

Running head: WHAT HAS BECOME OF GRIEF COUNSELING?

What Has Become of Grief Counseling?

An Evaluation of the Empirical Foundations of the New Pessimism

© American Psychological Association

Accepted for publication in [\*Professional Psychology: Research and Practice\*](#).

This article may not exactly replicate the final version published in the APA journal. It is not the copy of record. The TIDE Analysis: Technical Considerations section at the end of the paper will be available through the APA online archive when the article is published. Please do not copy or distribute without permission of the authors.

Dale G. Larson

Santa Clara University

William T. Hoyt

Department of Counseling Psychology, University of Wisconsin-- Madison

Abstract

A pessimistic view of grief counseling has emerged over the last seven years, exemplified by R. A. Neimeyer's (2000) oft-cited claim that "such interventions are typically ineffective, and perhaps even deleterious, at least for persons experiencing a normal bereavement" (p. 541). This negative characterization has little or no empirical grounding, however. The claim rests on two pieces of evidence. The first is an unorthodox analysis of *deterioration effects* in 10 outcome studies in B. V. Fortner's (1999) dissertation, usually attributed to Neimeyer (2000). Neither the analysis nor Fortner's findings has ever been published or subjected to peer review, until now. This review shows that there is no statistical or empirical basis for claims about deterioration effects in grief counseling. The second piece of evidence involves what we believe to be ill-informed summaries of conventional meta-analytic findings. This misrepresentation of empirical findings has damaged the reputation of grief counseling in the field and in the popular media, and offers lessons for both researchers and research consumers interested in the relationship between science and practice in psychology.

Keywords: grief therapy; grief counseling; treatment deterioration; scientist-practitioner model; bereavement

What Has Become of Grief Counseling? An Evaluation of the Empirical Foundations of  
the New Pessimism

"It ain't what you don't know that gets you into trouble. It's what you know for sure that just ain't so."

Attributed to Mark Twain

This well-known aphorism provides a humorous reminder of the perils of unwarranted certainty. In this article, we consider an emerging consensus in the literature on grief or bereavement (we use these terms interchangeably) counseling that, in our view, has arisen independent of the scientific literature on this subject, but is nonetheless prevalent (indeed, almost omnipresent) in scientific journals and professional summaries of research designed to guide policy and practice. We describe the evolution of this new pessimism regarding the efficacy of bereavement counseling, with attention to two central claims that ground this reading of the literature. We consider these claims in detail, in an effort to demonstrate that neither of them can be substantiated relative to traditional scientific standards, and offer a different summary of findings on grief counseling that is more faithful to the actual research literature. Finally, we consider the implications of this unwarranted certainty in the research community for practitioners striving to conform to the ideal of empirically-grounded practice, and we remind practitioners of several key guidelines that may be helpful to nontechnical readers as they evaluate conclusions offered in scientific journals.

The New Pessimism

Strong consensus has emerged over the past seven years in the bereavement research community: Grief counseling is at best weakly effective, and at worst harmful. This new pessimism about the effectiveness of grief interventions is widely reflected in reviews of literature published in major research journals and also in the 2003 *Report on Bereavement and Grief Research* published by the Center for Advancement of Health (CFAH, 2003). The CFAH national report concluded that

the evidence from well-conducted studies of interventions ...challenges the efficacy and effectiveness of grief interventions for those experiencing uncomplicated bereavement. This evidence also indicates that concerns are warranted about the potential of interventions to cause harm to some individuals. (CFAH, 2003, p. 72)

Such conclusions are regularly reiterated in journals targeted at grief researchers and practitioners. For example, in *Death Studies*, Paletti (2005) cautioned that “the alignment of research and practice is a critical issue in bereavement care; recent evaluations of a number of intervention programs have indicated that such care might be ineffective, and worse, counterproductive” (p. 669). Also in *Death Studies*, Shapiro (2005) warned: “As bereavement practitioners, we have recently been challenged by careful reviews (Jordan & Neimeyer, 2003) suggesting that one-size-fits all approaches to “grief work” mandating exploration of distressing feelings can intensify distress for a significant number of mourners” (p. 262).

Claims about deterioration effects form perhaps the most persuasive argument for restraint in offering interventions for bereaved individuals and are certainly shocking, when taken at face value, to bereavement practitioners. Our purpose in this article is to

trace this new pessimism among bereavement researchers to its empirical roots and to evaluate the basis for claims of iatrogenic effects in particular, and for the more global (though less alarming) claim that grief interventions generally are much less efficacious than psychosocial interventions for other emotional and behavior problems.

#### Deterioration Effects in Grief Counseling

Probably the major impetus for the rash of cautionary messages regarding bereavement interventions has been the striking finding, first published by Neimeyer (2000), that such treatments may be harmful. Neimeyer reported on the findings of a meta-analysis of grief interventions using a “novel procedure [designed] to estimate *treatment-induced deterioration*, which represents the proportion of participants who are worse off after treatment than they would have been if they had been assigned to the control group” (p. 544).

The findings for the treatment-induced deterioration effects (TIDE) were staggering: Averaging over all studies providing the necessary information, Neimeyer concluded that “nearly 38% of recipients of grief counseling theoretically would have fared better if assigned to the no-treatment condition” (p. 545). In other words, more than 1/3 of grief counseling clients are worse off at the end of treatment than they would have been in the absence of treatment! Neimeyer reported that the results of subsidiary analyses focusing on “normal” grievors (as opposed to those who were “traumatically bereaved”) were even more alarming: In these “normal or unselected samples...nearly one in two clients suffered as a result of treatment” ( p. 546).

These findings are now treated as established scientific fact and have been widely cited in major professional journals, including *American Psychologist* (Bonanno, 2004),

*Journal of Personality and Social Psychology* (Bonanno, Moskowitz, Papa, & Folkman, 2005; Bonanno et al., 2002), *Psychiatric Annals* (Reissman, Klomp, Kent, & Pfefferbaum, 2004), *Journal of Clinical Psychiatry* (Harkness, Shear, Frank, & Silberman, 2002), *Suicide and Life-Threatening Behavior* (Jordan & McMenemy, 2004), *Applied and Preventative Psychology* (Bonanno, Papa, & O'Neill, 2002), and *Death Studies* (Jordan & Neimeyer, 2003; Ott & Lueger, 2002; Sikkema, Hansen, Kochman, Tate, & Difrancisco, 2004; Wolf & Jordan, 2000). The findings are also reported in great detail in both the *Handbook of Bereavement Research* (Stroebe, Hansson, Stroebe, & Schut, 2001), and the CFAH (2003) *Report on Bereavement and Grief Research*. They have begun to be cited in popular books (Sommers & Satel, 2005) and newspapers ("Grieving is an individual process," 2003), and even in a report on grief and bereavement services for the Department of Human Services in Victoria, Australia (Nucleus Group, 2004).

*TIDE Findings: In Search of the Evidence*

Certainly, if the TIDE findings are valid, pessimism about the efficacy of bereavement interventions and extreme caution in their application are warranted. But what is the basis for these claims? This question became a pressing concern for the first author (DGL) when he was invited to participate in a scientific panel titled "Grief Counseling: Can It Be Harmful?" at a national professional meeting. Preparing for the panel, he dutifully reviewed the literature on iatrogenic effects of grief counseling, turning first to Neimeyer's (2000) article because it was frequently cited in journals, presented at conferences, and discussed among colleagues. He began to review the article, assuming that it contained the original empirical work. However, a cursory

reading of Neimeyer (2000) revealed that this was not an empirical study, but rather a summary of findings of past research, both published and unpublished. Neimeyer reported the TIDE findings as part of a summary of the results of an unpublished meta-analysis of grief interventions attributed to himself, Barry Fortner, Adam Anderson, and Jeff Berman and presented at a conference (Fortner & Neimeyer, 1999). Neimeyer (p. 543) explained in a footnote that the only publicly available report of the meta-analysis was in a dissertation completed by Fortner (1999).

When requests for published or unpublished reports of the research confirmed that the Fortner dissertation was, and still is, the only available record of this study, the dissertation became the focus of attention. The dissertation is relatively brief—just 28 pages of text, excluding references and tables. In the Method section, Fortner (1999) described a “statistical method for determining the theoretical proportion of participants who were worse off after treatment than they would have been if they had been assigned to the control group” (p. 14). This is the method that yielded the 38% TIDE estimate. Obviously, the validity of this statistical technique is crucial to evaluating the validity of the TIDE findings, so we review this procedure in some detail.

The TIDE statistic has never been published, nor is any statistical justification for this approach available in unpublished form. Fortner cites two sources for this technique: a presentation by Anderson, Berman, and Fortner (Anderson, Berman, & Fortner, 1998) to the Annual Meeting of the Society for Psychotherapy Research, and a master’s thesis by Adam Anderson (Anderson, 1988) completed at the University of Memphis. Print copies of these sources were not available, though it appears that each of them reports on another application of the TIDE technique (to a broad sample of psychotherapy outcome

studies), rather than presenting a rationale for the technique itself. Thus, the critique presented here must be based on the brief description provided by Fortner (1999).

*TIDE Findings: Pattern of Citations*

The first question raised by the TIDE findings involved the pattern of citations. As Figure 1 shows, this result was cited regularly in the seven years from 2000 (when Neimeyer summarized the findings in *Death Studies*) to 2006 (when the first author sought to understand the basis for the TIDE claims). What is interesting in Figure 1 is that the only publicly available (but not peer-reviewed) account of the data behind these findings (Fortner, 1999), was cited only once—in the Neimeyer (2000) footnote. Most authors have simply cited Neimeyer’s (2000) summary as the source of this finding, with no reference to the actual (Fortner, 1999) study. For example, in a review article on loss and trauma published in *The American Psychologist*, Bonanno (2004) offered the following summary of Fortner’s findings: “In one of these analyses, an alarming 38% of the individuals receiving grief treatments actually got worse relative to no-treatment controls, whereas the most clear benefits were evidenced primarily with bereaved individuals experiencing chronic grief (Neimeyer, 2000)” (p. 22). Similarly, in a multiple case study published in the *Journal of Clinical Psychiatry*, Harkness, Shear, Frank, and Silberman (2002) asserted that “a recent meta-analysis found a strikingly low effect size (0.15) [sic] for supportive grief interventions and a significant minority worsened with these treatments...(Neimeyer, 2000)” (p. 1120).

At best, such citations represent inadequate scholarship, in that they fail to point readers to the empirical evidence that justifies these summary claims. This failure to cite the Fortner dissertation is also troubling in that it implies that these authors have not read

the empirical report, and therefore were not in a position to evaluate the validity of its conclusions.

More recently, authors have cited a subsequent review paper by Jordan and Neimeyer (2003), or even other writers (who themselves cited Neimeyer, 2000) as the basis for the TIDE claim. For example, in a research review (targeted at clinicians) published in *Clinical Neuroscience Research*, Craighead and Neeroff (2005) ignored both Fortner (1999) and Neimeyer (2000): “In summarizing ‘grief counseling’ for bereavement, Jordan and Neimeyer [2003] reported that many of the participants (up to 38% in one study) would have had a better outcome by having been assigned to the no-treatment control group as compared to the grief counseling group” (p. 409). In a review article published in *Psychiatric Annals*, Reissman, Klomp, Kent, and Pfefferbaum (2004) cited both Neimeyer (2000) and Bonanno (2004) (who in turn cited Neimeyer) as the bases for the conclusion that grief work may be “ineffective, and perhaps even deleterious” (p. 631).

This pattern of citations can be misleading in two ways. First, attributing the finding to Neimeyer’s (2000) published article, rather than Fortner’s (1999) unpublished dissertation, can lead readers to believe that the empirical data and statistical analyses underlying the TIDE results have been subjected to peer review, which is not the case. Second, the citation of more recent published articles, either instead of or in addition to Neimeyer (2000), may give naïve readers the impression that this finding has been corroborated in additional investigations, which again is not the case.

*TIDE Findings: A Post Hoc “Blind” Peer Review*

The TIDE technique represents an unorthodox statistical analysis—one which has never been subjected to peer review. Recognizing that the pattern of citations may have masked this fact, and that an evaluation of the validity of this technique would be beyond his own statistical competence, the first author (DGL) contacted Gary R. VandenBos, the APA Publisher and the Managing Editor of the *American Psychologist*, for guidance about the citation and evaluation of non-peer reviewed work. Based on the apparent high impact of the TIDE findings, and their unconventional citation pattern (including numerous citations in journals published by the American Psychological Association), VandenBos agreed to conduct a post-hoc blind peer review of both the TIDE statistic and the findings reported in Fortner (1999), following the procedures that any APA journal would use. The reviewers, both national methodological and statistical experts, were asked to consider whether the formula for TIDE generates a stable, reliable, valid, and meaningful number that can be interpreted as claimed (i.e., as a deterioration effect). Neither the TIDE statistic nor Fortner's conclusions passed this peer review. The experts conclusively agreed that the TIDE statistic is seriously flawed and that there is no valid basis for the claim that 38% of grief counseling clients suffered deterioration (nor for Neimeyer's statement that one in two normally-bereaved clients suffered).

*TIDE Analysis: Technical Considerations*

Meanwhile, the second author (WTH) joined the search for the basis for the TIDE claims. With a background in meta-analytic work, he was intrigued by the claim that summary data (i.e., means and standard deviations for treatment and control groups) could be used to derive an estimate of the proportion of clients whose condition deteriorated as a result of treatment. He was not familiar with any commonly accepted

approach to this problem in meta-analysis and was curious about the statistical justification for this conclusion. In an on-line appendix to this article [insert URL here], we show how the statistic used by Fortner is computed and raise questions about its interpretation as an index of TIDE.

### *Summary*

After searching out the original report of the TIDE finding, reading it carefully, and consulting with statistical experts, we came to an astonishing conclusion: The 38% finding, which has swept over the death studies field like a veritable TIDE-al wave, appears to have no basis in fact. It is based on a meta-analysis that has never been published, using a statistic which itself has never been published or reviewed by the methodological community, and which seems on its face to be obviously flawed. A blind post-hoc peer review of Fortner's (1999) study by national experts confirmed these startling conclusions. This is good news for grief counselors who may have been concerned, based on this "finding," that they may have been unknowingly harming their clients. There appears to be no basis for this concern (Larson & Hoyt, April, 2006).

But recent reviewers have not confined themselves to the 38% TIDE finding as a basis for their negative views of grief counseling. What about the second basis for pessimism: the claim that research shows that grief therapy is ineffective, or at any rate so weakly effective that it is not warranted, at least for so-called "normal" griever? This conclusion is based on findings of meta-analyses--quantitative reviews of existing research. In our quest to understand the pessimistic turn among grief researchers, we next examined the basis for this claim.

### *Analysis of Meta-Analyses and Reviews*

*Meta-Analyses of Grief Counseling Outcome Studies**Allumbaugh and Hoyt (1999)*

The most thorough published quantitative review of research on grief interventions was conducted by Allumbaugh and Hoyt (1999). These authors conducted an extensive search including both published and unpublished literature. They used an innovative meta-analytic technique developed by Becker (1988) to allow for incorporation of both pre- and post-treatment scores on the dependent variable(s), from both treatment and control groups. This method is advantageous, when most studies collect both pre- and posttreatment data, because it treats change over time (rather than simple posttreatment status) as the outcome of interest, and because change over time is estimated separately for treatment and control groups and then compared. This feature is especially interesting relative to the presenting problem of bereavement, as it is expected that symptoms will abate over time to some extent as a natural process (Bonanno et al., 2002), and Becker's method quantifies the extent to which this natural recovery was observed in the studies reviewed. It also allows inclusion of data from studies with no control group, because pre-post gains for the treatment group can be compared meta-analytically with corresponding gains from control group participants in other studies (Becker, 1988).

Data from 35 independent investigations of grief counseling yielded an aggregate effect size of  $d = 0.43$ ,<sup>1</sup> which is smaller than effect sizes typically observed when counseling and psychotherapy are applied to other psychological or emotional problems. (Such effect sizes typically average  $d = 0.8$ ; see Lipsey & Wilson, 1993; Wampold, 2001.) However, *homogeneity tests* indicated significant variability in effect sizes

beyond that expected due to chance (i.e., to sampling error). Such heterogeneity among study effect-size estimates can arise when studies differ on characteristics that are systematically related to treatment effectiveness—that is, on *moderator variables*.

Allumbaugh and Hoyt (1999) found two moderator variables that appeared particularly important for understanding the relatively low aggregate effect size. These were study characteristics on which participants in the modal study in this review differed from typical clients in grief counseling: *motivation for counseling* and *time since loss*.

*Motivation for counseling.* Clients in most of these studies were recruited into treatment rather than electing to seek grief counseling on their own. These research participants may not be typical of clients in actual practice settings. Furthermore, effect sizes for the handful of studies involving self-referred clients were quite large (ranging from  $d = 1.17$  to  $d = 3.05$ ), suggesting a substantial benefit for this client group.

*Time since loss.* Also, the mean time between the loss of a loved one and the start of the grief intervention for these studies was 27 months. This may be related to the participant recruitment procedures just discussed. Researchers recruiting participants who had experienced a particular type of loss (e.g., spouse, child) may have felt compelled to appeal to participants whose losses were relatively remote, if they were unable to recruit a sufficient number of recently bereaved participants. On the basis of the moderator findings for time since loss, Allumbaugh and Hoyt estimated that, holding all other study characteristics constant, the aggregate effect size for a similar set of studies whose participants were recently bereaved ( $d = 0.70$ ) would be comparable to that typically observed in psychotherapy outcome research.

*Conclusions.* Allumbaugh and Hoyt concluded that the relatively small aggregate effect size ( $d = 0.43$ ) for grief interventions in these studies “may say more about the nature of the studies than about the effectiveness of grief treatment per se” (p. 378). The findings in most of these studies were based on participants who (a) were actively recruited by researchers, rather than seeking help on their own, and (b) had been bereaved, on average, several years before treatment. Studies of samples more typical of actual client populations in bereavement counseling settings “are likely to be as effective [as] or possibly even more effective than psychotherapy in general” (p. 378).

#### *Other Meta-Analyses*

Two other meta-analyses have been cited frequently in subsequent literature reviews in grief studies.

*Kato and Mann (1999).* Kato & Mann (1999) published a detailed review of a small number of studies ( $k = 11$ ) of psychological interventions for the bereaved. Presumably the sample was small because the search procedures were limited to electronic searches using the keywords bereaved and bereavement. Thus, studies that used related terms such as “grief,” “griever,” or “grieving” would not have been included. In addition, the quantitative portion of Kato and Mann’s review used problematic procedures. For example, when a study included more than one dependent measure (as most of these studies did), a separate effect size was computed for each measure. These estimates were then treated as independent effect sizes in the meta-analysis, although they must certainly be correlated (i.e., statistically dependent). This procedure violates the independence assumption of the statistical tests in meta-analysis and (perhaps more important, for our purposes) can bias study findings, because effect sizes from studies

with larger numbers of measures are arbitrarily given greater weight in computing the aggregate  $d$ . So, although Kato and Mann obtained a substantially lower aggregate effect size ( $d = 0.11$ ) than Allumbaugh and Hoyt (1999), this smaller effect size was not based on a full review of existing studies, nor were accepted meta-analytic procedures used to analyze data from the studies that were identified.

*Fortner (1999)*. The meta-analytic findings that have been most widely cited in the bereavement literature are those of Fortner's (1999) unpublished dissertation.<sup>2</sup> Fortner used conventional meta-analytic procedures, as well as the TIDE procedure, to quantify the effects of grief interventions, arriving at a very weak aggregate effect size of  $d = 0.13$  for 23 studies<sup>3</sup> Neimeyer (2000) summarized both the conventional findings and the TIDE findings in detail, and subsequent reviewers citing the TIDE findings often cited the very weak effect size as additional evidence for a pessimistic view of the effectiveness of grief counseling.

Evaluating the merits of Fortner's (1999) meta-analytic conclusions is difficult. First, this report has never been published, and it is unclear whether it has ever been submitted for publication. In any case, the findings presented in the dissertation, though they have undergone scrutiny by dissertation committee members, have not been vetted by the more rigorous peer review process necessary for publication in a scientific journal. Analytic procedures are described sketchily, and homogeneity tests, which are important for interpreting findings and which provide the warrant for conducting moderator analyses, are not reported in the Results section.

Interpretation of moderator findings is made even more difficult because coding procedures for the moderator variables are not described, nor are statistics reported

reflecting the agreement among multiple raters in the coding of studies. For these reasons, it is difficult to know what to make of the finding that grief interventions were significantly more effective for clients experiencing complicated bereavement than for so-called normal grievers. This finding has been much cited in the literature, along with the even more alarming (and inaccurate) TIDE interpretation of iatrogenic effects for nearly 1 in 2 normally bereaved clients during grief counseling (Neimeyer, 2000). However, there is no information on how complicated bereavement was coded for these studies or on whether it was reliably coded. Allumbaugh and Hoyt (1999) compared normal and “high risk” grievers (using designations by the authors of the primary research reports) and found no difference in intervention effectiveness for these two groups. Yet, on the strength of Fortner’s (1999) conclusion and Neimeyer’s (2000) reiteration, it is now common to suggest that counseling is inappropriate, and perhaps even dangerous, for normally bereaved individuals.

Still, it is puzzling that the aggregate effect sizes reported by Allumbaugh and Hoyt (1999) and Fortner (1999) are rather different, especially given that search procedures for both studies appear to have been thorough and quite closely comparable.<sup>4</sup> In an effort to understand this discrepancy, we examined the reference lists of the two meta-analyses, and made a surprising discovery. The two reviews examined a total of 55 research reports, with only three studies in common between them. Thus, their samples of studies were almost completely different. This difference is partly attributable to differences in the two studies’ exclusion criteria. Fortner excluded unpublished studies and studies not using an experimental design (i.e., random assignment of participants to treatment and control groups). Allumbaugh and Hoyt excluded studies that did not

measure pre-treatment status on at least one outcome variable. On the face of it, it seems unlikely that any of these differences in study characteristics would lead to systematic differences in study effect sizes. Given the different sets of studies included in the two reviews, however, it seems that the definitive chapter on the findings of grief intervention research will require a more inclusive analysis considering both sets of studies (as well as research conducted after 1998, at least some of which is methodologically superior to much of what came before).

*Grief Counseling Meta-Analyses: Patterns of Citations*

Although, for the reasons just cited, an impartial reviewer might conclude that although the jury is still out on the effectiveness of grief counseling, there is clearly reason for optimism. One might expect that such a reviewer would give the greatest weight to the Allumbaugh and Hoyt (1999) findings. These appeared in a respected journal published by the American Psychological Association and included the largest sample of studies, with carefully documented and sophisticated statistical analyses. Fortner's (1999) findings are certainly important, especially to a reader who looks closely enough to realize that Fortner examined a substantially different set of research reports than Allumbaugh and Hoyt. However, one might prefer not to assign the same weight to his conclusions, given the lack of peer review and the omissions in reporting of results discussed above. Kato and Mann's (1999) article includes valuable reflections on the nature of bereavement and on the strengths and weaknesses of individual studies. However, their quantitative findings would not likely be given much weight by a reviewer familiar with meta-analytic principles, because their literature search was far from exhaustive, and their analyses were problematic statistically.

Given this set of findings, how have reviewers found support for the pessimistic view of grief counseling so prevalent in the literature in recent years? Such an interpretation could only arise through distorting, dismissing, or even outright ignoring Allumbaugh and Hoyt's (1999) review. Neimeyer (2000) dismissed their findings with the explanation that Allumbaugh and Hoyt "included numerous uncontrolled one-group studies that could have inflated [effect size] estimates" (p. 543), and went on to present Fortner's (1999) findings as authoritative. This critique represents a misunderstanding of Becker's (1988) approach to including pretreatment as well as posttreatment measures in meta-analysis.

Another approach to de-emphasizing the importance of Allumbaugh and Hoyt's (1999) findings has been distortion or misrepresentation of their conclusions. For example, Bonanno (2004) began his summary of grief intervention research by citing "two recent meta-analyses [that] independently reached the conclusion that grief-specific therapies tend to be relatively inefficacious (Kato & Mann, 1999; Neimeyer, 2000)" (pp. 21-22), then stated that "a third meta-analytic study" (Allumbaugh & Hoyt, 1999) concluded that "grief therapies can be effective but generally to a lesser degree than ... other forms of psychotherapy" (p. 22). Compare this with Allumbaugh and Hoyt's own summary of their findings, cited in full above, stating that this moderate aggregate effect size reflects unrepresentative sampling procedures associated with most of these studies, and that more ecologically valid studies tend to find that grief interventions are "as effective [as] or possibly even more effective than psychotherapy in general" (p. 378). Lalande and Bonanno (2006) have taken this approach one step further by implying that the three meta-analyses have equivalent (and equivalently discouraging) findings:

"Unfortunately, recent meta-analyses have shown that grief treatments have generally been ineffective and can possibly do harm to those who participate in them (Allumbaugh & Hoyt, 1999; Jordan & Neimeyer, 2003; Kato & Mann, 1999; Neimeyer, 2000)" (p. 303).

A final strategy for defending a pessimistic view of grief counseling has been to ignore Allumbaugh and Hoyt's (1999) findings entirely. For example, in a review intended as a definitive synthesis to guide policy and practice in the bereavement area, the *Report on Bereavement and Grief Research* (CFAH, 2003) does not cite Allumbaugh and Hoyt at all, and instead offers a detailed summary of Fortner's (1999) findings (which are attributed not to Fortner, but to Neimeyer, 2000).

#### Summary: Rebuttal of Evidentiary Basis for the New Pessimism

The recent pessimism about the effects of grief counseling is based on three interrelated claims, each of which has doubtful empirical grounding.

*First claim: A large proportion of clients are harmed.* To our knowledge, no valid method exists to quantify treatment deterioration from summary data provided in psychotherapy outcome studies. The TIDE findings for grief counseling, usually attributed to Neimeyer (2000), are actually based on a student dissertation (Fortner, 1999) that has never been published and possibly never subjected to peer review outside the dissertation committee. The TIDE statistical procedure used by Fortner is attributed to another student's master's thesis (Anderson, 1988), which is unpublished and unavailable for examination. Reviewers appointed by Gary VandenBos (APA Publisher and the Managing Editor of the *American Psychologist*) agreed that there is no valid basis for the claim that the 38% figure represents a percentage of clients deteriorating as a result of

treatment and that the statistic is seriously flawed. In the on-line Appendix [insert URL again here] we have described Fortner's method in some detail and show that it cannot provide a valid TIDE estimate.

*Second claim: Effect sizes for grief counseling outcome studies are zero or trivial.* The most thorough and methodologically rigorous meta-analysis to date (Allumbaugh & Hoyt, 1999) paints a relatively optimistic portrait of grief counseling outcomes, while acknowledging limitations in the literature on which these conclusions are based. Still unresolved are the reasons for discrepancies between Allumbaugh and Hoyt's findings and those of Fortner (1999), who used a largely different set of studies. Unfortunately, Fortner's (1999) findings have never been published, and the unpublished dissertation lacks some of the information that would be helpful to address this question.

*Third claim: Grief counseling is especially ineffective (and especially harmful) for "normal" grievers.* The contention that "normal" (uncomplicated) bereavement is unresponsive to grief counseling is based on a moderator analysis in Fortner's (1999) dissertation, as summarized by Neimeyer (2000). As noted above, Fortner (1999) does not report homogeneity tests (which normally provide the warrant for conducting moderator analyses in meta-analysis). He also does not explain how he distinguished "normal" from "complicated" grief in coding this moderator variable, nor does he report the level of interrater agreement (to show that this study characteristic, as he defined it, could be coded reliably from the information in the research reports). All of these factors make the validity of this conclusion difficult to gauge. Allumbaugh and Hoyt (1999), who coded a similar study characteristic ("normal" versus "high risk" grievers), found no evidence of a difference in effect size for these two groups.

The related claim (Neimeyer, 2000) that nearly 1 in 2 normally bereaved clients is actually worse off following counseling is based on an application of the TIDE statistic to Fortner's (1999) moderator analysis comparing normal and complicated bereavement. Because the TIDE statistic has been shown not to yield valid deterioration estimates, this claim lacks any empirical grounding.

#### Lessons from the Propagation of the TIDE Findings

Researchers are only human, and are susceptible to human foibles including cognitive biases and sloppy reasoning in the service of cherished beliefs. As Donald Campbell (1986/1999) noted, science is a reliable method of drawing valid conclusions from empirical data only to the extent that scientists evolve social structures that tend to maximize scientific validity. The widespread acceptance of the TIDE findings among researchers and practitioners interested in issues of grief and loss represents a breakdown in this validity-enhancing system, and suggests lessons for both groups in the interest of creating a scientific research culture that will inform, rather than misinform practice.

#### *Recommendations for Researchers, Editors, and Reviewers*

##### *Cite Your Sources*

Giving credit to the originators of theories or specific knowledge claims is a hallmark of scholarly conduct in all academic disciplines. In empirical disciplines, such as psychology, writers are asked to clarify whether the source is cited as an acknowledgment of an intellectual debt (to a previous theorist who has espoused a similar or identical position) or as a source of empirical evidence in support of the proposition (American Psychological Association, 2001, p. 28). The former type of citation acknowledges intellectual priority; the latter type claims an empirical basis for the

author's assertion, and points the reader to the evidence so that he or she can evaluate its quality and relevance.

One of the challenges for readers seeking to evaluate the validity of the TIDE findings was that authors cited Neimeyer (2000) rather than the actual reference for the research report (Fortner, 1999). This type of scholarly lapse is certainly fair game for reviewers or editors (some or all of whom will have strong familiarity with the research area) as they provide comments on manuscripts submitted for publication.

#### *Read the Sources You Cite*

Authors citing a source as empirical support for theoretical claims imply that they have read and evaluated the cited article, and believe that it makes a plausible case for the veracity of these claims. If author C cites publication A without examining it, simply because author B (whom C trusts) has also cited A as evidence for a similar proposition, then C is participating not in the conventions of scientific scholarship, but rather in a sort of grown-up version of the children's game of "telephone." In this game, receivers later in the chain of communication never hear directly from the source, but only from those just preceding them in the chain. In the grown-up version of the game, just as in the children's version, there is every likelihood of serious distortion of the original message.

#### *Know Your Limitations*

A pernicious factor contributing to the spread of the TIDE findings was the use of a "novel" (in fact unpublished) statistical procedure. It seems evident that neither Fortner (1999) himself nor those citing his work (or Neimeyer's summary of his work) had sufficient grasp of meta-analytic principles to recognize the limitations of the TIDE statistic, and neither Fortner (1999) nor Neimeyer (2000) addressed these limitations so

that readers could take them into consideration. The fact that the statistic itself had never been reviewed by methodologists was critical information that was omitted from all published reports of the TIDE findings. Readers, including editors and reviewers, could have noticed that the citation for the TIDE procedure was an unpublished master's thesis, but this would have required tracing the finding back to Fortner's (1999) work.

### *Recommendations for Readers*

According to Campbell (1986/1999) scientific validity depends on “disputatious communities of truth-seekers” (p. 191) who keep one another honest by challenging assumptions and identifying plausible rival explanations for research findings (which qualify current findings and suggest directions for future research). On this view, readers have a role to play in maintaining a healthy research climate. A skeptical attitude and a willingness to question where one does not understand (or where one needs more information to verify one's understanding) are important traits for editors and reviewers of scientific manuscripts, and are also important for consumers of research findings. At the heart of the scientist-practitioner model of training is the goal of empowering applied psychologists to become active and critical evaluators of the research literature.

### *Where Do We Go From Here?*

Although they are less than 10 years old, the erroneous TIDE findings have had a major impact on both the theory and practice of grief counseling. For practitioners, the widespread acceptance of these findings has been demoralizing. As the TIDE claims have permeated mental health agencies, hospices, and other funding sources, these parties have become skeptical of the value of grief counseling (and concerned about potential harm to clients). The TIDE claims have also been repeated in the popular media

("Grieving," 2003; Sommers & Satel, 2005, "Therapy", 2005), and the general pessimism about the efficacy of grief counseling (which we believe has been fueled in large part by alarm about these claims) has been reflected in major media articles as well (Brody, 2004). These echoes in the mainstream media of the unwarranted pessimism that has flourished in the scientific journals could foster negative impressions of these interventions in the general public and discourage distressed bereaved persons from seeking needed counseling.

To begin to reverse the damage, and to minimize the future impact of these claims, we recommend three corrective actions. First, to restore the integrity of the scientific study of grief counseling, it would be desirable for the principal journals responsible for propagating these findings to publish retractions in print from proponents of the unfounded conclusions. Public retractions, pointing authors and reviewers to criticisms of the TIDE statistic, will help to prevent future citations of the TIDE claims in professional journals. Second, the TIDE episode provides an opportunity to reflect on scientific standards in psychology. It is likely that the TIDE claims would not have attained credibility without laxness on the part of authors and editors in standards for critically evaluating and referencing cited work. We have suggested several lessons for authors, editors, reviewers, and even readers of scientific reports, designed to reduce the probability that erroneous findings will be published and repeatedly cited in the future. Finally, to counteract the potential misinformation about TIDE findings from the popular media, it will be important to find mechanisms to educate the general public about the problematic nature of these claims.

In conclusion, let us emphasize the good news growing out of our investigation of the basis for pessimistic claims about the efficacy of grief counseling: There is no empirical or statistical basis for these claims. There is no evidence that bereaved clients are harmed by counseling, nor that clients who are “normally” bereaved are at special risk. And there is not even any strong evidence that grief counseling, as typically practiced, is less efficacious than other forms of counseling and psychotherapy (which have been shown to have strong positive outcomes for many forms of psychological distress; Lipsey & Wilson, 1993; Wampold, 2001). More work can be done to clarify the implications of existing studies (some of which have been conducted since the publication of the most recent meta-analysis of grief counseling), but findings to date indicate that cautious optimism, rather than the recently fashionable dire pessimism, is the attitude most congruent with empirical findings on grief counseling outcomes.

References

- Allumbaugh, D. L., & Hoyt, W. T. (1999). Effectiveness of grief therapy: A meta-analysis. *Journal of Counseling Psychology, 46*(3), 370-380.
- Anderson, A. S. (1988). *Does psychotherapy make some clients worse? A reanalysis of the evidence for treatment-induced deterioration*. University of Memphis, Memphis, TN.
- Anderson, A. S., Berman, J. S., & Fortner, B. V. (1998). *The evidence for deterioration induced by treatment*. Paper presented at the Annual Meeting of the Society for Psychotherapy Research, Snowbird, UT.
- American Psychological Association (2001). *Publication manual of the American Psychological Association* (5th ed.). Washington, DC: Author.
- Becker, B. J. (1988). Synthesizing standardized mean-change measures. *British Journal of Mathematical and Statistical Psychology, 41*(2), 257-278.
- Bonanno, G. A. (2004). Loss, trauma, and human resilience: Have we underestimated the human capacity to thrive after extremely aversive events? *American Psychologist, 59*(1), 20-28.
- Bonanno, G. A., Moskowitz, J. T., Papa, A., & Folkman, S. (2005). Resilience to loss in bereaved spouses, bereaved parents, and bereaved gay men. *Journal of Personality and Social Psychology, 88*(5), 827-843.
- Bonanno, G. A., Papa, A., & O'Neill, K. (2002). Loss and human resilience *Applied and Preventative Psychology, 10*, 193-206.

- Bonanno, G. A., Wortman, C. B., Lehman, D. R., Tweed, R. G., Haring, M., Sonnega, J., et al. (2002). Resilience to loss and chronic grief: A prospective study from preloss to 18-months postloss. *Journal of Personality and Social Psychology*, 83(5), 1150-1164.
- Brody, J. E. (2004, January 27). Often, time beats therapy for treating grief. *New York Times*, p. F7.
- Campbell, D. T. (1986/1999). Sociology of scientific validity. In D. T. Campbell & M. J. Russo (Eds.), *Social experimentation* (pp. 179-202). Thousand Oaks, CA: Sage.
- CFAH. (2003). *Report on bereavement and grief research*. Washington, DC: Center for Advancement of Health.
- Craighead, W. E., & Neeroff, C. B. (2005). The impact of early trauma on response to psychotherapy. *Clinical Neuroscience Research*, 4, 405-411.
- Currier, J. M., Holland, J. M., & Neimeyer, R. A. (2006). Sense-making, grief and the experience of violent loss: Toward a mediational model. *Death Studies*, 30, 403-428.
- Fortner, B. V. (1999). *The effectiveness of grief counseling and therapy: A quantitative review*. The University of Memphis, Memphis, TN.
- Fortner, B. V., & Neimeyer, R. A. (1999). *The effectiveness of grief counseling and therapy: A quantitative review*. Paper presented at the annual meeting of the Association for Death Education and Counseling, San Antonio, TX.
- Grieving is an individual process. (2003, April 6). *Baltimore Sun*, Star Edition Section: Lifestyle, p.3.

- Groopman, J. E. (2004, January 26). The grief industry: How much does crisis counseling help--or hurt? *New Yorker* 79, 30-38.
- Harkness, K. L., Shear, M. K., Frank, E., & Silberman, R. A. (2002). Traumatic grief treatment: Case histories of 4 patients. *Journal of Clinical Psychiatry*, 63(12), 1113-1120.
- Jordan, J. R., & McMenemy, J. (2004). Interventions for suicide survivors: A review of the literature. *Suicide and Life-Threatening Behavior*, 34(4), 337-349.
- Jordan, J. R., & Neimeyer, R. A. (2003). Does grief counseling work? *Death Studies*, 27, 763-786.
- Kato, P. M., & Mann, T. (1999). A synthesis of psychological interventions for the bereaved. *Clinical Psychology Review*, 19(3), 275-296.
- Lalande, K. M., & Bonanno, G. A. (2006). Culture and continuing bonds: A prospective comparison of bereavement in the United State and the People's Republic of China. *Death Studies*, 30, 303-324.
- Larson, D. G., & Hoyt, W. T. (2006, April). Deterioration effects in grief counseling: In search of the evidence. In S. Connor (Symposium Chair), Grief counseling: Can it be harmful? Paper presented at the 7<sup>th</sup> annual Clinical Team Conference and Scientific Symposium of the National Hospice and Palliative Care Organization, San Diego.
- Lehman, K. (2005). A variation on the story of the mustard seed [Review of the book *Responding to loss: A resource for caregivers*]. *Death Studies*, 29(6), 573-682.

- Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *American Psychologist*, 48(12), 1181-1209.
- McNally, R. J., Bryant, R. A., & Ehlers, A. (2003). Does early psychological intervention promote recovery from posttraumatic stress? *Psychological Science in the Public Interest*, 4(4), 45-79.
- Neimeyer, R. A. (2000). Searching for the meaning of meaning: Grief therapy and the process of reconstruction. *Death Studies*, 24, 541-558.
- Nucleus Group (2004). *Final Report: Review of Specific Grief and Bereavement Services*. Melbourne, Victoria, Australia: Metropolitan Health and Aged Care Services Division, Department of Human Services.
- Ott, C. H., & Lueger, R. J. (2002). Patterns of change in mental health status during the first two years of spousal bereavement. *Death Studies*, 26, 387-411.
- Paletti, R. (2005). Childhood bereavement intervention: A phenomenological path to resilience [Review of the book *Children, bereavement and trauma: Nurturing resilience*]. *Death Studies*, 29, 665-680.
- Reissman, D. B., Klomp, R. W., Kent, A. T., & Pfefferbaum, B. (2004). Exploring psychological resilience in the face of terrorism. *Psychiatric Annals*, 33(8), 627-632.
- Shapiro, E. (2005). When and how do helpers harm [Review of the book *Helping the helpers not to harm: Iatrogenic damage and community mental health*]. *Death Studies*, 29, 261-273.

- Shear, K. & Shair, H. (2005). Attachment, loss, and complicated grief. *Developmental Psychobiology*, 47(3), 253-267.
- Sikkema, K. J., Hansen, N. B., Kochman, A., Tate, D. C., & Difrancesco, W. (2004). Outcomes from a randomized controlled trial of a group intervention for HIV positive men and women coping with aids-related loss and bereavement. *Death Studies*, 28(3), 187-209.
- Sommers, C. H., & Satel, S. (2005). *One nation under therapy: How the helping culture is eroding self-reliance*. New York: St. Martin's Press.
- Stroebe, M. S., Hansson, R. O., Stroebe, W., & Schut, H. (Eds.). (2001). *Handbook of bereavement research: Consequences, coping, and care*. Washington, DC: American Psychological Association.
- Therapy is not always a remedy (2005, July 19). *The Irish Times*, Health: Your Lifestyle, p.2.
- Wampold, B. E. (2001). *The great psychotherapy debate: Models, methods, and findings*. Mahwah, NJ: Erlbaum.
- Wolf, B., & Jordan, J. R. (2000). Ramblings from the trenches: A clinical perspective on thanatological research. *Death Studies*, 24, 569-584.

Author Note

DALE G. LARSON received his PhD in clinical psychology at the University of California, Berkeley. He is a professor in the Department of Counseling Psychology at Santa Clara University. His research and scholarly interests include end-of-life care, interdisciplinary teams, coping with loss, grief counseling, and self-concealment.

WILLIAM T. HOYT received his PhD in counseling psych from Virginia Commonwealth University. He is an associate professor in the Department of Counseling Psychology at the University of Wisconsin-Madison. He has methodological interests in data analysis and measurement, and substantive interests in interpersonal processes in close relationships, and more generally in social determinants of psychological well-being.

We gratefully acknowledge the contributions of Dr. Gary VandenBos, the APA Publisher and the Managing Editor of the *American Psychologist*, and the American Psychological Association. We thank Kristy Bourret for her assistance on this project..

Correspondence concerning this article should be addressed to Dale G. Larson, Department of Counseling Psychology, Santa Clara University, Santa Clara, California 95053-0201. E-mail: dlarson@scu.edu

## Footnotes

<sup>1</sup>Technically, the effect sizes for Allumbaugh and Hoyt (1999) were reported as Becker's (1900)  $\Delta$ , not Cohen's (1988)  $d$ . However, interpretation of these two parameters is similar, and the two analyses will yield identical results when treatment and control groups have equivalent scores on the outcome measures prior to treatment. (For non-equivalent groups designs, Becker's method is more accurate.) Studies included in the other two meta-analyses randomly assigned participants to treatment and control groups (which is expected to equalize pre-treatment means). So, for simplicity here, we report Allumbaugh and Hoyt's findings as  $d$  statistics, to avoid burdening readers with this technical issue and to emphasize their comparability to the effect sizes from the other meta-analyses.

<sup>2</sup>As in the case of the TIDE findings, these citations have attributed Fortner's conventional meta-analytic findings to Neimeyer (2000) or later authors, not to Fortner (1999) himself.

<sup>3</sup>Although Fortner (1999) located and analyzed 23 studies for the conventional meta-analysis, only 10 of these studies provided the data needed to compute the TIDE statistic, as described earlier.

<sup>4</sup>One difference is that Allumbaugh and Hoyt sought out unpublished as well as published studies, whereas Fortner ignored unpublished research. This difference does not account for the effect size discrepancy, however, because Allumbaugh and Hoyt's effect size for the subset of published studies was  $d = 0.39$  (see Table 1, p. 373), still much larger than that obtained by Fortner.

<sup>5</sup>Occasionally, reviewers have alluded to an additional line of evidence on the ineffectiveness of early intervention for trauma. For example, Bonanno et al.(2005) cite McNally, Bryant, and Ehlers' (McNally, Bryant, & Ehlers, 2003) review of psychological debriefing interventions for trauma survivors (as well as Bonanno, 2004, and Jordan & Neimeyer, 2003) for the claim that “indiscriminately encouraging bereaved or traumatized individuals to participate in grief counseling or other forms of crisis intervention is not only ineffective but can actually be harmful” (p. 841). This type of argument seems to conflate two very different populations (bereaved individuals seeking counseling and persons who have suffered trauma and may be mandated by employers—see Groopman (2004), e.g.—to participate in a “stress debriefing”) and different interventions. We do not believe that research on trauma debriefing as a preventative against development of posttraumatic stress disorder (PTSD) is informative about the effectiveness of grief counseling, as typically practiced.

# Citation Patterns for TIDE Findings

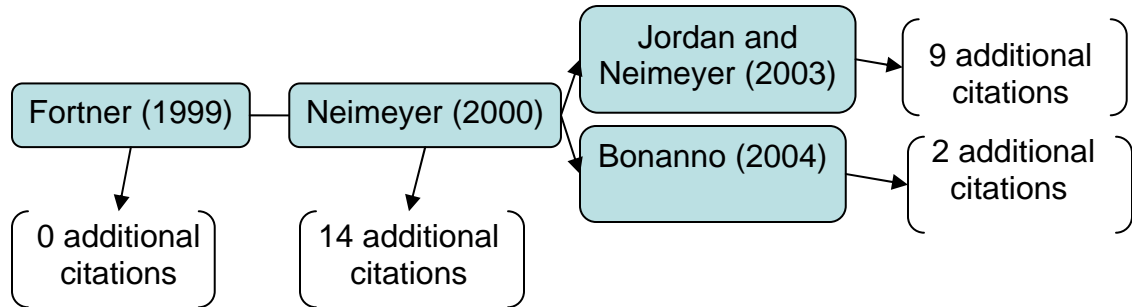


Figure Caption

*Figure 1.* References in boxes have been cited in support of the finding of treatment-induced deterioration effects (TIDE) in grief counseling. Arrows originate at the reference, and point to the study citing that reference. “Additional citations” refer to citing articles (not itemized here) that have not themselves been cited in support of TIDE claims. Publications citing Neimeyer (2000): (Bonanno, Papa, & O’Neill, 2002); (Bonanno et al., 2002); (CFAH, 2003); (Grieving, 2003); (Harkness, Shear, Frank & Silberman, 2002); (Jordan & McMenemy, 2004); (Lalande & Bonanno, 2006); (Lehman, 2005); (Nucleus Group, 2004); (Reissman, Klomp, Kent, & Pfefferbaum, 2004); (Sikkema, Hansen, Kochman, Tate, & Difrancisco, 2004); (Sommers & Satel, 2005); (Therapy, 2005); (Wolf & Jordan, 2000). Publications citing Jordan and Neimeyer (2003): (Bonanno, Moskowitz, Papa, & Folkman, 2005); (CFAH, 2003); (Craighead & Nemeroff, 2005); (Currier, Holland, & Neimeyer, 2006); (Lalande & Bonanno, 2006); (Paletti, 2005); (Shapiro, 2005); (Shear & Shair, 2005); (Sommers & Satel, 2005). Articles citing Bonanno (2004): (Reissman, Klomp, Kent, & Pfefferbaum, 2004); (Bonanno, Moskowitz, Papa, & Folkman, 2005).

## TIDE Analysis: Technical Considerations

In this Technical Appendix, we consider the validity of B. V. Fortner's (1999) analysis of treatment-induced deterioration effects (TIDE) in grief counseling. The meta-analytic TIDE findings were based on the following ratio,

$$z = \frac{M_T - M_C}{sd_T - sd_C} \quad (\text{Fortner, 1999, p. 14}),$$

which can be computed from summary data in published research reports. Fortner claimed that this ratio can be interpreted as a  $z$  score (i.e., a deviation score, calibrated in standard deviation units, from the mean of a normal distribution):

In this formula,  $M_T$  and  $M_C$  are the means of the scores on the outcome variable for the treatment and control groups, respectively;  $sd_T$  and  $sd_C$  are the standard deviations for these groups. Fortner notes that the outcome variables are scaled so that “a positive [mean difference] indicates that participants receiving intervention fared better than control participants” (p. 13). This will be the case if higher scores on the outcome variable reflect better (healthier) functioning. One assumes (although Fortner does not state this explicitly) that outcome variables that assess poorer functioning (e.g., a depression scale, on which higher scores reflect greater depression) were reverse-scored prior to analysis.

The logic of this procedure is a familiar one in meta-analysis. It is common to standardize group mean differences by dividing this difference by  $sd_C$ , with the resulting ratio (usually referred to as Cohen's  $d$  statistic—see Cohen, 1988) interpreted as the difference between the group means that has been expressed in standard deviation units (relative to the standard deviation of the scores of control group participants). The claim

that the TIDE ratio can be treated as a  $z$  score is puzzling, however, because it is not clear that the denominator ( $sd_T - sd_C$ ) represents the standard deviation of some meaningful distribution. This claim needs justification, because the interpretation of this ratio as a  $z$  score is critical to computation of the deterioration percentage, as seen below.

This conceptual problem is reinforced by a statistical consideration: In general, we expect the variance of outcome scores in the treatment group to be quite similar to the variance in the control group. (Equivalently, we expect  $sd_T$  to be quite close numerically to  $sd_C$ .) Fortner (1999, p. 18) reports that this expectation held in the studies he reviewed: The mean variance ratio (averaging across studies) was very close to 1.0, which indicates that the variances of treatment and control groups were quite similar. This means that the denominator of the formula for the TIDE ratio will ordinarily be quite close to zero. This means that one can expect the “ $z$ ” scores to be quite large, yielding very high deterioration percentages for some studies, and very low percentages for others. (We show how Fortner computed these percentages in the example below.)

This property of the TIDE ratio becomes even more disturbing when we realize that, if treatment and control groups in the population have equal variances (and therefore equal  $sds$ ), then when we take many *samples* from each of these populations, about half the time (by chance)  $sd_T$  will be larger than  $sd_C$ . The other half of the samples will have  $sd_C$  larger than  $sd_T$ . That means that for any single study, given this assumption of equal population variances, when we compute the TIDE statistic, the sign of the denominator will be arbitrary—exclusively a function of sampling error.

When we consider these two facts—a denominator that is very close to zero, and a ratio the sign of which is arbitrary—in conjunction, we may conclude that the TIDE

ratio as described by Fortner will yield extreme “ $z$ ” scores of arbitrary sign. That is, whether the “ $z$ ” score is large and positive or large and negative will be solely a function of sampling error (i.e., solely a function of whether  $sd_T$  in this sample happens to be larger or smaller than  $sd_C$ , even though the population values estimated by these sample statistics are identical). It is hard to see how a statistic with this property could be a valid estimate of deterioration in the treatment group.

Given this interpretation, however, Fortner goes on to describe how the TIDE estimate for each of the 10 studies in his analysis was derived, using a table of probability values for the normal ( $z$ ) distribution:

1. When variance is greater in the treatment group than in the control group [i.e., when the denominator of “ $z$ ” is positive], the proportion of treatment deteriorated participants is the portion of the normal distribution falling below the identified  $z$  score.
2. When variance is smaller in the treated group [i.e., when the denominator of “ $z$ ” is negative], the proportion of treatment deteriorated participants is the portion of the normal distribution falling above the  $z$  score.
3. When variance is equal in the two groups [in which case the denominator is zero, so the “ $z$ ” ratio is undefined] and the mean of the treated group is worse than the mean of the control group, 100% of the treatment participants are identified as treatment deteriorated. (pp. 14-15)

The rationale for these decision rules is unclear. To give an idea of how they work on data from an actual study, we tried them on the summary data from one of the outcome variables in the Treatment of Depression Collaborative Research Project (Imber et al., 1990)). The TDCRP served as a model for multi-site, collaborative research on psychotherapy efficacy. Only well-validated treatments were used. Interventions were fully manualized and monitored for fidelity, and they were delivered by experienced therapists who were well trained in the intervention models. We compared means on the Approval subscale of the Dysfunctional Attitudes Scale (DAS) for the cognitive-behavior therapy group ( $M_T = 32.9$ ,  $sd_T = 9.9$ ) and the placebo control group ( $M_C = 37.5$ ,  $sd_C = 9.7$ ). (This is one of the few outcome variables on which statistically significant group differences were observed for this study.)

To compute the TIDE estimate for this study, we first needed to reverse-score the outcome variable. Higher scores on the Approval subscale represent more dysfunctional attitudes; higher scores on the reverse-scored variable represent more functional (healthier) attitudes. This was accomplished simply by reversing the sign of the means for the treatment and control groups:

$$z = \frac{M_T - M_C}{sd_T - sd_C} = \frac{-32.9 - (-37.5)}{9.9 - 9.7} = \frac{4.6}{0.2} = 23.0.$$

As expected, given our analysis above, the obtained "z" score is very large. Interpreting this finding according to Decision Rule 1, above (because  $sd_T > sd_C$ ), we would conclude that *virtually all patients in the treatment group deteriorated*. (The proportion who did *not* deteriorate, according to this rule, is about  $2.3 \times 10^{-18}$ , or only two out of every billion *billion* patients!)

The absurdity of the claim that virtually all clients in the treatment group deteriorated (despite the fact that the post-treatment mean for this group was significantly higher than that for the control group on this variable) led us to consider a final noteworthy property of the TIDE index described by Fortner (1999). This characteristic is evident from examination of the numerator, which represents the difference between the means for the treatment and control groups. A moment's reflection reveals that the larger the absolute value of this difference (assuming numerator and denominator have the same sign), the larger the value of the "z" score, and therefore the higher the estimated deterioration percentage. But this mean difference is also the treatment effect (in raw score units). This means that, if numerator and denominator have the same sign, the larger the treatment effect, the greater will be the proportion of clients who are claimed to have deteriorated. Clearly, if this were a valid index of treatment deterioration, we would expect just the opposite—more effective treatments would be associated with fewer adverse outcomes.

Thus, a careful consideration of the method described by Fortner (1999, pp. 14-15) indicates that it cannot yield a valid index of treatment-induced deterioration. The rationale for the statistic has never been published, nor has its interpretation as a z score (representing the location on the normal curve of the first non-deteriorated participant) been vetted by knowledgeable methodological reviewers. Analysis of the likely value of this ratio under typical conditions indicates that it will produce "z" scores that are quite extreme; indeed, application of Fortner's procedure to summary data from the TDCRP study yielded patently nonsensical results. We believe that the most sensible conclusion from this analysis is that the TIDE statistic, at least as described by Fortner, is not a valid

index of deterioration in intervention groups, and that the 38% TIDE figure for grief counseling has no valid statistical or empirical basis.