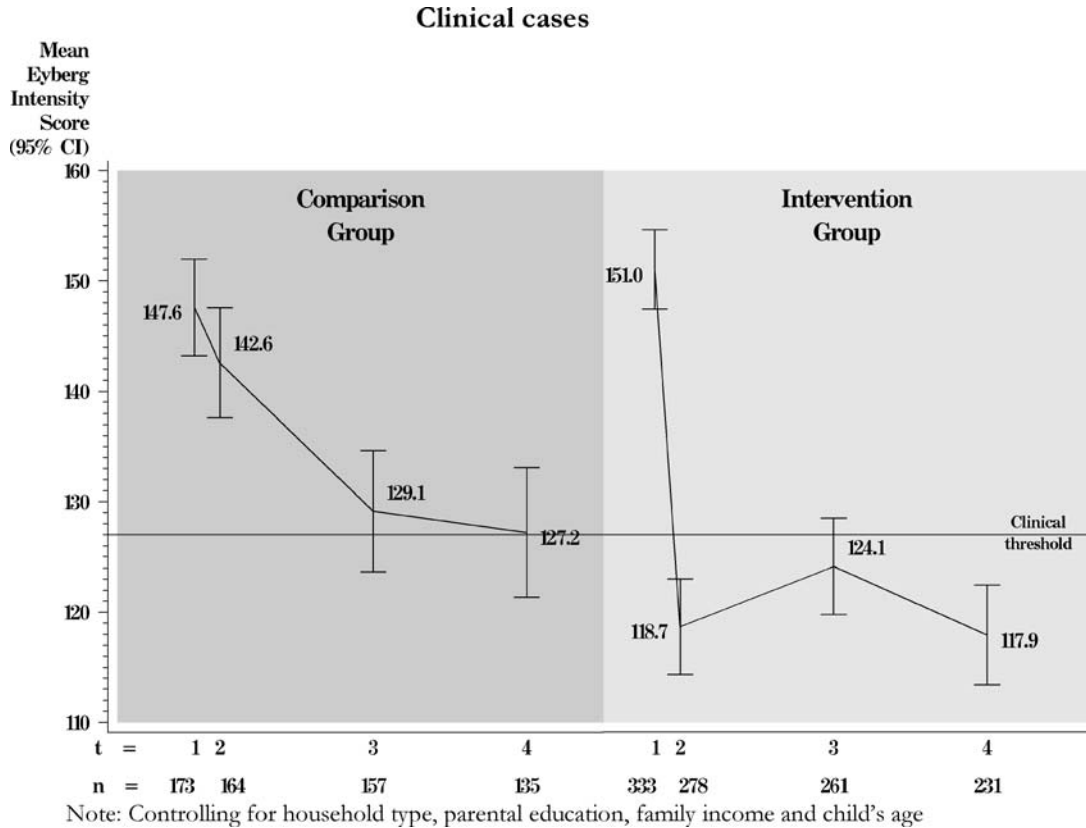


Fig. 2. Estimated Eyberg intensity score (ECBI) at pre-test (1), immediate post-test (2), twelve (3) and twenty-four months (4).

with fewer problem behaviors being reported by parents as children grow older.

#### *Differential Referral of Families*

Ethical considerations governing the evaluation necessitated referring families that met all the following criteria to additional treatment: 1) The child's behavior concerned the parent and they requested help, 2) the ECBI was in the clinical range ( $\geq 127$ ); 3) parent(s) reported five or more (out of 16) problems on the PPC, and 4) the primary caregiver reported high levels of depression ( $\geq 14$ ) on the DAS. Thirty-nine families met all these criteria and received additional help from the clinical psychologist in the period following the intervention, all from the intervention group.

To determine the extent to which this additional help inflated the estimated effects of the Triple-P intervention we re-estimated the linear mixed model of the ECBI excluding these 39 families from the

analysis. The new estimates changed by little at each assessment and not all in the same direction: 23.23 vs. 22.43; 11.31 vs. 11.29; and 12.68 vs. 12.92 for the immediate, 12-, and 24-month follow-up assessments, respectively. Thus, it would appear that referral and extra treatment of a select number of families with serious problems following Triple-P intervention has not introduced bias favoring the outcomes of Triple-P.

#### *Sample Attrition*

Finally nonrandom attrition from both the intervention and the comparison groups over the 2-year period introduces a potential source of bias. To assess the impact of sample attrition we re-estimated the intervention effects on ECBI, using only parents who had provided data at all four measurement points ( $N=1,213$ ). Again the new estimates changed by very little, and not all in the same direction: 22.54 vs. 22.43; 11.11 vs. 11.29; 12.62 vs. 12.92 at the immediate, 12-, and 24-month assessments.

## DISCUSSION

There is increasing evidence that parenting behaviors are not *inconsequential* to child behavior and developmental outcomes (Collins *et al.*, 2000, p. 226). Quite powerful effects have been observed in longitudinal studies of children, where these effects can be differentially observed in families of varying composition and type (Lipman *et al.*, 1998). While the causal pathways including these effects remain yet to be fully understood, the current evidence has been compelling enough for many agencies to implement programs that seek to support and develop better parenting practices or directly intervene to change the repertoire of parenting behavior (Spoth *et al.*, 1998). What happens when these programs are implemented in the real world? Who attends them, for how long and what, if any, are their effects? These were the questions that this evaluation study sought to answer.

This study highlights some of the conceptual, methodological, and analytical issues that are central to the current debate concerning prevention efficacy and prevention effectiveness research (Weisberg *et al.*, 2003). There are two considerations in assessing the value of a new preventive intervention. First, whether the preventive intervention is responsible for reported behavior changes or whether they may have other possible explanations. This efficacy is ideally assessed in large randomized trials where all known and unknown determinants of the behavior (with the exception of the intervention) are anticipated to be equally distributed between groups being compared. The second consideration is whether interventions with proven efficacy in clinical trials, will be similarly effective in real-world practice (Bickman, 1996; Weisz *et al.*, 1992). Effectiveness trials demonstrate whether a program's benefits generalize to an actual population when delivered under "real world" conditions by measuring the impact on outcomes of both program efficacy and other factors which can affect outcomes. These include subject compliance, program access, and community and professional acceptance. Effectiveness trials may also address efficiency issues by establishing longer-term functional and health benefits and weighing these against the implementation and other resource utilization costs (Klein & Smith, 1999).

The effect sizes observed in the present evaluation ranged from large (.83) to moderate (.47). Ideally one would like to compare these effect sizes to other evaluations with a more rigorous design im-

plemented in universal primary care settings with relatively large samples. Unfortunately most published studies are of small samples in predominately clinical or preschool nursery settings [see (Barlow & Stewart-Brown, 2000)]. These studies show effect sizes ranging from 0.6 to 2.9. From a review of 117 studies on behavioral parenting training, Serketich and Dumas (1996) selected 26 studies that met their inclusion criteria for meta-analysis. These studies had an average effect size of .86 (*SD* 0.36) for disruptive behavior in the home (Serketich & Dumas, 1996), and are congruent with the observed effect sizes in the present study. None of the sources of bias identified in the present study made systematic, clinically, or statistically significant differences to the estimated size of the effect of Triple-P on child behavior.

In population terms the effect sizes observed in this evaluation have a significant implication for the reduction of the burden of behavior problems in the target population of 3–4-year olds (Rose, 1995). For example, in Australia, the prevalence of disruptive behavior problems in 4-year-old children is approximately 16.9% (95% *CI* 12.7, 21.8) (Sawyer *et al.*, 2000; Zubrick *et al.*, 1997). Thus, in theory, a shift of the ECBI population mean by .47 of a standard deviation at 24 months postintervention means if behavioral family intervention were to be carried out in Australian settings as a universal population intervention, and if it reached all eligible families, it would reduce, 2 years later, the proportion of all children in the ECBI clinical range by about 55.4% (95% *CI* 47.4, 63.4). The assumptions underlying this observation are important: namely, that the intervention could be implemented on a population-wide basis and that this could be done cost-effectively.

In this implementation, the program reached about 66% of the estimated eligible children in the targeted postcode areas with high retention throughout the program. This would *in theory* achieve a reduction in the total proportion of children in the clinical range of about 36.5%. While this probably overstates the effect of the program given the constraints on the evaluation, these estimates are important information for funding authorities in their assessment of where to invest prevention effort. As a recent Western Australian survey showed that less than 2% of children with identified mental health disorders received any specialized mental health care (Zubrick *et al.*, 1995), we would regard this as a promising intervention effect to reduce the population mental health burden.

The effect of the intervention on other parental outcomes also showed both immediate and longer term improvement of caregiver-reported depression, anxiety and stress, marital adjustment, and parental conflict over child rearing. This may reflect the effects of participating in parent groups where social support and other care are made available. We estimate that the relative benefits of Triple-P generally diminish with the passage of time. However, this seems to result from a tendency for scores in the comparison group to improve as their children age, rather than deterioration toward pre-treatment levels in the scores of the intervention group. It may be that even a 1-year or 2-year improvement in these parental characteristics may contribute to better outcomes for some children, however just why these effects diminish with time requires further study.

The group Triple-P as implemented here was successful in attracting and retaining a preponderance of high-risk families as measured by the levels of child behavior disturbance and frequency of dysfunctional parenting at intake. This is an important feature of the intervention and helps underpin its use by the health authority. The loss of some of the higher risk families from the intervention group is, of course, a matter of concern, and common to most intervention programs (Karoly *et al.*, 1998). Certainly there is scope to modify program characteristics (e.g., daycare facilities, time of day, language of instruction) to ensure that programs are accessible and culturally relevant to those families most in need. However, our observations of parents who did not complete the intervention suggest that for some, the very capacity for sustained participation is impaired. Poverty, family disorganization and conflict, and parental mental health problems can impede participation. An important feature of this program is that it attracted these parents initially, thus providing self-identification and offering potential opportunities for individual follow-up and engagement with other services. Some of these parents may need to be targeted with home-visiting programs prior to involvement in a group in order to promote engagement, and increase motivation and commitment to remain involved.

### Limitations of the Evaluation

The limitations imposed upon the design of the evaluation by the requirement for universal access in an area of high need are likely to frequently con-

front prevention scientists in settings where policy makers, service agencies, and providers require convincing demonstrations of program effectiveness, not simply efficacy. Thus, it was our view that a quasi-experimental design can be informative if carefully conducted and analyzed. We have tried particularly to attend to hazards that arise in interpreting findings owing to potential selection bias (Larzelere *et al.*, 2004).

The initial inequality of groups being compared was addressed using linear mixed modelling and post-stratification on clinical status to allow for the effects of any regression to the mean. The further possibility of bias through referral of the subset of families with serious or chronic problems postintervention was addressed but their inclusion made no difference to the effect measure of the intervention, and their identification is another benefit of universal implementation. Non-random attrition over time is clearly evident but limiting analyses only to those completing all assessments did not alter the effect measure of the intervention.

The exclusive reliance on self-report methods to assess parenting and child behaviors without more independent or objective measures requires particular comment.

Multi-modal, multi-informant assessment protocols are considered the gold standard in measurement of child social competence and conduct problems (Webster-Stratton & Lindsay, 1999). It was a requirement of the administering health authority that the intervention was to be available and made accessible to all families of preschoolers in the health region as their interest was to trial the implementation of the program and recruitment process at a population level. Direct observation in home-based assessments of parenting and family functioning was considered too expensive to conduct, potentially intrusive, and likely to affect implementation participation rates. For this study we considered this was not viable due to the evaluation's large scale and its central purpose as an effectiveness trial. Moreover, the health authority was not interested in imposing any barriers to program participation or any processes beyond those that would be used in routine service delivery. In our experience the health authority's view is a common one, and it poses significant implications for prevention science with respect to wider dissemination and evaluation of the effectiveness of empirically supported interventions in "real world" service settings (Webster-Stratton & Taylor, 1998).

In addition, we would also note that there are some subjective phenomena that can only be reported by individuals themselves. This would include self-reported parental anxiety, stress and depression, and self-ratings of marital adjustment. While it could be argued that some of these may be inferred to some extent from external observation, it is also the case that observational methods are not a panacea and have their own limitations—particularly in assessing low frequency events.

Finally, there is evidence that suggests that parent ratings of their own parenting behavior show modest correlations with those of other observers (Webster-Stratton & Lindsay, 1999). While these correlations are by no means perfect, this gives some support to our reliance on parent's scores on well-designed parenting measures as reflective of objective differences in their own parenting behavior (Lovejoy, 1991; Lovejoy *et al.*, 1999).

Another potential limitation of this study is the possibility that demand effects could be present in the treatment group. Given the desire on the part of the parents in the treatment group to change child behavior as a result of the parent's participation in the program, they may perceive greater improvements in child behavior than are warranted by actual changes in behavior. However, similar effects could also be present in the comparison group. The motivation for parents to volunteer in the child behavior survey (i.e., comparison group) was to support a research survey being undertaken by the local children's hospital to inform planning and possibly secure future government services to support families in the area. These possibilities cannot be investigated by the use of multivariate mixed modelling procedures. We do note though, that there were also significant improvements in the treatment group, but not in the comparison group, on other variables that influence parent-child interaction. These included maternal depression, anxiety and stress, marital adjustment, and parental conflict over parenting. These were not targeted directly in the intervention, suggesting that the parent reports of child behavior are unlikely to be simply due to "demand characteristics" of the participating parents.

### CONCLUSION

With these limitations in mind, the present evaluation suggests that there are measurable and enduring effects attributable to a structured program

of behavioral family intervention made available universally through regular health services. These effects included clinically meaningful improvements in parent-reported child behavior and lower levels of parent-reported coercive parenting. The program was popular and attracted and retained a diverse range of families including those in high-risk categories. Moreover, it offered the opportunity of identifying those families who, for a variety of reasons, were not able to complete the program and might benefit from appropriately targeted services.

We believe these results show that a program of carefully monitored, measured, and delivered behavioral family intervention is one strategy that could be effective in changing population rates of adverse behavioral outcomes in children (Commonwealth Department of Health and Aged Care, 2000). Such an intervention is not the only opportunity that should be made available. Like other researchers before us we encourage the avoidance of a "hit and run" approach to prevention and instead recommend a careful selection of evidence-based strategies across the universal, selective, and indicated spectrum with evaluation built in as was done here. The evidence presented here for behavioral family intervention suggests that it should be seriously considered among such strategies.

### ACKNOWLEDGMENTS

The Telethon Institute for Child Health Research worked in close collaboration with the Public Health and Mental Health Divisions of the Health Department of WA, the Western Australian Department of Family and Children Services, the Education Department of WA, and the Parenting and Family Support Centre, School of Psychology at the University of Queensland. Thanks to Dr. Andrew Penman for initiating this project and to Barbara Hughes and to the staff of the Eastern Metropolitan Public and Community Health Service, the Kalamunda Health Service, and the Swan Health Service who implemented it. Peta Serna recruited parents and collected the comparison group data. Finally, particular thanks to our secretary Helen Howells and to Professors Nick deKlerk, Carol Bower, Fiona Stanley, Wendy Hall, D'Arcy Holman, Nadine Henley, and Dr. Davina French who read and commented on drafts of the manuscript.



## REFERENCES

- Arnold, D. S., O'Leary, S. G., Wolff, L. S., & Acker, M. M. (1993). The Parenting Scale: A measure of dysfunctional parenting in discipline situations. *Psychological Assessment, 5*, 137–144.
- Australian Bureau of Statistics. (1997). *Census of population and housing: CDATA96, Western Australia* (2019.5.30.001). Canberra: Australian Bureau of Statistics.
- Barlow, J., & Stewart-Brown, S. (2000). Behavior problems and group-based parent education programs. *Developmental and Behavioral Pediatrics, 21*(5), 356–370.
- Bickman, L. (1996). The evaluation of a children's mental health managed care demonstration. *Journal of Mental Health Administration, 23*, 7–15.
- Brestan, E. V., & Eyberg, S. M. (1998). Effective psychosocial treatments of conduct disordered children and adolescents: 29 years, 82 studies, and 5,272 kids. *Journal of Consulting and Clinical Psychology, 27*, 180–189.
- Burns, G. L., & Patterson, D. R. (1990). Conduct problem behaviors in a stratified random sample of children and adolescents: New standardization data on the Eyberg Child Behavior Inventory. *Psychological Assessment, 2*(4), 391–397.
- Burns, G. L., Patterson, D. R., Nussbaum, B. R., & Parker, C. M. (1991). Disruptive behaviors in an outpatient pediatric population: Additional standardisation data on the Eyberg Behavior Inventory. *Psychological Assessment, 3*(2), 202–207.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd Ed.). Hillsdale, NJ: Erlbaum.
- Coie, J. D. (1996). Prevention of violence and antisocial behavior. In R. D. Peters & R. J. McMahon (Eds.), *Preventing childhood disorders, substance abuse, and delinquency* (pp. 1–18). Thousand Oaks, CA: Sage.
- Collins, W. A., Maccoby, E. E., Steinberg, L., Hetherington, E. M., & Bornstein, M. H. (2000). Contemporary research on parenting: The case for nature and nurture. *American Psychologist, 55*(2), 218–232.
- Commonwealth Department of Health and Aged Care. (2000). *Promotion, prevention and early intervention for mental health*. Canberra, Australia: Commonwealth of Australia.
- Crawford, J. R., & Henry, J. D. (2003). The Depression Anxiety Stress Scales (DASS): Normative data and latent structure in a large non-clinical sample. *British Journal of Clinical Psychology, 42*, 111–131.
- Dadds, M. R., & Powell, M. B. (1991). The relationship of interparental conflict and global marital adjustment to aggression, anxiety, and immaturity in aggressive non-clinic children. *Journal of Abnormal Child Psychology, 19*, 553–567.
- Dishion, T., Kavanagh, K., Schneiger, A., Nelson, S., & Kaufman, N. (2002). Preventing early adolescent substance use: A family-centered strategy for the public middle school. *Prevention Science, 3*(3), 191–201.
- Doll, R. (1996). Weak associations in epidemiology: Importance, detection, and interpretation. *Journal of Epidemiology, 6*, s11–s20.
- Eyberg, S. M., & Pincus, D. (1999). *Eyberg Child Behavior Inventory and Sutter-Eyberg Student Behavior inventory—Revised: Professional manual*. Odessa, FL: Psychological Assessment Resources.
- Eyberg, S. M., & Ross, A. W. (1978). Assessment of child behavior problems: The validation of a new inventory. *Journal of Clinical Child Psychology, 7*, 113–116.
- Karoly, L. A., Greenwood, P. W., Everingham, S. S., Hoube, J., Kilburn, M. R., Rydell, P., Sanders, M., & Chiesa, J. (1998). Investing in our children: What we know and don't know about the costs and benefits of early childhood interventions (Vol. MR-898-TCWF). Santa Monica, CA: Rand.
- Klein, D., & Smith, L. (1999). Organisational requirements for effective clinical effectiveness studies. *Prevention and Treatment, 2*(Article 0002a).
- Larzelere, R. E., Kuhn, B. R., & Johnson, B. (2004). The intervention selection bias: An unrecognised confound in intervention research. *Psychological Bulletin, 130*(2), 289–303.
- Lipman, E. L., Boyle, M. H., Dooley, M. D., & Offord, D. R. (1998). *Children and lone-mother families: An investigation of factors influencing child well-being* (W-98-11E). Quebec, Canada: Applied Research Branch, Strategic Policy, Human Resources Development Canada.
- Lochman, J. E. (1990). Modification of childhood aggression. In M. Hersen & R. M. Eisler (Eds.), *Progress in behaviour modification* (Vol. 25, pp. 47–85). Newbury Park, CA: Sage.
- Loeber, R., & Farrington, D. P. (1998). Never too early, never too late: Risk factors and successful interventions for serious and violent juvenile offenders. *Studies on Crime and Crime Prevention, 71*(1), 7–30.
- Long, P., Forehand, R., Wierson, M., & Morgan, A. (1994). Does parent training with young non-compliant children have long term effects. *Behavior Research and Therapy, 32*, 101–107.
- Lovejoy, M. C. (1991). Maternal depression: Effects on social cognition and behavior in parent-child interactions. *Journal of Abnormal Child Psychology, 19*, 693–706.
- Lovejoy, M. C., Weis, R., O'Hare, E., & Rubin, E. C. (1999). Development and initial validation of the Parent Behavior Inventory. *Psychological Assessment, 11*(4), 534–545.
- Lovibond, P. F., & Lovibond, S. H. (1995a). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behavior Research and Therapy, 33*, 335–343.
- Lovibond, S. H., & Lovibond, P. F. (1995b). *Manual for the Depression Anxiety Stress Scales* (2nd Ed.). Sydney, NSW, Australia: Psychology Foundation of Australia.
- Markie-Dadds, C., Turner, K. M. T., & Sanders, M. R. (1997). *Every Parent's Group Workbook*. Brisbane, Queensland, Australia: Australian Academic.
- Marshall, J., & Watt, P. (1999). *Child behaviour problems*. Perth, Western Australia, Australia: Interagency Committee on Children's Futures.
- McNeil, C. B., Eyberg, S. M., Eisenstadt, T. H., Newcomb, K., & Funderburk, B. W. (1991). Parent child interaction therapy with behavior problem children: Generalisation of treatment effects to the school setting. *Journal of Clinical Child Psychology, 20*, 140–151.
- Murray, C. J. L., & Lopez, A. D. (Eds.). (1996). *The global burden of disease*. Boston: Harvard University Press.
- National Crime Prevention. (1999). *Pathways to prevention*. Canberra, Australia: National Crime Prevention, Attorney-General's Department.
- Patterson, G. R. (1982). *Coercive family process*. Eugene, OR: Castalia.
- Rose, G. (1995). *The strategy of preventive medicine*. Oxford: Oxford University Press.
- Sanders, M. R. (1992). *Every parent: A positive approach to children's behaviour*. Sydney, NSW, Australia: Addison-Wesley.
- Sanders, M. R. (1999). *Triple P-Positive Parenting Program: Towards an empirically validated multilevel parenting and family support strategy for the prevention of behaviour and emotional problems in children*. *Clinical Child and Family Psychology Review, 2*(2), 71–90.
- Sanders, M. R., & Dadds, M. R. (1982). The effects of planned activities and child management procedures in parent training: An analysis of setting generality. *Behavior Therapy, 13*, 452–461.
- Sanders, M. R., Markie-Dadds, C., Tully, L. A., & 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 onset conduct*

- problems. *Journal of Consulting and Clinical Psychology*, 68(4), 624–640.
- Sanders, M. R., Markie-Dadds, C., & Turner, K. M. T. (1996a). Positive parenting. Brisbane, Queensland, Australia: Families International.
- Sanders, M. R., Turner, K. M. T., & Markie-Dadds, C. (1996b). Families: Every Parent's Guide to Preschoolers [Video]. Brisbane, Australia: Families International.
- SAS Institute Inc. (2000). *SAS/STAT User's Guide, Version 8* (Vol. 1–3). Cary, NC.
- Sawyer, M. G., Arney, F. M., Baghurst, P. A., Clark, J. J., Graetz, B. W., Kosky, R. J., Nurcombe, B., Patton, G. C., Prior, M. R., Raphael, B., Rey, J., Whaites, L., & Zubrick, S. R. (2000). Child and adolescent component of the National Survey of Mental Health and Well Being: The mental health of young people in Australia. Canberra, Australia: Mental Health and Special Programs Branch, Commonwealth Department of Health and Aged Care.
- Serketich, W. J., & Dumas, J. E. (1996). The effectiveness of behavioral parent training to modify antisocial behavior in children: A meta analysis. *Behavior Therapy*, 27, 171–186.
- Sharpley, C. F., & Rogers, H. J. (1984). Preliminary validation of the abbreviated Spanier Dyadic Adjustment Scale: Some psychometric data regarding a screening test of marital adjustment. *Educational and Psychological Measurement*, 44, 1045–1049.
- Silburn, S. R., Zubrick, S. R., Garton, A. F., Gurrin, L., Burton, P., Dalby, R., Carlton, J., Shepherd, C., & Lawrence, D. (1996). Western Australian Child Health Survey: Family and community health (Catalogue 4304.5). Perth, Western Australia, Australia: Australian Bureau of Statistics.
- Spanier, G. B. (1976). Measuring dyadic adjustment: New scales for assessing marital quality. *Journal of Marriage and the Family*, 38, 15–28.
- Spoth, R., & Redmond, C. (2002). Project Family prevention trials based in community–university partnerships: Toward scaled-up preventive interventions. *Prevention Science*, 3(3), 203–221.
- Spoth, R., Redmond, C., & Shin, C. (1998). Direct and indirect latent-variable parenting outcomes of two universal family-focused preventive interventions extending a public health-orientated research base. *Journal of Consulting and Clinical Psychology*, 66(2), 385–399.
- Stormshak, E., Kaminski, R., & Goodman, M. (2002). Enhancing the parenting skills of Head Start families during the transition to kindergarten. *Prevention Science*, 3(3), 223–234.
- Taylor, T. K., & Biglan, A. (1998). Behavioral family interventions for improving child rearing: A review of the literature for practitioners and policy makers. *Clinical Child and Family Psychology Review*, 1, 41–60.
- Turner, K. M. T., Markie-Dadds, C., & Sanders, M. R. (1998). Facilitator's Manual for Group Triple P. Brisbane, Australia: Family International.
- Webster-Stratton, C. (1989). Systematic comparison of consumer satisfaction of three cost-effective parent training programs for conduct problem children. *Behavior Therapy*, 20, 103–115.
- Webster-Stratton, C. (1998). Preventing conduct problems in head start children; strengthening parenting competencies. *Journal of Consulting and Clinical Psychology*, 66(5), 715–730.
- Webster-Stratton, C., & Lindsay, D. W. (1999). Social competence and conduct problems in young children: Issues in assessment. *Clinical Child Psychology*, 21(1), 25–43.
- Webster-Stratton, C., & Taylor, T. K. (1998). Adopting and implementing empirically supported interventions: A recipe for success. In A. Buchanan & B. L. Hudson (Eds.), Parenting, schooling and children's behaviour—Interdisciplinary approaches. Hampshire, UK: Ashgate.
- Weisberg, R., Kumpfer, K., & Seligman, M. (2003). Prevention that works for children and youth. *American Psychologist*, 58(6/7), 425–432.
- Weisz, J. R., Weisz, B., & Donenberg, G. R. (1992). The lab versus the clinic: Effects of child and adolescent psychotherapy. *American Psychologist*, 47, 1578–1585.
- Zubrick, S. R., Silburn, S. R., Garton, A. F., Dalby, R., Carlton, J., Shepherd, C., Lawrence, D., & Burton, P. (1995). Western Australian Child Health Survey: Developing health and well being in the '90s (Catalogue 4303.5). Perth, Western Australia, Australia: Australian Bureau of Statistics.
- Zubrick, S. R., Silburn, S. R., Teoh, H.-J., Carlton, J., Shepherd, C., & Lawrence, D. (1997). Western Australian Child Health Survey: Education, health and competency (Catalogue 4305.5). Perth, Western Australia, Australia: Australian Bureau of Statistics.