

Rebuttal: The Black Swan Fallacy in Evaluating Psychological Interventions for Distress in Cancer Patients

James C. Coyne, Ph.D.

Abramson Cancer Center of the University of Pennsylvania

Stephen J. Lepore, Ph.D.

Temple University

ABSTRACT

Isolated clinical trials in which psychosocial interventions to reduce distress among cancer patients appear superior to control conditions are insufficient to establish the overall efficacy of this class of interventions. We note that Andrykowski and Manne (1) depend on reviews that include nonrandomized trials in their case for the efficacy of these interventions and the one exception is a review that provides a pessimistic assessment of their efficacy. Four of the five intervention trials Manne and Andrykowski (2) cite as the best available evidence failed to provide the requisite treatment by time interaction needed to demonstrate efficacy. Even this best foot forward set of arguments fails to provide even a modest case for the efficacy of psychosocial interventions to reduce distress among typical cancer patients. We note the need to consider more carefully possible adverse effects and other costs of these interventions.

(*Ann Behav Med* 2006, 32(2):115–118)

In the late 1700s, Dr. John Latham published one of the first scientific descriptions of a black swan, effectively dispelling the widespread belief among Europeans at the time that all swans were white (3). However, the average European would probably still never encounter a black swan, and the best bet is that an as yet unseen swan will not be black because almost all swans in the world are white.

Like Latham, our colleagues have been able to find an occasional black swan. In examining reviews of the literature, Andrykowski and Manne (1) are still able to find declarations that cancer patients will benefit from psychological interventions to reduce distress. However, as studies accumulate and the quality of reviews markedly improves, sightings of such statements will become so infrequent that we might consider placing them on some sort of endangered list. Moreover, we show that Andrykowski and Manne had to abandon their own explicit commitment to a high standard of evidence to put forth their claims. Similarly, Manne and Andrykowski (2) identify what they term “our best intervention trials” (p. 99), which they claim are supportive of the efficacy of psychological interventions for distress. We show that four of the five identified best studies do

not provide evidence of the efficacy of psychological interventions and the fifth by Nezu, Nezu, Felgoise, McClure, and Houts. (4) should be considered a black swan. It would be an instance of the black swan fallacy to conclude from this single study that psychological interventions are generally efficacious. There are far too many counterexamples, and we already have shown (5) that some of the best candidates—studies from research programs funded by millions of dollars of federal grant funds and appearing in top ranked journals—turn out on close inspection to be more ordinary white swans; that is, they do not demonstrate that the interventions would be efficacious if provided to the typical cancer patient.

We first present our critique of Andrykowski and Manne’s argument, which relies on literature reviews, and then critique Manne and Andrykowski’s argument, which relies on evidence from a highly select set of recently published randomized controlled trials (RCTs). We show that, even if unintentionally, both articles perpetuate the confirmatory bias that we argue characterizes the field (5).

Andrykowski and Manne (1) announce that they judge the evidence on efficacy using the highest standards, namely, the Priority Symptom Management (PRISM) Level 1 evidence (6). Had they done this, we doubt we would disagree with their conclusions. The PRISM Level 1 evidence is based on qualitative systematic reviews or quantitative systematic reviews of multiple, well-designed, RCTs of adequate quality. However, Andrykowski and Manne draw on five reviews, and only one, Newell, Sanson-Fisher, and Savolainen (7), meets this criterion (8).

Two of the reviews considered by Andrykowski and Manne (1) are meta-analyses (9,10). The highest quality review, Newell et al. (7), considered and rejected meta-analysis as inappropriate, given the nature and quality of the available studies, stating: “Wide variations in the nature of interventions, outcome measures, length of follow up periods, and presentations of trials’ results prohibited us from using meta-analysis” (p. 561). The Devine and Westlake (9) meta-analysis had liberal inclusion criteria and included nursing and informational interventions that we would agree with Manne and Andrykowski (2) should be excluded when evaluating psychological interventions to reduce distress. Indeed, we believe that one of the crucial evaluation and policy issues for cancer patient supportive services is the choice between specialized interventions to reduce distress that require trained mental health professionals to deliver them and considerable investment of time and resources by cancer patients versus supportive and informational interventions that

Reprint Address: J. C. Coyne, Ph.D., Department of Psychiatry, University of Pennsylvania School of Medicine, 3400 Spruce Street 11 Gates, Philadelphia, PA 19104. E-mail: jcoyne@mail.med.upenn.edu

© 2006 by The Society of Behavioral Medicine.

provide patients with better access to oncology clinicians and information about the cancer experience (5). Devine and Westlake also included nonrandomized studies in their review, which is not Level 1 PRISM evidence.

Sheard and McGuire (10) employed stricter inclusion criteria than Devine and Westlake (9) and were consequently left with a much smaller pool of studies and a more negative assessment. Nonetheless, the bulk of studies considered by Sheard and McGuire had too few patients to warrant inclusion in a meta-analysis. This is an important part of our basis for agreeing with Newell et al. (7) that this literature presents a situation in which it is unwise to proceed with a meta-analysis. Meta-analyses that rely too heavily on small trials should be treated with skepticism, even when the overall effect is statistically significant (for cogent analyses of the problems of small trials, see 11–13). Further, Sheard and McGuire included nonrandomized studies in their review, again contrary to PRISM Level 1 criteria. Sheard and McGuire reported medium-sized effects of interventions on depression (.36) and anxiety (.42), but these effect sizes dropped precipitously when appropriate quality controls were applied (e.g., removal of lower quality studies, some small studies removed, and outliers). According to Sheard and McGuire, when quality controls were applied, the effect size was negligible for depression (.19) and moderate for anxiety (.36). However, we examined their Figure 1, which identifies the studies they reviewed, and we found that if one excludes 4 studies that were not randomized and an additional 13 studies that had less than 35 patients per condition, only 2 studies examining anxiety remain. The requirement of a minimum of 35 patients is an arbitrary but liberal response to the problems of small trials, notably confirmatory bias and otherwise false positive results (11–13). A study with cell sizes of 35 is adequately powered to detect only effects that are much larger than the existing literature would suggest are likely. In medicine, it is well known that positive results of meta-analyses that incorporate a considerable number of positive small trials fail to replicate in subsequent large-scale single studies (14). To address this issue, Kraemer et al. (11) proposed considering studies with 20 to 80 patients per conditions as too small.

The two adequate trials examining anxiety are Maguire, Tait, Brooke, Thomas, and Sellwood (15), which had null findings, and Greer et al. (16). Like Nezu et al. (4), Greer et al. is one of those rare black swan studies that included only cancer patients with significant clinical distress. The study thus does not share the assumptions of the preponderance of trials that, as a group, cancer patients need and benefit from a particular intervention. Furthermore, Greer et al. examined multiple outcome measures at multiple time points, and most were not significant, hardly robust evidence of efficacy.

Qualitative reviews by Trijsburg, van Knippenberg, and Rijpmma (17) and Barsevick, Swenney, Haney, and Chung (18) also included nonrandomized trials and did not exclude low-quality RCTs, inconsistent with the PRISM Level 1 criteria. Barsevick et al. integrated 48 trials, six reviews, and one treatment guideline. Obviously, covering such a large sweep of the literature involves accepting lower quality trials, and Barsevick

et al.'s inclusion of Sheard and McGuire (10) and Devine and Westlake (9) is without the caveats we have introduced. Barsevick et al. also includes the nursing and educational interventions that we agree with Manne and Andrykowski (1) should be excluded.

We are left with a single review (7) that concluded that the evidence on efficacy is inconclusive at best. All the trials examining dozens of different psychological interventions were reviewed, and few obtained significant effects on distress in more than half of the outcomes examined, hardly the “preponderance of evidence” (p. 93) that Andrykowski and Manne (1) sought. Indeed, Newell et al. (7) found that the most highly touted and popular psychological interventions, such as cognitive-behavioral therapy and individual-therapist-delivered therapy, were effective at reducing distress outcomes less than 50% of the time. Newell et al. concluded that only two kinds of interventions appeared to be effective at least 50% of the time but cautioned that these interventions had only been tested in trials once (music therapy) or twice (counseling). Moreover, we gave a close read to Newell et al.'s Table 3, which summarizes the studies that were retained for systematic review. A number of the studies have 10 or fewer patients in a treatment condition, only five have 35 or more patients per condition, and of these, only two provided positive findings for half of the outcomes considered by Newell et al. One of them evaluates music therapy, and the other is that black swan again, the previously discussed Greer et al. (15) study. Critics (see 19) have charged that Newell et al. were too harsh in their judgment of the field, but our assessment is that to complete their review, they had to accept evidence of a quality that would not be acceptable in biomedicine.

For their case for the efficacy of psychological interventions, Manne and Andrykowski (2) rely on individual RCTs, identified as the “highest quality randomized clinical trials published within the last 5 years” (p. 99). They do not reveal their search strategy or selection criteria or from what size pool of studies (and therefore the denominator for the numerator 5) they drew, but for the sake of debate, we accept that these trials as the best that can be mustered. Manne and Andrykowski apparently depended on findings of a main effect for an intervention as evidence of efficacy, which all of these trials obtained. However, Vickers (20) showed that such main effects are “uninteresting and irrelevant” and potentially misleading as indicators of the efficacy of an intervention: “What we are interested in, and why we conducted the randomized trial, is whether the change over time is different between groups. This is technically known as the ‘group by treatment interaction.’” (p. 654).

Manne and Andrykowski (2) cite a trial by Manne and colleagues (21) as one of their five high-quality trials, but this trial found no Treatment \times Time interaction despite multiple endpoints. Andersen et al. (22) is identified as another positive trial, but, here too, the Treatment \times Time interaction was nonsignificant for the primary endpoint, total mood disturbance. Examination of five other measures of mood and distress revealed an effect only for anxiety. We thus regard Manne et al. and Andersen et al. as null trials.

A third trial by Scott, Halford, and Ward (23) is modestly sized and does not report a Treatment \times Time interaction. A fourth trial by Boesen et al. (24) reported significant effects for total mood disturbance, apparently explained by changes in fatigue and vigor, but no significant effects on depression or anxiety. The study also appeared to suffer from selective retention, with 16 of the patients assigned to the intervention, but only 2 in the control group dropping out without providing follow-up data. The dropouts had twice the anxiety and depression scores as what was obtained for the total sample at baseline. This degree of selective retention precludes evaluation of this clinical trial, but it is possible that retrieval of follow-up data from the dropouts would have revealed that there were deleterious effects for the patients assigned to the intervention group. The remaining trial nominated by Manne and Andrykowski (2) is our familiar black swan, Nezu et al. (4), again, so designated because it does not presume that typical cancer patients will benefit from intensive intervention but rather only those who are distressed. Overall, despite having set out to refute our position concerning the lack of efficacy of psychological interventions, Manne and Andrykowski identify a set of studies that nicely complement those selected by Coyne et al. (5) in failing to demonstrate the value of such interventions for the typical cancer patient.

Andrykowski and Manne (1) suggest that we should not be too troubled by a lack of strong effects for psychological interventions because they are likely to do no harm. Yet cancer patients are caught up in a struggle with a life-altering and possibly life-threatening condition, and they undoubtedly consider their personal resources already taxed and their time precious. It would seem cruel or at least unfair to mislead patients into investing in a treatment that would likely prove inert. Furthermore, there has been a lack of attention to possible adverse effects of psychological interventions (25). Nonetheless, evidence exists that group therapy including patients who are obviously in pain, who are suffering recurrences, and who die in the course of the psychological treatment has a deleterious effect on the mood and well-being of the group as a whole (25,26). Clearly, we should avoid the premature conclusion that psychological interventions are uniformly benign or worth the investment required of patients, given a lack of demonstrated efficacy.

Andrykowski and Manne (1) also provide reassurance that cancer patients want and can access the kinds of interventions that are being evaluated in the RCTs. We do not revisit the contrary evidence we previously discussed (5), but we add to it. Greer et al. (15) argue that resistance to randomization did not figure into their difficulties recruiting patients to their trial but that the demands of cancer and its treatment seriously compromised patients' ability to accept treatment and receive an adequate exposure to it, with most patients assigned to the intervention group getting inadequate exposure. We also asked Dr. Jennifer Scott (personal communication, August 2005), one of the investigators whose trial is discussed in this article, about the logistical challenges involved in a study of a home-based couples intervention for cancer patients. She replied, "Difficulty recruiting couples is one major issue, and taking steps that enhance this process can be like a second project in itself. Over the

course of the project I drove 140,000 kilometers, and a further two therapists drove around 20,000 kilometers each." She noted the many features of the study that contributed to the relatively high rate of recruitment in her study: social marketing strategies to publicize the study, use of chart reviews to identify women potentially eligible for the study who could be personally approached and contacted by study staff, enlistment of oncologists in expressing enthusiasm for patients' participation in the study, and the tailoring of the home-based intervention to the schedules and preferences of the women and their husbands. These are heroic efforts, perhaps crucial to the completion of the project, but are unlikely to occur in routine community cancer care.

When we organized and prepared for this debate, we did not anticipate praising the accomplishments of psychological interventions in reducing distress, but we were not interested in burying them either. This was not intended simply as an exercise in floccinaucinihilipilification (27). However, with the hindsight provided by our extensive review of the literature, we are quite troubled by the failure of the field to produce even minimal quality evidence that it is worthwhile for patients and institutions to invest in psychological interventions. We had hoped to provoke the field to refocus its efforts and produce evidence that patients with clinically significant distress benefit from psychological interventions. We believed this was a reasonable task with an attainable goal. At this juncture, we have a profound sense of crisis and feel the need for such evidence is all the more urgent. Even though we do not give the field a failure grade of "E," we do think it has an outstanding "Incomplete" in the course of demonstrating that psychosocial interventions are worth the investment of institutional and patient resources. Our colleagues' uncritical acceptance of the confirmatory bias that is endemic in the field highlights the barriers to achieving a comprehension of the unfortunate state of the field. We hope that this debate serves to raise awareness of the enormity of the task of developing interventions that are efficacious and accessible for cancer patients who have some likelihood of benefiting from them. Furthermore, if most cancer patients do not have a clinical need for specialty quasi-mental health interventions, we must reopen the question of what behavioral scientists can best do to assist cancer patients in meeting all the demands they face as well as preserve or recover functioning in the face of the challenges posed by their predicament.

REFERENCES

- (1) Andrykowski MA, Manne SL: Are psychological interventions effective and accepted by cancer patients? I. Standards and levels of evidence. *Annals of Behavioral Medicine*. 2006, 32:93-97. [this issue]
- (2) Manne SL, Andrykowski MA: Are psychological interventions effective and accepted by cancer patients? II. Using empirically supported therapy guidelines to decide. *Annals of Behavioral Medicine*. 2006, 32:98-103. [this issue]
- (3) Oliver WRB: *New Zealand Birds* (2nd Ed.). Wellington, New Zealand: Reed, 1955.
- (4) Nezu AM, Nezu, CM, Felgoise SH, McClure KS, Houts PS: Project Genesis: Assessing the efficacy of problem-solving

- therapy for distressed adult cancer patients. *Journal of Consulting & Clinical Psychology*. 2003, 71:1036–1048.
- (5) Coyne JC, Lepore SJ, Palmer SC: Efficacy of psychosocial interventions in cancer care: Evidence is weaker than it first looks. *Annals of Behavioral Medicine*. 2006, 32:104–110. [this issue]
 - (6) Ropka ME, Spencer-Cisek P: PRISM: Priority Symptom Management Project Phase I: Assessment. *Oncology Nursing Forum*. 2001, 28:1585–1594.
 - (7) Newell SA, Sanson-Fisher RW, Savolainen NJ: Systematic review of psychological therapies for cancer patients: Overview and recommendations for future research. *Journal of the National Cancer Institute*. 2002, 94:558–584.
 - (8) Lepore SJ, Coyne JC: Psychological interventions and distress in cancer patients: A review of reviews. *Annals of Behavioral Medicine*. 2006, 32:85–92. [this issue]
 - (9) Devine EC, Westlake SK: The effects of psychoeducational care provided to adults with cancer: Meta-analysis of 116 studies. *Oncology Nursing Forum*. 1995, 22:1369–1381.
 - (10) Sheard T, Maguire P: The effect of psychological interventions on anxiety and depression in cancer patients: Results of two meta-analyses. *British Journal of Cancer*. 1999, 80:1770–1780.
 - (11) Kraemer HC, Gardner C, Brooks JO, Yesavage JA: Advantages of excluding underpowered studies in meta-analysis: Inclusionists versus exclusionists viewpoints. *Psychological Methods*. 1998, 3:23–31.
 - (12) Muncer S, Taylor S, Craigie M: Power dressing and meta-analysis: Incorporating power analysis into meta-analysis. *Journal of Advanced Nursing*. 2002, 38:274–280.
 - (13) Piantadosi S: Hazards of small clinical trials. *Journal of Clinical Oncology*. 1990, 8:1–3.
 - (14) LeLorier J, Gregoire G, Benhaddad A, Lapierre J, Derderian F: Discrepancies between meta-analysis and subsequent large randomized, controlled trials. *New England Journal of Medicine*. 1997, 337:536–542.
 - (15) Maguire P, Tait A, Brooke M, Thomas C, Sellwood R: Effect of counselling on the psychiatric morbidity associated with mastectomy. *British Medical Journal*. 1980, 281:1454–1456.
 - (16) Greer S, Moorey S, Baruch JDR: Adjuvant psychological therapy for patients with cancer: A prospective randomised trial. *British Medical Journal*. 1992, 304:675–680.
 - (17) Trijsburg RW, van Knippenberg FC, Rijpma SE: Effects of psychological treatment on cancer patients: A critical review. *Psychosomatic Medicine*. 1992, 54:489–517.
 - (18) Barsevick AM, Swenney C., Haney E, Chung E: A systematic qualitative analysis of psychoeducational interventions for depression in patients with cancer. *Oncology Nursing Forum*. 2002, 29:73–87.
 - (19) Bredart A, Cayrou S, Dolbeault S: Re: Systematic review of psychological therapies for cancer patients: overview and recommendations for future research. *Journal of the National Cancer Institute*. 2002, 94:1810–1812.
 - (20) Vickers AJ: Analysis of variance is easily misapplied in the analysis of randomized trials: A critique and discussion of alternative statistical approaches. *Psychosomatic Medicine*. 2005, 67:652–655.
 - (21) Manne SL, Ostroff JS, Winkel G: Couple-focused group intervention for women with early stage breast cancer. *Journal of Consulting & Clinical Psychology* (in press).
 - (22) Andersen BL, Farrar WB, Golden-Kreutz DM: Psychological, behavioral, and immune changes after a psychological intervention: A clinical trial. *Journal of Clinical Oncology*. 2004, 22:3570–3580.
 - (23) Scott JT, Halford K, Ward B: United we stand? The effects of a couple-coping intervention on adjustment to early stage breast or gynecological cancer. *Journal of Consulting & Clinical Psychology*. 2004, 72:1122–1135.
 - (24) Boesen EH, Ross L, Frederiksen K: Psychoeducational intervention for patients with cutaneous malignant melanoma: A replication study. *Journal of Clinical Oncology*. 2005, 23:1270–1277.
 - (25) Chow E, Tsao MN, Harth T: Does psychosocial intervention improve survival in cancer? A meta-analysis. *Palliative Medicine*. 2004, 18:25–31.
 - (26) Kissane DW, Love A, Hatton A: Effect of cognitive-existential group therapy on survival in early-stage breast cancer. *Journal of Clinical Oncology*. 2004, 22:4255–4260.
 - (27) Popay J: Moving beyond floccinaucinihilipilification: Enhancing the utility of systematic reviews. *Journal of Clinical Epidemiology*. 2005, 58:1079–1080.