

Finding What Is Not There

Unwarranted Claims of an Effect of Psychosocial Intervention on Recurrence and Survival

Michael E. Stefanek, PhD¹; Steven C. Palmer, PhD²; Brett D. Thombs, PhD^{3,4}; and James C. Coyne, PhD⁵

In a recently published randomized trial, Andersen et al. claimed to observe a reduced risk of recurrence and improved survival among women with early stage breast cancer who were assigned to a psychological intervention versus an assessment-only group. Anderson et al. concluded that "psychological intervention, as delivered and studied here, can improve survival." The current commentary challenges that conclusion on methodological and statistical grounds, noting that the study by Andersen et al. was not designed to assess survival and used methods that capitalized on chance, making it highly unlikely that their claims could be replicated. No other study designed to assess whether psychosocial intervention provides a survival benefit for cancer patients has ever demonstrated such an effect; and, currently, there is no support for the hypothesis that survival benefits can be attributed to psychosocial intervention. The authors of this commentary argue that much needs to be learned at the more basic biobehavioral level about the impact of stress or psychological factors on tumor biology before even considering whether large clinical trials are warranted. In addition, making the claim that psychological interventions improve survival is not evidence-based, is potentially harmful, and detracts from the potential significant benefits of psychological interventions related to quality of life. **Cancer 2009;000: 000-000. © 2009 American Cancer Society.**

KEY WORDS: psychological intervention, survival, recurrence, breast cancer.

In a recently published randomized trial, Andersen and colleagues claimed that patients with breast cancer who had local progression and were exposed to 39 hours of psychosocial intervention over 12 months demonstrated fewer recurrences and deaths and achieved longer recurrence-free and survival intervals over a median follow-up of 11 years compared with women who were randomized to no intervention.¹ If these claims are valid, then they contrast markedly with recent findings that *no* trial of psychosocial intervention in which survival was an a priori and primary endpoint and that was free of serious medical cointervention has demonstrated a survival effect among cancer patients.²

Assertions about recurrence-free intervals and improved survival are medical claims and should be evaluated using the same standards of evidence that are used for other medical claims. These standards include, but

Corresponding author: Michael E. Stefanek, PhD, Vice President, Behavioral Research and Director, Behavioral Research Center, American Cancer Society, 250 Williams Street, NE, Atlanta, GA 30303; Fax: (404) 728-0856; mstefanek@bellsouth.net

¹American Cancer Society, Behavioral Research Center, Atlanta, Georgia; ²LIVESTRONG Survivorship Center of Excellence, Cancer Control, and Outcomes, Abramson Cancer Center, University of Pennsylvania, Philadelphia, Pennsylvania; ³Department of Psychiatry, McGill University, Montreal, Quebec, Canada; ⁴Institute of Community and Family Psychiatry, Sir Mortimer B. Davis-Jewish General Hospital, Montreal, Quebec, Canada; ⁵Behavioral Oncology Program, Abramson Cancer Center and Department of Psychiatry, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania

See also pages 000-000, this issue.

Received: January 12, 2009; **Accepted:** March 3, 2009

Published online in Wiley InterScience (www.interscience.wiley.com)

DOI: 10.1002/cncr.24671, www.interscience.wiley.com

are not limited to, a priori identification of primary endpoints, fixed observation periods or a prespecified number of events, and prespecification of plans for data analysis. The trial presented by Andersen et al.¹ fails to meet these modest standards. Briefly, it appears that survival was not a primary endpoint, the observation period was not specified beforehand, and the time-to-event prediction that was the basis for their power analysis was too optimistic to be credible. Furthermore, the analyses presented appear to be post hoc and do not allow for a straightforward interpretation of the trial outcome.

If this trial had examined a pharmacologic or medical treatment rather than a psychosocial intervention, then the general consensus would be that the trial produced null results. There were no differences in unadjusted rates of recurrence or survival between the intervention and control groups. The differences that were claimed by Andersen et al.¹ were based on analyses that capitalized on chance and overfit the data to such an extent that the results were not interpretable. Moreover, across previous articles that reported the results from this trial,^{3,4} there was an inconsistent specification of endpoints without the identification of which results were primary, and there was selective reporting of results to build the case that the intervention was largely effective despite considerable evidence to the contrary.

Andersen et al.¹ do not report standard, unadjusted outcomes, such as a Kaplan-Meier estimate of the survival function. Their data reveal that the proportion of women experiencing a cancer recurrence did not differ significantly between the intervention condition (25.4%) and the control condition (29.2%; odds ratio, 0.83; confidence interval [CI], 0.46-1.48; $P = .525$). Moreover, there was no difference between the proportion of women who died in the intervention group (21.1%) versus the control group (26.5%; odds ratio, 0.74; CI, 0.40-1.36; $P = .332$); similar results are obtained if only those deaths caused by breast cancer are examined. Thus, the unadjusted analyses suggest that this is a null trial with respect to disease recurrence and survival.

The difference in median time to recurrence was small (6 months) given the range of time to event in the intervention group (range, 0.9-11.8 years) and the control group (range, 0.2-12.0 years). Similarly, differences in overall survival are small given the range of observation

periods reported. It is noteworthy that no rationale is given for limiting the study follow-up observation period to 7 to 13 years rather than shorter or longer periods. It is not possible to evaluate the claims of statistical significance made by Andersen et al. with respect to time to events without access to the data. However, any effects appear to be small at best given the broad range of observation times specified and the absence of effects on the rate of events.

The authors report results of multivariate analyses without reporting the straightforward, unadjusted analyses typically included in reports of randomized controlled trials. This strategy makes an intervention in what is likely a null trial appear to delay the time to recurrence and improve survival. Simply put, when straightforward analyses produce null effects, the inappropriate use of multivariate statistical analyses can “find” effects that do not really exist. This may be particularly true when assumptions are unmet or data are overfit. It is puzzling that Andersen et al. base their claims on analyses that controlled for initial group differences given the remarkable equivalence in baseline characteristics between groups. Indeed, the authors report “no significant differences between study arms in sociodemographics, disease, prognostic factors, type of surgery received, or adjuvant treatments scheduled to begin or eventually received.”¹ Regardless, 8 to 10 of these factors (as well as performance status and mood) were retained as covariates through a backward elimination process in each of the 3 main multivariate analyses. These multivariate procedures allowed the authors to claim a significant effect for intervention when none was demonstrated with more straightforward analyses among a well randomized sample.

The authors justify the use of covariates on the minimization method that was used to assign patients to groups, indicating that “adjustment should always be made for the minimization factors when analyzing data from a trial using this method.”¹ The authors of the review that was cited⁵ to support the analysis, however, presented a more nuanced discussion of the topic, noting that adjustment can lead to concerns about the validity of results. We note that Andersen et al. used 4 minimization factors, far fewer than the 8 to 10 variables they included as covariates. In addition, the authors did not find it necessary to covary these minimization factors when presenting data from the same trial in 2 earlier reports,^{3,4} begging

the question of precisely which criteria entered into the decision to adjust findings in this report, rather than other reports of the same trial, given the absence of significance for unadjusted results.

In any case, the number of covariates used by Andersen et al. was inappropriate, whatever the rationale. The number of covariates was simply too high relative to the number of events being explained (ie, recurrences, breast cancer deaths, other deaths). The general rule is that predictors should not be added to the equation if the ratio of outcome events to predictor variables is not at least 10:1.⁶ Anderson et al. greatly exceeded that ratio. The final model for recurrence-free survival included 11 predictors for 62 events: a ratio of 5.6:1. Babyak provides a clear example of how over fitting a regression model can result in large amounts of “predicted” variance even when all predictors are random numbers.⁷ Similarly, in survival models, over fitting produces coefficient estimates that are greatly biased relative to true values,⁶ leading to spurious results and making successful generalization of the findings to other samples unlikely. The procedure used by Andersen et al. of selecting covariates through analyses of their relation to the outcome and backward elimination procedures amplifies this problem. In the context of a lack of differences in unadjusted findings with groups that were well matched through randomization, Andersen et al.¹ opted to perform procedures that increase the probability of spurious results rather than specifying a small number of covariates ahead of time based on the prior literature. The result was a “statistically significant” multivariate finding that, based on what we know about the sort of exploratory analytic strategies used by Andersen et al., will fail to generalize.

In an earlier report of this trial, Andersen et al. demonstrated the exploratory nature with which this trial was designed, reporting results for a variety of psychosocial and immunologic outcomes.³ Emotional distress was assessed in terms of the 6 subscales of the Profile of Mood States plus a composite negative mood score. In addition, the Impact of Events Scale, which is used commonly as a measure of cancer-specific distress, was administered. There were at least 15 measures of immune function as well as 4 measures of social adjustment and 8 measures of health behavior, including a composite food behavior measure that summed 5 components, and measures of

smoking and exercise. Finally, there were 4 measures of adherence to chemotherapy.

In that report, only 1 of the 8 mood time \times treatment interactions was significant, and it would not have been significant had appropriate type I error controls been applied. Similarly, results from analyses of covariance were not significant for CD3, CD4, or CD8 counts, cell counts, or 6 assays of natural killer cell lysis. Significant differences were observed for each of 6 phytohemagglutinin and concanavalin A-induced dilutions; but, again, it is unlikely that statistical significance would be sustained in all analyses if appropriate adjustments for multiple comparisons of correlated data were applied. The composite measures of dietary behavior produced a significant time \times treatment interaction, as did the measure of smoking behavior, although no significant effect was reported for exercise or adherence to chemotherapy. Again, however, many if not all of these effects would have reverted to nonsignificance if appropriate Type-I error rate controls had been applied. Overall, the trial was characterized by weak, inconsistent findings with a pattern that likely would not have been predicted a priori and that are difficult to make sense of post hoc. Nonetheless, a subsequent article⁴ declared that the intervention yielded biobehavioral and health effects that were robust.

In the current study, were recurrence and survival the primary endpoints? It is clear that survival was not a primary endpoint, although it is unlikely that this would be gleaned from the title of the article. In the initial article concerning outcomes of this trial,³ it was reported that the trial tested the hypothesis that a biobehavioral intervention with multiple components would have an impact on *the incidence of and time to recurrence* in women with regional stage breast cancer, not an impact on *survival*. Was the study powered to assess these outcomes in a plausible manner? Andersen et al.¹ powered their study for the ability to detect a doubling of time to events, which would require 27 events per group for 80% power. In contrast, power analyses for a trial of trastuzumab plus adjuvant chemotherapy for operable HER-2-positive breast cancer indicated that interim analyses should be conducted after 394 events (recurrence, second primary cancer, or death before recurrence) had accumulated.⁸ Do Andersen et al.¹ believe that a psychosocial intervention consisting of a mixture of relaxation training, problem solving, and health behavior promotion should be so much more

potent that it could be tested adequately with a sample size and event accrual less than one-seventh those of the trastuzumab trial? Like other studies that examine survival outcomes in cancer patients, small samples that generate a small number of events lead to unreliable results that do not replicate.² Indeed, it seems very likely that a medical intervention for patients with breast cancer that claims to improve the time to recurrence based on 29 events within the intervention group would be viewed with healthy skepticism by the community at large.

Other studies demonstrating null or minimal results are not cited in the Andersen et al. literature review. For example, a recent, well designed study that examined the impact of psychological group interventions on early stage breast cancer⁹ reported no survival differences between the intervention and control groups. Even the source in the article by Anderson et al. supporting the role of psychological factors in cancer survival¹⁰ identified a very modest hazard ratio of 1.13 (CI, 1.05-1.21). In addition, the authors of that review reported significant heterogeneity, suggesting a strong likelihood of positive publication bias and the influence of unreported or unmeasured confounds.

What can we conclude from what has been presented in this and other reports of the Andersen et al. trial? This trial was successful in attracting and retaining a moderate sample of women with breast cancer. The randomization scheme largely was successful in equalizing baseline characteristics. Women in the intervention group were satisfied with their experience, completed homework assignments, and found the groups cohesive. There were some moderate effects from the intervention on a subset of health behaviors with some but less than robust effects on mood or immunologic function. However, the trial did not generate evidence that reasonably suggests decreased recurrence or improved survival.

So, where does this leave us? We agree that there is much work to be done investigating potential pathways that link psychological stress to tumor biology.¹¹ There is a need to understand a great deal more about the influence of stress response systems on tumor growth and progression. The endocrine and neurobiologic molecules that mediate effects of stressors on tumor biology are unknown. It is not clear which, if any, human tumor types and tumor genomes are sensitive to stress to the extent necessary to demonstrate effects that are relevant to recur-

rence and survival. These areas of investigation are important but require the cooperation of numerous disciplines performing very basic biobehavioral research. This research should be done before moving into large clinical trials. Such clinical trials involve huge investments of time, money, and professional and patient resources and, at best, are *very* premature.

Aside from a commitment to being empirically based in treatment recommendations, why should there be concern that a psychosocial intervention trial is misinterpreted as conveying recurrence and survival benefits? After all, the intervention appears to be benign and satisfying to participants. A primary reason for concern centers on patient decisions about how best to manage their disease and the expectations they have about their treatments. Andersen et al. strongly suggest that psychological intervention programs for breast cancer patients can do more than help patients to handle their stress and function more effectively.¹² Such overstatements are misleading regarding the strength of evidence and are unhelpful in assisting patients with decision making; in addition, they are disrespectful of patient time, resources, and the effort that may be expended pursuing authoritative but unwarranted medical claims.

The requirement that the designs, primary endpoints, and analytic strategies of biomedical trials be pre-registered and available for comparison with the claims made in subsequent reports of results¹³ should be extended to psychosocial trials. This is particularly the case for trials that make what amount to medical claims about physical health benefits and survival. Such registration would go far in resolving inconsistent reporting of the results of psychosocial trials, cherry-picking of outcomes, and use of analytic strategies that allow results to be misused in unwarranted claims by investigators.

Psychological interventions have much to offer to cancer patients, but statements that breast cancer patients can substantially improve their prognosis with a regimen of relaxation and basic health education are not empirically based. Indeed, such messages can lead to the blaming of patients who do not avail themselves of such interventions or who relapse or die despite them. Patients with cancer may benefit from psychological interventions, but the provision of these services should stand on the evidence that they can improve quality of life, decrease emotional distress, and assist survivors in

changing harmful health behaviors. The strategy of claiming increased survival may influence public policy and opinion, but if the claims are false, then such strategies can backfire. We reiterate our previous statement that there is no good reason to reject a priori the assumption that interventions that reduce prolonged distress or distress that functionally impairs the patient in other contexts will benefit individuals with cancer.² No evidence, however, has indicated that these benefits include a lower risk of mortality and longer survival.

Conflict of Interest Disclosures

Dr. Thombs is supported by a New Investigator Award from the Canadian Institutes of Health Research and by an Établissement de Jeunes Chercheurs award from the Fonds de la Recherche en Santé Québec.

References

1. Andersen B, Yang H, Farrar W, et al. Psychologic intervention improves survival for breast cancer patients. *Cancer*. 2008;113:3450-3458.
2. Coyne J, Stefanek M, Palmer S. Psychotherapy and survival in cancer: the conflict between hope and evidence. *Psychol Bull*. 2007;133:367-394.
3. Andersen B, Farrar W, Golden-Kreutz D, et al. Psychological, behavioral, and immune changes after a psychological intervention: a clinical trial. *J Clin Oncol*. 2004;22:3570-3580.
4. Andersen B, Shelby R, Golden-Kreutz D. RCT of a psychological intervention for patients with cancer: 1. Mechanisms of change. *J Consult Clin Psychol*. 2007;75:927-938.
5. Scott N, McPherson G, Ramsay C, Campbell M. The method of minimization for allocation to clinical trials: a review. *Control Clin Trials*. 2002;23:662-674.
6. Peduzzi P, Concato J, Feinstein A, Holford T. Importance of events per independent variable in proportional hazards regression analysis II. Accuracy and precision of regression estimates. *J Clin Epidemiol*. 1995;48:1503-1510.
7. Babyak M. What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. *Psychosom Med*. 2004;66:411-421.
8. Romond E, Perez E, Bryant J, et al. Trastuzumab plus adjuvant chemotherapy for operable HER2-positive breast cancer. *N Engl J Med*. 2005;353:1673-1684.
9. Kissane D, Love A, Hatton A, et al. Effect of cognitive-existential group therapy on survival in early stage breast cancer. *J Clin Oncol*. 2004;22:4255-4260.
10. Chida Y, Hamer M, Wardle J, Steptoe A. Do stress-related psychosocial factors contribute to cancer incidence and survival? *Nat Clin Pract Oncol*. 2008;5:466-475.
11. Antoni M, Lutgendorf S, Cole S, et al. The influence of biobehavioral factors on tumor biology: pathways and mechanisms. *Nat Rev Cancer*. 2006;6:240-248.
12. Nexcura, Inc. Psychotherapy May Improve Breast Cancer Survival. Seattle, Wash: Nexura Inc.; 2008. Available at: <http://www.cancerfacts.com>. Accessed November 17, 2008.
13. Dickersin K, Rennie D. Registering clinical trials. *JAMA*. 2003;290:516-523.