A META-ANALYSIS OF INTERVENTIONS FOR BEREAVED CHILDREN AND ADOLESCENTS

RITA ROSNER, JOACHIM KRUSE, and MARIA HAGL
Department of Psychology, Ludwig Maximilians University of Munich, Munich, Germany

The main objective of this review was to provide a quantitative and methodologically sound evaluation of existing treatments for bereavement and grief reactions in children and adolescents. Two meta-analyses were conducted: 1 on controlled studies and 1 on uncontrolled studies. The 2 meta-analyses were based on a total of 27 treatment studies published before June 2006. Hedges's $g$ and Cohen's $d$ were used as measures of effect size and a random-effects model was applied. Results yielded small to moderate effect sizes. Interventions for symptomatic or impaired participants tended to show larger effect sizes than interventions for bereaved children and adolescents without symptoms. Promising treatment models were music therapy and trauma/grief-focused school based brief psychotherapy.

Experiencing loss and bereavement in childhood is a painful and sometimes traumatic experience that can be associated with grief symptoms, overall negative affect, and later depression or anxiety (Lutzke, Ayers, Sandler, & Barr, 1997). Because of the impact of childhood loss and bereavement, there is an abundance of therapeutic literature on how to help bereaved children and adolescents. In contrast to the amount of literature, however, there is a lack of empirical support for most of the suggested interventions. Newly published reviews usually include narrative descriptions of four or five intervention studies (e.g., Cohen & Mannarino, 2004; Forte, Hill, Pazder, & Feudtner, 2004) and until recently,
no published meta-analysis concerning treatment effects on children and adolescents has been available.

In their recently published meta-analysis on effectiveness of 13 controlled studies of treatment for childhood bereavement, Currier, Holland, and Neimeyer (2007) reported a small, non-significant effect size of 0.14. This result corresponds to the effect sizes of two meta-analyses based on controlled studies with bereaved adults ($d = 0.11$; Kato & Mann, 1999; $d = 0.13$; Fortner, 2000; see Neimeyer, 2000). A third meta-analysis of treatment of adult bereavement, which included uncontrolled studies, yielded a moderate treatment effect of $d = 0.43$ (Allumbaugh & Hoyt, 1999). Thus, the overall picture on grief treatment seemed to be rather discouraging. Yet Larson and Hoyt (2007) argued that existing evidence is still not sound enough to reach clear conclusions.

Just recently, Currier, Neimeyer, and Berman (2008) presented a comprehensive quantitative review, including both studies on children and adults in this new meta-analysis on grief treatment efficacy. They reported a small, but significant effect for randomized studies ($d = 0.16$) and a moderate effect for non-randomized studies ($d = 0.51$). Thus, Currier et al. (2008) corroborated a small to moderate benefit of grief treatment in general, while looking at all age groups simultaneously. Still, focusing exclusively on the field of childhood bereavement, the database on treatment effects remains scarce and further analyses are warranted. Due to the fact that our initial data collection and effect size estimation preceded the first meta-analysis by Currier et al. (2007), which focused on child and adolescent intervention groups, the results presented below provide an independent estimation of effects.

Our analysis differs from Currier et al. (2007) in several ways: First, although 11 controlled studies are the same as in Currier et al., we include two additional studies that were not analyzed by Currier et al. On the other side, Currier et al. included two unpublished studies that we were unable to locate. Thus, the database overlap is 73%. Second, we did a separate analysis with uncontrolled studies as well, for a closer look at the matter. Controlled trials have been criticized for a number of threats to validity, including differential attrition rates (Kazdin, 1994). Participants seem to drop out of waiting list or placebo conditions more often, because they are not blind to treatment condition or do not want to wait any longer. Pfeffer, Jiang, Kakuma, Hwang, and
Metsch (2002), for example, reported dropout rates of 75% for children in the control condition as compared with 18% for those in the intervention group. Due to these high attrition rates in some controlled studies, it is worthwhile to also look at the effect sizes in uncontrolled studies, because they may help to judge the generalizability of controlled studies. Third, we provide separate effect-size estimations for distinct symptom areas such as depression, PTSD, and social adjustment, because of our interest in differential efficacy of treatment for different domains of outcome. For example, Kato and Mann (1999) reported an effect size of $d = 0.27$ in somatic symptoms, whereas symptoms of grief and depression remained unchanged ($d = 0.05$). Finally, we considered a number of possible moderators, that is, age, gender, amount of treatment, time since loss, treatment rationale, publication status, and, most importantly, symptom severity.

**Symptom Severity**

Allumbaugh and Hoyt (1999) did not report significant differences related to symptom severity in adults. But in his unpublished meta-analysis, Fortner (2000; see Neimeyer, 2000) reported that traumatically bereaved participants suffering from complicated grief benefited more from treatment than bereaved participants who were not functionally impaired (or at least not assessed for impairment or symptoms). Yet, the coding for complicated grief in this study merely reflected an estimation of the severity of grief and other symptoms and did not follow the diagnostic criteria outlined by Prigerson, Vanderwerker, and Maciejewski (2008). Nevertheless, this points to the possibility that grief treatment could be particularly helpful for people suffering from some form of complicated or prolonged grief (Neimeyer, 2000; Jordan & Neimeyer, 2003). Both the results of Currier et al. (2008) and the results of a meta-analysis focusing on grief treatments for adults (Rosner, Kruse, & Hagl, 2005; see Rosner & Hagl, 2007) strongly supported such a differential indication. Focusing on children and adolescents, Currier et al. (2007) found that studies including children regardless of diagnostic status as well as studies excluding overly distressed children tended to produce poorer outcomes than studies focusing on symptomatic children.
Age

According to Brent, Speece, Lin, Dong, and Yang (1996), the concept of death can be divided in three distinct components: (a) universality (i.e., the understanding that all living things must die); (b) irreversibility (i.e., the understanding that once a person dies, his or her physical body cannot ever be reanimated); and (c) nonfunctionality (i.e., the understanding that a deceased body can no longer perform any activities like a living body can). Although most kindergarten-aged children have a mature understanding of universality, the majority of them do not understand the other components until the age of 10. Given this difference concerning the concept of death, one might assume that children of different age groups may need specific interventions or show different reactions to treatment (cf. Öltjenbruns, 2001). Accordingly, Pfeffer et al. (2002) reported that younger children (6-year-olds) and young adolescents (14-year-olds) showed less reduction in symptom scores than school-age children (11-year-olds) after a manual-based bereavement group intervention. Another study showed a differential treatment effect insofar that parents reported decreased behavior problems for older, but not for younger children (Sandler et al., 1992).

Gender

Research on the role of children’s gender concerning adjustment after the loss is still scarce. Whereas Worden (1996) concluded that bereaved girls are particularly at risk for distressing symptoms, Kalter et al. (2002) found that adolescent boys showed more adjustment problems than girls, at least when evaluated by parents. However, gender might also play a role in treatment response. In a counseling program with bereaved adults, Schut, Stroebe, and van den Bout (1997) found gender-related treatment effects: Whereas men had greater benefits from emotion-focused interventions, women profited more from problem-focused treatment. Within childhood treatment studies Sandler et al. (2003) reported improvement in follow-up for girls only.

Amount of Treatment

Usually, the number of treatment sessions is linearly related to the outcome (Lambert & Ogles, 2004). Allumbaugh and Hoyt’s (1999)
results parallel this finding, with each additional treatment session increasing the effect size by .07 units. Currier et al. (2008), on the other side, found no significant correlation of session number or total intervention time and treatment outcome.

Average Time Since Loss

This variable may be important in two ways: On the one hand, for normal bereavement, a longer amount of time allows for grief symptoms to diminish naturally. On the other hand, the presence of grief symptoms many years after the bereavement may indicate a complicated grieving process. Currier et al. (2007) found that children whose losses were more distant in time showed poorer treatment outcomes. The authors explain their results by indicating that treatments offered long after the loss may have been “too ‘weak’ to produce measurable effects, perhaps after children and families had already accommodated the loss” (Currier et al., 2007, p. 257). For adults, Allumbaugh and Hoyt (1999) also reported a significant effect of time since loss, indicating that interventions taking place closer to the time of the loss are more effective than delayed interventions. In their meta-analysis based on children as well as adults, Currier et al. (2008) did not find a significant association between average time since loss and outcome.

Therapeutic Confrontation

Confrontation as a part of the intervention rationale has not been subject to investigation so far. However, the bereavement literature and clinical knowledge indicate two different approaches to treatment: clients are either allowed to freely choose what they want to talk about or the therapist provides some kind of structure. Within the structured treatments, some manuals include sessions or exercises on especially painful aspects related to the bereavement, such as talking about the circumstances when the loved one died and/or the situation at the funeral. Talking about the most painful aspects in a structured manner impedes avoidant forms of coping. Therapeutic confrontation therefore seems to be a further variable worth exploring.
Publication Type

Since Rosenthal (1979), it seems common knowledge that effect sizes in unpublished studies are smaller than in published studies. Contrary to this, Allumbaugh and Hoyt (1999) reported a reverse effect: in their meta-analysis unpublished studies reported higher effects than published studies. In order to examine this issue, we categorized studies on whether they were published in a journal or a book or whether they were unpublished theses or “gray literature” (such as unpublished agency reports).

Method

Identification of Studies

To obtain all relevant studies, we conducted a multiphase literature research. The following databases were searched using the keywords grief, grieving, bereavement, bereaved, and mourning: PsychINFO, MEDLINE, EMBASE, BIOSIS, ISI Current contents/Web of Science/Social Services Abstracts, ERIC, PILOTS, OAIster, and the Cochrane Library (DARE). To obtain theses, we searched Dissertation Abstract International, ProQuest, Theses Canada, British Theses Service, and TheO (German). Our search strategy was further refined by using truncation whenever possible or other database specific search strategies. Additionally, two German literature databases, PSYNDEXplus and SOMED, were also searched using the keyword Trauer [grief]. A search for controlled studies was conducted in the following databases: Cochrane Library (CENTRAL), PsiTri, and Current Controlled Trials. The inclusion of German literature databases produced several additional hits. Nevertheless, after checking for inclusion criteria, only one study could be added for analysis (Möhlen, 2005; see Möhlen, Parzer, Resch, & Brunner, 2005). Additionally, reference lists in primary studies and reviews were scanned. Over 40 journals, including German ones, were hand searched, especially journals on child and adolescent psychiatry, journals relating to the topics of bereavement and traumatic loss, as well as high impact journals covering clinical psychology in general. Furthermore, a Google search identified research groups focusing on bereavement.
Letters and e-mails were sent to colleagues and organizations asking for unpublished studies or studies in progress. Our database search covered all years available by the respective databases (i.e., from the first year onward) and was conducted in summer 2006. Documentation of search terms, databases and the complete list of hand-searched journals can be obtained from the first author.

Selection Criteria

Criteria for inclusion in the meta-analysis were as follows: (a) participants were bereaved children and adolescents (less than 18 years old), (b) treatment had started following the loss (instead of before), (c) the publications offered quantitative outcome measures, and (d) sufficient data were reported to allow for the computation of an effect size. Studies reporting only posttreatment or follow-up data were included as long as no other additional treatment took place between pre- and post-follow-up data collection.

A total of 91 studies and articles were obtained and screened for inclusion. All retrieved studies were rated independently by two raters for inclusion criteria. Rater concordance was 97%. In the case of disagreement, the two raters discussed the studies until reaching an agreement. If data for the effect size estimation was insufficient, the authors of the respective studies were contacted for additional information. As the studies varied considerably concerning methodological standards, two separate groups of analyses were performed, one for studies with a control group (Analysis 1) and another for those without (Analysis 2).

Studies with a Control Group

A total of 16 studies fulfilled the control-group criterion. One of these studies (Tarakoff, 1992) found no difference between treatment conditions and therefore reported only pre–post data for both groups combined. Consequently, we treated this study as uncontrolled. Of the remaining 15 studies (see Table 1), 13 had untreated control groups, which were either in a waiting-list condition or in a no-treatment condition. In one study (Sandler et al., 2003), the control group consisted of a self-study condition. As it is debatable whether this is comparable to no-treatment conditions, we report
<table>
<thead>
<tr>
<th>Study</th>
<th>Relationship to deceased</th>
<th>Type of intervention(^a)</th>
<th>Age</th>
<th>Setting</th>
<th>Intervention(^b)</th>
<th>Design/Type of control</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adams, 1995</td>
<td>Family member or friend</td>
<td>Prevention</td>
<td>5–11</td>
<td>Group</td>
<td>“Talking through,” good and bad memories, skills training</td>
<td>Random, waiting list</td>
<td>69</td>
</tr>
<tr>
<td>Black &amp; Urbanowicz, 1987</td>
<td>Parent</td>
<td>Prevention</td>
<td>0–16</td>
<td>Family</td>
<td>Promote mourning, circumstances of death, intensification, communication about death</td>
<td>Random, no treatment</td>
<td>83</td>
</tr>
<tr>
<td>Dalton &amp; Krout, 2005</td>
<td>Family member or friend</td>
<td>Prevention</td>
<td>12–18</td>
<td>Group</td>
<td>Music therapy based on tasks of grieving</td>
<td>No randomization, waiting list</td>
<td>20</td>
</tr>
<tr>
<td>Goenjian et al., 1997/Goenjian et al., 2005</td>
<td>Family member</td>
<td>Psychotherapy</td>
<td>10–13</td>
<td>Group and individual</td>
<td>Psychoeducation, confrontation, working on maladaptive cognitions, interplay between trauma and grief, grief management, resume developmental progression</td>
<td>No randomization, no treatment</td>
<td>64</td>
</tr>
<tr>
<td>Study</td>
<td>Participants</td>
<td>Type</td>
<td>Age</td>
<td>Intervention</td>
<td>Randomization</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>-----------------------</td>
<td>--------------</td>
<td>---------------</td>
<td>-----</td>
<td>-----------------------------------------------------------------------------</td>
<td>---------------</td>
<td>----</td>
<td></td>
</tr>
<tr>
<td>Hilliard, 2001</td>
<td>Loved one</td>
<td>Psychotherapy</td>
<td>6–11</td>
<td>Group Cognitive-behavioral music therapy: Circumstances of death, normalization, education, and expression of grief and anger</td>
<td>No randomization, waiting list</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>Huss, 1997</td>
<td>Primary family member</td>
<td>Prevention</td>
<td>10–12</td>
<td>Group Review stages of change, focus on emotions and memories, intensification of emotions</td>
<td>Random</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>Loy, 2000</td>
<td>Primary family member and grandparents</td>
<td>Prevention</td>
<td>10–17</td>
<td>“Bereavement” camp and two parent sessions Normalizing of bereavement, social sharing</td>
<td>Random, waiting list</td>
<td>61</td>
<td></td>
</tr>
<tr>
<td>Pfeffer, Jiang, Kakuma, Hwang, &amp; Metsch, 2002</td>
<td>Primary family member</td>
<td>Prevention</td>
<td>6–15</td>
<td>Group for children and parents separately Attachment theory, responses to loss through suicide, coping skills</td>
<td>No randomization, no treatment</td>
<td>41</td>
<td></td>
</tr>
<tr>
<td>Poijula, Dyregov, Wahlberg, &amp; Jokelainen, 2001</td>
<td>Friend (Suicide of classmate)</td>
<td>Prevention</td>
<td>14–17</td>
<td>Group Talking through and debriefing</td>
<td>No randomization, no treatment</td>
<td>87</td>
<td></td>
</tr>
</tbody>
</table>

(Continued)
<table>
<thead>
<tr>
<th>Study</th>
<th>Relationship to deceased</th>
<th>Type of intervention</th>
<th>Age</th>
<th>Setting</th>
<th>Intervention $^b$</th>
<th>Design/Type of control</th>
<th>$N$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ryan, 1982</td>
<td>Primary family member and grandparents</td>
<td>Prevention</td>
<td>7–17</td>
<td>Group</td>
<td>Support group: Normalization, talking about death, expression of emotions</td>
<td>Random, waiting list</td>
<td>21</td>
</tr>
<tr>
<td>Sandler et al., 2003</td>
<td>Parent or caregiver</td>
<td>Prevention</td>
<td>8–16</td>
<td>Group</td>
<td>Improvement of caregiver–child relationship, coping and problem solving skills</td>
<td>Random, alternative treatment: Self-study program</td>
<td>230</td>
</tr>
<tr>
<td>Sandler et al., 1992</td>
<td>Parent</td>
<td>Prevention</td>
<td>7–17</td>
<td>Group</td>
<td>Improvement of parental demoralization, fostering parental warmth and family cohesion, support</td>
<td>Random, no treatment</td>
<td>54</td>
</tr>
<tr>
<td>Study</td>
<td>Primary Family Member</td>
<td>Prevention</td>
<td>Group</td>
<td>Circumstances of death, positive and negative memories, discussion of feelings of guilt and anger</td>
<td>Randomization, Waiting List</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>-----------------------</td>
<td>------------</td>
<td>-------</td>
<td>---------------------------------------------------------------------------------------------------</td>
<td>-----------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tonkins &amp; Lambert, 1996</td>
<td>Primary family member</td>
<td>Prevention</td>
<td>7–11 Group</td>
<td>No randomization, waiting list</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wilson, 1995</td>
<td>Primary family member and grandparents</td>
<td>Prevention</td>
<td>5–11 Group</td>
<td>Pilot study: Alternative treatment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Pilot study</td>
<td></td>
<td></td>
<td>Play therapy (normalization, caring, and acceptance) vs. Counseling: Tasks of grieving, expression of needs, setting of goals, circumstances of funeral</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Main study</td>
<td></td>
<td></td>
<td>Main study: Random, waiting list</td>
<td>22</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*To evaluate symptom severity as moderator, studies were coded as prevention when participants were either non-symptomatic, or information on symptoms was missing, or coded as psychotherapy when participants were symptomatic, functionally impaired, or reliably diagnosed as suffering from complicated grief.

*As described in the respective publication.
results with and without the study of Sandler et al. (2003). One other study (the pilot study from Wilson, 1995) used an alternative treatment as control group and could only be used in pre–post test comparisons for controlled groups. For another study (Goenjian et al., 1997; Goenjian et al., 2005), there was no assessment immediately after therapy. Hence, this study was only included in the follow-up analysis. Altogether 13 (Sandler et al., 2003, included) studies reported complete data for comparison of treatment and control group. For these, the number of outcome measures varied from 1 to 15 per study, and overall 65 single effect sizes were obtained for each of these measures. Eight studies reported sufficient data for pre–post comparisons. For these pre–post analyses of controlled studies 50 effect sizes based on different measures could be obtained.

Only two studies provided follow-up data, yet they differed substantially in terms of the follow-up interval: while one study reported data from about two weeks after treatment, the other study reported data from 16 and 40 months, respectively. The follow-up analyses were based on 26 single effect sizes; 6 of these effect sizes were based on data dealing with changes from pre-treatment to follow-up only.

Studies Without a Control Group

A total of 12 uncontrolled studies were identified. Since one study reported only means and no other statistics, this study was excluded from the analysis (Johnson-Schroetlin, 2000). Another study (Tarakoff, 1992) had a controlled design, but only reported data for both groups combined and was therefore treated as uncontrolled. Thus, the second analysis was based on 12 uncontrolled studies that reported only pre–post data (see Table 2). The number of the studies’ outcome measures varied from 1 to 32. Overall, 81 effect sizes were computed. None of the uncontrolled studies provided follow-up data.

Coding and Data Reduction

Studies used a wide range of settings and designs for example, from home based family sessions to group workshops including several families or from bereavement camps lasting several days
### Table 2: Description of Studies Included in Meta-Analysis 2: Uncontrolled Studies (Analysis 2)

<table>
<thead>
<tr>
<th>Study</th>
<th>Relationship to deceased</th>
<th>Type of intervention</th>
<th>Age</th>
<th>Setting</th>
<th>Intervention</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carver, 2000</td>
<td>Not defined</td>
<td>Prevention</td>
<td>6–12</td>
<td>Support group and parent group</td>
<td>Support group</td>
<td>11</td>
</tr>
<tr>
<td>Cohen, Mannarino, &amp; Knudsen, 2004</td>
<td>Family member</td>
<td>Psychotherapy</td>
<td>6–17</td>
<td>Individual and family</td>
<td>Psychoeducation, relaxation, coping skills, trauma narrative, trauma reminders</td>
<td>22</td>
</tr>
<tr>
<td>Layne, Pynoos, et al., 2001</td>
<td>Family member or friend</td>
<td>Psychotherapy</td>
<td>15–20</td>
<td>Group</td>
<td>Psychoeducation, confrontation, working on maladaptive cognitions, interplay between trauma and grief, grief management, resume developmental progression</td>
<td>55</td>
</tr>
<tr>
<td>Möhlen, 2005</td>
<td>Family member or friend</td>
<td>Prevention</td>
<td>10–16</td>
<td>Group, individual and family</td>
<td>Psychoeducation, relaxation, art therapy techniques, coping</td>
<td>10</td>
</tr>
</tbody>
</table>

(Continued)
<table>
<thead>
<tr>
<th>Study</th>
<th>Relationship to deceased</th>
<th>Type of intervention&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Age</th>
<th>Setting</th>
<th>Intervention&lt;sup&gt;b&lt;/sup&gt;</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Opie et al., 1992</td>
<td>Family member</td>
<td>Prevention</td>
<td>9–15</td>
<td>Group</td>
<td>Group discussion of loss, perception and feelings, coping skills</td>
<td>8</td>
</tr>
<tr>
<td>Quarmby, 1993</td>
<td>Parent</td>
<td>Prevention</td>
<td>12–15</td>
<td>Group</td>
<td>Counseling</td>
<td>6</td>
</tr>
<tr>
<td>Ross &amp; Hayes, 2004</td>
<td>Family member or friend</td>
<td>Prevention</td>
<td>7–11</td>
<td>Group</td>
<td>Emotional awareness, discussion of difficult feelings, narratives, solution focused and future oriented thinking</td>
<td>5</td>
</tr>
<tr>
<td>Salloum, Avery, &amp; McClain, 2001</td>
<td>Family member or friend</td>
<td>Prevention</td>
<td>11–19</td>
<td>Group</td>
<td>Psychoeducation concerning PTSD and grief, coping skills, safety, identification and sharing of feelings, revenge and anger</td>
<td>37</td>
</tr>
<tr>
<td>Author(s)</td>
<td>Role(s)</td>
<td>Type of Intervention</td>
<td>Age Range</td>
<td>Setting</td>
<td>Techniques</td>
<td></td>
</tr>
<tr>
<td>---------------------------------</td>
<td>------------------------------</td>
<td>----------------------</td>
<td>------------</td>
<td>---------</td>
<td>-----------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Saltzman, Pynoos, Layne, Steinberg, &amp; Aisenberg, 2001</td>
<td>Family member or friend</td>
<td>Psychotherapy</td>
<td>11–14</td>
<td>Group</td>
<td>Psychoeducation, confrontation, working on maladaptive cognitions, grief management, problem solving skills</td>
<td></td>
</tr>
<tr>
<td>Schilling, Koh, Abramovitz, &amp; Gilbert, 1992</td>
<td>Parent or caregiver</td>
<td>Prevention</td>
<td>6–12</td>
<td>Group</td>
<td>Psychodynamic</td>
<td></td>
</tr>
<tr>
<td>Tarakoff, 1992</td>
<td>Family member</td>
<td>Prevention</td>
<td>11–14</td>
<td>Group</td>
<td>Tasks of grieving, expression of feelings, development of a narrative</td>
<td></td>
</tr>
</tbody>
</table>

*a*To evaluate symptom severity as moderator, studies were coded as prevention when participants were either non symptomatic or information on symptoms was missing, or coded as psychotherapy when participants were symptomatic, functionally impaired, or reliably diagnosed as suffering from complicated grief.

*b*As described in the respective publication.
(Loy, 2000) to more traditional group meetings, lasting two hours. The number of treatment hours was therefore used to obtain an estimation of treatment dosage. In order to compare studies using confrontation as part of their intervention strategy to those that did not, descriptions of the interventions were rated as follows: if the intervention included a structured discussion or exercise relating to the actual event of loss or aspects thereof (e.g., talking about the day of the funeral), then it was rated as “confrontation” and if the intervention only offered the opportunity to share one’s feelings, then it was rated as “no confrontation.” Further coding of therapeutic strategies was not possible, as the descriptions of the interventions were not always detailed enough.

In order to allow for an estimation of differential treatment effects regarding severity of symptoms, we categorized grief treatments as either preventive or psychotherapeutic interventions (see Tables 1 and 2). Participants in the prevention category had to be bereaved but were either non-symptomatic or information on symptoms was missing, or participants were heterogeneous regarding symptom status. On the other hand, participants in the psychotherapy category had to be symptomatic, functionally impaired, or reliably diagnosed as suffering from complicated grief. Inclusion criteria adhered to a strict differentiation: only studies requiring the criterion of psychopathological symptoms and/or heightened grief scores in all participants were coded as psychotherapy. This type of grouping resulted in a heterogeneous “prevention” group, with some studies in this group excluding participants who showed symptoms of any kind (e.g., Huss, 1997) and other studies including some children with diagnoses of manifest syndromes also (e.g., Möhlen, 2005).

Data reduction was performed in three steps. In the first step, each measure was assigned to a symptom area (namely grief, depression, anxiety, posttraumatic symptoms, social adjustment, well-being, somatic symptoms/health, and other). In the second step, the respective effect sizes were pooled for each individual study. Finally, these effect sizes were averaged across studies.

**Effect Size Calculation**

In order to compute effect sizes, we used Hedges’s $g$, which is the difference between conditions’ means divided by their pooled
standard deviations, corrected for small sample bias (Hedges & Olkin, 1985). If means and standard deviations were not available, the algebraically equivalent effect sizes of t and F values, or exact p values based on t tests, were computed (Ray & Shadish, 1996) or estimated from proportions of group successes as suggested by Lipsey and Wilson (2001). Computations are based on formulae provided by Ray and Shadish and Lipsey and Wilson, respectively. Furthermore, to ensure comparability to other meta-analyses, we also report Cohen’s d.

Of the 65 effect sizes originating from controlled studies, 50.8% were calculated from means and standard deviations and 16.9% from t statistics (i.e., 67.7% of effect sizes were calculated directly). In 13.9%, effect sizes were estimated from F ratios and in 18.5% from other sources (dichotomous proportions, chi-square tests, U values). A sensitivity analysis revealed that there was no significant difference between effect sizes derived from different estimation procedures, $Q(3, 12) = 10.00, p = .62$. Therefore, different calculation procedures for effect sizes did not appear to have introduced a systematic bias.

Within the uncontrolled studies 95.8% of 65 effect sizes were calculated either from means and standard deviations or t statistics; 4.2% were estimated from p values. For an estimation of the file drawer problem, fail-safe N was computed (i.e., the number of unpublished, non-significant studies that would be necessary to render the results of our meta-analysis non-significant with $p < .05$). Rosenthal (1991) argued that it is unlikely that file drawers contain more than five times as many studies as the studies published. Therefore, the “critical number” of studies (i.e., the number of unpublished studies one could at most expect according to Rosenthal) is usually also reported along with the fail-safe N.

In order to derive effect sizes for changes between pre- and posttest we computed the standardized mean gain as developed by Becker (1988) and slightly modified by Lipsey and Wilson (2001). For this measure, the correlation of pretest and posttest values is necessary. Because this correlation was only given in one of the studies we used a global estimate of .70 as an approximation for the average test–retest reliability of the instruments used.

Effect sizes were computed for each assessment scale of the respective studies and then averaged across all dependent variables. Effect sizes were weighted for n, because sample sizes
sometimes varied between the different subscales. Within studies, effect sizes are stochastically dependent on each other, and Gleser and Olkin (1994) discussed various approaches to combine such effect sizes. The dependency of the assessment scales themselves could not be accounted for because intercorrelations were unknown. Therefore, our approach resulted in more conservative averages. A similar averaging procedure was used to differentiate effects for symptom areas (i.e., grief, depression, anxiety, etc.).

Individual effect sizes were weighted by their inverse variance (Hedges & Olkin, 1985) to account for sample size differences. Hence, studies with larger sample sizes have a stronger impact on our results than smaller studies. In order to compute the average effect sizes across studies or functional domains, we adopted the random-effects model (Lipsey & Wilson, 2001). This model assumes variability in the population of effects in addition to variability in sampling error. The fixed-effects model, on the other hand, assumes that the entire variability between effect sizes is due to sampling error. Because studies differed in many characteristics such as therapists, therapy settings, reasons for grief, etc., the added variance component of the random-effects model accommodates these differences and allows for a greater generalization of results. For example, one can generalize these results to treatment conditions that do not exactly resemble those of the analyzed studies. In order to test for moderator effects, we used a weighted, random-effects regression analysis procedure as proposed by Lipsey and Wilson.

**Results**

The analyses were based on an overall $N=1073$, with $n=812$ from controlled studies and $n=261$ from uncontrolled studies. The percentage of studies coded as *psychotherapy* (vs. *prevention*) is 25% within the uncontrolled studies as compared with 13% within the controlled studies.

**Controlled Studies**

**OVERALL EFFECT OF INTERVENTIONS**

Table 1 displays controlled studies included in Analysis 1. Eight studies used a randomized design. No study reported an
intent-to-treat analysis. Therefore, the effect sizes were based on completer analyses only. Furthermore, the reporting of drop-outs varied substantially across studies, making it impossible for us to reliably report on differential drop-out rates. Table 3 displays the effect sizes for the respective areas of outcome.

A specific characteristic for studies with a large number of outcome measures such as the study by Sandler et al. (2003) needs to be pointed out: The authors of that study reported a very large number of proximal and distal outcomes due to the elaborate theory-driven approach of their study. For the calculation of effect sizes, all distal outcomes and a number of proximal outcomes (i.e., negative thoughts about the stressful event, self-esteem, control-related beliefs, positive coping, and active inhibition) were taken into account. In the case of many outcome measures (which of course, did not all reflect positive change), the overall effect size for one study usually became smaller, because the respective outcomes were averaged.

The overall effect size for all 13 controlled studies was 0.35 ($p < .01$, confidence interval ranging from 0.15 to 0.57). If Cohen’s $d$ instead of Hedges’s $g$ was used as the effect size estimate, an overall effect size of 0.37 resulted. Excluding Sandler et al. (2003), who used a self-study control group that may have resulted in a downward biased effect size, the effect size ($g$) of the 12 remaining studies was 0.39 ($p < .01$, confidence interval ranging from 0.16 to 0.63). The fail-safe $N$ for this effect size was 326 studies (255 excluding Sandler et al., 2003). Taking into account a critical number of 75 studies (70 excluding Sandler et al., 2003), it is highly unlikely that there are enough unpublished studies to contradict these results.¹

Seven of the controlled studies reported pre–post measures. Their weighted average effect size was 0.46 ($z = 2.65, p < .01$). Fail-safe $N$ and critical number were 241 and 45, respectively. For those two controlled studies that reported follow-up data (Goenjian et al., 1997; Wilson, 1995) a weighted average effect size of 0.52 ($z = 1.60, ns$) resulted, based on 20 individual effect sizes. Fail-safe $N$ is 18 compared with a critical number of 20 studies. Due to the

¹Newer and more sophisticated methods to test for publication bias include funnel plots of effect size by standard error (or total $N$) and the trim-and-fill method (Duval & Tweedie, 2000). Using a random-effects model, the trim-and-fill method identified no studies to trim and therefore yielded no adjusted estimate for the effect size. Statistical tests for funnel plot symmetry were not significant.
<table>
<thead>
<tr>
<th>Study</th>
<th>Total</th>
<th>Grief</th>
<th>Depression</th>
<th>Anxiety</th>
<th>Posttraumatic symptoms</th>
<th>Social adjustment</th>
<th>Well-being</th>
<th>Somatic symptoms</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adams, 1995</td>
<td>0.17</td>
<td>0.44</td>
<td></td>
<td>0.06</td>
<td></td>
<td>0.02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black &amp; Urbanowicz, 1987</td>
<td>0.72</td>
<td>0.78</td>
<td></td>
<td>0.76</td>
<td></td>
<td>0.56</td>
<td>1.07</td>
<td>0.59</td>
<td></td>
</tr>
<tr>
<td>Dalton &amp; Krout, 2005</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.63</td>
</tr>
<tr>
<td>Hilliard, 2001</td>
<td>0.99</td>
<td>1.12</td>
<td>0.23</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.31</td>
</tr>
<tr>
<td>Huss, 1997</td>
<td>0.32</td>
<td>0.40</td>
<td></td>
<td></td>
<td></td>
<td>0.26</td>
<td></td>
<td></td>
<td>0.44</td>
</tr>
<tr>
<td>Loy, 2000</td>
<td>0.05</td>
<td>0.09</td>
<td>0.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pfeffer et al., 2002</td>
<td>0.53</td>
<td>0.37</td>
<td>1.22</td>
<td>0.10</td>
<td></td>
<td>0.44</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poijula et al., 2001</td>
<td>0.03</td>
<td>0.05</td>
<td></td>
<td>0.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ryan, 1982</td>
<td>0.40</td>
<td></td>
<td>0.32</td>
<td></td>
<td></td>
<td>0.50</td>
<td></td>
<td></td>
<td>0.47</td>
</tr>
<tr>
<td>Sandler et al., 1992</td>
<td>0.45</td>
<td>0.54</td>
<td></td>
<td></td>
<td></td>
<td>0.36</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sandler et al., 2003</td>
<td>0.18</td>
<td>0.11</td>
<td></td>
<td></td>
<td></td>
<td>0.25</td>
<td></td>
<td></td>
<td>0.12</td>
</tr>
<tr>
<td>Tonkins &amp; Lambert, 1996</td>
<td>0.06</td>
<td>0.67</td>
<td>0.11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.27</td>
</tr>
<tr>
<td>Wilson, 1995 (main study)</td>
<td>0.19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.22</td>
</tr>
<tr>
<td>Total</td>
<td>0.35</td>
<td>0.59</td>
<td>0.22</td>
<td>0.47</td>
<td>0.05</td>
<td>0.28</td>
<td>0.96</td>
<td>0.59</td>
<td>0.20</td>
</tr>
</tbody>
</table>
limited power because of the small sample size this result should be considered a rough estimate. Pre-follow-up analyses (including as a third study the results of a pilot study by Wilson, 1995) resulted in a weighted average effect size of 0.74 (z = 2.56, p < .05).

CHANGES IN SYMPTOM AREAS

Concerning the areas of outcome, the averaged effect sizes varied from 0.05 for posttraumatic symptoms to 0.96 for general well-being (see Table 3). Yet within the controlled studies only two studies reported on posttraumatic symptom change. A test for heterogeneity of effect sizes over areas of outcome was significant (Q = 15.95, df = 7, p < .05, I² = 56.1).

MODERATOR EFFECTS

As possible moderators, we considered symptom severity, age, gender, amount of treatment, mean length of time since loss, therapeutic confrontation as part of the treatment rationale, and publication type. The test for heterogeneity was highly significant (Q = 167.70, df = 12, p < .001, I² = 92.84) and therefore warranted the search for moderators. However, since not enough studies reported different statistics relating to gender, this particular analysis had to be omitted.

Symptom severity. Comparing studies of patients who showed some level of distress, impairment, or a diagnosis (categorized as psychotherapy), with studies in which participants either were non-symptomatic or heterogeneous regarding symptom status or information on symptoms was missing (categorized as prevention), the inverse variance weighted regression showed only a trend towards a significant moderator effect, Q(12) = 2.88, β = .43, p < .10; v = .03: The two studies with patients at risk for complicated grief (psychotherapy, Goenjian et al., 1997; Hilliard, 2001) yielded a weighted average effect size of 0.68 (z = 1.47, ns) as compared to 0.34 (z = 3.15, p < .01) for the 12 remaining studies (prevention).

Age. The age ranges reported in the studies varied considerably. While some studies reported mean age, others reported the age range. Therefore, it was impossible to analyze the effect of age using regression analysis. Nevertheless, the data allowed differentiating with sufficient accuracy between two age groups: one
younger than 12 years of age, the other 12 years or older. However, the following results should be interpreted with caution.

Comparing these two age groups for the controlled studies, a weighted average effect size of 0.34 \( (z = 3.20, p < .01) \) for children younger than 12 years of age was found versus an effect size of 0.41 \( (z = 1.84, p < .10) \) for the older age group. An inverse-variance weighted regression showed no significant effect for age, \( (Q = 0.001, df = 1, \text{ns}; v = .06) \).

**Amount of treatment.** In order to compare the effect of treatment dosage, three groups were established varying in the amount of therapy sessions: one short-term intervention group (1 to 9 hours of treatment), an average term intervention group (10 to 16 hours of treatment), and a long-term intervention group (17 and more hours of treatment). Table 4 displays the respective results. Due to negative random error variance components, the analysis was based on a fixed-effects model.\(^2\)

Although short- and mid-term treatments seemed to show larger effect sizes as compared to long-term treatments, this difference was not statistically significant, \( Q(12) = 1.55, \text{ns} \). Because the results concerning designs with higher treatment dosage were only based on two studies (Sandler et al., 2003; Loy, 2000) they need to be interpreted with caution. Furthermore, one of these studies

\(^2\)The random-effects model is based on a random variance component. A negative variance is not defined and if the variance component is negative, no further computations are possible. If the chi-square test for heterogeneity of effect sizes is not significant, as indicated by the \( Q \) statistic, a fixed-effects model is feasible (see Friedman, 2000).

<table>
<thead>
<tr>
<th>Amount of treatment</th>
<th>Number of studies(^a)</th>
<th>Hedges's (g)</th>
<th>(z)</th>
<th>95% confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>1–8 hours</td>
<td>7 (25)</td>
<td>0.43</td>
<td>3.53***</td>
<td>0.19–0.66</td>
</tr>
<tr>
<td>10–16 hours</td>
<td>4 (27)</td>
<td>0.42</td>
<td>2.39*</td>
<td>0.08–0.77</td>
</tr>
<tr>
<td>17 and more</td>
<td>2 (13)</td>
<td>0.14</td>
<td>0.94**</td>
<td>-0.15–0.42</td>
</tr>
</tbody>
</table>

\(^a\)Number of individual effect sizes in parentheses.
\(*p < .05. ***p < .001.\)

**Table 4** Effect Sizes of Amount of Treatment: Fixed-Effects Model with Weighted Average Effect Size, \(z\) Statistic and Confidence Interval (Analysis 1)

Note. Fixed-effects model due to negative variance component (see Footnote 2).
(Sandler, 2003) featured a self-study control group, which may have a general downward bias in effect sizes.

**Time since bereavement.** The average length of time between bereavement and intervention varied considerably across studies. Three groups were differentiated: less than 6 months since bereavement, 6 to 12 months, and longer than 12 months (up to 3 years) since bereavement. Differences were not statistically significant, $Q(12) = 1.05$, *ns*, see Table 5.

**Therapeutic confrontation.** Four studies with confrontational elements (Black & Urbanowicz, 1987; Hilliard, 2001; Tonkins & Lambert, 1996; Poijula, Dyregrov, Wahlberg, & Jokelainen, 2001) yielded a weighted average effect size of 0.38 ($z=1.63$, *ns*), whereas nine studies without confrontation yielded an average effect size of 0.35 ($z=2.75$, *p* < .01), hence no statistical difference was found, $Q(12) = 12.13$, *ns*. The study by Goenjian and colleagues (Goenjian et al., 1997) also included confrontational elements as part of the intervention but was not included in this analysis, because no postintervention data were available.

**Publication type.** Weighted average effect size for eight journal articles was 0.41 ($z=4.12$, *p* < .001) vs. 0.18 ($z=1.21$, *ns*) for five unpublished theses. Those were the only two publication types for controlled studies in our meta-analysis. The effect size calculation in this sub-analysis was based on the fixed-effects model due to negative variance components (see Footnote 2). No statistical difference was confirmed, $Q(12) = 1.53$, *ns*.

<table>
<thead>
<tr>
<th>Time since bereavement</th>
<th>Number of studies$^a$</th>
<th>Hedges’s $g$</th>
<th>$z$</th>
<th>95% confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;6 months</td>
<td>3 (23)</td>
<td>0.34</td>
<td>1.47$^+$</td>
<td>−0.11 to 0.79</td>
</tr>
<tr>
<td>6–12 months</td>
<td>7 (32)</td>
<td>0.21</td>
<td>2.72$^{**}$</td>
<td>0.06 to 0.36</td>
</tr>
<tr>
<td>&gt;12 months</td>
<td>3 (10)</td>
<td>0.81</td>
<td>1.94$^*$</td>
<td>−0.01 to 1.63</td>
</tr>
</tbody>
</table>

$^a$Number of individual effect sizes in parentheses.

$^*$ *p* < .05, one-tailed. $^{**}$ *p* < .01. $^+$ *p* < .10, one-tailed.
PROMISING TREATMENT MODELS

Additionally, we were interested whether especially promising treatment models could be identified. Therefore, we explicitly searched for studies applying the same treatment rationale or type of intervention, but which were conducted by independent researchers. We found that the most successful interventions were two music therapy interventions (Hilliard, 2001; Dalton & Krout, 2005; see Table 3). An inverse variance weighted one-way analysis of variance (ANOVA) showed a significant difference between music therapy interventions with $g = 1.36$ versus $g = 0.27$ for other interventions, $Q(1, 11) = 10.15, p < .01; v = .00$. As these two studies yielded comparably large effect sizes (0.99 and 1.63, respectively), their outcomes contributed substantially to the overall effect size for controlled studies. With the exception of these two, all studies within the group of controlled studies applied differing treatment models.

Uncontrolled Studies

OVERALL EFFECT OF INTERVENTIONS

In order to further explore treatment efficacy, we computed a second meta-analysis on the basis of the retrieved uncontrolled studies. Table 2 describes the respective studies (Analysis 2) and Table 6 displays the studies’ effect sizes. As these studies were uncontrolled, effect sizes are solely based on pre–post data. The overall weighted average effect size for 12 uncontrolled studies (pre–post) was 0.49 ($z = 3.57, p < .001$). Fail-safe $N$ was 1,154 studies compared with a critical number of 70 studies. However, the fact that these studies were uncontrolled precludes firm conclusions regarding the effectiveness of these treatments, as changes observed might reflect natural attenuation of grief-related distress over time.

CHANGES IN SYMPTOM AREAS

Concerning the areas of outcome, the largest change was observed in grief and PTSD symptoms and the smallest effect in depression and somatic symptoms. However, compared to the controlled studies where averaged effect sizes varied between 0.05 for PTSD symptoms and 0.96 for well-being, effect sizes for uncontrolled studies varied only between 0.36 for depressive
<table>
<thead>
<tr>
<th>Study</th>
<th>Total</th>
<th>Grief</th>
<th>Depression</th>
<th>Anxiety</th>
<th>Posttraumatic symptoms</th>
<th>Social adjustment</th>
<th>Well-being</th>
<th>Somatic symptoms</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carver, 2000</td>
<td>-0.13</td>
<td>-0.13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cohen et al., 2004</td>
<td>1.08</td>
<td>0.98</td>
<td>0.93</td>
<td>1.27</td>
<td>1.46</td>
<td>0.64</td>
<td>0.88</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Layne, Pynoos, et al., 2001</td>
<td>0.95</td>
<td>0.95</td>
<td>0.77</td>
<td></td>
<td></td>
<td>1.12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Möhlen, 2005</td>
<td>0.61</td>
<td>0.45</td>
<td>0.36</td>
<td>0.61</td>
<td></td>
<td>1.01</td>
<td></td>
<td></td>
<td>0.55</td>
</tr>
<tr>
<td>Opie et al., 1992</td>
<td>0.36</td>
<td></td>
<td>0.18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quarmby, 1993</td>
<td>0.51</td>
<td>0.36</td>
<td></td>
<td></td>
<td></td>
<td>0.51</td>
<td>0.65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ross &amp; Hayes, 2004</td>
<td>0.55</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.05</td>
<td>0.42</td>
<td></td>
<td>1.18</td>
</tr>
<tr>
<td>Salloum et al., 2001</td>
<td>0.59</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.59</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Saltzman et al., 2001</td>
<td>0.43</td>
<td>0.66</td>
<td>0.11</td>
<td></td>
<td></td>
<td>0.46</td>
<td></td>
<td></td>
<td>0.50</td>
</tr>
<tr>
<td>Schilling et al., 1992</td>
<td>0.48</td>
<td></td>
<td>0.16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.80</td>
</tr>
<tr>
<td>Tarakoff, 1992</td>
<td>0.19</td>
<td>0.23</td>
<td>0.22</td>
<td></td>
<td></td>
<td>0.12</td>
<td></td>
<td></td>
<td>0.32</td>
</tr>
<tr>
<td>Webb-Ferebee, 2003</td>
<td>0.20</td>
<td>0.33</td>
<td>0.25</td>
<td></td>
<td></td>
<td>0.26</td>
<td>0.18</td>
<td></td>
<td>0.07</td>
</tr>
<tr>
<td>Total</td>
<td>0.49</td>
<td>0.89</td>
<td>0.36</td>
<td>0.42</td>
<td>0.83</td>
<td>0.39</td>
<td>0.61</td>
<td>0.55</td>
<td>0.50</td>
</tr>
</tbody>
</table>
symptoms and 0.89 for grief symptoms (see Table 6). A test for heterogeneity of effect sizes over areas of outcome was not significant ($Q = 6.10$, $df = 7$, ns, $I^2 = 0$).

**MODERATOR EFFECTS**

A test for heterogeneity of effect sizes was highly significant ($Q = 29.08$, $df = 11$, $p < .01$, $I^2 = 62.17$) and therefore warranted the search for moderators. Again, since not enough studies reported different statistics relating to gender, we had to omit this analysis.

**Symptom severity.** We compared three studies (Saltzman et al., 2001; Layne, Pynoos, et al., 2001; Cohen, Mannarino, & Knudsen, 2004) with participants showing a risk for developing complicated grief (categorized as *psychotherapy*) with the remaining nine studies. The weighted average effect size for psychotherapy clients resulted in 0.82 ($z = 3.25$, $p < .01$), while the prevention studies yielded an overall weighted effect size of 0.35 ($z = 2.16$, $p < .05$). Here, the psychotherapy category showed significantly better results, $Q(12) = 8.91$, $\beta = .67$, $p < .01$; $v = .02$.

**Age.** In our comparisons of the two age groups (younger than 12 years of age vs. 12 years and older) the pre–post effect size was slightly larger for the older age group: Average weighted effect size was 0.52 ($z = 3.01$, $p < .01$) for children 12 years and older versus 0.43 ($z = 1.95$, $p < .10$) for those younger than 12 years of age. Nevertheless, a weighted random-effects regression showed no significant effect for age as moderator ($Q = 0.1$, $df = 1$, ns; $v = .09$).

**Amount of treatment.** Studies with patients in long-term treatment (17 hours and more) had an average weighted effect size of 0.69 ($z = 2.78$, $p < .01$) while studies with patients receiving average-term treatment (10 to 16 hours) had an overall effect size of 0.58 ($z = 2.72$, $p < .01$). The uncontrolled studies from the short-term group (9 and less hours of treatment) showed an overall effect size of only 0.14 ($z = 0.54$, ns). An inverse variance weighted regression was highly significant ($Q = 7.91$, $df = 1$, $\beta = .64$, $p < .01$; $v = .02$), confirming the larger effect size for those in long-term treatment.
**Time since bereavement.** The more time had passed since bereavement, the larger were the effect sizes. Uncontrolled studies revealed overall effect sizes of 0.26 ($z = 1.30, \text{ ns}$), 0.67 ($z = 2.68, p < .01$), and 0.70 ($z = 2.50, p < .05$) for the respective three groups (less than 6 months, 6 to 12 months, more than 12 months). These effect sizes partially corroborate the non-significant result for the controlled studies. Moreover, here the inverse variance weighted regression was highly significant: $Q(12) = 9.08, p < .01; \beta = .64, p < .01; v = .01$.

**Therapeutic confrontation.** For uncontrolled studies, there was a pronounced effect of structured confrontation as intervention strategy. Four studies applying confrontation showed a weighted average effect size (pre–post) of 0.79 ($z = 3.35, p < .001$), while the studies without confrontation had an average effect size of 0.34 ($z = 2.01, p < .05$). The inverse variance weighted regression showed that this effect is significant, $Q(12) = 8.80, \beta = .67, p < .01; v = .02$.

**Publication type.** The comparison of nine studies published in journals versus three unpublished theses yielded an effect size of 0.64 ($z = 3.97, p < .001$) for studies published in journals contrasted with an effect size of 0.10 ($z = 0.39, \text{ ns}$) for the unpublished ones. A weighted random-effects regression showed a highly significant effect, $Q(12) = 12.78, \beta = .75, p < .001; v = .01$.

**PROMISING TREATMENT MODELS**

Although treatment models differed substantially, we were able to retrieve four studies using the same approach, referred to as “trauma/grief-focused school based brief psychotherapy” (Goenjian et al., 1997; Layne, Pynoos, et al., 2001; Möhlen, 2005; Saltzman et al., 2001). While Goenjian et al. used a control group design, the other three did not. Study participants were multiply traumatized adolescents. This treatment model yielded a weighted average effect size of 0.78 ($z = 4.67, p < .001$) as compared with all other studies with an average effect size of 0.42 ($z = 4.77, p < .001$). An inverse variance weighted one-way ANOVA showed the difference was approaching significance, $Q(1, 11) = 3.76, p = .05; v = .07$. 
Some narrative reviews published in the area of childhood grief suggest that there are not enough studies warranting a meta-analysis. However, we located 15 studies with a control group design and 12 studies with an uncontrolled design that provided enough information to be analyzed. Altogether, these studies represent a total of 1,073 children and adolescents. Along with the meta-analysis of Currier et al. (2007), which is based on controlled studies only and represents 783 children, our meta-analysis is therefore the first attempt in systematically evaluating the overall effect of treatment of childhood grief. Yet, Currier et al. included different studies and applied a slightly different statistical procedure than we did. This led to partially different results, which will be discussed as follows.

**Overall Effect Size**

The omnibus effect sizes in our meta-analysis of controlled studies and uncontrolled studies are 0.35 and 0.49, respectively. Our results thus indicate a small to moderate treatment effect. Because our study is based on the results of two separate meta-analyses, it is possible to compare results across the two analyses (controlled vs. uncontrolled studies): If the results support each other, this points to the robustness of the results and allows a more thorough interpretation.

Compared to the meta-analysis by Currier et al. (2007), our results favor treatment for bereaved children and adolescents. The difference between our results and those of Currier et al. (2007), who reported an effect size of 0.14, can be explained by the inclusion of different studies: Currier et al. (2007) included two unpublished studies (Housely, 1996; Zebrowski, 2000), both dissertations, which we were unable to locate. On the other hand, our meta-analysis on controlled studies included two additional published studies (Dalton & Krout, 2005; Poijula et al., 2001). Database overlap for Currier et al. (2007) and our meta-analysis on controlled studies was 73%.

Results indicate that treatment effects persist over time. However, the small number of studies providing follow-up data limits the power of that conclusion considerably. Follow-up
intervals differed considerably, with one study reporting data up to 40 months after treatment. During such a long time, many factors could have diminished the treatment results and remission effects might have occurred. More data on long-term effects of treatments is therefore warranted. It is noteworthy, however, that in their comprehensive review on grief treatment, Currier et al. (2008) did not find an enduring overall treatment benefit at follow-up compared to non-treated controls.

_Treatment Effects Across Symptom Areas_

Given that the evaluation of grief treatments for children and adolescents is still at a very early stage, it is too early to reduce the multiple area outcomes reported for any one study to one general outcome regardless of symptom area. Furthermore, an outcome profile fosters the evaluation of specific treatments and may later allow a therapist to select a treatment for a patient characterized by a specific form of distress (e.g., internalizing vs. externalizing symptoms). Notwithstanding this, within the multiple outcome areas, the most important result for studies on grief and bereavement are outcomes directly relating to grief symptoms. In this area the studies—controlled and uncontrolled—rendered moderate to large effect sizes (0.59 and 0.89, respectively). However, not all studies applied a grief measure (i.e., six studies among the controlled studies reported no data based on a grief measure), instead reporting results based on the assessment of depression, PTSD symptoms, or social adjustment. In these studies, which applied only measures for symptoms other than grief, it remains unclear whether participants showed reduced grief as well.

_Moderator Effects_

Several possible moderators of treatment effects were considered: symptom severity, age, amount of treatment, time since bereavement, confrontation, and publication bias. Results tended to show significant heterogeneity for uncontrolled studies, whereas controlled studies yielded mostly non-significant effects. Yet trends usually showed the same direction between the results of uncontrolled and controlled studies. Despite a significant overall test for heterogeneity, no significant moderators and only one trend
could be identified within the controlled studies—a result that needs to be looked at in greater detail in the future.

Concerning age, there was no significant heterogeneity, neither in controlled nor in uncontrolled studies. However, in both designs, older children seemed to profit more from treatment than younger children. As Currier et al. (2007) did not report on it, any interpretation would be premature. The same holds for possible effects of amount of treatment, where the moderator effect was not significant for controlled studies. For uncontrolled studies, there was a significant effect in line with treatment research in general, that is, the longer the treatment, the higher the impact (within the limits of 25 sessions).

Concerning publication bias, a possible moderator effect was not significant for controlled studies. However, in both designs, published studies seemed to show better effects, and this result was highly significant for uncontrolled studies. This is contrary to the result of Allumbaugh and Hoyt (1999), but in line with treatment research concerning Rosenthal’s (1979) file drawer problem.

Concerning symptom severity as a possible moderator, we assumed that participants with manifest symptoms and functional impairment or even a diagnosis of complicated grief would benefit more from treatment. However, controlled studies showed only a trend related to symptom severity as a moderator, albeit in the expected direction. Within uncontrolled studies, a significant effect confirmed our assumptions. Previous meta-analyses yielded inconsistent results. While Fortner (2000; see Neimeyer, 2000) reported differences for adults, Allumbaugh and Hoyt (1999) did not find any significant differences, and Currier et al. (2007) reported a trend for children. On the other side, we did find support for the role of symptom severity in our meta-analysis on bereaved adults (Rosner et al., 2005; see Rosner & Hagl, 2007). Furthermore, Currier et al. (2008) reported strong support, when categorizing studies into three groups according to the targeted population: studies targeting populations with manifest syndromes showed a significant moderate effect size that was stable until follow-up. Effect sizes for interventions on a universal basis (interventions for bereaved regardless of risk factors or symptom status) varied around zero.

This corresponds well with results evaluating preventive treatment for other types of psychopathology such as depression.
For example, Horowitz and Garber (2006) found that symptomatic (depressed) children benefited more from interventions than children without any symptoms. Considering all results reported so far, symptom severity is a topic that shows some promise for the future (Currier et al., 2008; Rosner & Hagl, 2007). However, since results are inconsistent, further studies, especially with symptomatic children, are warranted.

One possible reason for the slightly differing results between controlled and uncontrolled studies could be the percentage of studies based on symptomatic children (coded as psychotherapy in contrast to prevention studies), which is higher within the uncontrolled studies (25%) than within the controlled studies (13%). This is not surprising given that the standard procedure in the course of clinical treatment evaluation asks researchers to do uncontrolled pilot trials before moving on to randomized controlled designs (as an example for this type of pilot study, see Cohen et al., 2004). If symptom severity represented a relevant moderator, the small percentage of studies dealing with highly symptomatic participants would hamper the corroboration of the impact of not only symptom severity itself, but also related moderators like time since loss.

With regard to time since bereavement, results were similar to symptom severity. We found no significant difference within controlled studies and a significant effect in uncontrolled studies. Studies with a longer time interval between bereavement and intervention yielded larger effects than studies with a shorter interval. The latter result is in contrast to Currier et al. (2007) who reported a trend in the sense that those whose losses were more distant experienced poorer outcome. As results are inconsistent, one might hypothesize that time since bereavement may be a correlate of symptom severity in the sense that recruited participants with losses longer ago perhaps more readily agreed to participate if they still experienced grief symptoms. This would finally lead to a sampling bias. The differences with the results of Currier et al. (2007) might again be explained by the composition of included studies.

In terms of confrontation as part of the treatment rationale, results are only partially promising: mean effect sizes between treatment rationales including confrontation versus no confrontation yielded pronounced differences in favor of confrontation.
within the uncontrolled studies (0.79 vs. 0.34), but this was not the case in controlled studies. Our coding of this moderator variable—confrontation—needs to be described again for a better understanding of this result. Confrontation was not defined in the strict sense of behavior therapy (as, for example, provided for anxiety disorders). Instead, it was understood as addressing the most distressing aspects of the loss as a topic of a structured discussion. We believe that this result may be worth following up in future analyses when more studies are available.

**Promising Treatment Models**

Interventions for children and adolescents were very diverse. Therefore an exploratory grouping beyond analyzing the effect of therapeutic confrontation was possible for only two types of treatment models: music therapy and the trauma/grief treatment manual by Goenjian et al. (1997).

Looking at the single study effect sizes, the two most successful studies turned out to be music therapy interventions (Hilliard, 2001; Dalton & Krout, 2005). Given the enormous part music plays in contemporary youth culture, music might be a promising venue for grief intervention, especially with adolescents, and should be explored further. A second grouping, although mostly based on uncontrolled studies, was possible for four studies using the same treatment model “trauma/grief-focused school based brief psychotherapy” (Goenjian et al., 1997; Layne, Pynoos, et al., 2001; Möhlen, 2005; Saltzman et al., 2001). In these studies, the authors used a specific treatment model (see Goenjian et al., 1997), whereas most other interventions either lack a description of their underlying model or remain very vague about it. The effect size for this specific treatment model is large. Furthermore, the only measure of childhood grief that was used in a number of studies included in our analyses was developed by this group of authors (Expanded Childhood Grief Inventory; Layne, Savjak, Saltzman, & Pynoos, 2001). Yet, since all studies of this grouping included severely and usually multiple traumatized children and adolescents, the intervention is therefore only empirically supported for a very specific subgroup of youths suffering from grief with comorbid PTSD. So, the intervention has to prove its efficacy in further controlled studies.
Limitations

Quantitative reviews are only as good as the primary studies on which the secondary analysis is based. As the literature basis of primary studies on the topic investigated here is still relatively small compared to other areas of psychopathology, even a small number of further controlled studies might change the results of the current evaluation. Nonetheless, the present study represents a state-of-the-art review concerning childhood and adolescent grief treatments. However, both the quality of research as well as the report of study results varied considerably across the included studies. Some studies only provided indirect estimations of group means, so that a number of effects had to be estimated. Although most of the estimation procedures are well established in the meta-analytic literature, a calculation of effect sizes based on means and standard deviations alone would have certainly been preferable.

Recommendations for Future Research

During the rating procedure it became evident that measures for childhood grief are extremely diverse. To improve research in the field, reliable and valid measures for different age groups need to be developed. The most promising measure, the Expanded Childhood Grief Inventory by Layne, Savjak, et al. (2001) was generally used in samples of severely traumatized children and adolescents. It needs to be applied and evaluated in children and adolescents with other types of bereavement and without comorbid PTSD.

Future studies should adhere to the usual reporting procedures in intervention research. Across studies, drop-out rates were either not accounted for at all or reported inconsistently. Therefore, a detailed analysis of drop-out rates was impossible. Moderator analyses for two crucial moderators, namely gender and age, were impeded by either missing or inconsistent reports.

Furthermore, most studies lacked external validity concerning participant recruitment. In almost all studies, participants were actively recruited by researchers. Studies of youths and family members who proactively seek therapy are rare. Because treatment motivation is a central aspect of therapy, one might assume
that effect sizes would increase if participants were patients, who seek treatment themselves.

Finally, research on bereaved youth who show some kind of clinically significant distress in terms of developing clinical symptoms beyond an acute grief reaction is very important. Clear definitions of complicated grief in childhood would help to identify children with a special need for support and thus improve health care in general. A common finding is that effect sizes are usually smaller for preventive interventions than for psychotherapy. Counseling (or in our terminology: preventive interventions) only partially aims at preventing future symptoms; it also provides support and comfort in difficult times.

Therefore, outcome measures for counseling interventions should reflect this different goal. Hopefully, the data basis will increase in the near future so that a differential indication for either counseling (i.e., preventive interventions) or psychotherapy can be determined as already indicated in our study and other research. This would in turn allow the differential merits of the two approaches to emerge. In a nutshell, and in line with Larson and Hoyt (2007), we do not share “the new pessimism” (p. 347) about grief counseling.

References

References marked with an asterisk (*) indicate studies included in Analysis 1 (controlled studies), and references marked with a dagger (†) indicate studies included in Analysis 2 (uncontrolled studies).


Efficacy of Interventions After Bereavement


