Psychological Bulletin
2007, Vol. 133, No. 3, 367–394

0033-2909/07/$12.00 DOI: 10.1037/0033-2909.133.3.367

Psychotherapy and Survival in Cancer: The Conflict Between Hope and Evidence

James C. Coyne
Abramson Cancer Center of the University of Pennsylvania

Michael Stefanek
American Cancer Society

Steven C. Palmer
Abramson Cancer Center of the University of Pennsylvania

Despite contradictory findings, the belief that psychotherapy promotes survival in people who have been diagnosed with cancer has persisted since the seminal study by D. Spiegel, J. R. Bloom, H. C. Kramer, and E. Gottheil (1989). The current authors provide a systematic critical review of the relevant literature. In doing so, they introduce some considerations in the design, interpretation of results, and reporting of clinical trials that have not been sufficiently appreciated in the behavioral sciences. They note endemic problems in this literature. No randomized clinical trial designed with survival as a primary endpoint and in which psychotherapy was not confounded with medical care has yielded a positive effect. Among the implications of the review is that an adequately powered study examining effects of psychotherapy on survival after a diagnosis of cancer would require resources that are not justified by the strength of the available evidence.

Keywords: metastatic breast cancer, randomized clinical trial, supportive–expressive, depression, CONSORT

The belief that psychological factors affect the progression of cancer has become prevalent among the lay public and some oncology professionals (Doan, Gray, & Davis, 1993; Lemon & Edelman, 2003). An extension of this belief is that improvement in psychological functioning can prolong the survival after a diagnosis of cancer. Were this true, psychotherapy could not only benefit mood and quality of life but increase life expectancy as well. Indeed, there is some lay acceptance of this notion, as a substantial proportion of women with breast cancer attending support groups do so believing they may be extending their lives (Miller et al., 1998).

Two studies (Fawzy et al., 1993; Spiegel et al., 1989) have been widely interpreted as providing early support for the contention that psychotherapy promotes survival. Neither study, however, was designed to test this hypothesis. Provocative claims have been made that women with metastatic breast cancer who received supportive–expressive group psychotherapy survived almost twice as long as women in the control group (Spiegel et al., 1989). Claims have also been made that group cognitive–behavioral therapy provided persons with malignant melanoma with a sevenfold decrease in risk of death at 6-year follow-up and a threefold decrease in risk of death at 10 years (Fawzy, Canada, & Fawzy, 2003; Fawzy et al., 1993).

Yet studies yielding null findings include a large-scale, adequately powered clinical trial attempting to replicate the Spiegel et al. (1989) intervention, on which Dr. Spiegel served as a consultant (Goodwin et al., 2001). Three meta-analyses have also failed to find an overall effect of psychotherapy on survival (Chow, Tsao, & Harth, 2004; Edwards, Hailey, & Maxwell, 2004; Smedslund & Ringdal, 2004). More positive assessments of the literature have been made on the basis of box scores derived from diverse studies of interventions with people with cancer (Sephton & Spiegel, 2003; Spiegel & Giese-Davis, 2004). Before the publication of an additional null trial (Kissane et al., 2004), Spiegel and Giese-Davis (2004) concluded that "5 of 10 randomized clinical trials demonstrate an effect of psychosocial intervention on survival time" (p. 275). They proposed a variety of mechanisms by which psychological factors might affect disease progression. Similarly, Sephton and Spiegel (2003) declared, "If nothing else, these studies challenge us to systematically examine the interaction of mind and body, to determine the aspects of therapeutic intervention that are most effective and the populations that are most likely to benefit" (p. 322).

Enumerating the mechanisms by which a phenomenon might occur increases confidence that there is actually a phenomenon to explain (Anderson, Lepper, & Ross, 1980), and repeating claims that psychotherapy promotes survival may lend more credibility than is warranted by the evidence. Consensus appears to be growing that the evidence for a benefit to survival attributable to

James C. Coyne and Steven C. Palmer, Department of Psychiatry, Abramson Cancer Center of the University of Pennsylvania; Michael Stefanek, Behavioral Sciences, American Cancer Society, Atlanta, Georgia.

This article was inspired in large part by the original critiques of Spiegel, Bloom, Kraemer, and Gottheil's (1989) study provided by Bernard H. Fox (1995, 1998, 1999). Special thanks are extended to Lydia R. Temoshok for her explanation of Dr. Fox's key points.

Correspondence concerning this article should be addressed to James C. Coyne, Department of Psychiatry, University of Pennsylvania School of Medicine, 3535 Market Street, Philadelphia, PA 19104. E-mail: jcoyne@mail.med.upenn.edu

psychotherapy is, at best, "mixed" (Lillquist & Abramson, 2002, p. 65), "controversial" (Schattner, 2003, p. 618), or "contradictory" (Greer, 2002, p. 238). However, ambiguity as to the implications of such assessments remains (Blake-Mortimer, Gore-Felton, Kimerling, Turner-Cobb, & Spiegel, 1999; Palmer & Coyne, 2004; Ross, Boesen, Dalton, & Johansen, 2002), and it is unclear what would be required to revise a claim, based on a recent meta-analysis that found no effect of psychotherapy on survival, that "a definite conclusion about whether psychosocial interventions prolong cancer survival seems premature" (Smedslund & Ringdal, 2004, p. 123).

Can we move beyond the unsatisfying ambiguity of an appraisal of the available evidence as mixed, controversial, or contradictory? It is the nature of science that provocative findings from a well-conducted study can unseat a firmly established conclusion. In that sense, the claim that "further research is needed" can always be made. However, important decisions need to be based on the existing evidence: Namely, what priority should be given to further studies examining survival and psychotherapy, and more immediately, what advice should be given to patients contemplating psychotherapy as a means of extending their lives? These decisions take on more importance in the face of scarce research funding and restricted coverage for psychotherapy from third-party payers.

An evaluation of this literature has broad implications. For instance, disagreement over whether Spiegel et al. (1989) and Fawzy et al. (1993) demonstrated a genuine effect of psychotherapy on survival figured centrally in a great debate over whether psychosocial interventions improve clinical outcomes in physical illness (Relman & Angell, 2002; Williams & Schneiderman, 2002). Some of the valuation of psychosocial interventions in cancer care has been based on the presumption that they might promote survival, not only reduce distress or improve quality of life (Cunningham & Edmonds, 2002; Greer, 2002). If this presumption remains a cornerstone of the argument that patients should be provided with psychosocial care, the credibility of a range of interventions and justification for the role of mental health professionals in cancer care will depend on psychotherapy contributing to survival. In addition, as Lesperance and Frasure-Smith (1999) noted in another context, "Prevention of mortality has always been one of the most important factors in determining the allocation of funding for research and clinical activities" (p. 18).

There are, however, risks to promoting survival as the crucial endpoint in studies of psychotherapy among people with cancer, particularly when an effect has not been established and when such a focus can be construed as deemphasizing the importance of improvements in quality of life and psychosocial functioning. Lesperance and Frasure-Smith (1999) recognized this, and their opinion is noteworthy because their initial studies provided part of the justification for efforts to demonstrate that psychotherapy for depression would reduce mortality in persons who had recently suffered a myocardial infarction—an effort that ultimately proved unsuccessful (Berkman et al., 2003). They cautioned that "although the prevention of death is a powerful tool to influence many of our medical colleagues . . . death is not everything" (Lesperance & Frasure-Smith, 1999, p. 19). Staking the main claim for the importance of psychosocial intervention on survival distracts from more readily demonstrable effects on psychosocial well-being and quality of life. Moreover, if claims about the effects of psychotherapy on survival are advanced and then abandoned, it becomes an undignified retreat to claim importance for psychosocial interventions based on their "mere" psychosocial benefits. An unwarranted strong claim could thus undercut the credibility of what has always been a reasonable claim.

The argument has also been made that there are no deleterious effects for people with cancer of participating in psychotherapy (Spiegel & Giese-Davis, 2004). Yet the mean change scores for mood measures of women with metastatic breast cancer who have received supportive–expressive therapy are often dwarfed by the variance in these scores (e.g., Goodwin et al., 2001), allowing for considerable adverse reactions on an individual basis, and there has been no systematic effort to determine whether participation is benign for all individuals (Chow et al., 2004). That psychotherapy can have negative as well as positive effects is well established (Hadley & Strupp, 1976), and there is some evidence of negative effects of participation in peer support groups for women with breast cancer, including declines in self-esteem and body image and increased preoccupation with cancer (Helgeson, Cohen, Schulz, & Yasko, 1999, 2001). If nothing else, attendance of weekly sessions for a year or more (as in Spiegel et al., 1989, or Goodwin et al., 2001) places considerable demands on ill and dying patients that are difficult to justify when therapy is sought with the expectation that it will prolong life.

On the other hand, if the evidence suggests that psychotherapy does not extend survival, people with cancer might lose confidence in their ability to influence the course and outcome of their disease. This belief contributes to morale and promotes effective coping regardless of its validity. Yet it would be disrespectful of patient autonomy to knowingly provide patients with illusions, even if it were with the intention of improving adaptation. Proponents of a survival effect (e.g., Spiegel, 2004) and other psycho-oncologists (e.g., Holland & Lewis, 2001) have actively discouraged the implication that the attitudes of persons with cancer are responsible for their disease progression. Nonetheless, a spoof article in the parody newspaper *The Onion* headlined "Loved Ones Recall Man's Cowardly Battle With Cancer" comes too close to the sense of some people with cancer that a judgment is being made that "brave and good people defeat cancer and that cowardly and undeserving people allow it to kill them" (Diamond, 1998, p. 52). If psychotherapy does not prolong survival, recognition of this would remove one basis for blaming persons with cancer for progression of their disease, however unfair such negative views are in the first place.

Rationale

The process of critically examining the evidence could have important benefits for people who have been diagnosed with cancer, for psycho-oncology, and for behavioral medicine more generally. Critical evaluation involves recognizing a number of underlying assumptions that have not been well articulated in the behavioral medicine literature. These assumptions will undoubtedly be confronted in other contexts, and it is desirable to be better prepared to recognize them when they recur. Namely:

1. *Claims that psychotherapy extends life after a diagnosis of cancer are claims about medical effects.* Claims for possible medical benefits of psychotherapy need to be evaluated with the usual scrutiny to which medical claims are subject. The standards

of evidence should not be lowered when the intervention is psychosocial, nor should we accept as evidence methodology that would not be acceptable when evaluating other medical claims. Much of the evidence for a survival benefit comes from two trials with small sample sizes in which survival was not an a priori primary endpoint (Fawzy et al., 1993; Spiegel et al., 1989). Unexpected benefits for survival in modest scale studies are intriguing, but they require the balance between interest and skepticism that ultimately guides hypothesis-driven research.

2. *Claims that psychotherapy prolongs the life after a diagnosis of cancer are based on the results of randomized clinical trials, and interpretation of these results is not a straightforward task*. The methodologies used in the conduct of randomized clinical trials involve a number of assumptions that differ from those of the particular experimental tradition in which many behavioral and social scientists are trained. Even in fields more familiar with randomized clinical trials, interpretation of results is based on the transparency with which methodological decisions are reported. In medicine, recognition that many randomized clinical trials were not being reported in a manner that allowed independent evaluation led to calls for reform, culminating in the original (Begg et al., 1996) and revised (Altman et al., 2001) Consolidated Standards of Reporting Clinical Trials checklist (CONSORT; see Appendix) as a means of reforming the reporting of randomized clinical trials and making methodology transparent. Recently some psychology journals, led by *Annals of Behavioral Medicine*, *Journal of Pediatric Psychology*, and *Health Psychology* and followed later by *Journal of Consulting and Clinical Psychology*, joined the over 200 medical journals in endorsing CONSORT, but the checklist, its rationale, and its application are not widely understood in the behavioral and social sciences. There is an indication that, as judged by CONSORT standards, the reporting of the results of randomized clinical trials in psychology journals has been substandard generally (J. M. Cook, Palmer, Hoffman, & Coyne, in press; Stinson, McGrath, & Yamada, 2003), just as the reporting of psychosocial interventions for people with cancer in particular has been (Coyne, Lepore, & Palmer, 2006). CONSORT can be used to evaluate the quality of reports of randomized clinical trials relevant to claims about psychotherapy prolonging life. This exercise can serve to illustrate for more general purposes what is entailed in adhering to CONSORT.

Well-conceived and well-reported randomized clinical trials are, presumably, well-conceived and well-reported experiments. Yet, as seen in the rationale for the National Institute of Health's annual Summer Institute on Design and Conduct of Randomized Clinical Trials and the organizing of the Society of Behavioral Medicine's Evidence-Based Medicine Working Group, there are specialized bodies of knowledge needed for conducting, reporting, and interpreting randomized clinical trials. This knowledge cannot be inferred from an understanding of conventional experimental design in the social and behavioral sciences alone. Some of this knowledge is technical, but some is practical and ethical. Examining how these issues arise in studies deemed relevant to psychotherapy and survival can serve as an example of how these issues need to be addressed more broadly in behavioral medicine.

3. *Claims about survival benefits are often made using statistical techniques and interpretations that are unfamiliar to social and behavioral scientists*. Survival curves, slopes analysis, and proportional-hazard modeling are not typically addressed in social science graduate training. Although these techniques are often applied appropriately, their interpretation should seldom be taken at face value, and social and behavioral scientists may be less than well equipped to evaluate these interpretations without additional training. For example, Fawzy et al.'s (2003) statement that melanoma patients receiving psychoeducational intervention had a sevenfold decrease in relative risk of death after 6 years may seem to be a declaration of an exceptionally strong effect. The curious reader, however, may discover that reclassification of a single patient would remove the statistical significance of the effect, and that a number of patients in the intervention group who were unlikely to show a benefit of treatment had been excluded from analysis (Fox, 1995; Palmer & Coyne, 2004). Statistical issues such as this are likely to continue to arise in behavioral medicine, and we hope to provide some examples of how they can be explored.

4. *Evaluating claims that psychotherapy prolongs life after a diagnosis of cancer involves integrating the results of trials that differ in their quality, primary outcomes, recruitment criteria, and sample sizes and in the interventions being evaluated.* Integrating these disparate data is a difficult task, and there are no simple solutions. Commentators have variously relied on narrative review, box scores, and meta-analysis, but the studies typically considered have been described as a mixture of "apples and oranges" (Smedslund & Ringdal, 2004, p. 123; Spiegel, 2004, p. 133).

How does one select relevant studies and integrate their findings in a way that takes into account their broad-ranging differences? For example, how does one reconcile or weigh evidence when the two studies offering the strongest support for a survival effect—Spiegel et al. (1989) and Fawzy et al. (1993)—were not designed with this as an a priori hypothesis, whereas studies for which this was the express hypothesis have not found an effect? Should the latter studies be given more weight? Without adequate reporting of results, how are we, as a field, to disentangle conflicting outcomes? Spiegel (2002) acknowledged that there is an implausibility to the hypothesis of a survival effect. How do we take into account that some unknown proportion of investigators of psychosocial interventions for people with cancer agree with this assessment and therefore do not undertake a post hoc follow-up of their study participants?

Although analogous questions about how to integrate the findings of diverse studies are routinely confronted in psychology and the behavioral sciences, there has been much less skepticism expressed about the wisdom of integrating diverse studies than has occurred in clinical epidemiology and medicine (Chalmers, 1991; Feinstein, 1995; LeLorier, Gregoire, Benhaddad, Lapierre, & Derderian, 1997; Smith & Egger, 1998). A critical review of the literature concerning psychotherapy and survival of cancer patients provides an opportunity to confront some of the differences in how studies are identified, evaluated, weighed, and integrated across disciplines.

Purpose and Organization of the Article

We have undertaken this review in order to address a topic of pressing scientific and clinical importance. Yet our review is also intended to raise issues of broader relevance, with the goal of improving the standards of the field and with implications for the

subsequent design and interpretation of clinical trials in behavioral medicine. Our strategy will be to (a) proceed from a critical narrative review of the individual trials reporting data that have been deemed relevant to the hypothesis that psychological interventions promote survival in people with cancer; (b) provide a more systematic evaluation of the adequacy with which these trials have been reported through an application of the CONSORT criteria; (c) examine attempts to integrate these trials that have formed global conclusions using box scores and meta-analysis; and (d) end with an integrative summary and commentary that provides clinical and public policy implications and a look to the future.

The Key Studies

Spiegel (2001) and Spiegel and Giese-Davis (2003) included 10 studies in their box score evaluation of whether psychotherapy improved survival (see Table 1), and it is clear that the Kissane et al. (2004) study would have been added had it been published at the time of their reviews. Kissane et al. provided survival data for a randomized clinical trial evaluating cognitive–existential group psychotherapy for persons who had been diagnosed with cancer, and in this case survival was an a priori outcome. Spiegel and colleagues were not entirely clear on their criteria for selecting these particular studies to the exclusion of others. All but one of the studies they discussed are randomized clinical trials, which are considered the strongest form of evidence for efficacy (Higgins & Green, 2005). The one study that is not a randomized clinical trial (J. L. Richardson, Shelton, Krailo, & Levine, 1990) has a quasi-experimental, sequential cohort design, but this study has tended to be treated by commentators as a randomized clinical trial (Smedslund & Ringdal, 2004, is an exception), and perhaps Spiegel (2001; Spiegel & Giese-Davis, 2003) simply failed to note that it was not a randomized clinical trial. Spiegel (2001; Spiegel & Giese-Davis, 2003) excluded without comment a large randomized clinical trial (Grossarth-Maticek, Frentzel-Beyme, & Becker, 1984) claimed by its investigators to have demonstrated an effect on survival. However, elsewhere, Spiegel (1991) dismissed the results claimed for this trial as too strong to be credible, and this is an opinion shared by others (Fox, 1999; Ross et al., 2002).

Smedslund and Ringdal (2004) conducted a thorough search of the literature and failed to uncover additional randomized clinical trials examining survival as an endpoint. Some reviewers have accepted Spiegel's (2001) and Spiegel and Giese-Davis's (2003) entire list (Goodwin, 2004), whereas other reviewers have excluded some of the studies (Chow et al., 2004; Ross et al., 2002; Smedslund & Ringdal, 2004). Chow et al. excluded one study (McCorkle et al., 2000) cited by Spiegel as supporting an effect of psychotherapy on survival, because of nursing and medical components to the intervention, and Ross et al. excluded the same trial without commenting why. Smedslund excluded one trial (Linn, Linn, & Harris, 1982) from meta-analysis counted by Spiegel because the requisite hazards ratio was not provided. Smedslund and Ringdal included three additional trials (Bagenal, Easton, Harris, Chilvers, & McElwain, 1990; Gellert, Maxwell, & Siegel, 1993; Shrock, Palmer, & Taylor, 1999), although none of them were randomized, as well as a fourth study (Ratcliffe, Dawson, & Walker, 1995) for which they could not determine whether treatment was by random assignment.

For the purposes of the present review, we are accepting the 10 studies entered into Spiegel's (2001) box score plus Kissane et al. (2004) because it seems to meet the criteria for inclusion. We will revisit the issue of J. L. Richardson et al. (1990) not being a fully randomized clinical trial but accept the view of Spiegel and others that the earliest trial (Grossarth-Maticek et al., 1984) is not a credible addition to the literature. (Readers interested in further discussion on the status of Grossarth-Maticek et al. are encouraged to consult Volume 2 [1999], Issue 3 of *Psychological Inquiry*.) These studies are heterogeneous in terms of quality, patient populations sampled, and interventions being evaluated, and there is room for critical evaluation of how they were selected and whether or how they should be integrated. Of importance, we will consider whether this box score is an adequate means of summarizing the relevant literature. But it would be useful to first have narrative summaries of each, as there is at least some consensus among reviewers and commentators as to their individual relevance, and we wish for readers to be able to form judgments independent of our own.

Application of CONSORT

The CONSORT standards (Altman et al., 2001) provide a means of evaluating the adequacy of the reporting of randomized clinical trials. Although focusing on initial reporting of primary outcomes from two-arm parallel trials, it can be applied to other designs. The goal of CONSORT is to ensure transparency of reporting of clinical trials so that readers can assess the strengths and weaknesses of a trial and use this information to make informed judgments concerning outcomes. It is hoped that through greater transparency in reporting, the quality of trials themselves will be improved. CONSORT encompasses items (see Appendix) that cover adequacy of reporting in the title, abstract, introduction, method, results, and discussion sections. Item content is rated as present or absent, yielding an overall score and allowing one to examine reporting deficiencies.

Some caveats need to be kept in mind when interpreting CONSORT scores for published studies. Evaluations of the adequacy of trials as sources of efficacy data increasingly refer to CONSORT ratings (Coyne et al., 2006; Manne & Andrykowski, 2006), and noncompliance with some items is empirically associated with confirmatory bias (Schulz, Chalmers, Hayes, & Altman, 1995). Yet transparency of reporting is not equivalent to adequacy of methodology. Poor reporting sometimes represents inadequate description of adequately conducted trials (Soares et al., 2004). Furthermore, investigators who explicitly acknowledge methodological inadequacies in their conduct of a trial may score higher than those who fail to report that their trials were adequate in the same respect. Thus, reporting in a manner compliant with CONSORT needs to be seen as a necessary but not sufficient indicator of study quality. In applying CONSORT to the studies under review here, we will be getting some impressions of CONSORT ratings as indicators of study quality, as well as evaluating the studies themselves. Our effort will thus be one of the first examinations of the usefulness of CONSORT for this purpose.

There are some challenges in applying CONSORT to a literature such as this, with the most pressing concerning the time span over which these reports were published. Trials published before adoption of CONSORT cannot be expected to fully comply with

Table 1
Methodological Concerns and Consolidated Standards of Reporting Trials (CONSORT) Scores

Investigator	Methodological and analytical concerns	CONSORT points scored
Spiegel et al. (1989)	1. Survival not a priori endpoint 2. Possible cointervention confound 3. Study underpowered for survival analysis 4. Use of mean (vs. median) survival time 5. Integrity of intervention intensity 6. Possible bias in initial sampling	4, 12a, 12b, 13a, 13b, 15, 22
Fawzy et al. (1993)	1. Survival not a priori endpoint 2. Study underpowered for survival analysis 3. No intent-to-treat analysis 4. Inappropriate analysis and presentation of data	3a, 4, 12a, 12b, 14
J. L. Richardson et al. (1990)	1. Survival not a priori endpoint 2. Possible cointervention confound 3. Study underpowered for survival analysis 4. Quasi-experimental study design 5. Potential bias in death ascertainment 6. Survival curve presentation inconsistent with study design 7. Multivariate analysis overfitted 8. No explicit psychotherapy component	2, 3b, 4, 8b, 12a, 12b, 14, 18, 22
Kuchler et al. (1999)	1. Survival not a priori endpoint 2. Possible cointervention confound 3. Randomization not preserved	3a, 7a, 8b, 12a, 13a, 13b, 14, 15, 16, 18, 20, 22
McCorkle et al. (2000)	1. Randomization scheme unclear 2. Intervention explicitly medically focused 3. No survival effect in primary analyses (only in subgroup analyses)	3a, 4, 12a, 12b, 13a, 14, 15, 16, 21, 22
Linn et al. (1982)	1. Survival specifically rejected as a priori endpoint 2. No intent-to-treat analysis	3a, 5, 13a, 14, 22
Ilnyckyj et al. (1994)	1. Survival not a priori endpoint 2. Study underpowered for survival analysis 3. No intent-to-treat analysis 4. Significant attrition pre- and postrandomization 5. Interventions poorly described 6. Inconsistent levels of treatment exposure	1, 3a, 8b, 12a, 13a, 13b, 15
Edelman, Bell, & Kidman (1999)	1. Survival not a priori endpoint 2. Inconsistent levels of treatment exposure 3. Treatment integrity 4. Abbreviated follow-up period 5. Multivariate analysis overfitted	6a, 14, 15, 20, 22
Cunningham et al. (1998)	1. Study underpowered for survival analysis	1, 3b, 4, 8b, 9, 10, 12a, 12b, 15, 16, 20, 21, 22
Goodwin et al. (2001)	1. Possible cointervention confound 2. Treatment integrity	3a, 4, 5, 7a, 8a, 8b, 11a, 12a, 12b, 14, 15, 16, 18, 22
Kissane et al. (2004)	1. Rationale for sample (early-stage disease) unclear 2. Treatment integrity 3. Possible co-intervention bias 4. Integrity of intervention intensity	3a, 4, 7a, 8a, 8b, 12a, 12b, 13a, 14, 15, 16, 17, 18

Note. Scores on CONSORT range from 0 to 29, with higher scores indicating higher quality reporting of the design and analysis of trials.

current reporting standards. Yet another challenge is that survival was not originally designated as an outcome in many of the trials considered as relevant to the question of whether psychotherapy promotes survival, and trials not reporting original primary outcome variables are not specifically covered under CONSORT. Even within these limitations, CONSORT can be applied to allow us to determine the extent to which deficiencies in reporting and design of this set of trials should influence our evaluation of the claims that have been made from them.

Methods of Evaluation

In addition to a collaborative systematic narrative review of each article by the three authors, all articles were rated independently by two of the authors (James C. Coyne and Steven C. Palmer) in an unblinded fashion according to a modified CONSORT checklist (see Appendix). Although CONSORT is commonly described as comprising 22 items, some of the items are multifaceted and identified with both a number and letter (e.g., 6a,

6b; 7a, 7b), allowing possible scores on 29 items. As well, consistent with past applications of CONSORT (e.g., Stinson et al., 2003), items that were inapplicable to a given trial were scored as "absent." Although this solution is less than ideal, it allows our findings to be compared with other sets of studies to which CONSORT standards have been applied.

Disagreements between raters were resolved through consensus. Reliability was assessed using the kappa statistic (Cohen, 1960) for item-level analysis of individual articles and through interrater reliability at the level of composite item total scores across articles. Overall agreement on presence versus absence of CONSORT-consistent reporting was high (83%) at the item level within articles. Chance-adjusted interrater reliability was moderate, with kappas for the item-level ratings of articles ranging from .34 to .73 ($M = .57$). At the level of the collapsed 29 CONSORT items, interrater reliability was high ($r = .79$, $p < .01$).

On average, articles were compliant with fewer than one third of the CONSORT items ($M = 9.1$, $SD = 3.5$). Indeed, the most compliant articles (Cunningham et al., 1998 [13:29]; Goodwin et al., 2001 [14:29]; Kissane et al., 2004 [13:29]) met standards for fewer than 50% of the CONSORT items. Overall, 69% ($n = 20$) of the CONSORT items were adequately addressed by authors less than 50% of the time, and 49% ($n = 14$) were endorsed less than 25% of the time. Four items assessing reporting of enhancement of reliability (6b), stopping rules and interim analyses (7b), assessment of blinding (11b), and reporting of adverse events (19) received no endorsement. As well, six items assessing scientific background and rationale (2), identification of endpoints (6a), generation and implementation of the randomization scheme (9, 10), blinding (11a), and reporting of effect sizes and precision (17) were each endorsed by only 1 of the 11 studies. Clearly the transparency or clarity of reporting is less than ideal for allowing individuals to make informed judgments about the validity of claims made by authors regarding the relationship of psychotherapeutic intervention to survival. We believe, however, that brief summaries of the various strengths and weaknesses of the reporting in each study will allow the reader some insight into the difficulties faced when reconciling these diverse literatures.

Results

Spiegel et al. (1989)

Spiegel et al. (1989) reported the effects on survival of what they identified as a 1-year, structured group intervention delivered to women with metastatic breast cancer. The intervention was described in the original reports (Spiegel et al., 1989; Spiegel, Bloom, & Yalom, 1981) as focusing on discussions of coping with cancer and encouragement to express feelings. Content included redefining life priorities and detoxifying death, building bonds, management of physical problems and side effects of treatment, and self-hypnosis for pain management. The authors reported that the mean time from randomization to death was approximately twice as long in the active intervention group (36.6 months) as compared with the control group (18.9 months).

Primary endpoints. Survival was not an a priori primary endpoint in this study. The study was originally designed to examine the effect of group psychotherapy on psychosocial outcomes (Spiegel et al., 1981). The follow-up and survival analysis were undertaken post hoc, with the investigators initially favoring the null hypothesis of no effect on survival:

> We intended in particular to examine the often overstated claims made by those who teach cancer patients that the right mental attitude will help to conquer the disease. In these interventions patients often devote much time and energy to creating images of their immune cells defeating the cancer cells. (Spiegel et al., 1989, p. 890)

Intervention and cointervention. A cointervention confound refers to the differential provision of additional nonstudy treatments in a clinical trial (D. J. Cook et al., 1997), rendering the intended comparisons among treatment conditions more difficult to interpret. Thus, if medical patients assigned to a group psychotherapeutic intervention are encouraged to seek medical attention for any health problems observed by group leaders or members, it would be difficult to distinguish the effects of the psychotherapy being provided from this additional surveillance and care, particularly for medical outcomes such as survival. There is good reason to believe that psychotherapeutic intervention in Spiegel et al. (1989) was confounded with additional supportive care and enhanced medical surveillance. This presents problems for distinguishing the independent effects of psychotherapy on health outcomes and for specifying the mechanism by which any effects occurred.

More elaborated discussions of the intervention have suggested that it was longer, more intensive, and broader in focus than implied by the initial reports. For example, groups continued beyond a year (Kraemer & Spiegel, 1999). A report from Spiegel's replication study (Classen et al., 2001) noted one woman remaining in a group in that study for 8 years, but we have no indication of how long women remained in treatment in the original Spiegel et al. (1989) study. Spiegel (e.g., 1996) has emphasized that the groups differed from conventional group therapy in encouraging development of an active community that extended outside of the formal sessions. Members shared phone numbers and addresses and would have supplementary gatherings in the cafeteria after formal sessions. They also held meetings in the homes of dying members and accompanied one another to medical appointments (Spiegel & Classen, 2000). The implications of assignment to the group intervention for receipt of medical care have also become less clear. In talks, Spiegel (e.g., 1996) has mentioned encouraging group members to seek better pain management from their physicians. Discussing contact between therapists and the oncology treatment team in another study (Kuchler et al., 1999) Spiegel and Giese-Davis (2004) contended that consultation and coordination with medical care is routine in psychotherapy with medically ill patients. Regardless, likely cointervention bias would make it difficult to attribute any differences to the implementation of psychotherapy alone.

Analytic issues. Spiegel et al. (1989) reported that "the intervention group lived on average twice as long as did controls" (p. 889) on the basis of mean survival time. As well, there was a significant mean survival difference from first metastasis to death favoring the intervention group (58.4 months vs. 43.2 months), though no difference in survival from initial medical visit to death. Cox regression analyses controlling for stage remained significant.

A key issue concerns whether mean survival time is the best summary statistic for the effects of treatment. Given the skewness of most survival curves, median survival time is generally consid-

ered the better expression of central tendency because the median reduces the possible excessive influence of outliers (Motulsky, 1995). Sampson (2002) estimated that median survival times differ between Spiegel et al.'s (1989) intervention and control groups by only 2 months. Edwards et al. (2004) concurred that median survival did not differ between the intervention and control groups. Similarly, variability differed greatly between the groups, suggesting that outcomes were more inconsistent in one group than in the other. In this case, the intervention group had a variance 12 times that of the controls, suggesting that the at least some members of the intervention group experienced outcomes extremely different from those experienced by others assigned to the same intervention.

Exposure to intervention. The results reported were analyzed on an intent-to-treat basis: The outcomes of all randomized patients were included, regardless of exposure to the intervention. This is entirely appropriate (Lee, Ellenberg, Hirtz, & Nelson, 1991; Peto et al., 1977), and indeed, whether intent-to-treat analyses are available is one of the basic criteria by which adequacy of the reporting of randomized clinical trials is evaluated (Altman et al., 2001; Schulz, Grimes, Altman, & Hayes, 1996). Intent-to-treat analyses address the question of how effective the intervention would be if offered outside the clinical trial, and they preserve the baseline equivalence achieved by randomization (Lee et al., 1991; Peduzzi, Henderson, Hartigan, & Lavori, 2002).

However, much can be learned from "as treated" analyses that take exposure to treatment into account. Of the 50 patients assigned to the intervention in Spiegel et al. (1989), 14 were too ill to participate, 6 died before the group began, and 2 moved away. Another 15 died during the intervention period, and an undisclosed additional number did not receive the full course of intervention. Thus, an effect was found even though a considerable number of assigned patients received no exposure to intervention and *most* received substantially less than a full course. Overall, this suggests that the intervention would have to be even more powerful than would be implied from the intent-to-treat analysis, a point that becomes important when the question is raised of whether the results are too strong to reflect credible effects of psychotherapy on survival.

Power, sampling, and Type I error. Unanticipated strong findings invite scrutiny. Aside from the issue of exposure to treatment, the small group size meant that the study was underpowered to find anything but a large effect. Although low statistical power would not seem to be a basis for discounting an apparent strong effect, there are reasons to doubt the validity of an improbable result obtained with a small sample (e.g., Piantadosi, 1990). Indeed, when hypothesized, findings of small-to-moderate benefits in a large trial are more plausible than unexpectedly large benefits in a small trial. From a Bayesian perspective, such a finding in a trial with a low prior probability of finding an effect is likely to represent a false positive (Berry & Stangl, 1996; Peto et al., 1976). In keeping with this notion, it has been repeatedly found in medicine that summary positive findings from an accumulation of small trials are not replicated when a large-scale, appropriately powered study is undertaken (LeLorier et al., 1997).

Contributing to the likelihood of a false positive is the vulnerability of small samples to uncontrolled group differences, even when there has been no obvious breakdown in randomization procedures. With a small sample, either unmeasured variables or those for which there are no significant group differences can significantly influence outcomes, particularly when acting in a cumulative or synergistic fashion:

> In a RCT, the balance of pretreatment characteristics is merely one test of the adequacy of randomization and not proof that influential imbalances do not exist. Also, because such tabulations are invariably marginal summaries only (i.e., the totals for each factor are considered separately), they provide essentially no insight into the joint distribution of prognostic factors in the two treatment groups. It is simple to envision situations in which the marginal imbalances of prognostic factors are minimal, but the joint distributions are different and influential. (Piantadosi, 1990, p. 2)

With a few exceptions (Edelman, Craig, & Kidman, 2000; Edwards et al., 2004; Fox, 1995, 1998; Palmer & Coyne, 2004; Sampson, 1997, 2002; Stefanek, 1991; Stefanek & McDonald, in press), the over 900 citations of Spiegel et al. (1989) have tended to accept the investigators' interpretation of their results, even when noting that replication is needed. Sampson (2002) questioned the adequacy of the randomization, noting that the original report lacked details concerning randomization ratio and how individual patients were randomized. As seen in CONSORT, such details are now considered basic to the reporting of clinical trials. Sampson (2002) cited a 1997 personal communication from Dr. Spiegel indicating that straws were drawn for a 2:1 ratio favoring intervention. However, Sampson noted that the obtained 50:36 ratio is unlikely ($p = .06$) to result from a 2:1 strategy.

Regardless, anomalies in sampling may present difficulties for small trials. Until 2 years after randomization, survival curves for the intervention and control groups in Spiegel et al. (1989) were "almost superimposable" (Fox, 1998, p. 361). However, both Sampson (1997) and Fox (1995) observed an extraordinarily sharp drop-off in the survival of patients assigned to the control group 2 years after randomization, with Fox noting that of the 12 patients assigned to the control group who were still alive, all died by 1 day after the 4-year anniversary of randomization. Two factors make this pattern seem anomalous. First, it is inconsistent with typical survival curves for people with cancer, which are generally skewed owing to a few people surviving markedly longer than the rest. Second, patients were on average already 2 years past diagnosis at randomization, so this increased rate of death occurred relatively late.

Randomization. Speculation that the apparent efficacy of the intervention stemmed from the shortened survival of control patients gained more precision when Fox (1998) compared the Spiegel et al. (1989) findings with data obtained from the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) Program. Fox estimated that 32% of locale-matched women with metastatic breast cancer would be expected to be alive between 5 and 10 years after diagnosis. Yet Spiegel et al.'s control patients experienced a 4-year survival rate of only 2.8%. In contrast, the 4-year survival of patients randomized to intervention was 24%, substantially closer to the expected value in the absence of an effective intervention and suggesting bias in the initial sampling.

Spiegel, Kraemer, and Bloom (1998) argued that Fox (1998) underestimated the importance of randomization and questioned the expectation that persons with cancer participating in a randomized clinical trial of psychotherapy should be representative of the

more general patient population, noting that both groups survived shorter times relative to norms. Spiegel et al. also criticized Fox for his post hoc isolation of 12 patients to make a case that the apparent effect of the intervention was illusory, noting that investigators similarly isolating a subgroup of patients to argue that an apparently ineffective intervention had actually proven to be effective would be accused of having a confirmatory bias.

Responding, Fox (1999) essentially argued that although randomization provides some check on the influence of confounding factors, randomization is not foolproof. He clarified that he was not assuming that differences between participants and normative data invalidated a clinical trial, only that reference to norms might clarify anomalous results and allow evaluation of whether unmeasured group differences might account for the results. Goodwin, Pritchard, and Spiegel (1999) replied that randomization ensures balance with respect to all relevant factors, given large enough samples, and that comparison to groups outside of the clinical trial is irrelevant to evaluating the efficacy of an intervention, showing "a disregard for the fundamental scientific principles underlying clinical trials" (p. 275). Finally, Fox argued that acceptance of differences in survival as evidence of efficacy assumes that survival curves would have been identical had there been no intervention. In the case of the Spiegel et al. (1989) trial, the shape of the control group survival curve made this assumption less tenable, and comparison to population data provided only additional support for this hypothesis. In this important sense, the reference to the SEER Program was a means of evaluating the internal validity, the success of randomization in controlling extraneous sources of group differences in the trial, not its external validity.

CONSORT. Rated in terms of CONSORT (see Table 1), the Spiegel et al. (1989) trial received a score of 7:29. Strengths included adequate details of the intervention, a complete description of the statistical methods used, detailing of the flow of participants through the study and their baseline characteristics, and an interpretation of the results as they fit in the context of other evidence at the time. Weaknesses included a lack of detail regarding eligibility criteria, randomization scheme, sample size, and timing of analysis determination and an inadequate description of the background and scientific rationale for the investigation.

In summary, the Spiegel et al. (1989) study has received great attention with disproportionately little critical scrutiny. The crux of the controversy about this article hinges on basic differences about interpretation of clinical trials. Namely, how does one interpret unanticipated effects on outcomes that were not specified as primary in modest sized clinical trials? It is noteworthy that Fox and Spiegel seemed to share the view that unanticipated strong effects should be viewed with suspicion. In discussing results of their own trial, Spiegel et al. noted that the effect for the intervention was "consistent with, but greater in magnitude than those of Grossarth-Maticek et al. (1984)" (p. 890). However, like Fox (1991), Spiegel (1991) has rejected the results of the study reported by Grossarth-Maticek et al. as being too strong to be plausible and therefore as irrelevant to evaluating the effects of psychotherapy on the survival of people with cancer.

Regardless of which side one finds more persuasive, attention to the median differences in the survival curves of the intervention and control groups can provide another basis for resolving the significance of the Spiegel et al. (1989) results. Both Fox and investigators involved in the Spiegel et al. study agreed that an attempt at replication was warranted. If one accepts at face value Spiegel et al.'s claim that the intervention yielded nearly a doubling of survival time, then the expectation should be that null findings should be highly unlikely in subsequent clinical trials, if they are adequately conducted (Berry & Stangl, 1996; Brophy & Joseph, 1995). However, all of this becomes moot if we move from the mean to the more appropriate median to evaluate the group differences in this trial and find no significant effect.

Fawzy et al. (1993) and Fawzy et al. (2003)

Fawzy et al. (1993) reported effects on mood, coping strategies, and survival of a 6-week, 90-min, structured group intervention delivered to patients with malignant melanoma shortly after diagnosis and initial surgery. The intervention was a mixture of four components: education about melanoma and health behaviors; stress management; enhancement of coping skills; and psychological support from the group participants and leaders.

Primary endpoints. Survival was not originally identified as an outcome, and there was no provision made for long-term follow-up of patients (Fawzy et al., 1993). However, inspired by Spiegel et al. (1989), Fawzy et al. examined survival at 5–6 years (1993) and 10 years (2003) posttreatment. Fawzy et al. (2003) provided a provocative and seemingly compelling summary of the results for the intervention:

> When controlling for other risk factors, at 5- to 6-year follow-up, participation in the intervention lowered the risk of recurrence by more than 2 1/2 fold (RR = 2.66), and decreased the risk of death approximately 7-fold (RR = 6.89). At the 10-year follow-up, a decrease in risk of recurrence was no longer significant, and the risk of death was 3-fold lower (RR = 2.87) for those who participated in the intervention. (p. 103)

As with the Spiegel et al. (1989) trial, the unanticipated strong effect was based on a small sample (34 per group for survival analyses). However, as survival was not an a priori primary endpoint, the study was not powered to test for survival effects.

Close inspection suggests a number of issues, but before delving into these we should preface our discussion with some basic observations. Despite the way in which the 10-year follow-up results were presented, a log-rank test revealed no significant difference between groups in survival (Fawzy et al., 2003). At the initial follow-up, fewer patients randomized to intervention and retained for analysis had died (3/34) than patients randomized to control (10/34; $p = .03$). The small magnitude of this is highlighted in noting that differences would become nonsignificant with the reclassification of 1 patient (Fox, 1995; Palmer & Coyne, 2004). Despite the manner in which the results were depicted, they may be neither as striking nor as robust as they first appear.

Intention to treat, retention bias, and analytic issues. Fawzy et al.'s (1993, 2003) main analyses selectively excluded patients after randomization, introducing bias. Forty patients were each initially randomized to intervention and control conditions. In the intervention group, 1 patient was excluded owing to death, 1 owing to incomplete baseline data, and a 3rd owing to the presence of major depressive disorder. In the control condition, only 28 patients completed baseline and 6-month assessments. Although lack of complete data was a reason for exclusion from the intervention condition, survival data were included for those in the control

condition regardless of the completeness of their data. Thus, different decision rules were used in retaining patients across conditions. Arguably, the intervention patients selectively excluded from analysis were less likely to show an effect for treatment. Unfortunately, survival data were also unavailable for 3 of the individuals in the control condition. An additional 3 subjects per group were excluded by a later decision to focus only on individuals with Stage I melanoma.

Selective retention of patients was cited by Relman and Angell (2002) as reason for dismissing this study out of hand, with these authors concluding that the study was

> fatally flawed because the analysis is not by the intent-to-treat method, which should be standard epidemiologic practice. The authors did not report the results on all their randomized subjects, which would have been the proper, "intent-to-treat" procedure. The number of exclusions and losses to follow-up after randomization could easily have affected the outcome critically since their groups were relatively small and they report a relatively small number of deaths or recurrences. (pp. 558–559)

Sampson (2002) provided a more detailed critique, noting that at the time, 5-year survival of Stage I melanoma was approximately 92%, whereas the 5-year survival for patients from the control group retained for analysis was only about 72%. Sampson noted that the probability of a representative sample of 34 persons with Stage I melanoma having a 5-year survival rate this low is about .001.

Yet the claim that patients receiving the intervention had a two-and-a-half-fold decrease in likelihood of dying by 5–6 years and a sevenfold decrease by 10 years is impressive. Close examination, however, suggests that these figures reflect inappropriate interpretation of the data. Fawzy et al. (2003) treated the figures as if they represented reduction in the *relative risk* of death associated with the intervention. This involves the common mistake of interpreting the odds ratio in a multivariate logistic regression as if it were a relative risk (Sackett, Deeks, & Altman, 1996). Whereas odds ratios are useful in observational studies, when applied to results of randomized clinical trials, they are likely to overestimate the benefits of offering an intervention in clinical practice (Bracken & Sinclair, 1998; Deeks, 1998; Sinclair & Bracken, 1994).

As well, Fawzy et al. (1993) and Fawzy et al. (2003) used stepwise regression in which the inclusion of treatment group was forced but a range of possible control variables were tested and only significant predictors retained. This method capitalizes on chance and is biased toward finding a treatment effect. Thus, age, sex, Breslow depth, and site of tumor were entered, but only sex and Breslow depth were retained. Moreover, these variables were selected from a larger pool of candidates based on preliminary analyses. Under such conditions, the degrees of freedom are inflated if preselection of covariates is not taken into account (Babyak, 2004). However, the more basic problem may be that the regressions overfit the data (Babyak, 2004): Too many predictor variables were considered relative to the relatively modest number of deaths being explained. For instance, there were 20 deaths in the retained sample at 5–6 years, yielding far below any recommended minimum ratio of 10 to 15 events per covariate (Babyak, 2004; Peduzzi, Concato, Feinstein, & Holford, 1995; Peduzzi, Concato, Kemper, Holford, & Feinstein, 1996). The risk of spurious findings was thus high.

CONSORT. This study reported 5 of 29 CONSORT items. Its strengths included adequate reporting of eligibility, site descriptions, details concerning the intervention itself, description of the statistical methods, and details regarding the recruitment and follow-up period. As can be seen, the details that Fawzy et al. (1993) provided concerning the statistical analyses have been crucial to allowing others to evaluate the authors' claims. Primary weaknesses in reporting relate to a lack of specificity of primary outcomes and a priori hypotheses—which may reflect the post hoc nature of the report, a lack of information regarding methodological decisions, and a generally inadequate discussion of the results in the context of the evidence at the time.

McCorkle et al. (2000)

McCorkle et al. (2000) examined a specialized home nursing care protocol for older, postsurgical cancer patients. Patients were eligible if they were older than 60 years of age, diagnosed with a solid tumor prior to surgical excision, and likely to survive at least 6 months. Of 401 patients identified, 375 were recruited over a period of 35 months. The randomization scheme is unclear, although 190 participants were randomized to intervention and 185 to control.

Intervention consisted of standardized assessments of disease status, application of direct care through management guidelines, patient and family education about cancer, and assisting the participants in obtaining medical services when needed. Intervention nurses provided individualized care and support, consulted with physicians, and were available to participants on a 24-hr basis through a paging system. Intervention was delivered through three home visits and four telephone contacts over a 4-week period. Interventions were recorded and coded for content. Analysis suggested that education, monitoring of physical and emotional status, making referrals and activating community resources, and other activities were much more common (84% of the coded units) than provision of psychological support (16% of the coded units). Control participants received standard postoperative care.

Cointervention confound. The authors distinguish their trial from studies examining psychosocial interventions, stating, "this is the first [trial] to examine the impact of . . . nursing interventions on survival in cancer patients. Other studies have focused on patient's psychosocial status, including depressive symptoms, function, and the effects of support groups" (p. 1708). There was, however, a secondary aim to examine psychosocial and clinical predictors of survival.

Although the intervention consisted of both physical and psychosocial support, the authors identified monitoring of physical status and an offsetting of potentially lethal complications of surgery as key components: "We did what we did really because of the physical care. The deaths were related to major complications, sepsis, pulmonary embolus, etc. The nurses picked these things up and prevented the crisis" (R. McCorkle, personal communication, August 3, 2004). It is thus doubtful whether this intervention should be counted among studies examining the effects of psychotherapy on survival. Spiegel and Giese-Davis (2004) defended its inclusion, noting that education and monitoring of emotional status are key components of psychosocial interventions. Furthermore,

> If anything, McCorkle et al.'s (2000) account of the intervention minimizes attention to patients' physical needs in favor of intervening with patient and family to monitor emotional status and provide support, education, and to connect patients to their communities. They also comment that when they were able to solve physical problems, "this relieved psychological concerns" and that "the combination of psychosocial support with physical care in medically ill patients who are receiving cancer treatment may be essential" (p. 1712). (Spiegel & Giese-Davis, 2004, p. 62)

This argument misses the key point that there was an explicitly medical focus to the intervention. Even if psychosocial issues were addressed, there is strong confounding of this supportive aspect of the intervention with medical cotreatment: Patients in the intervention group got more of both medical and psychosocial care. There is no good reason to dismiss the medical aspects of care emphasized by McCorkle and attribute all effects on patient mortality to the psychosocial component. Thus, the McCorkle et al. (2000) study should be excluded from any box score or meta-analysis of survival effects, unless one is convinced that the medical intervention was immaterial because it was ineffective. One meta-analysis has excluded the McCorkle et al. study, stating, "The result may . . . reflect an effect of combined optimized medical treatment and psychosocial intervention" (Chow et al., 2004, p. 26).

Analytic issues. Analyses appear to have been performed on an intent-to-treat basis, but this is not stated explicitly by the authors. Initial unadjusted survival analyses revealed no significant differences between groups: Randomization to the intervention did not affect survival. However, subgroup analyses stratifying the sample by stage demonstrated a significant survival benefit for persons with later stage cancer in the intervention group. No intervention benefits were found for those with early stage cancer. Notably, although this study is counted as a positive result for psychotherapeutic intervention reducing mortality in Spiegel and Giese-Davis (2003), depressive symptoms did not predict survival in secondary analyses. This would seem to support the hypothesis that any observed improvement should be attributed to a skilled nursing intervention rather than psychotherapy.

It is important to note that survival effects were found only in post hoc analyses of subgroups, favoring late stage but not early stage patients. Although studies in the behavioral medicine literature have often emphasized subgroup analyses when they are positive in the face of negative primary analyses (Antoni et al., 2001; Classen et al., 2001; Schneiderman et al., 2004), this practice is uniformly criticized as inappropriate in the broader clinical trials literature (Pfeffer & Jarcho, 2006; Yusuf, Wittes, Probstfield, & Tyroler, 1991). The consensus is that unplanned subgroup analyses frequently yield spurious results (Assmann, Pocock, Enos, & Kasten, 2000; Senn & Harrell, 1997) and that "only in exceptional circumstances should they affect the conclusions drawn from the trial" (Brooks et al., 2004, p. 229).

CONSORT. With respect to CONSORT ratings, McCorkle et al. (2000) received a score of 10:29. Relative strengths included reporting of very detailed information regarding the intervention itself, the statistical analyses performed, and the methodology and adequate discussion of the generalizability of the results and how they fit in the context of existing research. Weaknesses included not stating specific hypotheses, a lack of clarity regarding the randomization scheme, and insufficient detail with respect to reporting of primary and secondary outcomes.

Kuchler et al. (1999)

In their box scores, Spiegel and Classen (2000) count a study conducted by Kuchler et al. (1999) as a positive finding concerning the effects of psychotherapy on survival. Kuchler et al. randomized 272 patients with a primary diagnosis of gastrointestinal cancer (esophagus, stomach, liver/gallbladder, pancreas, colorectum) to either routine care or inpatient individual psychotherapy, after stratifying by sex. A significant difference in survival was observed between groups after 2 years of follow-up ($p = .002$), with 49% of the intervention participants having died as compared with 67% of the control participants.

Primary endpoints. Kuchler et al. (1999) noted that the original primary endpoint in their study was quality of life, not survival, and sample size requirements were calculated on this basis. As with other studies in which survival was not an a priori endpoint (e.g., Spiegel et al., 1989), it is unclear whether as much weight should be placed on findings for an outcome for which there had not originally been a hypothesis. Because no effect had been hypothesized, the authors would not have had reason to publish a null finding for survival, and so there is a likely confirmatory bias in the availability of this report.

Cointervention confound. Kuchler et al. (1999) described their intervention as a "highly individualized program of psychotherapeutic support provided during the in-hospital period" (p. 323). Therapists provided ongoing emotional and cognitive support to foster "fighting spirit" and to diminish "hope- and helplessness" (p. 324). The investigators noted,

> Emphasis was placed on assisting the patient in forming questions for the other medical and surgical caregivers. The patient's overall well-being was routinely discussed with the surgical team. . . . The therapist was also present during the weekly surgical rounds and once a week at daily nursing rounds. The therapist often alerted other caregivers as to the psychological state of the patient. (pp. 324–325)

Thus, the intervention group seems to have received not only psychotherapy but increased medical monitoring and medical care. Consistent with this assessment, a review of descriptive information provided about the care patients received in the intervention versus control groups reveals some important differences. Although the length of hospital stay was approximately the same in the two groups, the intervention group received almost twice as much intensive care. Posttreatment, patients in the intervention group reported twice as much chemotherapy and three times as much "alternative treatment."

Palmer and Coyne (2004) argued that because psychotherapy was confounded with increased medical treatment, improved survival could not be attributed unambiguously to psychotherapy. Spiegel and Giese-Davis (2004) countered that such coordination of care is typical of psychotherapy with medically ill patients and necessary if psychotherapy is to be integrated with multidisciplinary care. However, it is reasonable to assume that better medical surveillance and more intensive medical care would contribute to longer survival, and certainly this hypothesis has wider empirical support than an attribution of effects on survival to the psychotherapy.

Analytic issues. Randomized assignment was not preserved in the Kuchler et al. (1999) trial. After randomization, 34 patients in the control group requested transfer to the intervention group, and 10 patients in the intervention group requested transfer to the control group. As an intent-to-treat analysis was used, the patients remained in their originally assigned groups for analysis purposes. Owing to the differential crossover, the actual difference associated with receiving the intervention was probably underestimated, although we cannot ascertain from the report whether there was any bias in these transfers.

CONSORT. Kuchler et al. (1999) received one of the higher CONSORT scores (12:29) for their reporting. Strengths included a strong emphasis on reporting of methodological decisions and execution and an adequate discussion of the results. The primary areas of weakness concerned the scientific rationale for the investigation, specification of primary and secondary outcomes, and information regarding the randomization procedure.

J. L. Richardson et al. (1990)

The study by J. L. Richardson et al. (1990) is counted by Spiegel and Giese-Davis (2004) as supporting an effect of psychotherapy on survival. In this study, sequential cohorts of patients with hematologic malignancies were assigned to either routine care or one of three interventions designed to increase adherence with medication taking and appointment keeping: (a) an educational package concerning hematologic malignancies, treatment and side effects, and the patient's responsibility for adherence and self-care, followed by a home visit; (b) a nurse-assisted slide presentation with a hospital-based adherence-shaping procedure; or (c) a combination of interactive slide show, home visit, and adherence shaping. The authors reported that assignment to the intervention condition was related to survival in multivariate analyses controlling for sex, severity of illness, Karnofsky score, number of appointments kept, and compliance with medication.

Randomization and study design. The basic design of the study appears to be quasi-experimental rather than randomized. A sequential cohort design was used in which all individuals entering treatment were assigned to either the control or one of the intervention conditions, whichever happened to be in effect during a given 2–3-month period. The exposure of patients to treatment or control groups in this design can depart considerably from what would occur in a randomized clinical trial. Staff are not blinded, and knowledge of the timing of transitions from intervention to control periods could influence the assignment of particular patients by influencing the timing of admission. As well, the visible withdrawal of special features of a program marking the end of a block of treatment can influence the treatment of the patients in the next period of routine care. Such breakdowns in study protocol can occur at the level of individual patients or for an entire patient cohort. It thus can be particularly difficult to maintain the integrity of complex medical interventions when they are embedded in an open-blind, programwise quasi-experimental design.

There may have been some bias in ascertaining patient death. Patients were considered deceased when contact was lost, and the patients in the control condition may have been more prone to lose contact in the absence of death because staff had never made a home visit.

Primary endpoints. It is not clear that survival was a primary endpoint in the original design of the study. The authors reported that participants were "entered into a control group or one of three different conditions designed to increase compliance" (p. 3576). An earlier report (Levine et al., 1987) made no mention of survival, only adherence. Furthermore, the trial is underpowered for examination of the effects of any one of the intervention packages on survival. The numbers of patients assigned to the control group and each of the three interventions were 25, 22, 23, and 24, respectively.

Analytic issues. Examination of survival curves was limited to a comparison of the control condition to a larger group combining all intervention participants. Such an analysis does not make use of there being three different interventions and is inconsistent with the design, if not simply post hoc. Univariate analyses revealed a survival benefit for assignment to intervention. The investigators then analyzed the effects of 25 other variables on survival, retaining 6 for multivariate analysis that included group assignment, which remained significant ($p < .03$).

The multivariate analysis in which this effect was demonstrated thus capitalized on chance and was overfitted in that the ratio of variables being considered to the number of deaths being explained was excessive (e.g., Babyak, 2004). As well, there are potential problems in assuming that appointment keeping and adherence to one medication are sufficient to eliminate effects of adherence on survival in a complex medical regimen. If these two variables do not account for all variation in pill-taking adherence and medical care, effects of adherence will be assigned to the intervention status variable. There is an illusion of statistical control in the assumption that including these two variables in the multivariate regression eliminates any causal role for differences in adherence in explaining improved survival (Christenfeld, Sloan, Carroll, & Greenland, 2004).

Construct validity of intervention. That group assignment remained significant after controlling for adherence and appointment keeping was taken by the investigators to indicate that the effects of the interventions were independent of adherence. They noted that interventions emphasized monitoring side effects and complications, improving communication with medical personnel, and receiving prompt attention for fever, bleeding, and other medical problems. The investigators acknowledged that improved patient actions in these areas may have increased survival. These activities suggest improvements in broader aspects of medical care that cannot be adequately addressed by the introduction of statistical controls for adherence to appointment keeping and one of many prescribed medications. The authors further speculated, "It is also possible that the programs, by training the patients to be responsible for their own care, allowed them a sense of greater control and resulted in less fear and anxiety" (J. L. Richardson et al., 1990, p. 363). This quotation has been cited as the basis for counting this study as evidence that psychotherapeutic interventions improve survival, independent of effects on adherence (Spiegel & Giese-Davis, 2004). Yet the intervention did not have an explicit focus on reducing fear and anxiety, and a related article from the project reported no changes in depression across the period of the interventions (J. L. Richardson et al., 1987).

We believe that the J. L. Richardson et al. (1990) study provides evidence that persons with cancer can derive benefit from the outreach of home visits and from basic measures to involve family

members, improve education, and encourage pill taking, appointment keeping, and appropriate use of medical services. Richardson stated,

> I would agree that our study was not psychotherapy. Our study was very behavioral in concept and delivery—teaching people how to manage the disease, the treatment and the health care system. I think you can go a long way with basic patient education, family education, and health care system manipulation strategies. (Personal communication, January 3, 2005)

Which, if any, of the various intervention components was decisive cannot be determined. Regardless, there was no explicit psychotherapeutic component, and it is unclear how educational contact with the nurse could be reasonably construed as psychotherapy.

CONSORT. Although we acknowledge that J. L. Richardson et al. (1990) is not a randomized clinical trial, we did perform a CONSORT-based analysis of the reporting. Richardson et al. received a score of 9:29. This score does not reflect adequate reporting in a specific section of the article (e.g., method) so much as adequate reporting of a number of issues throughout. Richardson et al. were the only authors to receive points for adequately reporting the scientific rationale for their investigation. As well, they adequately reported on the content of the interventions, the statistical analytic decisions, and the dates of recruitment and follow-up, and they addressed their findings in the context of the literature. Primary weaknesses included lack of specified primary endpoints, inadequate description of sample size determination, incomplete information concerning randomization protocol, and relatively poor description of statistical analyses.

Linn et al. (1982)

A study conducted by Linn and colleagues (1982) predates the Spiegel et al. (1989) study. The Linn et al. study is counted as a null finding in box scores (Sephton & Spiegel, 2003; Spiegel & Giese-Davis, 2004), but its inclusion raises some basic questions about the wisdom of such box score tallies.

Linn et al. (1982) randomized a mixed cancer-site sample of 120 male patients to individual psychotherapy or routine care. Patients were considered eligible if they presented with clinical Stage IV cancer and were judged by a physician and ward nurse to have more than 3 but less than 12 months to live. The sample was quite heterogeneous in terms of cancer site, but approximately half of the patients had lung cancer. A single counselor provided individual psychotherapy several times weekly, often at bedside. Therapy emphasized reducing denial while preserving hope, completing unfinished business, and taking an active role in treatment decisions, but "above all else, simply listening, understanding, and sometimes only sitting quietly with the patient" (Linn et al., 1982, p. 1048). Extension of life was explicitly rejected as a goal of therapy, and the authors reported considering that therapy that succeeded in providing a sense of life completion might actually shorten survival times. No significant differences in survival between intervention and control subjects were found, either for the sample as a whole or for the larger minority with lung cancer.

Primary endpoints. Improving survival was not a goal of this study. The authors reported that their primary hypothesis concerned psychotherapy improving "the quality but not the length of survival" (Linn et al., 1982, p. 1054) and that this hypothesis was supported. In fact, the authors' hypotheses concerning survival appear to hinge on an implicit mediational model in which psychotherapy improves quality of life, which in turn affects functional status, which then relates to increased survival times. Neither functional status nor survival differed between the groups, however. No differences were found for mean number of days from time of entry into the study to death, or from time of diagnosis to death, for the entire sample or for patients with lung cancer.

Analytic and design issues. A full intent-to-treat analysis was not conducted. Four patients moved or were lost to follow-up and 2 requested to be dropped from study, leaving complete data for 144 patients. One issue that was not adequately addressed concerned the restricted range of variability in survival that was available to be affected by intervention. Participants were selected partly because they were expected to survive between 3 and 12 months, but they were under active medical treatment during the intervention. Given this, the effect of psychotherapy would have to be substantially greater than what would be expected of medical intervention for there to be any noticeable effect on survival.

There seems little basis for considering this study as a test of the ability of psychotherapy to prolong survival. Lengthened survival would have been counter to the expectations of the investigators and is unlikely to have been communicated to the patients as a goal of their treatment. Although investigator allegiances and therapist expectancies might not be sufficient to prolong survival, it seems unreasonable to hypothesize that a psychotherapeutic intervention would promote survival when such allegiances and expectancies are absent or contradictory. Indeed, patients may have derived a sense of permission to die. There were none of the group processes possible that have been cited as important in Spiegel et al. (1989) and in attempted replications. Finally, the sample was heterogeneous, selected for being close to death, so that "advanced intervention [of any kind] has relatively little impact on survival" (Linn et al., 1982, p. 1054). Inclusion of this study in a tally of the effects of psychotherapy on survival seems to demonstrate the futility of undertaking such an overall assessment rather than the completeness with which the relevant studies have been assembled.

CONSORT. Linn et al. (1982) received a CONSORT score of 5:29, adequately reporting eligibility criteria and dates of recruitment and follow-up as well as examining their findings in the context of the existing literature. Primary weaknesses included a lack of rationale for the study, no clearly defined endpoints or description of sample size determination, a lack of specificity concerning the randomization protocol, and inadequate description of statistical analyses.

Ilnyckyj, Farber, Cheang, and Weinerman (1994)

Ilnyckyj et al. (1994) provided a post hoc survival analysis of follow-up data for patients who had participated in a trial 11 years earlier comparing three psychosocial interventions with a control condition. Inclusion criteria included diagnosis with any malignancy, and exclusion criteria included need for psychotherapy or overt evidence of psychosis. One of the intervention groups (n = 31) was led by a social worker and met for 6 months, and another (n = 30) met for 3 months with a social worker and for an additional 6 months without a professional leader. The third inter-

vention group initially enrolled 35 patients and was intended to meet for 6 months without professional leadership. However, this group suffered high attrition, and 21 new, nonrandomized patients assigned to it participated for only 3 months. The control group consisted of 31 patients who did not participate in any group meetings. Of 401 patients referred for the study, 127 consented to participate, but 26 withdrew before randomization. Another 4 patients died, and of these, 2 were too ill to participate before the first group meetings. Few details are provided concerning the structure, process, or conduct of the groups except that the professional leaders "were not instructed in any specific techniques" (Ilnyckyj et al., 1994, p. 93) but used a supportive and educational style to foster open sharing. In survival analyses, all intervention groups were combined and compared with the control condition. No significant differences were found.

Spiegel (2001) and Spiegel and Giese-Davis (2003) included this report as one of the null findings in calculating box scores. They cited its availability as evidence that there is enough interest in whether psychotherapy affects survival that it is not impossible to publish "negative" findings (Spiegel, 2004). The Ilnyckyj et al. (1994) report was prepared by a medical fellow who was not part of the original study team in response to the publication of Spiegel et al.'s (1989) findings (A. Ilnyckyj, personal communication, September 21, 2004). The only previous publication from the project had been a conference abstract more than a decade earlier focusing on null findings for psychological outcomes (Farber, Weinerman, Kuypers, & Behar, 1981). This study, however, raises interesting issues about the relevance of box score calculations that fail to take study quality into account.

Primary endpoints. Survival does not appear to have been an a priori endpoint for the initial investigation. Indeed, the authors stated that the "original intention of the randomized clinical trial was to evaluate the possible psychological benefit of participating in support groups" (Ilnyckyj et al., 1994, p. 93). Thus, the study was not originally powered to find an effect for survival, which may explain the extreme heterogeneity in the sample, and there is little rationale for the 11-year follow-up period.

Randomization. Although the study began as a randomized clinical trial, it did not remain so for long. Randomization broke down with the dropout of many members of the nonprofessionally-led support group and their nonrandom replacement with 21 new members. As well, exposure to treatment varied, as these 21 individuals were exposed to only 3 months of a 6-month protocol.

Analytic issues. Analyses were not performed on an intent-to-treat basis. Although a total of 148 individuals were randomized during the study, data are presented for only 127. As well, although the goal of combining intervention groups may have been to increase power, this post hoc combining of heterogeneous groups likely resulted in increased within-subject error, decreasing the likelihood of finding an effect but also the interpretability of any results.

CONSORT. Ilnyckyj et al. (1994) received a score of 7:29 using CONSORT criteria. It is interesting to note that relative strengths included the description of random assignment in the title and abstract, although a large number of participants were not randomly assigned. This brings up one of the difficulties with the CONSORT criteria, in that it assesses not the accuracy with which authors report pertinent information but simply that a report is made. Other relative strengths were descriptive in nature, concerning flow of participants through the study and reporting of baseline characteristics. Weaknesses centered on the description of scientific rationale for the study, inadequate details concerning the intervention itself and how sample size was determined, lack of information concerning the randomization scheme and statistical analyses, and insufficient discussion of the results.

Edelman, Lemon, Bell, and Kidman (1999)

A randomized clinical trial conducted by Edelman, Lemon, et al. (1999) evaluated group cognitive–behavioral therapy for persons with metastatic breast cancer. A block-randomization procedure was used with 124 patients to allow formation of 10-patient groups, with 10 patients randomized to the routine-care control group in the same block. The intervention was selected on the basis of demonstrated effectiveness in a pilot study (Cocker, Bell, & Kidman, 1994) and consisted of eight weekly sessions of cognitive–behavioral therapy supplemented by a family night and three monthly sessions (Edelman, Bell, & Kidman, 1999). Patients were further provided with a workbook, handouts, homework, and a relaxation tape. Survival analyses conducted 2–5 years after randomization demonstrated no significant effect of group status on survival.

Primary endpoints. It is unclear whether survival was an a priori primary endpoint in Edelman, Lemon, et al. (1999), but it seems unlikely. Psychosocial outcomes appear to have been the primary endpoints, as the authors reported in an earlier article that "improved mood state was a key outcome objective" (Edelman, Bell, & Kidman, 1999, p. 303) and no stratification of the sample based on medical or treatment variables was undertaken (which one might expect if survival were the primary outcome). Results of the psychosocial variables (Edelman, Bell, & Kidman, 1999) suggest an initial improvement on two measures of affect and self-esteem that was not maintained at a 3–6-month follow-up.

Exposure to treatment. A number of logistic problems led to inconsistent exposure to treatment. For the block-randomization scheme to work, 20 participants needed to be accrued at one time prior to initiation of treatment, and slow recruitment meant that some participants had to wait as long as 10 months from accrual to treatment initiation. The authors reported that by that time some participants had died or become too ill to participate, and that although groups were supposed to have 10 members each, some were reduced to 4 or 5 by the end of treatment. The illness burden of the sample was a barrier to participation, and 32 of the 134 participants were classified as "dropouts," with 16 dying before or during intervention, 10 dropping out owing to illness, 3 for "other reasons," and 3 once they were found not to have metastatic disease. Overall, a third of the patients assigned to the intervention group received either no treatment or only partial treatment.

Treatment integrity. The effects of disease and treatment of individual group participants affected not only attendance but the character of the groups themselves. For example, participants in two of the five intervention groups were substantially more ill than those in other groups, with 2 active participants dying during the intervention. These deaths resulted in "emotional challenges that were not experienced by the more 'healthy' groups" (Edelman, Bell, & Kidman, 1999, p. 303). As well, the Hospital Ethics Committee required that control participants be informed of peer

groups in the community, and some availed themselves of these. There were also problems with the family nights; a number had to be cancelled because family members, notably husbands, would not participate. Although these difficulties threaten the integrity of the evaluation of the intervention, they undoubtedly are inherent in clinical trials requiring repeated group sessions with patients with advanced cancer. Perhaps what is different about Edelman, Lemon, et al. is their frankness about having confronted these problems.

Analytic issues. Survival analyses utilized follow-up data obtained 2–5 years after enrollment and were conducted in an intent-to-treat fashion for all patients after the exclusion of the 3 who had been found not to have metastases. Thirty percent of the patients were alive at the end of the observation period. There was no evidence of the sudden drop-off in survival at 20 months postrandomization observed in the Spiegel et al. (1989) study. Primary analyses involved stepwise regression with group assignment and seven medical variables that have been shown in past research to predict survival. Although there was a trend for the control patients to have longer survival, group assignment was not retained as significant in the final equation. No group differences were observed in time from randomization to death or time of diagnosis of metastasis to death. Because performance status and date of first chemotherapy were predictive of survival, analyses were repeated with inclusion of these variables as covariates, but there was again no significant effect for group assignment. Forcing entry of group assignment into these stepwise multivariate regressions did not affect results. Finally, analyses taking into account participation in outside peer support groups still yielded no effect for group assignment. Overall, the follow-up period for ascertaining effects on survival was shorter than in some of the other studies, the size of groups was relatively small, and the multivariate regression was overfitted and capitalized, with too many variables being considered. Yet inspection of the survival curves gives little hint that a benefit for survival is being missed.

CONSORT. Edelman, Lemon, et al. (1999) received a score of 5:29 on the overall CONSORT checklist. Relative strengths included reporting of dates for recruitment and follow-up, providing adequate baseline characteristics, demonstrating an intent-to-treat analysis, and providing an interpretation of results and a statement of generalizability. Weaknesses included insufficient discussion of study rationale, lack of descriptions of treatment settings and administration of interventions, inadequate details of the randomization protocol, and absence of a statement of whether the primary outcome analysis was performed on an intent-to-treat basis.

Cunningham et al. (1998)

Cunningham et al. (1998) reported on the outcome of a randomized clinical trial of professionally led supportive–expressive and cognitive–behavioral psychotherapy compared with a home-study cognitive–behavioral package. The supportive–expressive component was based on the Spiegel et al. (1989) intervention and incorporated mutual support, encouragement to process emotion, and confronting the likelihood of death. The cognitive–behavioral component consisted of standard cognitive–behavioral homework assignments provided in workbook format. Patients were considered eligible if they were female, had a confirmed diagnosis of metastatic breast cancer with no known brain metastases, were fluent in English, and were under age 70. A total of 66 patients were randomized, and survival was assessed 5 years after the start of the study. Patients in both conditions received information and pamphlets on coping with cancer from the Canadian Cancer Society. The home-study control subjects also received standard care at the hospital, the cognitive–behavioral workbook, and two audiotapes. No significant difference in survival was found for the primary test examining survival at 5 years from randomization, a secondary analysis comparing survival curves from time of first metastasis, or a tertiary test examining survival from initial diagnosis to death.

Primary endpoints and sample size. Cunningham et al. (1998) is in the minority of studies for which survival was an a priori primary endpoint. Given this fact, it is odd that their study appears to have been underpowered and that the authors did not provide an explanation of how their modest sample size was determined. A post hoc power analysis suggests that 250 participants, rather than 66, would be needed to have .80 power to detect the small effect size found. Goodman and Berlin (1994) cautioned against attaching too much importance to such post hoc analyses, noting that power calculations based on null findings will always yield a larger required sample size than was available for the completed trial, and that assumptions about a similar effect size in the larger replication may not hold true. The Cunningham et al. (1998) sample size is consistent with earlier studies, approximating Spiegel et al.'s (1989) 36 patients in the control condition, Fawzy et al.'s (1993) 34 patients in the intervention condition, and J. L. Richardson et al.'s (1990) 25 patients in the control condition. Indeed, because all of the patients in the Cunningham et al. study received exposure to treatment, the effective sample size in that study was larger than for the Spiegel et al. study.

Given the limited previous literature, it is difficult to determine what would be a reasonable expectation for effect size and, therefore, sample size. However, if one views this study as an attempted replication of the large effects (i.e., a twice as long survival time for patients receiving the intervention) claimed by Spiegel et al. (1989), as the authors suggested, the sample is modest but not exceptionally small in comparison to any of these earlier studies except Kuchler et al. (1999).

Adequacy of intervention. Kraemer and Spiegel (1999) argued that substantive differences exist between the Cunningham et al. (1998) intervention and what was delivered in the original Spiegel et al. (1989) study and that these differences may play a role in negative findings. For example, it is possible that the attention paid to cognitive–behavioral homework may have interfered with emotional work, that the 35 weeks of intervention may have been insufficient in either intensity or duration, and that the active control condition may have provided too much intervention, thus diminishing effect sizes.

In the context of other trials, these criticisms appear to hold Cunningham et al. (1998) to unduly strict standards. The intervention combined elements of both Spiegel et al. (1989) and Fawzy et al. (1993), and the median number of attended sessions may have exceeded the median received by patients in the first year of Spiegel et al. owing to deaths in that study. There is currently no evidence that access to a cognitive–behavioral workbook prolongs survival. Thus, the control condition, though "active," is likely to have its putative survival effects attenuated and have only a minimal effect on effect sizes.

CONSORT. Cunningham et al. (1998) received a CONSORT score of 13:29. Of note, this is the one study in which the results were adequately discussed. Thus, the study receives all 3 points for the discussion section. Relative weaknesses, in this case, centered on the lack of specific objectives and hypotheses, clearly defined outcome measures, determination of sample size, description of the flow of participants through the trial, and reporting of effect sizes, multiplicity, and adverse events.

Goodwin et al. (2001)

Goodwin et al. (2001) attempted a replication of the Spiegel et al. (1989) findings, randomly assigning 235 women with metastatic breast cancer to weekly supportive–expressive therapy (n = 158) or a control group that received no support group intervention (n = 77). All participants received educational materials. The psychological intervention did not prolong survival; median survival in the intervention group was reported as 17.9 months, as compared with 17.6 months in the control group. Multivariate analyses incorporating the presence or absence of progesterone receptors and estrogen receptors, time from first metastasis to randomization, age at diagnosis, nodal stage at diagnosis, and use or nonuse of adjuvant chemotherapy identified no significant effect of the intervention on survival and no significant interactions with treatment and study center, marital status, or baseline total mood disturbance score.

Primary endpoint and sample size. Survival was the a priori primary endpoint in this trial and was used as the outcome variable in determining sample size. Power calculations were based on .85 power to identify 3-year survival of 15% in the control group compared with 30% in the intervention group with a Type I error rate of .05. As well, subsequent analysis suggested that the study maintained power of .99 to detect the 25% increase in 3-year survival reported by Spiegel et al. (1989).

Cointervention confound and treatment integrity. The authors stated that although the control group participants did not receive psychotherapy as part of the study, they were allowed to participate in peer support groups and therapist-led support groups that did not include supportive–expressive components, and they could receive necessary psychological care. The authors reported that 10.4% of those in the control condition availed themselves of outside intervention. Thus, it is possible that at least some women in the control condition exposed themselves to treatments similar in nature to supportive psychotherapy. As well, participants in the intervention condition were encouraged to interact and provide support outside of group sessions and to contact physicians for needed medical intervention, such as pain management. Thus, intervention group participants may have received increased medical monitoring and even medical care relative to controls.

The intervention group likely received both an adequate "dose" of psychotherapy and an intervention that was very similar to that performed by Spiegel et al. (1989). Ninety-five percent of women assigned to the intervention condition attended at least one session, and 81% remained involved throughout the first year. Interventionists were trained by Dr. Spiegel, receiving standardized training with the supportive–expressive therapy manual created by Spiegel and Spira (1991), including attending 2-day workshops conducted by the training team every 9–12 months, which included discussion of principles, videotape review, and feedback.

Analytic issues. Intent-to-treat analyses were performed to preserve the randomization, and interim analyses were neither planned nor performed, safeguarding against inflated familywise Type I error rates. The authors reported no substantial variations from recommended analytic procedures.

CONSORT. Goodwin et al. (2001) received a score of 14:29 using the CONSORT criteria. Throughout, the report provides adequate detail concerning intervention components and analytic decisions. It lost points primarily through deficits in the title and introduction; a lack of reporting about the allocation sequence, how it was implemented, and blinding; and inadequate discussion of the findings.

Kissane et al. (2004)

The Kissane et al. (2004) study is the latest to evaluate the hypothesis that psychological therapy can influence the survival of people with cancer. In this clinical trial, 303 women with early stage breast cancer receiving adjuvant chemotherapy were randomly assigned to either 20 sessions of weekly group therapy (cognitive–existential group therapy) plus three relaxation classes (n = 154) or a control condition of three relaxation classes (n = 149). The intervention did not extend survival, with median survival of 81.9 months in the intervention arm and 85.5 months in the control arm. The hazard ratio for death in the intervention arm versus control was 1.35 (95% confidence interval [CI] = 0.76–2.39, p = .31), with a multivariate Cox model identifying no significant effect of intervention on survival (hazard ratio = 1.37; 95% CI = 0.73–2.32, p = .37). Two medical variables were significantly associated with survival: favorable histology (Grade 1 or 2) and negative axillary node status.

Primary endpoints and sample size. Survival was the a priori primary endpoint of this trial and the variable on which decisions for sample size were based. The sample size was based on .80 power to detect a 15% difference between groups over a 5-year period with a Type I error rate of .05.

Study rationale. The rationale for the choice of women with early stage breast cancer is not clear. Kissane et al. (2004) noted that studies examining the psychological intervention–survival link have yielded "mixed results" and then stated that "a prospective trial of the impact of group therapy at a much earlier stage of breast cancer seems warranted" (p. 4255). However, the reasoning that links mixed results to the need to examine participants with earlier stage cancer is not obvious. In particular, why it should be expected that a psychosocial intervention could produce an effect on the survival of a population already expected to have excellent prospects for long-term survival is never addressed.

Cointervention confound and treatment integrity. The intervention was manualized, and therapist training was specified. As well, supervisors assessed treatment fidelity through the use of thematic checklists, although no audio- or videotapes were available and adherence to relaxation at home was not reported. As in the Goodwin et al. (2001) study, women in the intervention group were encouraged to meet informally outside of sessions. It is not clear whether this encouragement occurred in the control group. The degree to which intervention participants were encouraged to contact their physicians regarding additional medical care needs (e.g., pain, side effects of treatment) is not clear, although one intervention theme involved patient–physician interactions. Fi-

nally, exposure to treatment is unclear, although the authors reported that 12% of the sample failed to complete 6 of the 20 prescribed sessions and 94% received at least some exposure.

CONSORT. This study received a score of 13:29 using the CONSORT reporting criteria. This was the only study to receive points for describing results fully with the use of effect size statistics. Overall strengths included descriptions of the eligibility criteria, settings, and interventions; an adequate description of randomization and statistical analyses; and a very strong results section. Of interest, this study received no points relating to its discussion of results in the context of the existing data.

Summarizing Studies: Do Box Scores or Meta-Analyses Overcome the "Apples and Oranges" Problem?

The studies that are now the primary sources for evaluating whether psychotherapy improves survival in cancer patients have been termed "apples and oranges" (Smedslund & Ringdal, 2004, p. 123; Spiegel, 2004, p. 133). Even this analogy, however, fails to fully capture the range of differences among these studies and the methodological shortcomings from which they suffer. Kraemer, Gardner, Brooks, and Yesavage (1998) cautioned against optimism that combining flawed studies, particularly small studies (of 20–100 patients), can inform the literature, noting that such underpowered studies are likely to be at increased risk of producing false-positive results and thus more likely to be the source of inflated estimates of treatment effects when their end results are statistically significant.

Heterogeneity of Studies

A notable difference among the studies we have reviewed concerns initial design and whether survival was an a priori primary endpoint. Neither the original Spiegel et al. (1989) study nor the Fawzy et al. (1993) study was designed to evaluate the effect of psychotherapy on survival. Not until the Spiegel et al. study provided impetus for publishing survival data did the Ilnyckyj et al. (1994) report appear, and neither it nor the J. L. Richardson et al. (1990) study was designed with survival as a primary outcome; furthermore, in both of these studies, evaluation of effect depended on combining what were originally different interventions that were presumably intended to be compared with one another. Other studies (Cunningham et al., 1998; Edelman, Bell, & Kidman, 1999; Goodwin et al., 2001; Kissane et al., 2004) were designed as tests of the effects of psychotherapy on survival with survival as the primary outcome and as such ought to be given greater consideration.

The investigators in the J. L. Richardson et al. (1990) and McCorkle et al. (2000) studies deny that their interventions were conceived as psychotherapy, and, as with Kuchler et al. (1999), confounding of group assignment with medical care precludes examination of the independent effect of supportive aspects of the intervention on survival. It is difficult to compare these studies with studies in which there is no obvious medical cointervention confound (Cunningham et al., 1998; Edelman, Lemon, et al., 1999; Goodwin et al., 2001; Kissane et al., 2004).

Among those studies that examined psychotherapy, two consisted of individual therapies (Kuchler et al., 1999; Linn et al., 1982), whereas the others were group therapies. The group therapies included cognitive–behavioral therapy (Edelman, Lemon, et al., 1999; Fawzy et al., 1993), supportive–expressive therapy (Goodwin et al., 2001; Spiegel et al., 1989), integrative variants (Cunningham et al., 1998; Kissane et al., 2004), and supportive–educational approaches (Ilnyckyj et al., 1994).

A number of these studies, including the most positive (Fawzy et al., 1993; Spiegel et al., 1989), had modest sample sizes that were not determined by formal power analysis. In contrast, the Goodwin et al. (2001) and Kissane et al. (2004) studies were based on formal power analysis with survival as the endpoint. As we have noted, unexpected strong findings in a modest sized study should be greeted with suspicion. On the basis of the criteria of having an a priori hypothesis and formal power analysis, the Goodwin et al. and Kissane et al. studies should carry greater weight than the others.

Among the studies reviewed, different patient populations with different life expectancies were recruited, affecting the likelihood of an effect on survival being demonstrated. Studies of more ill populations already receiving adequate medical care may require an effect for psychotherapy that is greater than can be expected of additional medical interventions, whereas studies of less ill populations may have many fewer deaths to explain. Although many of the studies examined breast cancer (Cunningham et al., 1998; Edelman, Bell, & Kidman, 1999; Goodwin et al., 2001; Kissane et al., 2004; Spiegel et al., 1989), others examined melanoma (Fawzy et al., 1993), gastrointestinal tumors (Kuchler et al., 1999), hematologic cancers (J. L. Richardson et al., 1990), and mixed-site cancers (Ilnyckyj et al., 1994; Linn et al., 1982; McCorkle et al., 2000). As well, some sampled from early stage disease populations (Fawzy et al., 1993; Kissane et al., 2004), whereas others examined later stages (Cunningham et al., 1998; Edelman et al., 1999; Goodwin et al., 2001; Spiegel et al., 1989). Participants were recruited with the expectation that they would travel to weekly therapy sessions for at least a year (Goodwin et al., 2001; Spiegel et al., 1989) or because they were not expected to live a year (Linn et al., 1982).

Stopping rules for survival assessment differed among the studies, and end times were sometimes chosen after data were available for inspection, increasing the likelihood of Type I error, particularly when multiple unplanned analyses were carried out for varying time points. Spiegel et al. (1989) and the later follow-up by Fawzy et al. (2003) covered 10 years, and Ilnyckyj et al. (1994) covered 11 years. However, a number of other studies had 1- or 2-year follow-up periods (Kuchler et al., 1999; Linn et al., 1982; J. L. Richardson et al., 1990), a time frame within which the survival curves for Spiegel et al. were "almost superimposable" (Fox, 1998, p. 361).

Thus, the number of potentially crucial ways in which these studies differ approaches the number of studies available. Reliable answers to the primary question of "Does psychotherapy affect survival?" are unlikely to be gleaned from this group of studies, and more nuanced questions, such as "Is supportive–expressive therapy more effective than cognitive therapy?" and "Are effects more likely to be observed with earlier stage rather advanced stage patients?" are barred by confounding of these strata with other important differences among trials. What does seem to be consistent in this literature, however, is that those studies with superior methodology (Goodwin et al., 2001; Kissane et al., 2004) are more likely to produce null findings.

Does CONSORT Facilitate Evaluation of the Relative Merits of These Studies?

We are providing one of the first applications of the CONSORT criteria to the evaluation of already published trials of psychosocial interventions. How useful was this tool? We saw that overall, reporting of these trials met a minority of CONSORT criteria, on average only about a third, and that no trial met any of a number of important criteria. This could be seen as providing an important framing of our whole review. Transparency of reporting was important in facilitating evaluation of some trials. In the case of Fawzy et al. (1993), an acknowledged departure from intent-to-treat analyses suggested a fatal flaw (Relman & Angell, 2002) in the counting of this trial as evidence that psychotherapy promotes survival. Closer scrutiny provided further doubts that appropriate analyses would have yielded a significant effect on survival. Yet transparency in the reporting of what may have been a fatal flaw increased the CONSORT score for this study, thus highlighting the limitations of CONSORT as a direct indicator of trial quality.

Later trials with survival as an a priori endpoint received somewhat higher CONSORT ratings (Cunningham et al., 1998; Goodwin et al., 2001; Kissane et al., 2004). However, differences among the 11 studies were small, with only a minority of CONSORT items being endorsed for any of this collection of studies, and the substantive importance of such differences is unclear. Recall that noncompliance with some items has little or no implication for study quality; some are a matter of transparency of reporting and allowing adequate search terms whereas others have profound implications for quality. Yet all items are counted equally. Moreover, some of the most decisive factors in evaluating the trials that have been cited as evidence for an effect of psychotherapy on survival do not figure in CONSORT ratings. These include the use of mean rather than median survival time and the odd outcomes for the control group in Spiegel et al. (1989); the use of different rules for excluding intervention versus control patients and the inappropriate statistical analyses in Fawzy et al. (1993); and the definite confounding of psychosocial treatment with enhanced medical monitoring and care in J. L. Richardson et al. (1990), McCorkle et al. (2000), and Kuchler et al. (1999).

Would another rating scale have been more helpful? Over a decade ago, Moher et al. (1995) identified 25 different rating schemes for the quality of clinical trials, and undoubtedly, more have accumulated since. Although many of these schemes can be applied with adequate interrater reliability, they produce markedly inconsistent evaluations across studies because of differences in the criteria they invoke and the weight they attach to particular criteria (Juni, Witshi, Bloch, & Egger, 1999; Moher et al., 1998). There is a lack of rationale for emphasizing these particular criteria or weights or for choosing among competing schemes (Detsky, Naylor, Orourke, McGeer, & Labbe, 1992).

It is probably better to use explicit standards for deciding whether trials should be entered into consideration as acceptable evidence. Newell, Sanson-Fisher, and Savolainen (2002) rated 129 studies evaluating psychosocial interventions for persons diagnosed with cancer on 10 internal validity criteria, each rated on a 0–3 scale (0 = *not at all fulfilled*, 3 = *entirely fulfilled*). Requiring a minimum score of 11 excluded most (87, or 64%) trials from consideration. Although this effort has been criticized as too strict (Bredart, Cayrou, & Dolbeault, 2002), it still allowed consideration of many studies with serious methodological problems, including small cell size (Coyne & Lepore, 2006). Had this scheme been used, some of the most crucial features in our evaluation of trials relevant to the question of psychotherapy promoting survival would have been missed.

Our sense is that CONSORT was useful in characterizing the trials as a group and that the transparency that would result from compliance with CONSORT being a requirement for publishing results of trials would raise the quality of trials and the interpretability of their results. Yet, confronted with the heterogeneity we found in the studies we reviewed, we believe there is no substitute for a close read and careful application of a diverse range of critical appraisal skills.

An Appraisal of Box Scores as Summaries

Spiegel and colleagues (Sephton & Spiegel, 2003; Spiegel & Giese-Davis, 2004) used a box score approach to summarizing the first 10 studies relevant to the question of whether psychotherapy promotes survival. Results indicated that 5 studies demonstrated an effect and 5 did not. This tie was interpreted as an indication that the question was not settled. That there were any positive studies at all was deemed noteworthy and encouraging because of the improbability that psychotherapy could affect survival; the lack of studies demonstrating that psychotherapy had a deleterious effect on survival was also considered noteworthy (Spiegel, 2004).

Proponents of meta-analysis have long noted disadvantages to box score summaries (Cooper, 1989; Cooper & Hedges, 1994). Box scores give equal weight to all studies, regardless of size or quality; attach too much importance to significance levels that may partly reflect sample size; and fail to provide an estimate of effect size. Yet even more basic issues are left unaddressed by box scores. For example, to whom and across which interventions should box score summaries generalize? In the studies considered by Spiegel, the heterogeneity of patient populations and small number of studies argue against generalizing across cancer sites. Cointervention confounds in which psychotherapeutic intervention varies with quality and intensity of medical monitoring and care make it difficult to attribute outcomes to any specific therapeutic component. Moreover, the rejection by some investigators that their intervention constituted psychotherapy may be sufficient reason to exclude some studies that would have counted as positive scores.

There is also the concern that this set of studies may be both over- and underinclusive. That is, there are concerns about both the numerator and the denominator in "5 of 10 studies." The numerator depends on 2 key studies that may represent false positives given post hoc follow-up of a small number of patients and unexpected large effects and 3 studies for which cointervention confounds are likely. The validity of the denominator is dependent on capturing all relevant studies. If one accepts any unplanned retrospective analyses of survival as relevant, then there are potentially hundreds of psychotherapeutic, psychosocial, and nursing interventions that might be analyzed and included. Undoubtedly the investigators in the bulk of these studies did not collect survival data because they did not believe that a survival effect was likely. However, the investigators in most of the 10 studies included also did not initially contemplate a survival effect. In short, we do not have a good a priori reason for assuming that most psychosocial

intervention trials have an effect on survival, and certainly not 50% of them. It is therefore not clear what substance should be attached to the 10 in "5 of 10," particularly in light of the way in which these 10 studies were isolated from the larger pool of studies.

The Ilnyckyj et al. (1994) study provides a useful example of this problem. Given the difficulties publishing null findings and problems inherent in study design and implementation, it seems unlikely that the Ilnyckyj et al. study would have been published without the impetus provided by Spiegel et al. (1989). The initial report (Farber et al., 1981) found no significant effect of group assignment on psychosocial outcome variables, and there were major breakdowns in the implementation of the study. Furthermore, the report would have been difficult to locate before its citation by Spiegel (2001) and Spiegel and Giese-Davis (2003), as it was published in a journal that was not indexed in MEDLINE or the Institute for Scientific Information Web of Science. It is unlikely that this report could have been located had it not been cited by Spiegel (2001), leaving one to wonder how similarly nonindexed null findings are extant and providing little reassurance that all relevant findings have been retrieved for box scores and meta-analyses. Undoubtedly, there is a large but unknown number of studies targeting psychological outcomes whose flaws in design or execution or null findings for primary outcomes would discourage investigators from preparing manuscripts based on them or journal editors from accepting them.

What has been termed the "file drawer problem" (Rosenthal, 1979) represents the threat posed by potentially relevant but unpublished studies to the validity of summaries that rely on published results. The solution of estimating the number of studies with null findings that would be sufficient to revise a conclusion and the likelihood that these studies remain in desk drawers is problematic, however, in the context of small sample sizes and retrospective findings of unexpected effects. Although small sample size poses the threat that studies will lack statistical power, it also poses the threat of positive publication bias when there is an unexpected finding. Simon (1994) suggested that under the assumption that only 10% of trials are effective, with a Type I error rate of .05 and power of .80, over a third of claims of effectiveness are false. This proportion increases substantially in smaller trials and when there is no a priori expectation of effectiveness (Spiegelhalter, 2004). If a study is underpowered and does not yield an effect, particularly for an endpoint that was not specifically targeted, results are more likely to remain unpublished than if an unexpected positive finding is obtained for that outcome. Thus, weight is given to Kraemer et al.'s (1998) argument that when dealing with underpowered trials, we must guard against including false positives by excluding small trials, while at the same time being mindful of unpublished trials with null findings.

The adequacy of box scores as a meaningful way of summarizing the effects of psychotherapy on survival is thus questionable. Acceptance of the numerator in "5 of 10 studies" requires treating disparate studies as equivalent and ignoring the likelihood of false positives. There is no adequate way of evaluating the denominator, but it is potentially much greater than 10. In evaluating the box score, we have assumed that small-scale studies are particularly unreliable and likely to yield false positives. Finally, the retrospective identification of survival as an outcome and of interventions as psychotherapy poses additional serious problems for this enterprise.

Meta-Analysis As an Alternative to Box Scores

Three relevant meta-analyses have recently appeared. Chow et al. (2004) searched peer-reviewed journals from 1966 to 2002 for randomized clinical trials that involved psychosocial interventions for adults with cancer, specifically studies for which survival curves or tabular data were available and in which all participants received the same medical care. Chow et al. identified eight trials with 1-year data, and of these, six had 4-year data as well. Chow et al. concluded that there was no effect on 1- or 4-year survival either for the entire group of studies or for those examining group therapy specifically for women with breast cancer. They qualified their conclusion by noting that there were a small number of available trials, each with a small number of patients; that follow-up periods were relatively short; and that analyses depended on estimated event rates and end-of-trial event rates rather than actual deaths. "Moreover, the diversity of the psychosocial interventions and the lack of long-term follow-up data challenge the validity of our conclusion" (Chow et al., 2004, p. 30).

Smedslund and Ringdal (2004) identified 13 articles from 1989 to 2003, which together reported a total of 14 studies. Studies selected included nonrandomized clinical trials but excluded Linn et al. (1982) because it did not report the data necessary for calculating a log hazard ratio. Smedslund and Ringdal found no overall effect of group intervention on survival. However, they found a large effect for individual interventions, based on results from McCorkle et al. (2000), J. L. Richardson et al. (1990), and Kuchler et al. (1999), ignoring the confounding of medical care with psychosocial intervention in these studies.

Edwards et al. (2004) limited their search to randomized clinical trials of women with metastatic breast cancer. They identified five trials with available survival data, all of them involving group therapy, and noted that they had to accept analyses that did not use an intent-to-treat method. Edwards et al. concluded that there was no clear evidence for a benefit of group therapy for survival but that studies of cognitive therapy showed some benefits for survival in the control group at 1 year, whereas the reverse was true for supportive–expressive therapy. They cautioned, however, that this finding might be due to the anomalous results of Spiegel et al. (1989). Consistent with Chow et al. (2004), Edwards et al. noted that they could not rule out deleterious effects for some patients. They also expressed misgivings concerning the heterogeneity of even this subset of the trials, which have been identified as relevant to the question of whether psychotherapy promotes survival.

Taken together, these meta-analyses appear to give some precision to the judgment of a lack of evidence for an effect of psychotherapy on survival. Yet considerable compromises were involved in arriving at this conclusion, ranging from equating as-treated and intent-to-treat analyses, accepting investigators' choice of length of follow-up and point at which reported statistical tests were performed, and, in the case of Smedslund and Ringdal (2004), ignoring what we believe to be serious confounds. Rather than lending precision to an evaluation of the effects of psychotherapy on survival, these meta-analyses may represent application of this method of summarizing available data to a small literature that is too limited in quality and too heterogeneous to warrant such an effort. In short, meta-analysis may not be an appropriate tool for summarizing and evaluating the studies that have been identified as relevant to the question of whether psy-

chotherapy promotes survival in cancer patients, given the nature of the available evidence. We note that after a careful, comprehensive review of the available studies of psychosocial interventions for persons with cancer, Newell and colleagues (2002) came to a similar conclusion as to their appropriateness for meta-analysis.

Putative Mechanisms by Which Psychotherapy Could Affect Survival

Establishment of a plausible mechanism by which psychotherapy could promote survival is important for a number of reasons. Identification of a plausible mechanism is relevant to any reappraisal of an apparent effect on survival that Spiegel (2004) has termed as "inherently improbable" (p. 133) and an evaluation of the appropriate size of effect that has been sought when sample size has been determined with a formal power analysis. An identified mechanism by which psychotherapy could influence survival would take a positive study out of the realm of the improbable and should give some suggestion as to how strong of an effect could be expected and, therefore, the requisite sample size needed to reliably detect such an effect if it were present. A candidate mechanism might also encourage a persistent search for such an effect in the face of a pattern of weak or null findings. If there is a credible mechanism by which psychotherapy should influence survival, then perhaps disappointing results might reflect the relevant mechanism being missed or too weakly influenced. The adequacy of a test of whether psychotherapy affected survival would be determined by whether the intervention had the requisite effect on the mediator, the presumed mechanism of action. Spiegel et al. (1989) framed their original survival analysis as a test of whether having "the right mental attitude" (p. 890) could affect longevity, with the expectation that it would not. However, when analyses seemed to indicate prolonged survival, a range of putative mechanisms were posited.

One set of mechanisms related to improved adherence and health-related behaviors. Participants might have been activated to adhere more fully and keep appointments, improve their nutrition as a result of improved mood, or maintain health behaviors because of better pain control. Two of the studies identified in support of an effect of psychotherapy on survival (McCorkle et al., 2000; J. L. Richardson et al., 1990) have been construed by the investigator groups as primarily addressing adherence and access to medical care, and another (Kuchler et al., 1999) involved contact with medical staff that resulted in intensive medical care. Of the remaining trials identified as yielding a positive effect, neither Spiegel et al. (1989) nor Fawzy et al. (1993) provided any evidence of improved adherence.

Kissane et al. (2004) noted that there are not sufficient problems in the adherence to chemotherapy in metastatic breast cancer to warrant improved adherence as a goal for a broadly offered psychosocial intervention. Furthermore, if Spiegel et al. (1989) and Fawzy et al. (1993) had started with an express interest in improving adherence, many of the distinctive elements of the interventions in these two trials would not have been included, and indeed, much of the content of these interventions would be seen as superfluous.

A second set of putative mechanisms involve indirect effects of psychological benefits on neuroendocrine and immune function. Here, too, are post hoc speculation and few directly relevant data. Fawzy, Kemeny, et al. (1990) collected measures of immune function related to natural killer cells and T-lymphocyte activity. Although a 6-week follow-up revealed few differences between the intervention and control groups, some differences emerged by 6 months. Fawzy, Kemeny, et al. noted that neither the mechanisms by which the intervention might have affected the immune system nor the health consequences, if any, of these differences were known. When 6-year survival data became available, no relation was found between changes in immune function and recurrence or survival (Fawzy et al., 1993). Subsequent studies have consistently failed to find effects of psychosocial interventions on the immune functioning of persons with cancer (Andersen et al., 2004; Elsesser, van Berkel, Sartory, Biermanngocke, & Ohl, 1994; Hosaka, Tokuda, Sugiyama, Hirai, & Okuyama, 2000; Larson, Duberstein, Talbot, Caldwell, & Moynihan, 2000; M. A. Richardson et al., 1997; Van der Pompe, Duivenoorden, Antoni, Visser, & Heijnen, 1997).

Are Changes in Distress Necessary for Improved Survival?

Most of the proposed explanatory mechanisms for a role of psychotherapy in prolonging survival presume that interventions improve psychological functioning. Indeed, Spiegel (2004) argued that "it is hard to imagine that an intervention which does not benefit patients psychologically will extend survival time" (p. 254; see also Andersen et al., 2004). If a psychological intervention fails to have anticipated psychological effects, how can it be presumed to influence survival? Psychological effects have typically been defined in terms of mood or psychological distress. However, unambiguous demonstration of effects on mood is difficult when the patients under study are very ill and at risk of dying, and the types and effects of biases in available data may be different for intervention and control patients. Substantial missing data owing to death or illness preclude conventional intent-to-treat analyses, and the subgroup of patients for whom all or most data are available is likely to be biased. Thus, Spiegel et al. (1981) and Goodwin et al. (2004) obtained complete assessments from only 52% and 62% of participants, respectively, and Fawzy et al. (1993) collected psychological functioning data for a greater proportion of intervention than control patients.

Data are likely to be missing for different reasons in intervention and control patients. Completing mailed assessments rather than having to attend therapeutic meetings may lower the threshold for continued participation by ill control patients. On the other hand, intervention patients may be more motivated than control patients to continue to provide data despite being ill, as they perceive some benefit to their participation. Between-group differences in reasons for missing data may relate to biases in the data available for analysis (Bordeleau et al., 2003; Ross, Thomsen, Boesen, & Johansen, 2004).

The decline in health, functioning, and overall comfort level and quality of life seen in patients with advanced disease may render any psychological benefits of treatment temporary (Edelman, Bell, & Kidman, 1999). Patients' psychological well-being tends to be substantially lower as they approach death, with no differential effects associated with intervention or control group status (Butler et al., 2003; Ross et al., 2004). Spiegel and colleagues (Butler et

al., 2003) argued that such a decline in mood masks the true benefits of group participation and that an adjustment should be made in order to avoid a Type II error. In the first report of Spiegel's replication study, null findings in primary analyses were followed up with secondary analyses in which assessments were eliminated for patients who subsequently died within a year of the assessment (Classen et al., 2001). Such censoring of the data resulted in a steeper decline in negative mood for women in the intervention condition but a reversed slope for women in the control condition. Apparently, more negative mood scores were removed in the intervention condition than in the control condition (Ross et al., 2004). The difficulty obtaining complete psychological data from very ill persons with cancer, who typically experience increasing pain, fatigue, and other forms of distress as death approaches—thus yielding a "spike" in mood data (Butler et al., 2003, p. 416)—is more than a methodological and statistical issue. It represents barriers to the making of substantive, positive statements about the benefits of psychotherapeutic interventions with such populations. Basically, use of censored mood data shifts the question from "Does therapy benefit the mood of women with metastatic breast cancer?" to the very different question of "Does therapy benefit the mood of the subgroup of patients who in hindsight were not actively dying at the time their mood was assessed?" It would be misleading to accept the answer to the second question as a satisfactory answer to the first.

An additional barrier to demonstrating that these interventions affect psychological functioning is that these trials tend to attract patients who are not highly distressed and for whom it therefore may be difficult to demonstrate a reduction in distress. In none of the studies we have reviewed were patients purposefully selected for psychological distress; indeed, Fawzy et al. (1993) excluded one patient from analysis because of a diagnosis of major depression. Examination of mood data in Spiegel and colleagues' replication study (Classen et al., 2001) reveals that these women's baseline mood was more positive than that of female college student samples (McNair, Lorr, & Droppleman, 1971).

It may be that levels of distress and depression among persons with cancer have been overestimated (Coyne, Benazon, Gaba, Calzone, & Weber, 2000; Coyne, Palmer, Shapiro, Thompson, & DeMichele, 2004). Observational studies have sometimes found levels of distress among persons with cancer, particularly those with early stage disease or those who are posttreatment, comparable to those of college students, primary care patients, or the general population (Cassileth, Lusk, Walsh, Doyle, & Maier, 1989; Cella et al., 1989). Studies in which the effects of psychotherapy on survival were examined have generally involved samples with advanced disease, in which higher levels of distress might be anticipated. However, it could be that the requirement that patients be available for regularly scheduled sessions over a considerable time period selects for a less distressed sample.

A final source of doubt about changes in mood as the basis for improved survival is the poor performance of mood in predicting survival. Fawzy et al. (1993) found that more negative mood at baseline predicted longer survival, consistent with at least some observational studies (Brown, Butow, Culjack, Coates, & Dunn, 2000). Spiegel and Giese-Davis (2003) noted that the literature is at best mixed concerning depressed mood predicting cancer incidence, and efforts to demonstrate that depression predicts progression and mortality are challenging given the potential confounding of mood with physical symptoms. At the present time, there is considerable skepticism in the larger literature concerning whether a causal role for depression or emotional well-being in cancer progression can be demonstrated when appropriate controls are introduced for known biological prognostic indicators, physical symptoms, and side effects of treatment (Faller & Schmidt, 2004). Recent observational studies have failed to find that emotional well-being predicts survival in metastatic (Efficace, Biganzoli, et al., 2004) or early breast cancer (Efficace, Therasse, et al., 2004; Goodwin et al., 2001). These studies are part of a larger literature investigating whether patients' own self-assessments work as well as known biological prognostic factors, and although patients' judgments of their condition do have prognostic value, emotional well-being does not have value as an independent predictor of survival (Efficace, Therasse, et al., 2004).

Edwards et al. (2004) used meta-analysis to evaluate the mood effects of interventions tested to improve survival among women with metastatic breast cancer, and the authors confronted formidable barriers to meaningful integration of the data. They found that investigators would typically include multiple measures of similar constructs or would score the same instrument in multiple ways without controlling for the number of comparisons being made. Even when reviewing studies that used the same measure—the Profile of Mood States (POMS)—Edwards et al. had to contend with long versus short versions of the scale, varying timing of assessments, and seemingly conflicting results for very similar interventions (i.e., Goodwin et al., 2001; Spiegel et al., 1981). Edwards et al. nonetheless concluded that the evidence of improved psychological functioning was very limited and generally not maintained.

Data not included in Edwards et al. (2004) also fail to provide evidence of robust and reliable effects on mood. Spiegel and colleagues' replication study (Classen et al., 2001) revealed no effects of the intervention on POMS total mood score and no effect for self-reported depression as measured by the Center for Epidemiologic Studies—Depression Scale (C. Classen, personal communication, May 15, 2001) but an effect for cancer-specific distress on the Impact of Event Scale (Horowitz, Wilner, & Alvarez, 1979). Fawzy, Cousins, et al. (1990) found that patients in the intervention group had higher vigor at the end of the intervention period, but there were no group differences on six other POMS scales. However, differences in mood favoring the intervention group were found for five of the POMS scales at 6-month follow up. This pattern of a possible delayed mood benefit contrasts with the results of Cunningham et al. (1998) and Edelman, Bell, and Kidman (1999), in which postintervention mood effects were found but had dissipated by the first follow-up. Kissane et al. (2003) examined effects on 11 self-report measures, as well as the proportion of patients with a diagnosis of major or minor depression or anxiety disorder. When initial differences between the intervention and control group were taken into account, there were no group differences in psychological functioning.

We have thus far excluded from this part of our discussion three studies that appear to have confounded intervention with increased medical care (Kuchler et al., 1999; McCorkle et al., 2000; J. L. Richardson et al., 1990). The investigators in two of these studies contradict the classification of their interventions as psychotherapy (McCorkle et al., 2000; J. L. Richardson et al., 1990); two of the studies did not assess mood (Kuchler et al., 1999; McCorkle et al.,

2000), and the third failed to find an effect on depression (J. L. Richardson et al., 1990).

In summary, it is difficult to make the case that the interventions that have been examined for effects on survival have substantial impact on psychological functioning, particularly in patients with advanced stages of cancer. Claims of any enduring benefit depend on analyses of selective as-treated samples rather than intent-to-treat analysis. Results based on the availability of a complete set of assessments do not generalize to the full sample of patients initiating treatment. There has also been selective emphasis on positive findings among mixed findings with multiple measures of mood and selective ignoring of null effects on standardized measures of psychological functioning. Thus, although the original Spiegel et al. (1989) study has been cited as demonstrating positive effects on psychological functioning, complete data were lacking for almost half of the patients and no differences were found between intervention and control groups in depression, self-esteem, or denial.

It thus does not appear that a case can be made for the alleviation of psychological distress as the mechanism by which an intervention affects survival. We therefore lose a set of ready explanations for why psychotherapy should affect survival and are left without a means of distinguishing which intervention studies should be examined for unanticipated effects on survival. If we had found that interventions purporting to show an effect on survival also reliably affect psychological functioning, then we would have had at least some means of identifying which of the hundreds of psychosocial intervention studies (e.g., Newell et al., 2002) might be expected to demonstrate a survival effect, even those for which mortality data had not yet been examined (a factor that further complicates attempts to determine a denominator in calculating box score assessments).

Where Are We? Why Did It Take So Long to Get Here? Is Further Research Warranted?

As an overview, the idea that psychotherapy prolongs the survival of people with cancer remains "inherently improbable" (Spiegel, 2004, p. 133), despite an accumulation of more than 15 years of research. As we have shown, empirical support for the hypothesis that psychotherapy promotes survival depends on attaching considerable weight to two trials with modest samples sizes, no a priori hypotheses concerning survival, and less appropriate strategies for reducing, analyzing, and interpreting the resulting data. In each study, the investigators claimed a strong effect on survival. In support of this claim, the first trial (Spiegel et al., 1989) focused on mean survival times, rather than the more appropriate median, and had to accommodate evidence that the intervention affected survival because it warded off an anomalous increase in mortality among control patients 2 years after randomization. Making a strong claim on the basis of the second study (Fawzy et al., 1993) involves ignoring a host of problems: analyses that did not use an intent-to-treat method; selective exclusion of intervention patients who were unlikely to show a benefit from treatment; an anomalous level of death among controls; and a statistically significant effect that would be undone by reclassification of a single patient (in comparison to the multiple patients lost to follow-up in both groups). Results of these trials thus do not provide a basis for revising the assessment that survival effects for psychotherapy are inherently improbable. If the results of Spiegel et al. and Fawzy et al. are not sufficient to revise a negative appraisal of the evidence, we are not given further encouragement from recent null trials. Our conclusion is that given the limitations, there is not reason to assume that psychotherapy promotes survival. The lack of evidence for a mechanism by which psychotherapy should influence survival serves to strengthen this skepticism.

Much importance has been attached to the claim that psychotherapy promotes the survival of people with cancer, and abandoning this claim may have negative consequences for this field. It would be useful for the field's development to consider why it may have taken so long to recognize the lack of support for this claim. First, it appears that the field was excited by the positive interpretations given to the results of Spiegel et al. (1989) and Fawzy et al. (1993); if psychotherapy were to improve survival, a great deal of pain and suffering could be ameliorated and avoided. Second, interventions with little psychotherapeutic content or with substantial cointervention confound were presented as relevant by the leading researchers. Inclusion of these studies in box scores misspecified the constructs under investigation in the design of the interventions and created "bracket creep" (McNally, 2003) that allowed survival effects that might have been related to improved medical monitoring or more intensive medical care to be attributed to psychotherapy.

The problems with many studies cited as evidence of an effect of psychotherapy on survival are evident from a careful reading. However, we believe that a third factor in the persistent advocacy for a survival effect relates to differences in the training of behavioral scientists and medical trialists. The superiority of medians over means for summarizing survival data, given the characteristic distribution of length of patient survival, is well recognized in clinical epidemiology but seldom noted in behavioral medicine. Yet this recognition is crucial for critically appraising Spiegel et al. (1989). Similarly, the importance of intent-to-treat-analysis has not been appreciated in behavioral medicine until very recently, and the requisite acquisition of data from patients who do not complete treatment could even be seen as counterintuitive. Our discussion of the pitfalls of accepting unexpected strong results from trials with modest sample sizes also clashes with the common wisdom that significant results obtained with a small sample are more rather than less impressive. Additionally, the failure to appreciate the importance of cointervention confounds has hampered the ability of the field to interpret the relevance of other studies to the survival hypothesis. An evaluation of the available evidence for the effects of psychotherapy on survival (or any other effect based on data from randomized clinical trials) requires knowledge and skills that have been in short supply. Recognition of the inadequate response of the field to the quality of these data should serve as a call for higher standards and better education concerning the conduct, reporting, and interpretation of clinical trials. This effort has begun, as evidenced by the randomized clinical trial training sessions now offered by the National Institutes of Health Office of Behavioral and Social Science Research, but there remains much to do early in training as well.

We believe that claims that psychotherapy promotes survival have gone beyond the data that have been mustered in their support. Indeed, the reception of claims that psychotherapy promotes survival of persons who have been diagnosed with cancer is a striking instance of how social factors determine how empirical

data are filtered, interpreted, and accepted (Dopson & Fitzgerald, 2005). Initially, the claims that caught the attention of the media and a broad lay audience were that a psychotherapeutic intervention study demonstrated that women with cancer received a substantial survival benefit from intervention, and that this result was surprising even to the research team that carried out the study. This claim appears to have caused excitement in both professional and lay communities eager for an indication that patients could exert some direct control over their illness. Next, a study team that had completed an examination of the effects of group cognitive–behavioral therapy on psychosocial outcomes among melanoma patients produced a post hoc examination of their survival data, reporting an effect on survival and offering explanations of the mechanisms by which such an effect might have been obtained.

There were few outspoken skeptics of these trials (e.g., Fox, 1991, 1995; Sampson, 2002); their critiques had little effect on professional and lay opinions but were met with lively rebuttal (Goodwin et al., 1999; Kraemer & Spiegel, 1999). This polarization seemed to reify the findings such that what was originally presented as an unanticipated result that confirmed an improbable hypothesis came to be established as a secure finding, and the burden of proof shifted to failures of replication rather than the original data. As well, the limited effect of critiques may have been a matter of *Se non è vero è ben trovato* in the reception of the initial survival studies: Even if untrue, at least the claims were well crafted. These claims held promise for the field of psychooncology and behavioral medicine. Conversely, criticisms of the evidence could be seen as an undermining of the rationale for a promising new line of research and funding.

The claim of an effect on survival may have been consonant with larger sociocultural forces as well. At the time the initial survival studies were coming to light, cancer was being destigmatized and persons who had been diagnosed with cancer were being construed as survivors rather than victims. Cancer was being socially construed as a test of the will and a fight that could potentially be won by proper attitude and effort (Sontag, 1978). The potency of a "fighting spirit" (Greer, Morris, Pettingale, & Haybittle, 1990) was readily accepted, even if subsequent work failed to replicate its prognostic significance (Watson, Haviland, Greer, Davidson, & Bliss, 1999). In this context, skeptics were not granted the credibility of proponents, regardless of the quality of evidence. In short, one cannot understand the persistent enthusiasm for the claim that psychotherapy promotes survival among people with cancer without paying attention to its cultural context.

A Test of the Effect of Psychotherapy on Survival: Basic Parameters and Lack of Justification

We have noted that initial tests of the effects of psychotherapy on survival involved sample sizes so modest as to provide both inadequate statistical power and a basis for skepticism concerning an unexpected positive finding. In contrast, the sample sizes of the most recent studies have been determined by formal power analyses. Yet parameters for these power analyses were set by the unrealistically strong effects claimed for earlier studies, rather than taking into account the improbability that psychotherapy could substantially improve survival. Design of an adequate test of a survival effect requires a realistic appraisal of the size of effect that should be anticipated.

A number of the key studies have focused on women with metastatic breast cancer (Cunningham et al., 1998; Edelman, Bell, & Kidamn, 1999; Goodwin et al., 2001; Spiegel et al., 1989), and the hypothesis has been that psychotherapy improves survival obtained with routine care with first-line treatments. Such first-line treatments currently yield a 5-year survival rate of 23%, a figure remarkably difficult to improve on with additional available biomedical treatments (Gennari, Conte, Rosso, Orlandini, & Bruzzi, 2005; Vogel & Tan-Chiu, 2005). Only a small proportion of patients achieve long-term remission (Greenberg et al., 1996). As Bernard-Marty, Fatima Cardoso, and Piccart (2004) noted, "Despite more than 3 decades of research, metastatic breast cancer (MBC) remains essentially incurable and, after documentation of metastasis, the median survival time is approximately 2 years" (p. 617). It is unclear why we should expect psychotherapy to make a difference where a wide range of promising medical treatments have consistently failed. The virtually superimposable survival curves for intervention and control patients in Goodwin et al. (2001) would seem to give no basis for expecting an effect. The lack of consistent evidence for a mechanism would seem to provide further discouragement.

The appeal of a study with women with metastatic breast cancer can variously be seen as reflecting the precedence of Spiegel et al. (1989), the apparent inability of biomedical treatments to improve on established standards of care, and the pragmatic requirement of accumulating sufficient clinical events—that is, deaths—within the time constraints of what could be funded with available grant mechanisms. Yet metastatic breast cancer might be a particularly inappropriate context for demonstrating that psychotherapy improves survival because of the lack of evidence that *any* intervention confers improvement beyond standard care.

Does early breast cancer provide a more promising focus? In the United States, the 5-year survival rate for women with localized breast cancer is now 98% (American Cancer Society, 2006). This high rate of survival makes it difficult to demonstrate that any additional treatment would yield a clinically significant improvement. An integration of 28 trials with 16,513 women of whom 3,782 had died concluded that both tamoxifen and cytotoxic chemotherapy reduce 5-year mortality (Early Breast Cancer Trialists' Collaborative Group, 1988). Yet when trials were considered individually, only a single trial had an effect significant at $p < .01$.

Given these data, we question whether it would be ethical or practical to continue to undertake clinical trials examining whether psychotherapy prolongs the survival of women with early breast cancer. As Altman (1994) persuasively argued, sometimes the reflexive call that "further research is needed" needs to be countered with the notion that "we need less research, better research, and research done for the right reasons" (p. 283). Clearly another small, underpowered trial or more post hoc analyses of survival in trials for which survival was not originally designated as a primary outcome are not needed. Yet power analyses need to be justified with respect to a defensible estimate of effect size. As we noted in our analyses of the barriers to demonstrating an effect on survival of either early stage or metastatic breast cancer, an adequately powered trial would of necessity be a very large trial, larger than any to date, perhaps larger than the current strength of evidence would justify. Underpowered trials pose an ethical issue aside from the need to avoid the small-trial biases to which we have alluded in this article. One requirement is that trials be adequately powered

to yield a scientifically credible result in order to justify enrolling patients who would get no benefit from assignment to the control condition. Patients enrolled in underpowered trials are being asked to assume the burden and risks of participation without the opportunity to contribute to scientific knowledge (Halpern, Karlawish, & Berlin, 2002), a dubious ethical situation. An adequately powered study would require a much larger sample size than has been undertaken thus far. Another requirement for an ethical trial is that there exist a basis for informing patients that the intervention might provide some benefit. Existing data do not support the claim that psychotherapy prolongs survival, and there is an inadequate basis for specifying a mechanism by which such an effect would be produced. In the trials conducted to date with metastatic breast cancer patients, there has been no demonstration of a robust effect on mood, and so such side benefit cannot be promised on an empirical basis.

In short, we come to the conclusion that an adequate test of whether psychotherapy promotes survival is not justified by the available data. Certainly, in biomedicine, a large-scale trial would not be considered warranted for cases in which a hypothesis was interesting but improbable given the available data. At a time of limited resources for psychosocial studies among persons with cancer and cancer survivors, one must ask whether it would be justified to withhold funds from more promising lines of research to amass the enormous resources that an adequately powered study of survival would require.

This is particularly true when we, as a science, have better prospects for demonstrating that persons with cancer can be assisted in improving the quality, if not the quantity, of their lives. Yet here, too, claims have exceeded the strength of the evidence. When the same critical appraisal tools and methodological and statistical standards we have applied here are extended to the larger literature, the evidence that after a diagnosis of cancer people generally benefit from receiving psychosocial interventions is shown to be a lot weaker than it first appeared (Coyne & Lepore, 2006). A decade ago, Meyer and Mark (1995) declared on the basis of a meta-analysis that it would be a waste of resources to continue to research the question of whether persons with cancer benefit from intervention. More recently, there have been calls from influential groups such as the National Cancer Policy Board of the Institute of Medicine (Hewitt, Herdman, & Holland, 2004) and Central European Cooperative Group (Beslija et al., 2003) for the integration of psychosocial interventions into routine comprehensive care for cancer, as well as formulation of practice guidelines (Turner et al., 2005). Yet a recent review of available reviews concluded that as the sophistication of narrative and meta-analytic reviews improves, there is much less of "a compelling case for the value of these interventions for the typical person being treated for cancer. The more rigorous the review, the less likely it is to conclude there is evidence that psychological interventions are effective" (Lepore & Coyne, 2006, p. 85).

Aside from increasing awareness of the limitations of the quality of existing research, a major problem has been the prevailing assumption that persons with cancer are sufficiently distressed as to be able to register a clinically significant reduction in distress as a result of intervention (Coyne et al., 2006). When, in the unusual study, researchers break with this assumption and limit their samples to distressed persons with cancer, demonstrations of efficacy of intervention are more likely (Greer et al., 1992; Nezu, Nezu, Felgoise, McClure, & Houts, 2003).

There is no good a priori reason to reject the assumption that with appropriate tailoring to the demands of cancer and its treatment, interventions that reduce prolonged or functionally impairing distress in other contexts will benefit persons with cancer. However, we are concerned that the necessary retreat from the claim that all persons with cancer need or will benefit from formal psychosocial interventions becomes more awkward and embarrassing when it is accompanied by a delayed concession that such interventions do not extend survival.

References

Altman, D. G. (1994). The scandal of poor medical research: We need less research, better research, and research done for the right reasons. *British Medical Journal, 308,* 283–284.

Altman, D. G., Schulz, K. F., Moher, D., Egger, M., Davidoff, F., Elbourne, D., et al. (2001). The revised CONSORT statement for reporting randomized trials: Explanation and elaboration. *Annals of Internal Medicine, 134,* 663–694.

American Cancer Society. (2006). *Cancer facts and figures.* Atlanta, GA: Author.

Andersen, B. L., Farrar, W. B., Golden-Kreutz, D. M., Glaser, R., Emery, C. F., Crespin, T. R., et al. (2004). Psychological, behavioral, and immune changes after a psychological intervention: A clinical trial. *Journal of Clinical Oncology, 22,* 3570–3580.

Anderson, C. A., Lepper, M. R., & Ross, L. (1980). Perseverance of social theories: The role of explanation in the persistence of discredited information. *Journal of Personality and Social Psychology, 39,* 1037–1049.

Antoni, M. H., Lehman, J. M., Kilbourn, K. M., Boyers, A. E., Culver, J. L., Alferi, S. M., et al. (2001). Cognitive–behavioral stress management intervention decreases the prevalence of depression and enhances benefit finding among women under treatment for early-stage breast cancer. *Health Psychology, 20,* 20–32.

Assmann, S. F., Pocock, S. J., Enos, L. E., & Kasten, L. E. (2000). Subgroup analysis and other (mis)uses of baseline data in clinical trials. *Lancet, 355,* 1064–1069.

Babyak, M. A. (2004). What you see may not be what you get: A brief, nontechnical introduction to overfitting in regression-type models. *Psychosomatic Medicine, 66,* 411–421.

Bagenal, F., Easton, D. F., Harris, E., Chilvers, C. E. D., & McElwain, T. J. (1990). Survival of patients with breast cancer attending Bristol Cancer Help Center. *Lancet, 336,* 606–610.

Begg, C., Cho, M., Eastwood, S., Horton, R., Noher, D., Olkin, I., et al. (1996). Improving the quality of reporting of randomized controlled trials: The CONSORT statement. *Journal of the American Medical Association, 276,* 637–639.

Berkman, L. F., Blumenthal, J., Burg, M., Carney, R. M., Catellier, D., Cowan, M. J., et al. (2003). Effects of treating depression and low-perceived social support on clinical events after myocardial infarction: The Enhancing Recovery in Coronary Heart Disease Patients (ENRICHD) Randomized Trial. *Journal of the American Medical Association, 289,* 3106–3116.

Bernard-Marty, C., Fatima Cardoso, F., & Piccart, M. J. (2004). Facts and controversies in systemic treatment of metastatic breast cancer. *Oncologist, 9,* 617–632.

Berry, D. A., & Stangl, D. (1996). *Bayesian biostatistics.* New York: Marcel Dekker.

Beslija, S., Bonneterre, J., Burstein, H., Gnant, M., Goodwin, P., Heinemann, V., et al. (2003). For the Central European Cooperative Group: Consensus on medical treatment of metastatic breast cancer. *Breast Cancer Research and Treatment, 81*(Suppl. 1), S1–S7.

Blake-Mortimer, J., Gore-Felton, C., Kimerling, R., Turner-Cobb, J. M., & Spiegel, D. (1999). Improving the quality and quantity of life among patients with cancer: A review of the effectiveness of group psychotherapy. *European Journal of Cancer, 35,* 1581–1586.

Bordeleau, L., Szalai, J. P., Ennis, M., Leszcz, M., Speca, M., Sela, R., et al. (2003). Quality of life in a randomized trial of group psychosocial support in metastatic breast cancer: Overall effects of the intervention and an exploration of missing data. *Journal of Clinical Oncology, 21,* 1944–1951.

Bracken, M. B., & Sinclair, J. C. (1998). When can odds ratios mislead? Avoidable systematic error in estimating treatment effects must not be tolerated. *British Medical Journal, 317,* 1156.

Bredart, A., Cayrou, S., & Dolbeault, S. (2002). Re: Systematic review of psychological therapies for cancer patients: Overview and recommendations for future research. *Journal of the National Cancer Institute, 94,* 1810–1811.

Brooks, S. T., Whitely, E., Egger, M., Smith, G. D., Mulheran, P. A., & Peters, T. J. (2004). Subgroup analyses in randomized trials: Risks of subgroup-specific analyses; power and sample size for the interaction test. *Journal of Clinical Epidemiology, 57,* 229–236.

Brophy, J. M., & Joseph, L. (1995). Placing trials in context using Bayesian analysis. GUSTO revisited by Reverend Bayes. *Journal of the American Medical Association, 273,* 871–875.

Brown, J. E., Butow, P. N., Culjack, G., Coates, A. S., & Dunn, S. M. (2000). Psychosocial predictors of outcome: Time to relapse and survival in patients with early stage melanoma. *British Journal of Cancer, 83,* 1448–1453.

Butler, L. D., Koopman, C., Cordova, M. J., Garlan, R. W., DiMiceli, S., & Spiegel, D. (2003). Psychological distress and pain significantly increase before death in metastatic breast cancer patients. *Psychosomatic Medicine, 65,* 416–426.

Cassileth, B. R., Lusk, E. J., Walsh, W. P., Doyle, B., & Maier, M. (1989). The satisfaction and psychosocial status of patients during treatment for cancer. *Journal of Psychosocial Oncology, 7,* 47–57.

Cella, D. F., Tross, S., Orav, E. J., Holland, J. C., Silberfarb, P. M., & Rafla, S. (1989). Mood states of patients after the diagnosis of cancer. *Journal of Psychosocial Oncology, 7,* 45–55.

Chalmers, T. C. (1991). Problems induced by meta-analyses. *Statistics in Medicine, 10,* 971–979.

Chow, E., Tsao, M. N., & Harth, T. (2004). Does psychosocial intervention improve survival in cancer? A meta-analysis. *Palliative Medicine, 18,* 25–31.

Christenfeld, N. J. S., Sloan, R. P., Carroll, D., & Greenland, S. (2004). Risk factors, confounding, and the illusion of statistical control. *Psychosomatic Medicine, 66,* 868–875.

Classen, C., Butler, L. D., Koopman, C., Miller, E., DiMicelli, S., Giese-Davis, J., et al. (2001). Supportive–expressive group therapy and distress in patients with metastatic breast cancer: A randomized clinical intervention trial. *Archives of General Psychiatry, 58,* 494–501.

Cocker, K. I., Bell, D. R., & Kidman, A. D. (1994). Cognitive–behavior therapy with advanced breast-cancer patients: A brief report of a pilot study. *Psycho-Oncology, 3,* 233–237.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20,* 37–46.

Cook, D. J., Hebert, P. C., Heyland, D. K., Guyatt, G. H., Brun-Buisson, C., Marshall, J. C., et al. (1997). How to use an article on therapy or prevention: Pneumonia prevention using subglottic secretion drainage. *Critical Care Medicine, 25,* 1502–1513.

Cook, J. M., Palmer, S., Hoffman, K., & Coyne, J. C. (in press). Evaluation of clinical trials appearing in *Journal of Consulting and Clinical Psychology*: CONSORT and beyond. *The Scientific Review of Mental Health Practice.*

Cooper, H. (1989). *Integrating research: A guide for literature reviews* (2nd ed.). Newbury Park, CA: Sage.

Cooper, H., & Hedges, L. V. (Eds.). (1994). *The handbook of research synthesis.* New York: Russell Sage Foundation.

Coyne, J. C., Benazon, N. R., Gaba, C. G., Calzone, K., & Weber, B. L. (2000). Distress and psychiatric morbidity among women from high-risk breast and ovarian cancer families. *Journal of Consulting and Clinical Psychology, 68,* 864–874.

Coyne, J. C., & Lepore, S. J. (2006). Rebuttal: The black swan fallacy in evaluating psychological interventions for distress in cancer patients. *Annals of Behavioral Medicine, 32,* 115–118.

Coyne, J. C., Lepore, S. J., & Palmer, S. C. (2006). Efficacy of psychosocial interventions in cancer care: Evidence is weaker than it first looks. *Annals of Behavioral Medicine, 32,* 104–110

Coyne, J. C., Palmer, S. C., Shapiro, P. J., Thompson, R., & DeMichele, A. (2004). Distress, psychiatric morbidity, and prescriptions for psychotropic medication in a breast cancer waiting room sample. *General Hospital Psychiatry, 26,* 121–128.

Cunningham, A. J., & Edmonds, C. (2002). Group psychosocial support in metastatic breast cancer. *New England Journal of Medicine, 346,* 1247–1248.

Cunningham, A. J., Edmonds, C. V. I., Jenkins, G. P., Pollack, H., Lockwood, G. A., & Warr, D. (1998). A randomized controlled trial of the effects of group psychological therapy on survival in women with metastatic breast cancer. *Psycho-Oncology, 7,* 508–517.

Deeks, J. J. (1998). When can odds ratios mislead? *British Medical Journal, 317,* 1155–1156.

Detsky, A. S., Naylor, C. D., Orourke, K., McGeer, A. J., & Labbe, K. A. (1992). Incorporating variations in the quality of individual randomized trials into meta-analysis. *Journal of Clinical Epidemiology, 45,* 255–265.

Diamond, J. (1998). *Because cowards get cancer too: A hypochondriac confronts his nemesis.* New York: Random House.

Doan, B. D., Gray, R. E., & Davis, C. S. (1993). Belief in psychological effects on cancer. *Psycho-Oncology, 2,* 139–150.

Dopson, S., & Fitzgerald, L. (Eds.). (2005). *Knowledge into action?* New York: Oxford University Press.

Early Breast Cancer Trialists' Collaborative Group. (1998). Tamoxifen for early breast cancer: An overview of the randomised trials. *Lancet, 351,* 1451–1467.

Edelman, S., Bell, D. R., & Kidman, A. D. (1999). A group cognitive behaviour therapy programme with metastatic breast cancer patients. *Psycho-Oncology, 8,* 295–305.

Edelman, S., Craig, A., & Kidman, A. D. (2000). Can psychotherapy increase the survival time of cancer patients? A review. *Journal of Psychosomatic Research, 49,* 149–156.

Edelman, S., Lemon, J., Bell, D. R., & Kidman, A. D. (1999). Effects of group CBT on the survival time of patients with metastatic breast cancer. *Psycho-Oncology, 8,* 474–481.

Edwards, A. G. K., Hailey, S., & Maxwell, M. (2004). Psychological interventions for women with metastatic breast cancer (Cochrane Review). *Cochrane Database of Systematic Reviews, 2.*

Efficace, F., Biganzoli, L., Piccart, M., Coens, C., Van Steen, K., Cufer, T., et al. (2004). Baseline health-related quality-of-life data as prognostic factors in a Phase III multicentre study of women with metastatic breast cancer. *European Journal of Cancer, 40,* 1021–1030.

Efficace, F., Therasse, P., Piccart, M. J., Coens, C., Van Steen, K., Welnicka-Jaskiewics, M., et al. (2004). Health-related quality of life parameters as prognostic factors in a nonmetastatic breast cancer population: An international multicenter study. *Journal of Clinical Oncology, 16,* 3381–3388.

Elsesser, K., van Berkel, M., Sartory, G., Biermanngocke, W., & Ohl, S. (1994). The effects of anxiety management training on psychological variables and immune parameters in cancer patients: A pilot study. *Behavioral and Cognitive Psychotherapy, 22,* 13–23.

Faller, H., & Schmidt, M. (2004). Prognostic value of depressive coping

and depression in survival of lung cancer patients. *Psycho-Oncology, 13,* 359–363.

Farber, J., Weinerman, B., Kuypers, J., & Behar, K. (1981). A comparison of different support group formats in aiding cancer patients in coping with their disease and treatment. *Proceedings of the American Association for Cancer Research, 22,* 394.

Fawzy, F. I., Canada, A. L., & Fawzy, N. W. (2003). Malignant melanoma: Effects of a brief, structured psychiatric intervention on survival and recurrence at 10-year follow-up. *Archives of General Psychiatry, 60,* 100–103.

Fawzy, F. I., Cousins, N., Fawzy, N. W., Kemeny, M. E., Elashoff, R., & Morton, D. (1990). A structured psychiatric intervention for cancer patients: I. Changes over time in methods of coping and affective disturbance. *Archives of General Psychiatry, 47,* 720–725.

Fawzy, F. I., Fawzy, N. W., Hyun, C. S., Elashoff, R., Guthrie, D., Fahey, J. L., et al. (1993). Malignant melanoma: Effects of an early structured psychiatric intervention, coping, and affective state on recurrence and survival 6 years later. *Archives of General Psychiatry, 50,* 681–689.

Fawzy, F. I., Kemeny, M. E., Fawzy, N. W., Elashoff, R., Morton, D., Cousins, N., et al. (1990). A structured psychiatric intervention for cancer patients: I. Changes over time in immunological measures. *Archives of General Psychiatry, 47,* 729–735.

Feinstein, A. R. (1995). Meta-analysis: Statistical alchemy for the 21st century. *Journal of Clinical Epidemiology, 48,* 81–86.

Fox, B. H. (1991). Quandaries created by unlikely numbers in some of Grossarth-Maticek's studies. *Psychology Inquiries, 2,* 242–247.

Fox, B. H. (1995). Some problems and some solutions in research on psychotherapeutic intervention in cancer. *Supportive Care in Cancer, 3,* 257.

Fox, B. H. (1998). A hypothesis about Spiegel et al.'s 1989 paper on psychosocial intervention and breast cancer survival. *Psycho-Oncology, 7,* 361–370.

Fox, B. H. (1999). Clarification regarding comments about a hypothesis. *Psycho-Oncology, 8,* 366–367.

Gellert, G. A., Maxwell, R. M., & Siegel, B. S. (1993). Survival of breast-cancer patients receiving adjunctive psychosocial support therapy: A 10-year follow-up study. *Journal of Clinical Oncology, 11,* 66–69.

Gennari, A., Conte, P., Rosso, R., Orlandini, C. A., & Bruzzi, P. (2005). Survival of metastatic breast carcinoma patients over a 20-year period: A retrospective analysis based on individual patient data from six consecutive studies. *Cancer, 104,* 1742–1750.

Goodman, S. N., & Berlin, J. A. (1994). The use of predicted confidence intervals when planning experiments and the misuse of power when interpreting results. *Annals of Internal Medicine, 121,* 200–206.

Goodwin, P. J. (2004). Support groups in breast cancer: When a negative result is positive. *Journal of Clinical Oncology, 22,* 4244–4246.

Goodwin, P. J., Ennis, M., Bordeleau, L. J., Pritchard, K. I., Trudeau, M. E., Koo, J., et al. (2004). Health-related quality of life and psychosocial status in breast cancer prognosis: Analysis of multiple variables. *Journal of Clinical Oncology, 22,* 4184–4192.

Goodwin, P. J., Leszcz, M., Ennis, M., Koopmans, J., Vincent, L., Guther, H., et al. (2001). The effect of group psychosocial support on survival in metastatic breast cancer. *New England Journal of Medicine, 345,* 1719–1726.

Goodwin, P. J., Pritchard, K. I., & Spiegel, D. (1999). The Fox guarding the clinical trial: Internal vs. external validity in randomized studies. *Psycho-Oncology, 8,* 275.

Greenberg, P. A. C., Hortobagyi, G. N., Smith, T. L., Ziegler, L. D., Frye, D. K., & Buzdar, A. U. (1996). Long-term follow-up of patients with complete remission following combination chemotherapy for metastatic breast cancer. *Journal of Clinical Oncology, 14,* 2197–2205.

Greer, S. (2002). Psychological intervention: The gap between research and practice. *Acta Oncologica, 41,* 238–243.

Greer, S., Moorey, S., Baruch, J. D. R., Watson, M., Robertson, B. M., Mason, A., et al. (1992). Adjuvant psychological therapy for patients with cancer: A prospective randomised trial. *British Medical Journal, 304,* 675–680.

Greer, S., Morris, T., Pettingale, K. W., & Haybittle, J. L. (1990). Psychosocial response to breast cancer and 15-year outcome. *Lancet, 335,* 49–50.

Grossarth-Maticek, R., Frentzel-Beyme, R., & Becker, N. (1984). Cancer risks associated with life events and conflict solution. *Cancer Detection & Prevention, 7,* 201–209.

Hadley, S. W., & Strupp, H. H. (1976). Contemporary views of negative effects in psychotherapy: Integrated account. *Archives of General Psychiatry, 33,* 1291–1302.

Halpern, S. D., Karlawish, J. H. T., & Berlin, J. A. (2002). The continuing unethical conduct of underpowered clinical trials. *Journal of the American Medical Association, 288,* 358–362.

Helgeson, V. S., Cohen, S., Schulz, R., & Yasko, J. (1999). Education and peer discussion group interventions and adjustment to breast cancer. *Archives of General Psychiatry, 56,* 340–347.

Helgeson, V. S., Cohen, S., Schulz, R., & Yasko, J. (2001). Group support interventions for people with cancer: Benefits and hazards. In A. Baum & B. L. Andersen (Eds.), *Psychosocial interventions for cancer* (pp. 269–286). Washington, DC: American Psychological Association.

Hewitt, M., Herdman, R., & Holland, J. (2004). *Meeting psychosocial needs of women with breast cancer.* Washington, DC: National Academies Press.

Higgins, J. P. T., & Green, S. (2005). *Cochrane Handbook for Systematic Reviews of Interventions 4.2.5.* Chichester, England: Wiley.

Holland, J. C., & Lewis, S. (2001). *The human side of cancer: Living with hope, coping with uncertainty.* New York: HarperCollins.

Horowitz, M., Wilner, N., & Alvarez, W. (1979). Impact of Event Scale: A measure of subjective stress. *Psychosomatic Medicine, 41,* 209–218.

Hosaka, T., Tokuda, Y., Sugiyama, Y., Hirai, K., & Okuyama, T. (2000). Effects of a structured psychiatric intervention on immune function of cancer patients. *Experimental Clinical Medicine, 25,* 183–188.

Ilnyckyj, A., Farber, J., Cheang, M., & Weinerman, B. (1994). A randomized controlled trial of psychotherapeutic intervention in cancer patients. *Annals of the Royal College of Physicians and Surgeons of Canada, 272,* 93–96.

Juni, P., Witshi, A., Bloch, R., & Egger, M. (1999). The hazards of scoring the quality of clinical trials for meta-analysis. *Journal of the American Medical Association, 282,* 1054–1060.

Kissane, D. W., Love, A., Hatton, A., Smith, G., Clarke, D. M., Miach, P., et al. (2004). Effect of cognitive–existential group therapy on survival in early-stage breast cancer. *Journal of Clinical Oncology, 22,* 4255–4260.

Kissane, D. W., McKenzie, M., McKenzie, D. P., Forbes, A., O'Neill, I., & Bloch, S. (2003). Psychosocial morbidity associated with patterns of family functioning in palliative care: Baseline data from the Family Focused Grief Therapy controlled trial. *Palliative Medicine, 17,* 527–537.

Kraemer, H. C., Gardner, C., Brooks, J. O., & Yesavage, J. A. (1998). Advantages of excluding underpowered studies in meta-analysis: Inclusionist versus exclusionist viewpoints. *Psychological Methods, 3,* 23–31.

Kraemer, H., & Spiegel, D. (1999). Cunning but careless: Analysis of a non-replication. *Psycho-Oncology, 8,* 273–276.

Kuchler, T., Henne-Burns, D., Rappat, S., Holst, K., Williams, J. I., & Wood-Dauphinee, S. (1999). Impact of psychotherapeutic support on gastrointestinal cancer patients undergoing surgery: Survival results of a trial. *Hepato-Gastroenterology, 46,* 322–335.

Larson, M. R., Duberstein, P. R., Talbot, N. L., Caldwell, C., & Moynihan, J. A. (2000). A presurgical psychosocial intervention for breast cancer patients: Psychological distress and the immune response. *Journal of Psychosomatic Research, 48,* 187–194.

Lee, Y. J., Ellenberg, J. H., Hirtz, D. G., & Nelson, K. B. (1991). Analysis

of clinical trials by treatment actually received: Is it really an option? *Statistics in Medicine, 10,* 1595–1605.

LeLorier, J., Gregoire, G., Benhaddad, A., Lapierre, J., & Derderian, F. (1997). Discrepancies between meta-analyses and subsequent large randomized, controlled trials. *New England Journal of Medicine, 337,* 536–542.

Lemon, J., & Edelman, S. (2003). Perceptions of the "mind–cancer" relationship among the public, cancer patients, and oncologists. *Journal of Psychosocial Oncology, 21,* 43–58.

Lepore, S. J., & Coyne, J. C. (2006). Psychological interventions for distress in cancer patients: A review of reviews. *Annals of Behavioral Medicine, 32,* 85–92.

Lesperance, F., & Frasure-Smith, N. (1999). The seduction of death. *Psychosomatic Medicine, 61,* 18–20.

Levine, A. M., Richardson, J. L., Marks, G., Chan, K., Graham, J., Selser, J. N., et al. (1987). Compliance with oral-drug therapy in patients with hematologic malignancy. *Journal of Clinical Oncology, 5,* 1469–1476.

Lillquist, P. P., & Abramson, J. S. (2002). Separating the apples and oranges in the fruit cocktail: The mixed results of psychosocial interventions on cancer survival. *Social Work in Health Care, 36,* 65–79.

Linn, M. W., Linn, B. S., & Harris, R. (1982). Effects of counseling for late stage cancer patients. *Cancer, 49,* 1048–1055.

Manne, S., & Andrykowski, M. A. (2006). Are psychological interventions effective and accepted by cancer patients? II. Using empirically supported therapy guidelines to decide. *Annals of Behavioral Medicine, 32,* 98–103.

McCorkle, R., Strumpf, N. E., Nuamah, I. F., Adler, D. C., Cooley, M. E., Jepson, C., et al. (2000). A specialized home care intervention improves survival among older post-surgical cancer patients. *Journal of the American Geriatrics Society, 48,* 1707–1713.

McNair, D. M., Lorr, M., & Droppleman, L. F. (1971). *EdITS manual for the Profile of Mood States.* San Diego, CA: Educational and Industrial Testing Service.

McNally, R. J. (2003). Progress and controversy in the study of posttraumatic stress disorder. *Annual Reviews of Psychology, 54,* 229–252.

Meyer, T. J., & Mark, M. M. (1995). Effects of psychosocial interventions with adult cancer patients: A meta-analysis of randomized experiments. *Health Psychology, 14,* 101–108.

Miller, M., Boye, M. J., Butow, P. N., Gattellari, M., Dunn, S. M., & Childs, A. (1998). The use of unproven methods of treatment by cancer patients: Frequency, expectations and cost. *Supportive Care in Cancer, 6,* 337.

Moher, D., Jadad, A. R., Nichol, G., Penman, M., Tugwell, P., & Walsh, S. (1995). Assessing the quality of randomized controlled trials: An annotated bibliography of scales and checklists. *Controlled Clinical Trials, 16,* 62–73.

Moher, D., Pham, B., Jones, A., Cook, D. J., Jadad, A. R., Moher, M., et al. (1998). Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet, 352,* 609–613.

Motulsky, H. (1995). *Intuitive biostatistics.* London: Oxford University Press.

Newell, S. A., Sanson-Fisher, R. W., & Savolainen, N. J. (2002). Systematic review of psychological therapies for cancer patients: Overview and recommendations for future research. *Journal of the National Cancer Institute, 94,* 558–584.

Nezu, A. M., Nezu, C. M., Felgoise, S. H., McClure, K. S., & Houts, P. S. (2003). Project genesis: Assessing the efficacy of problem-solving therapy for distressed adult cancer patients. *Journal of Consulting and Clinical Psychology, 71,* 1036–1048.

Palmer, S. C., & Coyne, J. C. (2004). Examining the evidence that psychotherapy improves the survival of cancer patients. *Biological Psychiatry, 56,* 61–62.

Peduzzi, P., Concato, J., Feinstein, A. R., & Holford, T. R. (1995). Importance of events per independent variable in proportional hazards regression analysis: II. Accuracy and precision of regression estimates. *Journal of Clinical Epidemiology, 48,* 1503–1510.

Peduzzi, P., Concato, J., Kemper, E., Holford, T. R., & Feinstein, A. R. (1996). A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology, 49,* 1373–1379.

Peduzzi, P., Henderson, W., Hartigan, P., & Lavori, P. (2002). Analysis of randomized controlled trials. *Epidemiologic Reviews, 24,* 26–38.

Peto, R., Pike, M. C., Armitage, P., Breslow, N. E., Cox, D. R., & Howard, S. V. (1976). Design and analysis of randomized clinical trials requiring prolonged observation of each patient: I. Introduction and design. *British Journal of Cancer, 34,* 585–612.

Peto, R., Pike, M. C., Armitage, P., Breslow, N. E., Cox, D. R., Howard, S. V., et al. (1977). Design and analysis of randomized clinical trials requiring prolonged observation of each patient: II. Analysis and examples. *British Journal of Cancer, 35,* 1–39.

Pfeffer, M. A., & Jarcho, J. A. (2006). The charisma of subgroups and the subgroups of CHARISMA. *New England Journal of Medicine, 354,* 1667–1669.

Piantadosi, S. (1990). Hazards of small clinical trials. *Journal of Clinical Oncology, 8,* 1–3.

Ratcliffe, M. A., Dawson, A. A., & Walker, L. G. (1995). Eysenck Personality Inventory L scores in patients with Hodgkins disease and non-Hodgkins lymphoma. *Psycho-Oncology, 4,* 39–45.

Relman, A. S., & Angell, M. (2002). Resolved: Psychosocial interventions can improve clinical outcomes in organic disease (con). *Psychosomatic Medicine, 64,* 558–563.

Richardson, J. L., Marks, G., Johnson, C. A., Graham, J. W., Chan, K. K., Selser, J. N., et al. (1987). Path model of multidimensional compliance with cancer therapy. *Health Psychology, 6,* 183–207.

Richardson, J. L., Shelton, D. R., Krailo, M., & Levine, A. M. (1990). The effect of compliance with treatment on survival among patients with hematologic malignancies. *Journal of Clinical Oncology, 8,* 356–364.

Richardson, M. A., Post-White, J., Grimm, E. A., Moye, L. A., Singletary, S. E., & Justice, B. (1997). Coping, life attitudes, and immune responses to imagery and group support after breast cancer treatment. *Alternative Therapies in Health and Medicine, 3,* 62–70.

Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin, 86,* 638–641.

Ross, L., Boesen, E. H., Dalton, S. O., & Johansen, C. (2002). Mind and cancer: Does psychosocial intervention improve survival and psychological well-being? *European Journal of Cancer, 38,* 1447–1457.

Ross, L., Thomsen, B. L., Boesen, E. H., & Johansen, C. (2004). In a randomized controlled trial, missing data led to biased results regarding anxiety. *Journal of Clinical Epidemiology, 57,* 1131–1137.

Sackett, D. L., Deeks, J. J., & Altman, D. G. (1996). Down with odds ratios! *Evidence-Based Medicine, 1,* 164–166.

Sampson, W. (1997). Inconsistencies and errors in alternative medicine research. *Skeptical Inquirer, 21,* 35–59.

Sampson, W. (2002). Controversies in cancer and the mind: Effects of psychosocial support. *Seminars in Oncology, 29,* 595–600.

Schattner, A. (2003). The emotional dimension and the biological paradigm of illness: Time for a change. *QJM: An International Journal of Medicine, 96,* 617–621.

Schneiderman, N., Saab, P. G., Catellier, D. J., Powell, L. H., DeBusk, R. F., Williams, R. B., et al. (2004). ENRICHD investigators: Psychosocial treatment within sex by ethnicity subgroups in the Enhancing Recovery in Coronary Heart Disease clinical trial. *Psychosomatic Medicine, 66,* 475–483.

Schulz, K. F., Chalmers, I., Hayes, R. J., & Altman, D. G. (1995). Empirical evidence of bias: Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *Journal of the American Medical Association, 273,* 408–412.

Schulz, K. F., Grimes, D. A., Altman, D. G., & Hayes, R. J. (1996). Blinding and exclusions after allocation in randomised controlled trials: Survey of published parallel group trials in obstetrics and gynaecology. *British Medical Journal, 312,* 742–744.

Senn, S., & Harrell, F. (1997). On wisdom after the event. *Journal of Clinical Epidemiology, 50,* 749–751.

Sephton, S., & Spiegel, D. (2003). Circadian disruption in cancer: A neuroendocrine–immune pathway from stress to disease? *Brain, Behavior, & Immunity, 17,* 321–328.

Shrock, D., Palmer, R. F., & Taylor, B. (1999). Effects of a psychosocial intervention on survival among patients with Stage I breast and prostate cancer: A matched case-control study. *Alternative Therapies in Health and Medicine, 5,* 49–55.

Simon, R. (1994). Problems of multiplicity in clinical trials. *Journal of Statistical Planning and Inference, 42,* 209–221.

Sinclair, J. C., & Bracken, M. B. (1994). Clinically useful measures of effect in binary analyses of randomized trials. *Journal of Clinical Epidemiology, 47,* 881–890.

Smedslund, G., & Ringdal, G. I. (2004). Meta-analysis of the effects of psychosocial interventions on survival time in cancer patients. *Journal of Psychosomatic Research, 57,* 123–131.

Smith, G. D., & Egger, M. (1998). Meta-analysis: Unresolved issues and future developments. *British Medical Journal, 316,* 221–225.

Soares, H. P., Daniels, S., Kumar, A., Clarke, M., Scott, C., Swann, S., et al. (2004). Bad reporting does not mean bad methods for randomised trials: Observational study of randomised controlled trials performed by the Radiation Therapy Oncology Group. *British Medical Journal, 328,* 22–24.

Sontag, S. (1978). *Illness as metaphor.* New York: Farrar, Straus & Giroux.

Spiegel, D. (1991). Second thoughts on personality, stress, and disease. *Psychological Inquiry, 2,* 266–268.

Spiegel, D. (1996, February). *Living beyond limits: The role of group psychotherapy in treating cancer.* Presentation at the 53rd Annual Conference of the American Group Psychotherapy Association, San Francisco.

Spiegel, D. (2001). Mind matters: Coping and cancer progression. *Journal of Psychosomatic Research, 50,* 287–290.

Spiegel, D. (2002). Effects of psychotherapy on cancer survival. *Nature Reviews Cancer, 2,* 383–389.

Spiegel, D. (2004). Commentary on "Meta-analysis of the effects of psychosocial interventions on survival time and mortality in cancer patients," by G. Smedslund & G. I. Ringdal. *Journal of Psychosomatic Research, 57,* 133–135.

Spiegel, D., Bloom, J. R., Kramer, H. C., & Gottheil, E. (1989). Effect of treatment on the survival of patients with metastasic breast cancer. *Lancet, 2,* 888–891.

Spiegel, D., Bloom, J. R., & Yalom, I. (1981). Group support for patients with metastatic cancer. *Archives of General Psychiatry, 38,* 527–533.

Spiegel, D., & Classen, C. (2000). *Group therapy for cancer patients: A research-based handbook of psychosocial care.* New York: Basic Books.

Spiegel, D., & Giese-Davis, J. (2003). Depression and cancer: Mechanisms and disease progression. *Biological Psychiatry, 54,* 269–282.

Spiegel, D., & Giese-Davis, J. (2004). Examining the evidence that psychotherapy improves the survival of cancer patients: Reply. *Biological Psychiatry, 56,* 62–64.

Spiegel, D., Kraemer, H. C., & Bloom, J. R. (1998). A tale of two methods: Randomization versus matching trials in clinical research. *Psycho-Oncology, 7,* 371–375.

Spiegel, D., & Spira, J. (1991). *Supportive expressive group therapy: A treatment manual of psychosocial intervention for women with recurrent breast cancer.* Palo Alto, CA: Psychosocial Treatment Laboratory, Stanford University School of Medicine.

Spiegelhalter, D. J. (2004). Incorporating Bayesian ideas into health-care evaluation. *Statistical Science, 19,* 156–174.

Stefanek, M. E. (1991). Psychotherapy and cancer survival: A cautionary note [Letter]. *Psychosomatics, 32,* 237–238.

Stefanek, M., & McDonald, P. (in press). Brain, behavior and immunity in cancer. In S. Miller, D. Bowen, R. Croyle, & J. Rowland (Eds.), *Handbook of behavioral science and cancer.* Washington, DC: American Psychological Association.

Stinson, J. N., McGrath, P. J., & Yamada, J. T. (2003). Clinical trials in the *Journal of Pediatric Psychology*: Applying the CONSORT Statement. *Journal of Pediatric Psychology, 28,* 159–167.

Turner, J., Zapart, S., Pedersen, K., Rankin, N., Luxford, K., & Fletcher, J. (2005). Clinical practice guidelines for the psychosocial care of adults with cancer. *Psycho-Oncology, 14,* 159–173.

Van der Pompe, G., Duivenoorden, H. J., Antoni, M. H., Visser, A., & Heijnen, C. J. (1997). Effectiveness of a short-term group psychotherapy program on endocrine and immune function in breast cancer patients: An exploratory study. *Journal of Psychosomatic Research, 42,* 453–466.

Vogel, C. L., & Tan-Chiu, E. (2005). Trastuzumab plus chemotherapy: Convincing survival benefit or not? *Journal of Clinical Oncology, 23,* 4247–4250.

Watson, M., Haviland, J. S., Greer, S., Davidson, J., & Bliss, J. M. (1999). Influence of psychological response on survival in breast cancer: A population-based cohort study. *Lancet, 354,* 1331–1336.

Williams, R. B., & Schneiderman, N. (2002). Resolved: Psychosocial interventions can improve clinical outcomes in organic disease (pro). *Psychosomatic Medicine, 64,* 552–557.

Yusuf, S., Wittes, J., Probstfield, J., & Tyroler, H. A. (1991). Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. *Journal of the American Medical Association, 266,* 93–98.

(*Appendix follows*)

Appendix

The CONSORT Checklist

Paper section and topic	Item no.	Description	Reported on page no.
Title and abstract	1	How participants were allocated to interventions (e.g., "random allocation," "randomized," or "randomly assigned").	
Introduction			
Background	2	Scientific background and explantation of rationale.	
Methods			
Participants	3	Eligibility criteria for participants (a) and the settings and locations where the data were collected (b).	
Interventions	4	Precise details of the interventions intended for each group and how and when they were actually administered.	
Objectives	5	Specific objectives and hypotheses.	
Outcomes	6	Clearly defined primary and secondary outcome measures (a) and, when applicable, any methods used to enhance the quality of measurements (e.g., multiple observations, training of assessors) (b).	
Sample size	7	How sample size was determined (a) and, when applicable, explanation of any interim analyses and stopping rules (b).	
Randomization			
Sequence generation	8	Method used to generate the random allocation sequence (a), including details of any restriction (e.g., blocking, stratification) (b).	
Allocation concealment	9	Method used to implement the random allocation sequence (e.g., numbered containers or central telephone), clarifying whether the sequence was concealed until interventions were assigned.	
Implementation	10	Who generated the allocation sequence, who enrolled participants, and who assigned participants to their groups.	
Blinding (masking)	11	Whether or not participants, those administering the interventions, and those assessing the outcomes were blinded to group assignment (a). If done, how the success of blinding was evaluated (b).	
Statistical methods	12	Statistical methods used to compare groups for primary outcome(s) (a); Methods for additional analyses, such as subgroup analyses and adjusted analyses (b).	
Results			
Participant flow	13	Flow of participants through each stage (a diagram is strongly recommended). Specifically, for each group report the numbers of participants randomly assigned, receiving intended treatment, completing the study protocol, and analyzed for the primary outcome (a). Describe protocol deviations from study as planned, together with reasons (b).	
Recruitment	14	Dates defining the periods of recruitment and follow-up.	
Baseline data	15	Baseline demographic and clinical characteristics of each group.	
Numbers analyzed	16	Number of participants (denominator) in each group included in each analysis and whether the analysis was by "intention-to-treat." State the results in absolute numbers when feasible (e.g., 10/20, not 50%).	
Outcomes and estimation	17	For each primary and secondary outcome, a summary of results for each group, and the estimated effect size and its precision (e.g., 95% confidence interval).	
Ancillary analyses	18	Address multiplicity by reporting any other analyses performed, including subgroup analyses and adjusted analyses, indicating those pre-specified and those exploratory.	
Adverse events	19	All important adverse events or side effects in each intervention group.	
Discussion			
Interpretation	20	Interpretation of the results, taking into account study hypotheses, sources of potential bias or imprecision and the dangers associated with multiplicity of analyses and outcomes.	
Generalizability	21	Generalizability (external validity) of the trial findings.	
Overall evidence	22	General interpretation of the results in the context of current evidence.	

Note. CONSORT = Consolidated Standards of Reporting Trials.

Received March 7, 2006
Revision received August 9, 2006
Accepted August 21, 2006 ■