Examination of the Analytic Quality of Behavioral Health Randomized Clinical Trials

▼

Bonnie Spring

University of Illinois Chicago, Edward Hines, Jr. VA Hospital, and Northwestern University Feinberg School of Medicine

▼

Sherry Pagoto

University of Illinois Chicago, Edward Hines, Jr. VA Hospital, and University of Massachusetts Medical School

▼

Genell Knatterud

Maryland Medical Research Institute

▼

Andrea Kozak

University of Illinois Chicago, Edward Hines, Jr. VA Hospital, and Northwestern University Feinberg School of Medicine

▼

Donald Hedeker

University of Illinois Chicago

Adoption of evidence-based practice (EBP) policy has implications for clinicians and researchers alike. In fields that have already adopted EBP, evidence-based practice guidelines derive from systematic reviews of research evidence. Ultimately, such guidelines serve as tools used by practitioners. Systematic reviews of treatment efficacy and effectiveness reserve their strongest endorsements for treatments that are supported by high-quality randomized clinical trials (RCTs). It is unknown how well RCTs

WILEY **InterScience**®
DISCOVER SOMETHING GREAT

reported in behavioral science journals fare compared to quality standards set forth in fields that pioneered the evidence-based movement. We compared analytic quality features of all behavioral health RCTs ($n = 73$) published in three leading behavioral journals and two leading medical journals between January 2000 and July 2003. A behavioral health trial was operationalized as one employing a behavioral treatment modality to prevent or treat an acute or chronic physical disease or condition. Findings revealed areas of weakness in analytic aspects of the behavioral health RCTs reported in both sets of journals. Weaknesses were more pronounced in behavioral journals. The authors offer recommendations for improving the analytic quality of behavioral health RCTs to ensure that evidence about behavioral treatments is highly weighted in systematic reviews. © 2006 Wiley Periodicals, Inc. J Clin Psychol 63: 53–71, 2007.

Keywords: randomized controlled trial; statistical data interpretation; research design; systematic review; evidence-based medicine

Evidence-Based Health Care

In July 2005, the American Psychological Association (APA) approved a policy statement on evidence-based practice (APA, 2005). The statement defined evidence-based practice in psychology as "the integration of the best available research with clinical expertise in the context of patient characteristics, culture, and preferences." The definition resonates with the one originally set forth by the Institute of Medicine (IOM, 2001) and represents a joining of psychology to the larger evidence-based practice movement that includes numerous clinical disciplines. This is good news for psychology practitioners and researchers given that health care policies, including insurance reimbursement, are based increasingly upon systematic reviews of research evidence of the efficacy and effectiveness of treatments (Cookson, 2005; Fox, 2005). The continuing emphasis by psychologists on empirically supported treatments (Chambless & Ollendick, 2001) should bode well for how behavioral treatments will fare in research evaluation. Indeed, there is already considerable evidence that behavioral treatments reduce disease risk and chronic disease morbidity, while helping to contain medical costs (Chiles, Lambert, & Hatcher, 1999; Graves & Miller, 2003; Sobel, 1995). Systematic documentation of that evidence will further increase the credibility and coverage of behavioral treatments.

A systematic review involves use of a thorough, predefined search strategy to identify, critically appraise, and consolidate all research that meets pre-determined eligibility criteria. The systematic review is the tool used to synthesize evidence in evidenced-based medicine. For reviews of treatment, the search protocol usually specifies the intervention under consideration, relevant comparison conditions, population and outcomes of interest, and types of study. If studies are similar enough and of high enough quality, data can be combined and a treatment effect estimated statistically by meta-analysis. The optimal research design depends on the study question. In a systematic review of treatment efficacy or effectiveness, randomized controlled trials (RCTs) are usually considered to represent the gold standard of evidence because the RCT design minimizes threats to internal validity. Thus, the RCT is a well-controlled experiment that allows observed changes in a clinical outcome (the dependent variable) to be attributed to the treatment (the independent variable) and not to other causes (Campbell & Stanley, 1963). Because poor quality trials are known to estimate treatment effects unreliably (Schulz, Chalmers, Hayes, & Altman, 1995), systematic reviewers evaluate RCT quality using any of a number of

quality scoring systems (Moher et al., 1995; Verhagen, de Vet, de Bie, Boers, & van den Brandt, 2001). Findings from RCTs that are of poor quality (i.e., incompletely reported, poorly executed, or inadequately analyzed) can be excluded entirely from synthesis or weighted less heavily than evidence of higher quality. Such downgrading greatly reduces an RCT's potential impact on health policy and practice.

There is currently great interest and opportunity to perform systematic reviews of behavioral interventions for entities such as the new international Cochrane Collaboration Behavioral Medicine Field (Cochrane Collaboration, 2006) and the U.S. Preventive Services Task Force (Agency for Healthcare Research and Quality, 2006). It is important to determine to what degree our evidence about behavioral treatments meets quality standards routinely used by these groups.

Study quality has been defined as "the likelihood that the trial design generates unbiased results that are sufficiently precise and allow application in clinical practice" (Verhagen et al., 1998; p. 651). Recently, there has been increased demand to take RCT quality ratings into account quantitatively in systematic reviews (de Craen, van Vliet, & Helmerhorst, 2005). Of the three parameters of trial quality (internal validity, external validity, and analytic quality) (Verhagen et al., 2001), internal validity has received greatest attention. The most widely used quality scale, the Jadad scale, appraises only internal validity (Jadad et al., 1996). One of its three rated aspects (double-blinding of treatment assignment) is more challenging to establish in behavioral clinical trials than in drug trials, although blinding in the evaluation of the primary outcome can usually be implemented in behavioral trials. External validity has been of particular concern in the public health arena and has been addressed by a detailed quality rating system called *RE-AIM*. RE-AIM is steadily being adopted in behavioral science and public health effectiveness research (Glasgow, Davidson, Dobkin, Ockene, & Spring, 2006; Spring et al., 2005), but has not yet been widely adopted into systematic evidence reviews.

Relatively neglected but very important to valid interpretation of both behavioral and medical RCTs is analytic quality. Analytic quality refers to the validity of the procedures used in design and analysis of trial data. In this study, for a variety of reasons, we chose not to apply any single quality rating scale. At least 25 different scales are in use to rate clinical trial quality (De Craen et al., 2005; Juni, Witschi, Bloch, & Egger, 1999), and these differ in the emphasis they place on analytic quality. Because the use of composite quality summary scores has been shown to be problematic, consideration of individual quality elements is recommended instead (De Craen et al., 2005; Juni et al., 1999). Therefore, we appraised individual analytic elements that are considered important in major textbooks on clinical trial methodology (Friedman, Furberg, & DeMets, 1999) and included in various quality rating scales. Preserving randomization by employing the intent to treat policy (i.e., that all randomized cases were included in the analyses) (Hollis & Campbell, 1999), clearly identifying a single primary outcome, clearly reporting the denominator used in analyses, documenting loss to follow-up and missing data, and using procedures that account for missing data are central features of analytic quality. Other analytic elements include analyses of study-size estimation, treatment adherence, and treatment fidelity. Such features are considered important by experts in clinical trial assessment because they reduce biased conclusions generated by the analytic approach (Verhagen et al., 1998). Consequently, these were our focus in the present study.

The purpose of the present study was to assess aspects of the analytic quality of behavioral health RCTs reported in leading psychology journals. A behavioral health trial was defined as one employing a behavioral treatment modality to prevent or treat an acute or chronic physical disease or condition. We compared analytic features of behavioral health RCTs reported in three leading psychology journals to that reported in two

leading medical journals. Exemplary medical journals were selected as the comparison because they pioneered the EBP movement, have high impact ratings, represent a standard to strive towards, and provide a measuring stick to evaluate our progress toward publishing evidence that is likely to be highly weighted in systematic reviews. Behavioral health interventions were selected because such trials are reported in both psychology and medical journals, offering an opportunity for direct comparison. Although we expected the analytic quality of RCTs published in leading psychology journals to be less than that of comparable RCTs published in leading medical journals, we performed the comparison to identify among 14 criteria, specific weaknesses (and strengths) in the analytic quality of reported behavioral health RCTs. We hoped that doing so would aid psychology researchers, editors, and reviewers in the process of refining the quality of our evidence.

Methods

Inclusion criteria

All studies were required to meet criteria for a phase II or III randomized clinical trial (RCT). Those criteria include use of (a) randomization procedures to determine treatment assignment, (b) test of an intervention in humans against a standard treatment or control group, and (c) targeting clinical endpoints (Dickersin, Scherer, & Lefebvre, 1994). Because group randomized designs present unique analytic challenges, we included only those RCTs that randomized at the level of the individual, excluding those that randomized by group (e.g., family, site, school). In addition, the tested intervention was required to have employed a behavioral treatment modality. Further, to qualify as a behavioral health RCT, either the intervention or the primary endpoint needed to pertain to the prevention or treatment of an acute or chronic disease or condition. Psychological disorders (e.g., mood disorders, schizophrenia) and substance abuse were included only in so far as they contributed to medical disorders as an endpoint. Primary endpoints could be either biological or behavioral and the target population could be either healthy or ill.

Selection Procedures

We selected two leading medical journals (*Journal of the American Medical Association, New England Journal of Medicine*) and three leading psychological journals (*Journal of Consulting and Clinical Psychology, Annals of Behavioral Medicine, Health Psychology*). These journals were considered to represent the highest impact medical and psychology journals, respectively, that routinely publish clinical trials of behavioral health interventions. We manually searched every issue of each journal from January 2000 to June 2003 for behavioral health RCTs meeting the above-noted criteria. Eighty-five articles initially met inclusion criteria. During coding (see Appendix A), 12 articles were found ineligible and subsequently excluded, resulting in 73 articles reviewed. A complete list of reviewed articles appears in Appendix B.

Development of Coding Criteria

Four of the authors developed the rating system for each analytic quality element by coding the same 10 papers through progressive drafts of a coding scheme. On weekly conference calls, the authors discussed their coding discrepancies and identified ambiguous items. Features that yielded discrepant ratings were either eliminated or reworded to clarify their meaning. Each time an element's rating description was revised, all 10 arti-

Table 1
Analytic Quality Elements

Items	Possible responses
Was a rationale given for study size and/or power analysis?	Yes / No
Was the primary outcome specified?	Yes / No
What percent of randomized sample retained (complete outcome data)?	0–100%
Was the denominator at baseline reported?	Yes / No
Was drop-out on primary outcome reported?	Yes / No
Was drop-out less than or equal to 10%?	Yes / No
Was the denominator for primary analyses reported?	Yes / No
Were ITT procedures declared as analytic approach?	Yes / No
Were all randomized participants included in the analyses?	Yes / No
Who was included in the analyses of primary outcome?	1. All randomized
	2. Omits only those who were randomized but ineligible or did not start treatment
	3. Those who have data at all timepoints
	4. Those who have data for at least one time point
Did analyses account for missing data via analytic procedures or imputation?	Yes / No
Were data analyzed using both ITT and some other form of analyses (e.g., completer sample)?	Yes / No
Was treatment receipt reported?	Yes / No
Was treatment adherence reported?	Yes / No

cles were recoded. A feature's coding definition was considered sufficiently clear when its intercoder agreement across all 10 articles reached 90%. The final roster of analytic elements included the 14 items shown in Table 1. Twelve dichotomous items required a judgment of present or absent (or not applicable): study size estimation, identification of a primary endpoint, report of number of subjects at various points in the protocol, application of the intent to treat (ITT) principle in primary analyses, method of addressing missing data, report of ITT plus additional per protocol analysis of patients who completed treatment, and provision of information on treatment receipt and adherence. Two elements (retention and inclusion in primary analyses) required selection from multiple response options.

Coding Phase

The coding phase of the study occurred from August 2003 to June 2004. Five doctorally trained individuals (three clinical psychologists and two biostatisticians) coded 73 articles. Two of the coders were responsible for selecting eligible articles and distributing them to coding dyads with coding forms attached. As in other reviews of quality reporting (Berlin, 1997; Moher, Jones, & Lepage, 2001), it was not deemed necessary to mask the articles. Each article was reviewed and coded by two people, using all possible combinations of pairs of raters. To minimize drift, each dyad coded 2–4 articles before alternating coders. Average intercoder agreement across the 73 articles was 85% prior to resolving discrepant ratings. After initial ratings were made and submitted independently, one member of the coding dyad identified coding discrepancies. All discrepancies were resolved in person or by telephone discussion. When agreement on a discrepancy could

not be achieved within a coding dyad, the item was referred to the full authorship team to attain consensus. Monthly conference calls were held to resolve coding discrepancies and to discuss difficult or ambiguous articles. Once agreement was achieved for all elements rated for an article, a final consensus coding sheet was submitted to a data manager who entered it into a database.

Analytic Plan

The proportion of articles achieving criterion for each element of analytic quality was computed for each journal and journal type (medical or behavioral). All 12 dichotomously rated features were analyzed using the Cochran-Mantel-Haenszel test to examine the overall effect of journal type on analytic quality. Fisher's exact tests were then performed to test the effect of journal type on the individual dichotomously rated quality elements. Categorical variables were analyzed using Pearson chi square tests.

Results

Table 2 presents descriptive information about the articles coded as reporting RCTs of behavioral health interventions. A majority (68%) of the 73 trials appeared in psychology journals: 17 (23%) in the *Journal of Consulting and Clinical Psychology*, 14 (19%) in *Annals of Behavioral Medicine*, and 19 (26%) in *Health Psychology*. The remaining behavioral RCTs were published in medical journals: 15 (21%) in the Journal of the American Medical Association, 8 (11%) in the New England Journal of Medicine.

Dichotomous Quality Criteria

The Cochran-Mantel-Haenszel test revealed a significant overall effect of journal type on the 12 dichotomously reported elements of analytic quality, $\chi^2(1) = 25.27$, $p < .0001$. Taken collectively, the analytic quality features were reported more frequently in RCTs from leading medical journals than those from leading psychology journals. To identify

Table 2
Analytic Quality Elements by Journal

Journal	JAMA	NEJM	JCCP	ABM	HP	Medical	Psychology
# Articles	15	8	17	14	19	23	50
Study size rationale (%)	80	63	6	21	16	74	14
Define primary outcome (%)	93	100	59	79	42	95	58
Report missing baseline data (%)	83	75	71	67	50	80	62
Report drop-out (%)	100	100	88	100	89	100	92
Less than 10% drop-out (%)	47	38	23	21	32	44	26
Denominator reported (%)	100	100	71	64	79	100	72
ITT declared (%)	73	75	41	28	26	74	32
Analyze all randomized? (%)	53	63	41	28	32	57	34
Accounted for missing data? (%)	73	88	59	43	47	78	50
Used ITT and complete case (%)	47	25	12	21	11	39	14
Receipt (%)	58	28	23	33	71	47	44
Adherence (%)	75	71	92	67	71	73	77

Note. JAMA = Journal of the American Medical Association, NEJM = New England Journal of Medicine, JCCP = Journal of Consulting and Clinical Psychology, ABM = Annals of Behavioral Medicine, HP = Health Psychology, ITT = intent to treat.

Table 3
Fisher's Exact Tests of Association Between Journal
Type and Dichotomous Dependent Variables

Dependent variable	*p*-Value
Study size estimation	.0000
Define primary outcome	.0008
Denominator at baseline	.2179
Drop-out reported	.2989
Greater than 10% drop-out	.2826
Denominator of analyses	.0033
ITT declared	.0011
All randomized participants included in analyses	.0796
Accounted for missing data	.0392
Used ITT plus complete case	.0303
Treatment receipt	.2823
Treatment adherence	.4542

Note. ITT = Intent to treat.

specific areas of weakness, two-sided Fisher's exact tests performed on each rating revealed that RCTs published in leading medical journals surpassed those published in leading psychology journals on reporting study size estimation procedures ($p < .001$), specifying a primary outcome ($p < .001$), reporting the denominators used in primary outcome analyses ($p < .01$), declaring that primary analyses were performed on an ITT basis ($p < .01$), accounting for all missing data in analyses ($p < .05$), and analyzing data via both ITT plus per protocol approaches ($p < .05$) (see Table 3). In addition, behavioral health RCTs reported in leading medical as compared to psychology journals were marginally more likely to include all randomized cases in analyses ($p = .07$). Adjusting these individual tests for multiple comparisons using the Bonferroni-Holm method (Holm, 1979), maintained the significance of the above comparisons with *p*-values less than .01 (i.e., the first four listed comparisons). Thus, for one third of the elements of analytic quality, RCTs published in top-tier medical journals significantly surpassed those published in top-tier psychology journals.

Intent to Treat Policy

The analyses of 57% of behavioral health RCTs published in medical journals included all randomized participants (i.e., intent to treat), as compared to 34% of those published in psychology journals. Among RCTs that did not use ITT ($n = 43$), medical and psychology journals did not differ in the type of alternative analyses performed, $\chi^2(3) = 4.56$, $p = .20$. Per protocol or complete case analysis (retaining only those cases with data at all measured time points) was the most common alternative to ITT in psychology (28%) and medical journals (17%; Table 4).

Accuracy of Intent to Treat Declaration

Although authors declared having used ITT in a number of studies, ITT was not always implemented even when so declared. A Pearson chi square test demonstrated an association between journal type (medical vs. psychology) and declaring versus actually implementing ITT, $\chi^2(3) = 12.37$, $p < .01$, $N = 73$ (Table 5). The proportion of behavioral

Table 4

Participants Included in Analyses by Journal Type

Who's Included?	Medical *n* = 23 *n* (%)	Psychology *n* = 50 *n* (%)
Everyone randomized	13 (57)	17 (34)
Not everyone randomized	10 (43)	33 (66)
Only omits those who were randomized but ineligible or did not start treatment	3 (13)	2 (4)
Only includes those with a single data point or baseline data	1 (4)	6 (12)
Only includes those with data until a key point in treatment	2 (9)	11 (22)
Only includes those with complete data at all time points	4 (17)	14 (28)

health RCTs that both declared using ITT and actually analyzed all randomized cases was greater in medical journals (48%) than psychology journals (24%). On the other hand, reports of RCTs in medical journals (26%) were more likely than those in psychology journals (8%) to declare that they had used ITT when they had not actually done so (i.e., not included all randomized cases in the analyses).

Table 6 characterizes the procedures employed in the 33 studies that declared having implemented ITT. Only 23 of the 33 studies that declared implementing ITT actually analyzed all randomized cases. Among those studies in which ITT was declared, the difference between psychology versus medical journals in the likelihood of actually analyzing all randomized cases was not significant, $\chi^2(1) = .41$, $p = .52$. Among RCT reports that declared having implemented ITT analyses, 75% of those in psychology journals and 65% of those in medical journals actually analyzed all randomized cases.

Missing Data Procedures

Considering all RCTs (*n* = 30) that actually implemented ITT analyses (regardless of declaration), there was a marginally significant difference in the manner in which medical and psychology journals handled missing data, $\chi^2(5) = 9.76$, $p = .08$. Very similar minorities of behavioral health RCTs published in both kinds of journals (23–24%) imple-

Table 5

*Percentage of Behavioral Health RCTs (N = 73) That Declared and
Implemented ITT as a Function of Journal Type*

	Medical *n* = 23 *n* (%)	Psychology *n* = 50 *n* (%)
Declared and implemented ITT	11 (48)	12 (24)
Declared but did not implement ITT	6 (26)	4 (8)
Implemented ITT, but did not declare ITT	2 (9)	5 (10)
Neither declared nor implemented ITT	4 (17)	29 (58)

Note. ITT = Intent to treat.

Table 6

Characteristics of Randomized Clinical Trials (N = 33) That Declared Intent to Treat (ITT) Analyses

	Medical (n = 17) (%)	Psychology (n = 16) (%)
Analyzed all randomized cases	11 (65)	12 (75)
Used multiple analytic strategies (ITT + other)	7 (41)	4 (25)
Omitted only those randomized who were ineligible	3 (18)	0 (0)
Omitted only those randomized but who did not start treatment	0 (0)	1 (6)
Had >10% drop out	8 (47)	12 (75)
Handling of missing data:		
Had no missing data	0 (0)	1 (6)
Did not impute missing data	3 (18)	1 (6)
Used statistical technique that handled missing data	6 (35)	6 (38)
Used an imputation strategy for missing data:		
Deterministic (missing = dead)	2 (12)	3 (19)
Last value carried forward	5 (29)	2 (13)
Statistical imputation	0 (0)	3 (19)
Multiple imputations	1 (6)	0 (0)

mented statistical analyses that allow for incomplete data across time (e.g., survival analysis, generalized estimating equation (GEE); see Table 7). The proportion of articles using deterministic imputation that substitutes a single outcome (e.g., dead, relapsed) for all missing data was also similar (15–18%) in both journal types. However, two other simplistic imputation strategies (last value carried forward and mean substitution) were used to impute missing data for more than half of RCTs published in medical journals (54%) versus 11% of those in psychology journals. In contrast, regression or maximum likelihood analyses were used to impute missing values for 23% of behavioral health RCTs in psychology journals versus 0% in medical journals.

Discussion

The present findings revealed areas of weakness in elements of the analytic quality of behavioral health RCTs reported in both leading medical and behavioral journals. Weak-

Table 7

Missing Data Strategies Used for Randomized Clinical Trials (n = 30) That Used ITT Analyses (i.e., Analyzed All Randomized Cases) Regardless of Declaration

	Medical (n = 13) (%)	Psychology (n = 17) (%)	All (n = 30) (%)
No missing data	0 (0)	3 (18)	3 (10)
Statistical technique that allowed missing data	3 (23)	4 (24)	7 (23)
Deterministic imputation (e.g., missing = dead, missing = relapse)	2 (15)	3 (18)	5 (17)
Last value carried forward/mean imputation	7 (54)	2 (12)	9 (30)
Estimate based on regression/maximum likelihood estimation	0 (0)	4 (24)	4 (13)
Multiple strategies	1 (8)	1 (6)	2 (7)

nesses were more pronounced in behavioral journals. Specifically, reports in top psychology journals were less likely than those in top medical journals to specify a primary outcome, provide a rationale for estimating study size, state the denominators that were used in the analysis of the primary outcomes, declare using ITT analyses, account for missing data in analyses, and report both ITT and per protocol analyses. Only the first four (33% of 12 dichotomously scored criteria reviewed) remained statistically significant after controlling for multiple comparisons.

Several aspects of analytic quality were comparable for RCTs published in leading medical and psychology journals. On the criteria of reporting analytic denominators at baseline, clear reporting of drop-out, having less than 10% drop-out, and reporting of treatment receipt and adherence, there were no differences between behavioral and medical journals. However, although comparable across journal type, analytic quality was not always high for these elements. For example, top psychology and medical journals were comparably likely to clearly report the number of drop-outs. However, less than half of behavioral health RCTs published in both behavioral (26%) and medical (44%) journals reported drop-out less than 10%. High drop-out rates can severely limit study conclusions. Also, although treatment adherence was frequently reported in psychology (73%) and medical journals (73%), treatment fidelity was infrequently reported (38% and 47%, respectively). Reporting of analytic denominators at baseline was high across both psychology (72%) and medical journals (100%).

Even though we viewed these top-tier medical journals as a benchmark, their reports of behavioral health RCTs showed considerable room for improvement, and psychology can certainly learn from their shortcomings. Only 57% of behavioral trial reports in medical journals could be coded as actually having implemented ITT analyses, even though 74% claimed to have done so. The proportion of behavioral journal reports claiming to have implemented ITT was smaller (32%), but was more consistent with the proportion that had actually implemented ITT (34%). As an approach to imputing missing data, RCTs reported in medical journals relied heavily (54%) on two simplistic imputation strategies (last value carried forward and mean substitution). Comparable RCT reports in psychology journals used more sophisticated statistical approaches to imputation (e.g., maximal likelihood estimation) at least some of the time (23%). In contrast, these statistically sophisticated imputation approaches were completely absent from the reports in medical journals. (For an informed, accessible description of a variety of imputation approaches, see Schafer & Graham, 2002).

It can be hoped that implementation of the consolidated standards of reporting trials (CONSORT) statement by behavioral science journals will positively influence the quality of RCTs published in those journals in upcoming years (Altman et al., 2001). The CONSORT guidelines offer a checklist of items to be described when reporting a clinical trial, with the aim of achieving standardization and transparency of trial reporting. CONSORT was endorsed in 1996 by the editorial boards of most major medical journals, and more recently by some of the major journals covering behavioral and psychological health treatments: 2003 for *Health Psychology* (Stone, 2003), 2004 for *Annals of Behavioral Medicine* (Kaplan, Trudeau, & Davidson, 2004), and 2005 for *Journal of Consulting and Clinical Psychology* (La Greca, 2005). The introduction of CONSORT has been shown to improve although not completely eliminate variability in the reporting of clinical trials (Moher et al., 2001).

Routinely reporting comprehensive trial quality information, as required by CONSORT, may educate and sensitize researchers to important features of trial quality, secondarily improving the actual design, implementation, and analysis of clinical trials. It must be noted, though, that CONSORT addresses the quality of RCT reporting, rather

than directly addressing the quality of the conduct of clinical trials. It has been demonstrated that reporting quality and methodological quality are distinct dimensions (Huwiler-Muntener, Juni, Junker, & Egger, 2002). Well-conducted, well-analyzed trials can be reported poorly; similar quality of reporting can mask important differences in methodological quality. Implementation of CONSORT provides no guarantee that validity criteria will be correctly understood or accurately reported. In the present study, for example, 26% of behavioral health RCTs published in medical journals stated that they had performed ITT analyses when they had not actually done so. The CONSORT criteria would credit those papers for reporting whether they performed ITT analyses (even though their reporting was inaccurate). Methodological quality criteria would, conversely, offer no credit for the analytic element because ITT analyses were not performed. Ironically, another 9% of RCT reports that warrant methodological credit for performing ITT analyses, would not receive credit for the ITT CONSORT reporting item. The authors of those articles in fact conducted and reported analyses of all randomized cases, but they did not explicitly declare whether or not they performed ITT analyses.

Improving the analytic quality of behavioral RCTs goes beyond adopting CONSORT. Attaining the goal will likely require some improvements in the approach to training behavioral, allied health, and medical scientists and practitioners. Each discipline's approach to research training has different strengths and limitations. For example, psychology's strong emphasis on statistical training was evident from the greater quantitative sophistication of the missing data imputation approaches implemented in psychology versus medical journals. On the other hand, psychology graduate training programs seldom provide specific coverage of clinical trial methodology, despite strong offerings in research design and statistics. Our experience, mirrored by the present findings for behavioral science journals, is that few psychology trainees learn to apply the principle of intent-to-treat or master how to account for all missing data when analyzing results of a clinical trial. The low rate of use of ITT procedures in psychology journals is of concern because failing to preserve the randomization in a trial's analysis can introduce serious bias.

The present study also revealed a different flaw that may point to too cursory a coverage of trial methodology in medical training. The concept of ITT was often employed incorrectly in leading medical journals. Where ITT was declared but not employed, authors used an incomplete definition of ITT that referred only to analyzing cases by their original group assignment. That definition omits the most important aspect of ITT: including all randomized cases in the analyses, which requires using a valid approach to account for all missing data. The aim of educating individuals about EBP is to train professionals who can produce as well as consume and synthesize evidence about the efficacy and effectiveness of treatments. To generate such graduates, the training of scientist–practitioners in the behavioral, allied health, and medical sciences needs to afford education about the basic tools of EBP: research design and analysis, methodological and reporting quality, systematic evidence reviews, and clinical decision making.

Our study had some limitations. First, the research design was inherently correlational and intentionally descriptive. Second, we focused on only one aspect of methodological quality (analytic quality). As noted, we emphasized elements of analytic quality (rather than internal or external validity) because we think analytic quality has been relatively neglected and warrants attention. Future research might examine different aspects of the internal and external validity of RCTs in psychology journals to identify additional areas of strength and weakness. Third, the number of articles we examined and the time period sampled were necessarily limited. A larger sample of articles taken from psychology journals several years hence might reveal quality improvements. We plan to perform such a reassessment in the future to determine which aspects of quality have improved.

Finally, the RCTs we examined concerned behavioral interventions and outcomes in physical health because those trials permitted the most direct comparison between medical and psychology journals. Perhaps if mental health had been the outcome, the analytic quality of RCTs reported in psychology journals might have been superior because of the longer history of studying that content area in psychology.

Despite skepticism about whether the research evidence base is adequately developed or sufficiently relevant to guide policy decisions, research has increasingly become a basis for health care policy-making (Cookson, 2005; Fox, 2005). The Agency for Healthcare Research and Quality contracts with the Evidence-Based Practice Centers (EPC) to produce systematic evidence reviews. Those reviews guide the clinical practice recommendations of the U.S. Preventive Services Task Force, as well as reimbursement decisions by the Center for Medicare and Medicaid Services. Numerous, online continually updated systematic evidence reviews, like those produced by the International Cochrane Collaboration (Cochrane Collaboration, 2006), are being used to support clinical decision-making at the point of care delivery (Akl et al., 2004; Moyer, 2004). Practitioners of evidence-based behavioral health can use these resources efficiently to gain up-to-date-evidence about treatments that do and do not work. Scientists, in turn, can contribute to an evidence base that is wanted and needed to guide practice and policy.

The movement toward evidence-based health policy means that behavioral health treatments that are supported by a strong evidence base now have significant potential to be evaluated favorably, supported, and implemented widely. As developers of the nascent science of behavioral health, we should identify and correct our own weaknesses. Setting a high but attainable standard for our research will bolster support for behavioral health interventions in today's competitive health care market.

References

Agency for Healthcare Research and Quality. (2006). Retrieved August 15, 2006, from http://www.ahrq.gov/clinic/uspstfix.htm.

Akl, E. A., Izuchukwu, I. S., El-Dika, S., Fritsche, L., Kunz, R., & Schunemann, H. J. (2004). Integrating an evidence-based medicine rotation into an internal medicine residency program. Academic Medicine, 79, 897–904.

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 Psychological Association (APA). (2005). Retrieved August 15, 2006, from www.apa.org/practice/

Begg, C., Cho, M., Eastwood, S., Horton, R., Moher, 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.

Berlin, J. A. (1997). Does blinding of readers affect the results of meta-analyses? University of Pennsylvania Meta-analysis Blinding Study Group. Lancet, 350, 185–186.

Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research. Chicago: Rand McNally.

Chambless, D. L., & Ollendick, T. H. (2001). Empirically supported psychological interventions: Controversies and evidence. Annual Review of Psychology, 52, 685–716.

Chiles, J. A., Lambert, M. J., & Hatcher, J. W. (1999). The impact of psychological interventions on medical cost offset: A meta-analytic review. Clinical Psychology: Science and Practice, 6, 204–220.

Cochrane Collaboration. (2006). Retrieved August 15, 2006, from http://www.cochrane.org and http://www.cochranebehavmed.org

Cookson, R. (2005). Evidence-based policy making in health care: What it is and what it isn't. Journal of Health Services Research & Policy, 10, 118–121.

de Craen, A. J., van Vliet, H. A., & Helmerhorst, F. M. (2005). An analysis of systematic reviews indicated low incorporation of results from clinical trial quality assessment. Journal of Clinical Epidemiology, 58, 311–313.

Dickersin, K., Scherer, R., & Lefebvre, C. (1994). Identifying relevant studies for systematic reviews. British Medical Journal, 309, 1286–1291.

Fox, D. M. (2005). Evidence of evidence-based health policy: The politics of systematic reviews in coverage decisions. Health Affairs, 24, 114–122.

Friedman, L. M., Furberg, C. D., & DeMets, D. L. (1998) Fundamentals of clinical trials (3rd ed.). New York: Springer-Verlag.

Glasgow, R. E., Davidson, K. W., Dobkin, P. L., Ockene, J. K., & Spring, B. (2006). Practical behavioral trials to advance evidence-based behavioral medicine. Annals of Behavioral Medicine, 31, 5–13.

Graves, K. D., & Miller, P. M. (2003). Behavioral medicine in the prevention and treatment of cardiovascular disease. Behavior Modification, 27, 3–25.

Hollis, S., & Campbell, F. (1999). What is meant by intention to treat analysis? Survey of published randomised controlled trials. British Medical Journal, 319, 670–674.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics, 6, 65–70.

Huwiler-Muntener, K., Juni, P., Junker, C., & Egger, M. (2002) Quality of reporting of randomized trials as a measure of methodologic quality. Journal of the American Medical Association, 287, 2801–2804.

Institute of Medicine. (2001). Crossing the quality chasm: A new health system for the 21st century. Washington, DC: National Academy Press.

Jadad, A. R., Moore, R. A., Carroll, D., Jenkinson, C., Reynolds, D. J., Gavaghan, D. J., et al. (1996). Assessing the quality of reports of randomized clinical trials: Is blinding necessary? Controlled Clinical Trials, 17, 1–12.

Juni, P., Witschi, 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.

Kaplan, R. M., Trudeau, K. J., & Davidson, K. W. (2004). New policy on reports of randomized clinical trials. Annals of Behavioral Medicine, 27, 81.

La Greca, A. M. (2005). Editorial. Journal of Consulting & Clinical Psychology, 73, 3–5.

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., Jones, A., & Lepage, L. (2001). Use of the CONSORT statement and quality of reports of randomized trials: A comparative before-and-after evaluation. Journal of the American Medical Association, 285, 1992–1995.

Moyer, V. (2004). Evidence based medicine: Is it practical? Archives of Disease in Childhood, 89, 399–400.

Mullen, P. D., & Ramirez, G. (2006). The promise and pitfalls of systematic reviews. Annual Review of Public Health, 27, 81–102.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. Psychological Methods, 7, 147–177.

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.

Sobel, D. S. (1995). Rethinking medicine: Improving health outcomes with cost-effective psychosocial interventions. Psychosomatic Medicine, 57, 234–244.

Spring, B., Pagoto, S., Kaufmann, P. G., Whitlock, E. P., Glasgow, R. E., Smith, T. W., et al. (2005).

Invitation to a dialogue between researchers and clinicians about evidence-based behavioral medicine. Annals of Behavioral Medicine, 30, 125–137.

Stone, A. (2003). Editorial: Modification of instructions to authors. Health Psychology, 22, 331.

Verhagen, A. P., de Vet, H. C., de Bie, R. A., Boers, M., & van den Brandt, P. A. (2001). The art of quality assessment of RCTs included in systematic reviews. Journal of Clinical Epidemiology, 54, 651–654.

Verhagen, A. P., de Vet, H. C., de Bie, R. A., Kessels, A. G., Boers, M., Bouter, L. M., et al. (1998). The Delphi list: A criteria list for quality assessment of randomized clinical trials for conducting systematic reviews developed by Delphi consensus. Journal of Clinical Epidemiology, 51, 1235–1241.

Appendix A

Coding Sheet

Coder: Authors: Journal:

Title: Year:

	Yes	No	NA
1. Rationale given for study size and/or power analysis?	☐	☐	☐

Notes:

2. Primary outcome specified?

(*define:*   ) *(OK if implicit (e.g., hypothesized effects on up to 3 DVs) but not if a DV laundry list)*   ☐   ☐   ☐

Check type of variable: ☐ dichotomous,   ☐ categorical,   ☐ continuous

3. Was the denominator of <u>primary outcome</u> at baseline reported?   ☐   ☐   ☐

(N/A for time to event; <u>baseline reported</u> must include all randomized)

(*Note % complete baseline:*   )

4. Do they report drop-out on primary outcome?   ☐   ☐   ☐

(*Note % retention relative to randomization:*   )

5. Intermittent missing data on primary outcome reported?   ☐   ☐   ☐

(N/A for time to event) (*Note % complete outcome:*   )

6. Denominator reported for primary analyses (beyond df)?   ☐   ☐   ☐

7. Intent-to-treat analyses (ITT) declared for primary outcome?   ☐   ☐   ☐

8. Intent-to-treat not mentioned   ☐

9. Who's included in the analysis?

☐ Everyone randomized

☐ Only omits those who were randomized but ineligible

☐ All those with any single data point included

☐ Must have baseline/pre-test data to be included

☐ Must have data until a key point later in treatment (quit date, end of treatment)

☐ Must have final outcome data to be included

☐ Only those with complete data at all time points considered.

☐ Cross-sectional analysis stratified by time.

☐ Only omits those who were randomized but did not start treatment.

10. Is there imputation?

☐ No, not done but needed

☐ No, none required, no missing data

☐ No, none required, statistical technique handles missing data (e.g., GEE, survival)

☐ Yes, deterministic imputation (e.g., missing = dead/smoking)

☐ Yes, estimated

☐ substitute treatment group mean

☐ substitute control group mean

☐ last value carried forward

☐ average of individual's values (e.g., prior/following; all repeated measurements)

☐ estimate based on regression

☐ multiple imputations

☐ other (describe)

	Yes	No	NA
11. Were both completer and ITT analyses used?	☐	☐	☐
Describe which approaches:			
12. Treatment receipt or adherence reported?	☐	☐	☐
(*indicate which*)			
Notes:			

Appendix B

Psychology Journal Articles

Health Psychology

Carey, M. P., Braaten, L. S., Maisto, S. A., Gleason, J. R., Forsyth, A. D., Durant, L. E., et al. (2000). Using information, motivational enhancement, and skills training to reduce the risk of HIV infection for low-income urban women: A second randomized clinical trial. Health Psychology, 19(1), 3–11.

Domar, A. D., Clapp, D., Slawsby, E., Kessel, B., Orav, J., & Freizinger, M. (2000). The impact of group psychological interventions on distress in infertile women. Health Psychology, 19(6), 568–575.

Cruess, D.G., Antoni, M. H., Schneiderman, N., Ironson, G., McCabe, P., Fernandez, J. B., et al. (2000). Cognitive behavioral stress management increases free testosterone and decreases psychological distress in HIV-seropositive men. Health Psychology, 19(1), 12–20.

Wroe, A. L., Salkovskis, P. M., & Rimes, K. A. (2000). The effect of nondirective questioning on women's decisions whether to undergo bone density screening: An experimental study. Health Psychology, 19(2), 181–191.

Quinlan, K. B., & McCaul, K. (2000). Matched and mismatched interventions with young adult smokers: Testing a stage theory. Health Psychology, 19(2), 165–171.

Thoits, P. A., Hohmann, A. A., Harvey, M. R., & Fletcher, B. (2000). Similar-other support for men undergoing coronary artery bypass surgery. Health Psychology, 19(3), 264–273.

Schneider, T. R., Salovey, P., Apanovitch, A., Pizarro, J., McCarthy, D., Zullo, J., et al. (2001). The effects of message framing and ethnic targeting on mammography use among low-income women. Health Psychology, 20(4), 256–266.

Cohen, L. Reducing infant immunization distress through distraction. Health Psychology, 21(2), 207–211.

Rejeski, W. J., Focht, B. C., Messier, S. P., Morgan, T., Pahor, M., & Penninx, B. (2002). Obese, older adults with knee osteoarthritis: Weight loss, exercise and quality of life. Health Psychology, 21(5), 419–426.

Perri, M. G., Anton, S. D., Durning, P. E., Ketterson, T. U., Sydeman, S. J., Berlant, N. E. et al. (2002). Adherence to exercise prescriptions: Effects of prescribing moderate versus higher levels of intensity and frequency. Health Psychology, 21(5), 452–458.

de Moor, C., Sterner, J., Hall, M., Warneke, C., Gilani, Z., Amato, R., et al. (2002). A pilot study of the effects of expressive writing on psychological and behavioral adjustment in patients enrolled in a Phase II trial of vaccine therapy for metastatic renal cell carcinoma. Health Psychology, 21(6), 615–619.

Williams, G. C., Gagne, M., Ryan, R. M., & Deci, E. L. (2002). Facilitating autonomous motivation for smoking cessation. Health Psychology, 21(1), 40–50.

Dahlquist, L. M., Pendley, J. S., Landthrip, D. S., Jones, C. L., & Steuber, C. P. (2002). Distraction intervention for preschoolers undergoing intramuscular injections and subcutaneous port access. Health Psychology, 21(1), 94–99.

Rapoff, M. A., Belmont, J., Lindsley, C., Olson, N., Morris, J., & Padur, J. (2002). Prevention of nonadherence to nonsteroidal anti-inflammatory medications for newly diagnosed patietns with juvenile rheumatoid arthritis. Health Psychology, 21(6), 620–623.

Sandgren, A. K., & McCaul, K. D. (2003). Short-term effects of telephone therapy for breast cancer patients. Health Psychology, 22(3), 310–315.

Apanovitch, M., McCarthy, D., & Salovey, P. (2003). Using message framing to motivate HIV testing among low-income, ethnic minority women. Health Psychology, 22(1), 60–67.

Wardle, J., Williamson, S., McCaffery, K., Sutton, S., Taylor, T., Edwards, R., et al. (2003). Increasing attendance at colorectal cancer screening: Testing the efficacy of a mailed, psychoeducational intervention in a community sample of older adults. Health Psychology, 22(1), 99–105.

Rejeski, J. W., Brawley, L. R., Ambrosius, W. T., Brubaker, P. H., Focht, B. C., Foy, C. G., et al. (2003). Older adults with chronic disease: Benefits of group-mediated counseling in the promotion of physically active lifestyles. Health Psychology, 22(4), 414–423.

Margolin, A., Kelly Avants, S., Warburton, L. A., Hawkins, K. A., & Shi, J. (2003). A randomized clinical trial of a manual-guided risk reduction intervention for HIV-positive injection drug users. Health Psychology, 22(2), 223–228.

Annals of Behavioral Medicine

Toobert, D. J., Glasgow, R. E., & Radcliffe, J. L. (2000). Physiologic and related behavioral outcomes from the women's lifestyle heart trial. Annals of Behavioral Medicine, 22(1), 1–16.

McAuley, E., Blissmer, B., Katula, J., Duncan, T. E., & Mihalko, S. (2000). Physical activity, self-esteem, and self-efficacy relationships in older adults: A randomized controlled trial. Annals of Behavioral Medicine, 22(2), 131–139.

Li, F., Harmer, P., McAuley, M., Duncan, T. E., Duncan, S., Chaumeton, N., et al. (2001). An evaluation of the effects of Tai Chi exercise on physical function among older adults: A randomized controlled trial. Annals of Behavioral Medicine, 23(2), 139–146.

Anderson, E. S., Winett, R., Wojcik, J. R., Winett, S. G., & Bowden, T. (2001). A computerized social cognitive intervention for nutrition behavior: Direct and mediated effects on fat, fiber, fruits, and vegetables, self-efficacy, and outcome expectations among food shoppers. Annals of Behavioral Medicine, 23(2), 88–100.

Bock, B. C., Marcus, B. H., Pinto, B. M., & Forsyth, L. H. (2001). Maintenance of physical activity following an individualized motivationally tailored intervention. Annals of Behavioral Medicine, 23(2), 79–87.

Mann, T. (2001). Effects of future writing and optimism on health behaviors in HIV-infected women. Annals of Behavioral Medicine, 23(1), 26–33.

Wilson, D., Friend, R., Teasley, N., Green, S., Reaves, I. L., & Sica, D. A. (2002). Motivational versus social cognitive interventions for promoting fruit and vegetable and physical activity in African American adolescents. Annals of Behavioral Medicine, 24(4), 310–319.

Dobkin, P. L., DaCosta, D., Joseph, L., Fortin, P. R., Edworthy, S., Barr, S., et al. (2002). Counterbalancing patient demands with evidence: Results from a Pan-Canadian randomized clinical trial of brief supportive-expressive group psychotherapy for women with systemic lupus erythematosus. Annals of Behavioral Medicine, 24(2), 88–99.

Champion, V. L., Skinner, C., Menon, U., Sashadri, R., Anzalone, D. C., Rawl, S. M., et al. (2002). Comparisons of tailored mammography interventions at two months postintervention. Annals of Behavioral Medicine, 24(3), 211–218.

Blissmer, B., & McAuley, E. (2002). Testing the requirements of stages of physical activity among adults: The comparative effectiveness of stage-matched, mis-matched, standard care, and control interventions. Annals of Behavioral Medicine, 24(3), 184–189.

Marshall, A. L., Bauman, A. E., Owen, N., Booth, M. L., Crawford, D., & Marcus, B. H. (2003). Population-based randomized controlled trial of a stage-targeted physical activity intervention. Annals of Behavioral Medicine, 25(3), 194–203.

Patterson, T. L., Shaw, W. S., & Semple, S. (2003). Reducing the sexual risk behaviors of HIV+ individuals: Outcomes of a randomized controlled trial. Annals of Behavioral Medicine, 25(2), 137–146.

Perna, F. M., Antoni, M. H., Baum, A., Gordon, P., & Schneiderman, N. (2003). Cognitive behavioral stress management effects on injury and illness among competitive athletes: A randomized clinical trial. Annals of Behavioral Medicine, 25(1), 66–73.

Napolitano, M. A., Fotheringham, M., Tate, D., Schiamanna, C., Leslie, E., Owen, N., et al. (2003). Evaluation of an internet-based physical activity intervention: A preliminary investigation. Annals of Behavioral Medicine, 25(2), 92–99.

Journal of Consulting and Clinical Psychology

Antoni, M. H., Cruess, D. G., Cruess, S., Lutkendorf, S., Kumar, M., Ironson, G., et al. (2000). Cognitive-behavioral stress management intervention effects on anxiety, 24-hour urinary norepinephrine output, and T cytotoxic/suppressor cells over time among symptomatic HIV-infected gay men. Journal of Consulting and Clinical Psychology, 68(1), 31–45.

Higgins, S. T., Wong, C. J., Badger, J. G., Ogden, D. E., & Dantona, R. L. (2000). Contingent reinforcement increases cocaine abstinence during outpatient treatment and 1-year of follow-up. Journal of Consulting and Clinical Psychology, 68(1), 64–72.

Epstein, L. H., Paluch, R., Gordy, C. C., Saelens, B. E., & Ernst, M. (2000). Problem solving in child obesity. Journal of Consulting and Clinical Psychology, 68(4), 717–724.

Currie, S. R., Wilson, K. G., Pontefract, K. J., & deLaplante, L. (2000). Cognitive behavioral treatment of insomnia secondary to chronic pain. Journal of Consulting and Clinical Psychology, 68(3), 407–416.

Strom, L., Pettersson, R., & Andersson, G. (2000). A controlled trial of self-help treatment of recurrent headache conducted over the internet. Journal of Consulting and Clinical Psychology, 68(4), 722–727.

Killen, J. D., Fortmann, S. P., Schatzburg, A. F., Hayward, C., Sussman, L., Rothman, M., et al. (2000). Nicotine patch and paroxetine for smoking cessation. Journal of Consulting and Clinical Psychology, 68(5), 883–889.

Mohr, D. C., Likosky, W., Bertagnolli, A., Goodkin, D. E., Van Der Wende, J., Dwyer, P., et al. (2000). Telephone-administered cognitive behavioral therapy for treatment of depressive symptoms in multiple sclerosis. Journal of Consulting and Clinical Psychology, 68(2), 356–361.

Brandon, T. H., Collins, B. N., Juliano, L. M., & Lazev, A. B. (2000). Preventing relapse among former smokers: A comparison of minimal interventions through telephone and mail. Journal of Consulting and Clinical Psychology, 68(1), 103–113.

Renjilian, D. A., Perri, M. G., Nezu, A. M., McKelvey, W. F., Shermer, R. L., & Anton, S. D. (2001). Individual versus group therapy for obesity: Effects of matching participants to their treatment preferences. Journal of Consulting and Clinical Psychology, 69(4), 717–721.

Becona, E., & Vazquez, F. L. (2001). Effectiveness of personalized written feedback through a mail intervention for smoking cessation: A randomized controlled trial in Spanish smokers. Journal of Consulting and Clinical Psychology, 69(1), 33–40.

Brown, R. A., Kahler, C. W., Niaura, R., Abrams, D. B., Sales, S. D., Ramsey, S. E., et al. (2001). Cognitive behavioral treatment for depression in smoking cessation. Journal of Consulting and Clinical Psychology, 69(3), 471–480.

Lichstein, K. L., Riedel, B. W., Wilson, N. M., Lester, K. W., & Aguillard, N. R. (2001). Relaxation and sleep compression for late-life insomnia: A placebo-controlled trial. Journal of Consulting and Clinical Psychology, 69(2), 227–239.

Perkins, K. A.., Marcus, M. D., Levine, M. D., D'Amico, D., Miller, A., Broge, M., et al. (2001). Cognitive-behavioral therapy to reduce weight concerns improves smoking cessation outcomes in weight-concerned women. Journal of Consulting and Clinical Psychology, 69(4), 604–613.

Smith, S. S., Jorenby, D. E., Fiore, M. C., Anderson, J. E., Mielke, M. M., Beach, K. E., et al. (2001). Strike while the iron is hot: Can stepped-care treatments resurrect relapsing smokers? Journal of Consulting and Clinical Psychology, 69(3), 429–439.

Burling, T. A., Burling, A. S., & Latini, D. (2001). A controlled smoking cessation trial for substance-dependent inpatients. Journal of Consulting and Clinical Psychology, 69(2), 295–304.

St. Lawrence, J. S., Crosby, R. A., Brasfield, T. L., & O'Bannon, R. E. (2002). Reducing STD and HIV risk behavior of substance-dependent adolescents: A randomized controlled trial. Journal of Consulting and Clinical Psychology, 70(4), 1010–1021.

Mermelstein, R., Hedeker, D., & Wong, S. (2003). Extended telephone counseling for smoking cessation: Does content matter? Journal of Consulting and Clinical Psychology, 71(3), 565–574.

Medical Journal Articles

Journal of the American Medical Association

Gallagher, R. P., Rivers, J. K., Lee, T. K., Bajdik, C. D., McLean, D. I., & Coldman, A. J. (2000). Broad spectrum sunscreen use and the development of new nevi in White children: A randomized controlled trial. Journal of the American Medical Association, 283(22), 2955–2560.

Tate, D. F., Wing, R. R., & Winett, R. (2001). Using internet technology to deliver a behavioral weight loss program. Journal of the American Medical Association, 285(9), 1172–1177.

Edinger, J. D., Wohlgemuth, W. K., Radtke, R A., Marsh, G. R., & Quillian, R. E. (2001). Cognitive behavioral therapy for treatment of chronic primary insomnia. Journal of the American Medical Association, 285(14), 1856–1864.

Kramer, M. S., Chalmers, B., Hodnett, E. D., Sevkoskaya, Z., Dzikovich, I., Shapiro, S., et al. (2001). Promotion of breastfeeding intervention trial (PROBIT): A randomized trial in the Republic of Belarus. Journal of the American Medical Association, 285(4), 413–420.

Writing Group for the Activity Counseling Trial Research Group. (2001). Effect of physical activity counseling in primary care: The Activity Counseling Trial. A randomized controlled trial. Journal of the American Medical Association, 286(6), 677–687.

Burgio, K. L., Goode, P. S., Locher, J. L., Umlauf, M. G., Roth, D. L., Richter, H. L., et al. (2002). Behavioral training with and without biofeedback in the treatment of urge incontinence in older women: A randomized controlled trial. Journal of the American Medical Association, 288(18), 2293–2299.

Jula, A., Marniemmi, J., Huuponen, R., Virtanen, A., Rostas, M., & Ronemaa, T. (2002). Effects of diet and simvastatin on serum lipids, insulin, and antioxidants in hypercholestemic men. Journal of the American Medical Association, 287(5), 598–605.

Writing Committee for the ENRICHD Investigators. (2003). Effects of treating depression and low perceived social support on clinical events after myocardial infarction. Journal of the American Medical Association, 289(23), 3106–3136.

Esposito, K., Pontillo, A., DiPalo, C., Guigliano, G., Masella, M., Marfalle, R., et al. (2003). Effect of weight loss and lifestyle changes on vascular inflammatory markers in obese women: A randomized clinical trial. Journal of the American Medical Association, 289(24), 1799–1804.

Heshka, S., Anderson, J. W., Atkinson, R. L., Greenway, F. L., Hill, J. O., Phinney, S. D., et al. (2003). Weight loss with self-help compared with a structured commercial program: A randomized trial. Journal of the American Medical Association, 289(14), 1792–1798.

Writing Group of the Premier Collaborative Research Group. (2003). Effect of comprehensive lifestyle intervention on blood pressure control: Main results of the PREMIER clinical trial. Journal of the American Medical Association, 289(16), 2083–2093.

Rask-Nissila, L., Jokinen, E., Terho, P., Tammi, A., Lapinliemu, H., Rönnemaa, T., et al. (2003). Neurological development of 5 year old children receiving a low-saturated fat, low-cholesterol diet since infancy: A randomized controlled trial. Journal of the American Medical Association, 284(8), 993–1000.

O'Malley, P. G., Feuerstein, I. M., & Taylor, A. J. (2003). Impact of electron beam tomography, with or without case management, on motivation, behavior change, and cardiovascular risk profile: A randomized controlled trial. Journal of the American Medical Association, 289(17), 2215–2223.

Donta, S. T., Clauw, D. J., Engel, C. C., Guarino, P., Peduzzi, P., Williams, D. A., et al. (2003). Cognitive behavioral therapy and aerobic exercise for Gulf War veterans. Veterans' illnesses: Randomized controlled trial. Journal of the American Medical Association, 289(11), 1396–1404.

Tate, D. F., Jackvony, E. H., & Wing, R. R. (2003). Effects of internet behavioral counseling on weight loss in adults at risk for Type 2 diabetes: A randomized trial. Journal of the American Medical Association, 289(14), 1833–1836.

New England Journal of Medicine

Schatzkin, A., Lanza, E., Corle, D., Lance, P., Iber, F., Caan, B., et al. (2000). Lack of effect of a low-fat, high-fiber diet on the recurrence of colorectal adenomas. New England Journal of Medicine, 342(16), 1149–1155.

Tuomilehto, J., Lindstrom, J., Erikson, J. G., Valle, T. T., & Hamalanien, H. (2001). Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. New England Journal of Medicine, 344(18), 1343–1350.

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(24), 19–26.

Zhu, S., Anderson, C. M., Tedeschi, G. J., Rosbrook, B., Johnson, C. E., Byrd, M., et al. (2002). Evidence of real world effectiveness of a telephone quitline for smokers. New England Journal of Medicine, 347(14), 1087–1093.

Borghi, L., Schianchi, T., Meschi, T., Guerra, A., Allegri, F., Maggiore, U., et al. (2002). Comparison of two diets for the prevention of recurrent stones in idiopathic hypercalciuria. New England Journal of Medicine, 346(2), 77–84.

Diabetes Prevention Program Research Group. (2002). Reduction in incidence of type 2 diabetes with lifestyle intervention or metformin. New England Journal of Medicine, 346(6), 393–403.

Foster, G. D., Wyatt, H. R., Hill, J. O., McGucken, B. G., Brill, C., Mohammed, B. S., et al. (2003). A randomized trial of a low-carbohydrate diet for obesity. New England Journal of Medicine, 348, 2082–2090.

Samaha, F. F., Iqbal, N., Sashadri, P., Chicano, K. L., Daily, D. A., McGrory, J., et al. (2003). A low-carbohydrate as compared with a low fat diet in severe obesity. New England Journal of Medicine, 348, 2074–2081.