Does parent management training for children with externalizing problem behavior in routine care result in clinically significant changes?

CHRISTOPHER HAUTMANN1, PETRA STEIN2, CHARLOTTE HANISCH3, ILKA EICHELBERGER1, JULIA PLÜCK1, DANIEL WALTER1, & MANFRED DÖPFNER1

1Department of Child and Adolescent Psychiatry and Psychotherapy, University of Cologne, Cologne; 2Department of Sociology, University of Duisburg, Duisburg & 3Department of Psychology, University of Applied Sciences Düsseldorf, Düsseldorf; Germany

(Received 2 September 2008; revised 21 January 2009; accepted 25 January 2009)

Abstract
The clinical significance of a parent management training that is part of the Prevention Program for Externalizing Problem Behavior (PEP) was assessed in an effectiveness trial. Parent management training was offered under routine care conditions in a setting where a high proportion of children were expected to show clinically relevant symptoms of externalizing problem behavior. At the beginning of the study, 32.6% to 60.7% of children were classified as clinical cases (dysfunctional) on three outcome measures of child behavior problems. Three months after treatment, 24.8% to 60.4% of children were judged to be recovered. Parent management training can result in clinically significant changes in children with externalizing behavior problems.

Keywords: clinical significance; externalizing behavior; parent management training

The aim of the present study was to test whether a parent management training program for externalizing behavior problems in children, conducted under routine care conditions, produces clinically significant changes in different outcome measures.

Externalizing behavior comprises symptoms of attention-deficit/hyperactivity disorder (ADHD) as well as oppositional-defiant disorder (ODD) and conduct disorder (CD). ADHD and ODD have high prevalence rates in childhood. For example, in a survey of 10,367 children aged 4 to 17 years in the United States, 4.2% of males and 1.8% of females had clinically significant ADHD symptoms (Cuffe, Moore, & McKeown, 2005). In a similarly large nationally representative sample of children in the United Kingdom, ODD was seen in 3.2% of boys and 1.4% of girls (Maughan, Rowe, Messer, Goodman, & Meltzer, 2004). Externalizing problems are associated with many developmental costs, including poorer social skills, substance dependence disorders (alcohol, drug, and nicotine), low academic achievement, and aggression in intimate relationships (Biederman et al., 2006; Campbell, Spieker, Burchinal, Poe, & NICHD Early Child Care Research Network, 2006; Elkins, McGue, & Iacono, 2007; Fontaine et al., 2008; Nijmeijer et al., 2008; Spira & Fischel, 2005). Given the long-term consequences of these disorders, it is recommended that interventions start in the preschool years (Greenberg, Domitrovich, & Bumbarger, 2001).

Parent management training is a well-established treatment modality (Eyberg, Nelson, & Boggs, 2008; Kazdin, 2005; Nixon, 2002), particularly for children with ODD and CD (Brestan & Eyberg, 1998; Chronis, Jones, & Raggi, 2006; Farmer, Compton, Burns, & Robertson, 2002). The Prevention Program for Externalizing Problem Behavior (PEP; Plück, Wieczorrek, Wolff Metternich, & Döpfner, 2006) is a parent management training program that has a cognitive–behavioral basis and is founded on published prevention and treatment manuals for children with disruptive behavior problems (Barkley, 1997; Döpfner, Adrian, & Hanisch, 2007; McMahon & Forehand, 2003). PEP has many features that are common to other parent
management training programs. Parents are the primary agent of change and are taught how to alter their child's behavior. However, PEP especially addresses families with children aged 3 to 10 years who show externalizing problem behavior. Teacher training is another component of the program but was not investigated in the present study.

Clinical significance is an additional concept to those of statistical significance and effect size (Kirk, 2001). Tests of statistical significance assess whether an observed change during a treatment can be explained by chance. Effect size measures inform about the magnitude of effects, such as the percentage of variance explained by the treatment or the change in standard deviation units. Measures of clinical significance provide information about the practical meaning of the results. The information gathered by these three types of measure is complementary rather than mutually exclusive.

There is no single definition of clinical significance, and several different approaches have been developed to define it. Ogles, Lunnen, and Bonesteel (2001) provided three quantitative definitions of clinical significance. The first two definitions are (1) a return to normal functioning, such that the client reaches a level where he or she is indistinguishable from healthy people; and (2) a statistically reliable improvement, such that the change goes beyond that which can be attributed to measurement error. Different formulas can be used to ascertain statistically reliable improvements (Hageman & Arrindell, 1999; Hsu, 1999; Jacobson & Truax, 1991; Speer, 1992; Speer & Greenbaum, 1995). From their simulation study, Atkins, Bedics, McGlinchey, and Beauchaine (2005) concluded that the different methods are comparable. The third definition of a clinically significant change combines the two aforementioned criteria. Thus, the client has to not only return to normal functioning but also make statistically reliable improvements. This definition corresponds to the one proposed by Jacobson et al. (Jacobson, Roberts, Berns, & McGlinchey, 1999; Jacobson & Truax, 1991), whose method has been used to evaluate clinical significance in the present study.

In the present study, PEP was delivered in routine care settings by employees of different counseling and mental health services. This study can, therefore, be characterized as an effectiveness trial. In contrast to efficacy trials, effectiveness studies assess treatment effects under real-world conditions (Lutz, 2003; Nathan, Stuart, & Dolan, 2000; Weisz, Donenberg, Han, & Weiss, 1995). That is, effectiveness studies have high external validity, but this is often obtained at the expense of low internal validity.

To date, only a few outcome studies have been carried out as effectiveness trials (Glasgow, Lichtenstein, & Marcus, 2003). In general, the results of these effectiveness trials are less promising than those of efficacy trials. In effectiveness trials of traditional child psychotherapy, average effect sizes ranged from −0.08 (Weiss, Catron, Harris, & Phung, 1999) to 0.01 (Weisz & Jensen, 1999). However, in a recent effectiveness trial of parent management training in Norway, Ogden and Hagen (2008) reported that effect sizes of parent-rated behavior problem outcome measures ranged from 0.16 to 0.33.

PEP has already been tested in an efficacy trial, with positive results (Hanisch et al., 2006, 2008). That is, it has already been tested under laboratory conditions. Several features of the efficacy trial were distinct from the present effectiveness study. The efficacy trial had a randomized control group design, with clearly defined inclusion and exclusion criteria, and the treatment was delivered by research scientists. In contrast, the present study had a less rigorous within-subject control group design (Cotton, 1998), and less strict inclusion criteria, and the treatment was offered by experienced practitioners.

Research on the clinical significance of different treatment modalities in children with CD is sparse (Kazdin, 1997), and only a few studies have investigated the clinical significance of parent management training in externalizing behavior problems (Nixon, Sweeney, Erickson, & Touyz, 2003; Webster-Stratton & Hammond, 1997). Furthermore, only one of these studies (Nixon et al., 2003) assessed clinical significance using the method of Jacobson et al. (1999; Jacobson & Truax, 1991).

Nixon et al. (2003) tested Eyberg’s (1988) Parent–Child Interaction Therapy. At the posttreatment assessment, 59% of children were classified as recovered with respect to conduct problems; that is, a clinically significant change could be shown for about two thirds of the children. However, the sample size used to assess clinical significance was small (16 families).

The aim of the present study was to examine the clinical significance of a parent management training in a large sample using the method of Jacobson et al. (1999; Jacobson & Truax, 1991). Previous analyses of this study have shown statistically significant short-term effects on child externalizing problem behavior and parenting behavior; effect sizes were in the small to moderate range (Hautmann, Hanisch, Mayer, Plück, & Döpfner, 2008). The present study extends the analysis to include a 3-month follow-up and focuses on clinical significance. To our knowledge, this is the first study to assess the clinical significance of such interventions in real-world
settings as they are tested in effectiveness trials. Because the parent management training of the PEP was offered to patients referred to counseling institutions, it was expected that a high proportion of the children would exhibit clinically relevant externalizing problem behavior. Before treatment, there was a waiting period that served as a control period for each subject. Our hypothesis was that the training outcome could be considered clinically significant if there were more cases of clinically significant changes during the training period than during the waiting period.

**Method**

**Evaluation Procedure**

The parent management training of the PEP program was evaluated using a within-subject control group study (Cotton, 1998). Before the intervention, there were two assessment points: (1) pre1 (3 months before the training began) and (2) pre2 (immediately before the start of training). These were used to assess any alterations in behavior before starting the parent management training. Thus, this waiting period served as a control condition for each subject. The 3-month period was also necessary for organizational purposes, and this was communicated to the participating families. Treatment effects were assessed at two postintervention assessment points: (1) post (immediately after the training) and (2) FU1 (3 months after the intervention was finished).

Families included were those referred to one of 37 different counseling agencies, pediatric primary care centers, and psychotherapy services in Germany. Before the study, these institutions were asked whether they were interested in taking part in the study. Children were usually referred to these institutions because their parents or teachers were concerned about the behavior problems of the child and, therefore, asked for assessment and counseling or more extended treatment. The only inclusion criterion was a 3- to 10-year-old child with externalizing problem behavior according to clinical evaluation. Study inclusion did not depend on meeting a defined threshold of symptom severity. Trainers of the parent management training (n = 59) were employees of the 37 different participating institutions. On average, there were 1.61 (SD = 0.96) trainers per institution. The trainers decided which families were eligible for the study.

**Participants**

A total of 324 families were included at the pre1 assessment. Of these families, 265 supplied questionnaire data at the pre2 assessment, 210 at the post assessment, and 171 at the FU1 assessment. Families who did not attend the training were excluded from the analysis. Thus, conclusions about the effectiveness of treatment are valid only for families who attended at least one unit of the parent management training. After this correction, the numbers of families at each assessment point were as follows: pre1 (n = 270), pre2 (n = 248), post (n = 210), and FU1 (n = 171). These families comprised the sample analyzed in the present study.

We tested whether those families who did not participate in the training and were excluded from the analysis (n = 54) differed from those who attended the training (n = 270). There were no differences in any of the outcome measures at pre1 between these two groups.

We also tested whether families who participated in the training but who dropped out of the study either at post (n = 60) or at FU1 (n = 169) differed from those who attended the training and provided complete data. Participants with missing data at post did not differ in any of the outcome measures at pre1 from those whose data were available at post. The same was true for participants with missing and available data at FU1. In addition, patients with missing data at FU1 did not differ from those with full data sets at FU1 regarding their changes during treatment (pre2 to post) on any of the outcome variables.

Missing data were treated by expectation maximization (EM; McLachlan & Krishan, 1996) using the SPSS procedure (SPSS Inc., 2007). That is, missing data were imputed and analyses computed with full data sets (n = 270). The missing data handling strategy was considered justified because dropout from the study was hypothesized to be missing at random (Rubin, 1976).

Children in the 270 families analyzed were an average age of 6.53 years (SD = 2.02) and 214 (79.3%) were boys. In 171 (63.3%) families, both biological parents lived together, 68 (25.2%) parents were single parents, and children lived under other conditions in 31 (11.5%) cases. Of the 270 families, 43 (15.9%) had an immigration background (i.e., the parents or grandparents were immigrants to Germany mostly from southern and eastern European countries).

**Parent Management Training of PEP**

The parent management training of the PEP is delivered to parents in a group format, with a group typically consisting of four to eight families. In this study, there was an average of 4.66 families (SD = 1.37) per course. The parent management training
comprises 12 intervention sessions of 90 to 120 min each. There are six basic units and six additional units, as summarized in Table I. Each unit is individually tailored according to the needs of each participating family. Therefore, at the beginning of the training, the specific problem of each child is defined individually. The parents are taught to solve these specific problems by using the different interventions discussed in the units.

Overall, 59 trainers held parent management training courses. The trainers were introduced to the parent management training of the PEP by a 2-day course presented by the project members. The trainers were psychologists (37.3%), social or educational workers (23.5%), educators (15.7%), remedial teachers (15.7%), or members of other professions (7.8%).

In this study, trainers were required to deliver the six basic units, but the additional units were optional; that is, it was up to the trainers to choose which additional units they delivered. On average, trainers offered 7.88 units (SD = 1.43), and parents attended an average 4.63 lessons (SD = 1.56) of the six basic units. Of the 270 families included in this study, 250 mothers and 99 fathers attended at least one of the six obligatory basic units. The 250 mothers attended on average 4.45 of the six basic units and the 99 fathers attended on average 3.28 of the basic units.

**Dependent Measures**

All data were gathered using questionnaire booklets, which were completed by mothers and contained instruments that assess child behavior problems and parenting behavior.

### Table I. Basic and Additional Units of the Parent Component of the Parent Management Training of the Prevention Program for Externalizing Problem Behavior

<table>
<thead>
<tr>
<th>Type</th>
<th>Title</th>
<th>Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>AD</td>
<td>Constituent unit</td>
<td>Get-together of the group members and introduction to the program</td>
</tr>
<tr>
<td>BA</td>
<td>My child—happiness and distress</td>
<td>Identification of competencies and target problems of the child</td>
</tr>
<tr>
<td>BA</td>
<td>Vicious circle and play time</td>
<td>The coercive interaction process (vicious circle) as an explanatory model for the target problems of the child and the play time as a possibility to strengthen positive parent-child interactions</td>
</tr>
<tr>
<td>BA</td>
<td>Saving energy</td>
<td>General methods for parents to cope with stress in daily routine</td>
</tr>
<tr>
<td>BA</td>
<td>Rules and effective commands</td>
<td>Definition of firm rules for the target problems of the child and practice of effective commands</td>
</tr>
<tr>
<td>BA</td>
<td>Positive consequences</td>
<td>Methods for reinforcing the child if he or she complies with the rules</td>
</tr>
<tr>
<td>BA</td>
<td>Negative consequences</td>
<td>Methods for punishing the child adequately if he or she breaks the rules</td>
</tr>
<tr>
<td>AD</td>
<td>Problem behavior in public</td>
<td>Methods to cope with problem behavior in the public</td>
</tr>
<tr>
<td>AD</td>
<td>Enduring quarrels</td>
<td>Methods to cope with enduring quarrels between siblings/children</td>
</tr>
<tr>
<td>AD</td>
<td>Persevering play</td>
<td>Methods to strengthen persevering play of children</td>
</tr>
<tr>
<td>AD</td>
<td>Homework</td>
<td>Methods to strengthen attention and finishing homework</td>
</tr>
<tr>
<td>AD</td>
<td>Summary</td>
<td>Summary of the program contents</td>
</tr>
</tbody>
</table>

*Note. AD = additional unit; BA = basic unit*
Psychiatric Association, 1994) and ICD-10 (World Health Organization, 1990) for ODD and CD. For this study, only the nine items of the ODD subscale (SCL-ODD) were considered. Items are scored from 0 to 3, with higher values indicating more severe problems. Reliability and construct validity of this instrument have been shown (Döpfner et al., 2008). Internal consistency was Cronbach’s α = .91.

Parenting

Self-Efficacy Scale (SEFS). This is the German adaptation of Johnston and Mash’s (1989) Parenting Sense of Competence Scale and Coleman and Karra-ker’s (2000) Self Efficacy for Parenting Task Index. The SEFS comprises 15 items measuring parents’ perception of self-efficacy on a scale ranging from 0 to 3, with higher values indicating more competencies. Reliability and discriminant and convergent validity of this instrument have been shown (Hanisch et al., 2008). In this sample, one item was deleted because of low item total correlation. Internal consistency of the remaining items was Cronbach’s α = .85.

Problem Setting and Behavior Checklist (PSBC). The German adaptation of the PSBC developed by Sanders, Markie-Dadds, Tully, and Bor (2000) measures the perceived ability to solve difficult parenting situations. Items are scored on a scale ranging from 0 to 3; higher scores reflect a stronger ability to deal with difficult parenting situations. Reliability and discriminant and convergent validity of this instrument have been shown (Hanisch et al., 2008). Internal consistency for the overall score in our sample was Cronbach’s α = .91.

Clinical Significance

Clinical significance was evaluated using the method of Jacobson et al. (1999; Jacobson & Truax, 1991). This approach combines two criteria: (1) assessment of whether a person has returned to normal functioning and (2) judgment of whether the changes can be considered statistically reliable.

Jacobson et al. proposed three different cutoff scores for deciding whether a client is indistinguishable from healthy people: (1) The cutoff score is 2 SDs below the mean of the functional population (lower values representing less functionality), (2) the cutoff score is based on the distribution of scores in the dysfunctional population and is 2 SDs above the mean of the unhealthy population (higher values signify functionality), and (3) the cutoff score takes the distribution of the functional as well as the dysfunctional population into account and is the average of the means of the functional and dysfunctional populations.

Jacobson et al. suggested when it was appropriate to use the different cutoff scores. In the present study, the first cutoff score based on the distribution of the functional population was chosen because normative data were only available for this population.

The Reliable Change Index (RCI) of Jacobson et al. (1999; Jacobson & Truax, 1991) is defined as:

\[ \frac{X_2 - X_1}{\sqrt{2(S_{11}^1 - r_{XX})}} \]

where \( X_2 \) is the individual’s raw score at Time 2, \( X_1 \) is the individual’s raw score at Time 1, \( S_{11} \) is the standard deviation of the sample at Time 1, and \( r_{XX} \) is the reliability of the measure. If the RCI is greater than 1.96, the change is considered significant at the .05 level.

Using the twofold criterion to define clinical significance, a client’s state can be classified as (1) “improved but not recovered” if the client is still in a dysfunctional state but a positive change can be determined beyond that occurring by chance alone, (2) “recovered” when both criteria are fulfilled, (3) “deteriorated” when the parameter used to calculate the reliable change moves in the opposite direction, and (4) “unchanged” when the client is still in the dysfunctional range and no reliable change can be noted. If the patient is judged to be part of a normal and healthy population but does not show statistically reliable change, then the method cannot explain whether or not the change is clinically significant. Such cases are labeled (5) “not classifiable.”

Statistical Analysis

Five outcome measures were considered: three on the externalizing behavior of the child (CBCL-EXT, SCL-ADHD, SCL-ODD) and two on parenting (SEFS, PSBC). Bonferroni correction was applied such that alpha error was adjusted to 0.01 (0.05/5). Clinical significance was evaluated using the method of Jacobson et al. (1999; Jacobson & Truax, 1991).

Results

The means and standard deviations for the five outcome measures at each assessment point are given in Table II. All participants were classified as either dysfunctional (i.e., clinical cases) or functional with respect to each of the five outcome measures (Table III). For this classification, reference scores (means and standard deviations) were drawn from representative German community samples. For CBCL-EXT (Arbeitsgruppe Deutsche Child Behavior Checklist, 1998) as well as SEFS and PSBC.
data were reanalyzed to obtain reference scores. The reference scores of the community samples were as follows: CBCL-EXT, $M_{/C30} = 5.92$, $SD_{/C30} = 6.02$; SCL-ADHD, $M_{/C30} = 0.54$, $SD_{/C30} = 0.52$ (from Döpfner et al., 2008); SCL-ODD, $M_{/C30} = 0.59$, $SD_{/C30} = 0.56$ (from Döpfner et al., 2008); SEFS, $M = 2.20$, $SD = 0.39$; PSBC, $M = 2.17$, $SD = 0.40$. For CBCL-EXT, SCL-ADHD, and SCL-ODD, the cutoff values for dysfunctional behavior were 2 $SD$s above the mean of the reference samples. For SEFS and PSBC, the cutoff values were 2 $SD$s below the mean because for these two instruments higher values indicated positive parenting.

At beginning of the study (pre1), between 32.6% and 60.7% of the children were classified as dysfunctional (i.e., clinical cases) on the child behavior problem scales (see Table III). For parenting, the pre1 dysfunctional rates ranged from 6.7% to 15.2%.

In accordance with the method of Jacobson et al. (1999; Jacobson & Truax, 1991), we tested whether there were meaningful changes in clinical state over time. Table IV summarizes the frequency of the four different categories of clinical state (deteriorated, no change, improved but not recovered, recovered) for each outcome measure over three different transitions: trans 1 (transition from pre1 to pre2); trans 2 (transition from pre2 to post); and trans 3 (transition from pre2 to FU1). The fifth category of clinical state indicating that a person is not classifiable is not reported.

Two categories of clinical state were then created by collapsing the deteriorated and no change categories into one category and the improved but not recovered and recovered categories into one category. Binomial tests examined whether more participants fell in the improved but not recovered/recovered category at trans 2 and trans 3 versus trans 1 (Table V). For all child behavior problem variables, significantly more children improved/recovered during the training period than during the waiting period. This was not true for the parenting variables.

### Discussion

The present study evaluated the effects of parent management training under routine care conditions for children with externalizing behavior problems in terms of clinical significance at the end of the treatment and at 3-months follow-up. We found that between 32.6% and 60.7% of children were classified as dysfunctional at each assessment point for the five outcome measures (Table III).
classified as clinical cases at the beginning of the study, depending on the child behavior outcome measure. As expected, the rate of mothers with dysfunctional parenting was much lower (6.7/15.2%) because the study entry criterion was a child with externalizing problem behavior, not dysfunctional parenting. Parenting may contribute to child externalizing problem behavior but is not the only causal factor (Lahey, Waldmann, & McBurnett, 1999; Nigg, 2006). The low frequency of dysfunctional parenting at the beginning of the study may be a consequence of a problem with the community sample used for the reference scores (see later critique).

The proportion of participants who improved/recovered during treatment or during treatment and follow-up was significantly larger than during the waiting period for all outcome measures on child behavior problems but not for measures of parenting. Therefore, the effects of the intervention can be considered as clinically significant for child behavior problems. After treatment, the children’s behavior was classified as recovered in 24.8% to 60.4% of cases, depending on the outcome measure. This is comparable to the 59% of recovered cases reported by Nixon et al. (2003) when the Parent–Child Interaction Therapy was evaluated.

The limitations of the study are those generally associated with effectiveness studies. For example, effectiveness studies have high external validity but low internal validity. Moreover, the waiting period as...
a control condition is less rigorous than having a control group. However, the treatment has been proven to be effective in a more rigorous efficacy trial (Hanisch et al., 2006, 2008).

Another limitation is that the adherence of trainers to the treatment was not assessed. The parent training provided by the trainers may have diverged from that given in the manual. However, we consider this unlikely because the training material is highly structured through material such as overhead transparencies.

The normative data used to calculate the cutoff scores for the parenting measures (SEFS, PSBC) have to be questioned critically. The sample described by Heinrichs et al. (2005) was recruited by the offer to take part in a development study, and it is questionable whether the sample was sufficiently representative.

Missing data were handled by the EM algorithm of SPSS (SPSS Inc., 2007). Some authors have criticized this procedure because missing values are imputed without residual variation (von Hippel, 2004). We have also analyzed the study data using a different method of handling missing data: that is, with all available cases (listwise deletion). In general, the results were comparable and are available on request.

A further limitation of the study is that only mothers were the informants for the outcome measures. Because mothers were, in general, also the participants of the training, their expectations might have influenced their rating of the outcome measures. Additional observational data or an expert rating would have been useful. Such data were used in the efficacy trial (Hanisch et al., 2006, 2008), which demonstrated positive treatment effects on behavioral observations. Finally, long-term follow-up data are needed to reveal the further developmental course of children with externalizing problem behavior.

A substantial proportion of children recovered during the waiting period. It is not clear whether these changes are due to repeated assessment, spontaneous recovery, or concomitant treatment during the waiting period. In longitudinal studies, regression to the mean effect has to be considered (Nesselroade, Stigler, & Baltes, 1980). That is, high values at the first assessment point have a tendency to be reduced at the second assessment because of chance. Furthermore, the normative development of externalizing problem behavior is to decrease with age (Bongers, Koot, van der Ende, & Verhulst, 2003), but it is questionable whether this developmental trend can explain the clinically significant reductions seen during the short time frame of the present study. Furthermore, there were no restrictions on whether participants received concomitant treatment, which could have had an effect on outcomes. These factors influence changes not only during the waiting period but also during the training and follow-up periods. Nevertheless, the effectiveness of the treatment is not in doubt because we have shown there were stronger effects during the treatment period than during the waiting period.

In a previous report of this study, we showed that the parent management training of the PEP had statistically significant short-term effects and that the effect sizes were in the small to moderate range (Hautmann et al., 2008). These findings are extended by the present analysis, which indicates that the parent management training can also produce clinically significant results in terms of recovery from child behavioral problems immediately after the treatment has finished and at 3-months follow-up.

Acknowledgements

The study was funded by Deutsche Forschungsgemeinschaft Grant DO 620/2. The ethics committee of the University Hospital Cologne approved the study. We kindly thank Deirdre Elmhirst for carefully reading an earlier version of this article. We thank all the families who participated in the parent management training and all the institutions and trainers who offered the training.

References


Clinical significance of parent management training


