A Methodological Analysis of Randomized Clinical Trials of Computer-Assisted Therapies for Psychiatric Disorders: Toward Improved Standards for an Emerging Field

Brian D. Kiluk, Ph.D., **Dawn E. Sugarman, Ph.D.**, **Charla Nich, M.S.**, **Carly J. Gibbons, Ph.D.**, **Steve Martino, Ph.D.**, **Bruce J. Rounsaville, M.D.**, and **Kathleen M. Carroll, Ph.D.**
Department of Psychiatry, Yale University School of Medicine, West Haven, Conn.

Abstract

Objective—Computer-assisted therapies offer a novel, cost-effective strategy for providing evidence-based therapies to a broad range of individuals with psychiatric disorders. However, the extent to which the growing body of randomized trials evaluating computer-assisted therapies meets current standards of methodological rigor for evidence-based interventions is not clear.

Method—A methodological analysis of randomized clinical trials of computer-assisted therapies for adult psychiatric disorders, published between January 1990 and January 2010, was conducted. Seventy-five studies that examined computer-assisted therapies for a range of axis I disorders were evaluated using a 14-item methodological quality index.

Results—Results indicated marked heterogeneity in study quality. No study met all 14 basic quality standards, and three met 13 criteria. Consistent weaknesses were noted in evaluation of treatment exposure and adherence, rates of follow-up assessment, and conformity to intention-to-treat principles. Studies utilizing weaker comparison conditions (e.g., wait-list controls) had poorer methodological quality scores and were more likely to report effects favoring the computer-assisted condition.

Conclusions—While several well-conducted studies have indicated promising results for computer-assisted therapies, this emerging field has not yet achieved a level of methodological quality equivalent to those required for other evidence-based behavioral therapies or pharmacotherapies. Adoption of more consistent standards for methodological quality in this field, with greater attention to potential adverse events, is needed before computer-assisted therapies are widely disseminated or marketed as evidence based.

A great deal of excitement has been generated by the recent introduction of computer-assisted therapies, which can deliver some or all of an intervention directly to users via the Internet or a processor-based program. Computer-assisted therapies have a number of potential advantages. They are highly accessible and may be available at any time and in a broad range of settings; they can serve as treatment extenders, freeing up clinician time and offering services to patients when clinical resources are limited; they can provide a more consistently delivered treatment; they can be individualized and tailored to the user's needs and preferences; interactive features can link users to a wide range of resources and supports; and the multimedia format of many of these therapies can convey information and concepts in a helpful and engaging manner (1–3). Computer-assisted therapies may also greatly reduce the costs of some aspects of treatment (4–6). Among their most promising

features is the potential to provide evidence-based therapies to a broader range of individuals who may benefit from them (7), given that only a fraction of those who need treatment for psychiatric disorders actually access services (8–10), while 75% of Americans have access to the Internet (11).

However, the great promise of computer-assisted approaches is predicated on their effectiveness in reducing the symptoms or problems targeted. Recent meta-analyses and systematic reviews have suggested positive effects on outcome for a range of computer-assisted therapies, particularly those for depression and anxiety (2, 12, 13), but the existing reviews also point to substantial heterogeneity in study quality (2, 14–17). For example, a recent review of e-therapy (treatment delivered via e-mail) conducted by Postel et al. (14) noted that only five of the 14 studies reviewed met minimal criteria for study quality as defined by Cochrane criteria (14, 18).

Thus, this emerging field is in many ways reminiscent of the era of psychotherapy efficacy research prior to the adoption of current methodological standards for evaluating clinical trials (19) and prior to the codification of standards for evaluating a given intervention's evidence base (e.g., specification in manuals, independent assessment of outcome, and evaluation of treatment integrity) (20, 21). Moreover, as a novel technology, there are several methodological issues that are particularly salient to the evaluation of computer-assisted therapies, such as the level of prior empirical support for the (usually clinician-delivered) parent therapy, the level of clinician/therapist involvement in the intervention, the relative credibility of comparable approaches, and whether the approach is delivered alone or as an adjunct to another form of treatment.

Given the rapidity with which computer-assisted therapies can be adopted and disseminated, it is particularly important that this emerging field not only have a sound evidence base but also demonstrate safety, since heightened awareness of potentially negative or harmful effects of psychological treatments has increased the need for more stringent evaluation prior to dissemination (22–24). Although generally considered low risk (17), there are multiple potential adverse consequences of computer-assisted therapies. These may include providing less intensive treatment than necessary to treat severely affected or symptomatic individuals, ineffective approaches that may discourage individuals from subsequently seeking needed treatment, or inappropriate interpretation of program recommendations that could lead to harm (e.g., premature detoxification in substance users, worsening of panic attacks from exposure approaches that are too rapid or intensive), particularly in the absence of clinician monitoring and oversight. Systematic evaluation of possible adverse effects of computer-assisted therapies, as well as further evaluation of the types of individuals who respond optimally to computer-assisted versus traditional clinician-delivered approaches, is needed prior to their broad dissemination.

The present systematic analysis of randomized controlled trials of computer-assisted therapies for adult axis I disorders examines both methodological quality indices established to evaluate behavioral interventions (20, 21) and standards used to evaluate trials of general healthcare interventions (18, 25). This report differs from prior meta-analyses of computer-assisted therapies (2, 13, 17) not only in our focus (detailed analysis of study quality versus estimation of aggregate effect size) but also in our development of an instrument for evaluating the quality of clinical trials testing the effectiveness of these therapies. This instrument permitted exploration of several hypotheses regarding the quality and use of specific design features. For example, we hypothesized that as an emerging field, methodological quality would increase over time. We also expected that the likelihood of finding an effect favoring the computer-assisted approach would be greater in those studies with poorer methodological quality and in those using less stringent control conditions.

Method

Study Identification and Selection

Using PubMed, Scopus, and Psych Abstracts, as well as existing meta-analyses (2, 17, 26) and systematic reviews (6, 13, 26, 27), we identified reports that 1) were randomized controlled trials that included pre- and posttreatment evaluation of the target symptom or problem, 2) used a computer to deliver a psychotherapeutic or behavioral intervention, and 3) targeted adults (aged 18 years) with an axis I disorder or problem. Studies were excluded if they 1) were focused on prevention rather than intervention for an existing disorder, 2) consisted solely of a single-session assessment and feedback (excluded for reasons of inherent differences in the evaluation of compliance and integrity with intervention delivery), 3) did not report on symptom outcome or a defined target problem (e.g., studies that assessed knowledge acquisition or evaluated treatment process only), or 4) evaluated e-therapy (e.g., interventions delivered wholly by a clinician via e-mail rather than delivered at least in part by a computer program). E-therapy interventions, recently reviewed by Postel et al. (14), were excluded because their mode of delivery and methodology for evaluation differ from other computer-assisted therapies. Randomized trials published between January 1990 and January 2010 were included. As shown in Figure 1, of the 130 articles identified in the initial literature search, 55 were excluded, yielding 75 independent trials for rating in this review.

Development of Rating System

A methodological quality rating system was developed by reviewing and integrating standards from multiple existing systems. These included first those that were developed by the American Psychological Association (Division 12) for defining empirically validated behavioral therapies (20, 21) as well as criteria used in previous methodological analyses of the behavioral therapy literature (28–31). Second, commonly used standards for evaluating general medical randomized controlled trials (i.e., not necessarily for psychiatric disorders or behavioral therapies), such as Consolidated Standards of Reporting Trials (32), Cochrane criteria (18), and others (25, 33), were incorporated. This was done in order to encompass international standards and systems, since many of the trials of computer-assisted therapies have been developed and evaluated in the United Kingdom, European Union, Australia, and New Zealand (26, 34).

The criteria most relevant to the evaluation of trials of computer- assisted therapies for psychiatric disorders were then integrated into a single rating system (computer-assisted therapy methodology rating index), presented in Table 1, which shows a comparison between this system and other methodological quality rating systems. Items commonly included in other rating systems that were not included in the rating index for the present review were 1) specification of treatments in manuals (an item evaluating the level of empirical support for the parent therapy was included instead), 2) training of clinicians (a rating of the level of clinician involvement in the computerized treatment was included instead), and 3) reporting of adverse events (excluded since this was not reported for the studies examined in the present review).

The final rating scale included two general types of items (Table 1). First, the following seven items were used to evaluate basic features of general randomized clinical trials: identification of the method of randomization, with specification of baseline comparability of groups; inclusion of a posttreatment follow-up evaluation consisting of 80% of the intention-to-treat sample; independent assessment of outcome; adequacy of sample size/ power; appropriate statistical techniques; appropriate management of missing data; and whether the study was a replication of an independent trial. Second, seven additional items

evaluated features particularly relevant to trials of behavioral and computer-assisted interventions. These were as follows: use of a clinical versus general sample (since a number of computer-assisted therapies target subclinical levels of problems) (35); extent to which the computer-assisted approach was based on an existing manualized empirically validated approach; stringency of the comparison/control condition (i.e., wait-list versus attention placebo versus active treatment, such as a clinician-delivered version of the same intervention); level of validation of the outcome measures (i.e., whether validated measures or objective assessments of outcome were used); measurement of treatment adherence (in terms of the extent to which the participants accessed the assigned protocol intervention and completed the prescribed number of sessions or modules); comparability of the intervention conditions with respect to time or attention; and whether a measure of treatment credibility was included, which refers to participant perceptions and confidence in the likely efficacy of treatment intervention, since less credible comparison conditions or those that lack compelling rationales (e.g., "sham" websites used as control conditions) undermine internal validity (36). An item was rated 0 if it did not meet the methodological quality criterion; a rating of 1 indicated that the item met the basic quality standard; and a rating of 2 indicated that the item exceeded the criterion (i.e., gold standard methodology) (Table 2). For studies that used multiple comparison conditions, ratings were made for both the most potent (e.g., active treatment) and least potent (e.g., wait-list) comparison condition.

Several other features of computer-assisted therapies were rated for each trial to evaluate the relationship with methodological quality. These were as follows: whether the format of the intervention was web-based or resided on the computer, whether the intervention was delivered as a stand-alone treatment or as an addition to another form of treatment, the level of clinician involvement in the intervention, whether a peer-support feature was included, and whether the study sponsor had a financial interest in the intervention.

Ratings and Data Analysis

Raters were three experienced doctoral researchers and one master's-level statistician. All four raters initially rated a sample of six articles and reviewed the ratings through group discussion in order to achieve consensus and refine the detailed coding manual. All trial reports were then sorted according to the primary disorders addressed in the trials (depression, anxiety, nicotine dependence, alcohol/drug dependence), and two raters independently read and rated each article. Each item on the rating form was given a score of 0, 1, or 2 based on the scoring criteria, and item scores were summed to produce a general randomized controlled trial quality score (items 1–7), a computer-assisted therapy-specific quality score (items 8–14), and an overall quality score (total items 1–14). Kappas were calculated for each item to assess interrater reliability within each pair of raters. Next, all discrepant item ratings were identified and then resolved by the rater pairs by referring to the study report for clarification until consensus was achieved. Only the final consensus ratings were used for data analysis. An additional doctoral-level rater who was independent of our research group rated the study conducted by Carroll et al. (37) in order to reduce rater bias, and the study was not included in the evaluation of inter-rater reliability.

Frequency of item scores and mean item ratings for the 14 methodological quality items were calculated for the full sample of articles as well as for each disorder/problem area. The trial reports that addressed insomnia (N=2), gambling (N=1), eating disorders (N=1), and general distress (N=2) were grouped into an "other" category because of the low number of trials within these areas. Differences in mean quality scores were compared across the topic areas using analyses of variance (ANOVAs). Because scores were similar for the two general types of items (randomized controlled trial quality versus computer-assisted therapy-specific quality), with few differences across disorder types, only the overall quality scores were used in our analyses. Finally, exploratory ANOVAs were conducted to examine the

difference in mean quality scores according to the program format (e.g., web-based versus DVD), the potency of the comparison group (e.g., wait-list versus active), the level of clinician involvement, the type of sample chosen (e.g., clinical versus general community), and whether the study sponsor reported a potential conflict of interest.

Results

Reliability of Ratings

Results of the interrater concordance (using weighted kappas) for the randomized controlled trial and computer-assisted therapy-specific quality items, as well as the overall quality items, were computed for rated articles within each disorder type (see Table 1 in the data supplement accompanying the online version of this article). The overall concordance on all items for 74/75 rated articles (excluding the Carroll et al. article [37]) was a kappa of 0.52, which is consistent with moderate levels of interrater agreement (38, 39). Across all four main disorder types, concordance for randomized controlled trial items (kappa= 0.48) was comparable to that for computer-assisted therapy-specific items (kappa=0.44). Since kappas can be artificially low when sample sizes and prevalence rates are small (40), the percent of absolute agreement between the rater pairs was also calculated (range: 75%–78% overall).

Methodological Quality Scores

Frequencies and item scores for the full sample of articles (N=75) are presented in Table 2. The mean overall quality item score was 13.6 (SD=3.6), out of a maximum score of 28. Mean scores for randomized controlled trial and computer-assisted therapy-specific items were 6.5 (SD=1.9) and 7.1 (SD=2.7), respectively. In terms of the seven randomized controlled trial criteria, the majority of studies (73%) described the randomization method used and demonstrated equivalence of groups on baseline characteristics. Nearly all trials (91%) were judged as using appropriate statistical analyses. However, a substantial proportion (40%) was rated as not adequately describing the methods used to handle missing data or as reporting use of an inappropriate method (e.g., last value carried forward) rather than an intention-to-treat analysis. Very few trials (15%) obtained follow-up data on at least 80% of the total sample, and most (75%) relied solely on participant self-report for evaluation of outcome.

In terms of the seven computer-assisted therapy-specific items, nearly all trials (95%) indicated that the computerized approach was based on an existing manualized treatment with some prior empirical support. Slightly more than one-half of the studies (52%) used standardized diagnostic criteria as an inclusion criterion. When studies on nicotine dependence were excluded, the rate of use of standardized diagnostic criteria to determine eligibility rose to 62%. Forty-one percent of the trials relied on self-report or a cutoff score to determine participant eligibility, and 7% did not identify clear criteria for determining problem or symptom severity. In terms of control conditions, 23% of studies utilized wait-list conditions only; 41% used an attention/placebo condition; and 36% used active conditions. Most studies (65%) relied solely on participant self-report on a validated measure to evaluate primary outcomes. Only 25% of studies used an independent assessment (blind ratings or biological indicator), and 9% reported use of unvalidated measures. A number of studies (27%) did not report on participant adherence with the study interventions; 49% reported some measure of adherence; and only 24% measured compliance thoroughly and considered it in the analysis of treatment outcomes. Most studies (76%) did not include a measure of credibility of the study treatments. Comparison or control conditions were not equivalent in time or attention in the majority of studies (69%).

No study met basic standards (i.e., score of ⩾1) for all 14 items. Three trials met basic standards on 13 of the 14 items (see Figure 1 in the data supplement). Results of ANOVAs examining differences in mean overall quality scores indicated no statistically significant difference across disorder areas (depression, anxiety, nicotine dependence, drug/alcohol dependence).

Table 3 summarizes additional methodological features not included in the overall quality score. The majority of trials (72%) utilized a web-based format for treatment delivery, and most (72%) were stand-alone treatments rather than add-ons to an existing or standardized treatment. A notable issue across all studies reviewed was the lack of information provided on the level of exposure to study treatments. Only 16 studies (21%) reported either the length of time involved in treatment or number of treatment sessions/modules completed. The majority of computer-assisted therapies for alcohol and drug dependence (57%) had a study sponsor with a potential conflict of interest, whereas relatively few studies focusing on anxiety disorders (13%) had study sponsors with a potential conflict of interest. Most studies (72%) reported that the theoretical base of the computer-assisted approach was cognitive behavioral.

Intervention Effectiveness and Relationship With Quality Scores

In terms of the effectiveness of computer-assisted therapies relative to control conditions, 44% of the trials reported that the computer-assisted therapy was more effective than the most potent comparison condition with regard to effect on the primary outcome. Overall, the computer-assisted therapy was found to be more effective than 88% of the wait-list comparison conditions, 65% of the placebo/attention/education conditions, and 48% of the active control conditions (χ^2=6.7, p<0.05). In those studies where the control condition did not involve a live clinician, the computer-assisted therapy was typically found to be significantly more effective than the control condition (74% of trials). However, the computer-assisted therapies were less likely to be effective than comparison therapies when the control condition included face-to-face contact with a clinician (46% of studies) (χ^2=6.56, p<0.05). There were no differences in overall effectiveness of the computer-assisted therapy relative to the control conditions across the four major disorder/problem areas. However, as shown in Figure 2, studies that reported the computer-assisted therapy to be more effective than the most potent control condition had significantly lower overall methodological quality than studies where the computer-assisted therapy was found to have efficacy that was comparable to or poorer than the control condition (F=5.0, df=2, 72, p<0.01).

Relationship Between Methodological Quality Scores and Specific Design Features

Several exploratory analyses were conducted to evaluate relationships between specific methodological features and methodological quality scores (for these comparisons, the criterion in question was removed when calculating methodological quality scores, and thus 13 rather than 14 items were assessed). These analyses indicated higher methodological quality scores for studies that 1) used an active control condition relative to attention or wait-list conditions (active: mean=14.8 [SD=2.4]; placebo: mean=11.7 [SD=2.8]; wait-list: mean=10.4 [SD=2.5]; F=17.4, df=2, 72, p<0.001); 2) evaluated computer-assisted therapies based on a manualized behavioral intervention with previous empirical support (no prior support: mean=10.5 [SD=1.3]; some support: mean=11.1 [SD=3.4]; empirical support with manual: mean=13.5 [SD=3.0]; F=6.1, df=2, 72, p<0.01); 3) included at least a moderate level of clinician involvement (>15 minutes/week) with the computer-assisted intervention (at least moderate: mean=15.6 [SD=3.7]; little: mean=12.2 [SD=3.9]; none: mean=12.9 [SD=2.8]; F=6.1, df=2, 72, p<0.01); 4) clearly defined the anticipated length of treatment and measured the level of participant exposure/adherence with the computerized

intervention (treatment defined and adherence evaluated: mean=14.7 [SD=3.1]; treatment defined but adherence not evaluated: mean=14.2 [SD=3.5]; treatment not defined and adherence not evaluated: mean=10.4 [SD=2.7]; F=7.7, df=2, 72, p<0.001); 5) utilized a clinical sample rather than a general sample (clinical: mean=14.9 [SD=3.6]; general: mean=12.4 [SD=3.2]; F=10.8, df=1, 73, p<0.01); and 6) were replications of previous trials (replication: mean=14.5 [SD=3.2]; nonreplication: mean=12.3 [SD=3.3]; F=8.7, df=1, 73, p<0.05).

Overall, quality scores were somewhat higher for studies in which the authors reported a financial interest in the computerized intervention but not significantly different from studies with no apparent conflict of interest (mean: 14.1 [SD=3.7] versus 13.5 [SD=3.6], respectively). The control condition was more likely to be potent (active) in studies with authors who did not report a financial interest relative to studies with authors who did (40% versus 20%, not significant). There were no significant differences for overall methodological quality scores with respect to the format of the intervention (web-based versus DVD/CD), whether evaluated as a stand-alone intervention or delivered as an addition to an existing treatment, nor were there significant differences with regard to the geographic region where the study was conducted (United States versus United Kingdom versus European Union versus Australia/New Zealand). Finally, in contrast to our expectations of improving the quality of research methods over time, there was no evidence of substantial change in overall quality over time. Although the number of publications in this area increased yearly, particularly after 1998 (beta=0.62, t=3.01, p<0.01), the number of criteria met tended to decrease over time (beta=−0.08, t=2.10, p=0.06).

Discussion

This review evaluated the current state of the science of research on the efficacy of computer-assisted therapies, using a methodological quality index grounded in previous systematic reviews and criteria used to determine the level of empirical support for a wide range of interventions. The mean methodological quality score for the 75 reports included was 13.6 (SD=3.6) out of a possible 28 quality points (49% of possible quality points), with comparatively little overall variability across the four major disorder/problem areas evaluated (depression, anxiety, nicotine dependence, and alcohol and illicit drug dependence). Overall, this set of studies met minimum standards on only 9.5 of the 14 quality criteria evaluated.

Taken together, these findings suggest that much of the research on this emerging treatment modality falls short of current standards for evaluating the efficacy of behavioral and pharmacologic therapies and thus point to the need for caution and careful review of any computer-assisted approach prior to rapid implementation in general clinical practice. Relative strengths of this body of literature include consistent documentation of baseline equivalence of groups and inclusion of comparatively large sample sizes, with 43% of studies reporting at least 50 participants within each treatment group. Moreover, most of the computer-assisted interventions evaluated were based on clinician-delivered approaches with some empirical support. These strengths therefore highlight the broad accessibility of computer-assisted therapies and the relative ease with which empirically based treatments can be converted to digital formats (26, 41).

This review also highlights multiple methodological weaknesses in the set of studies analyzed. One of the most striking findings was that none of the 75 trials evaluated met minimal standards on all criteria, and only three studies met 13 of the 14 criteria. Three issues were particularly prominent. First, many of the studies used comparatively weak control conditions. Seventeen trials (23%) used waitlist conditions only, all of which relied

solely on self-report for assessment of outcome, without appropriate comparison for participant expectations or multiple demand characteristics, resulting in very weak evaluations of the computer-assisted approach. These trials also had lower overall quality scores relative to those that used more stringent control conditions.

The second general weakness was a striking lack of attention to issues of internal validity. Few of the studies (21%) reported the extent to which participants were exposed to the experimental or control condition or considered the effect of attrition in the primary analyses. For many studies (40%), it was impossible to document the level (in terms of either time or proportion of sessions/components completed) of intervention received by participants. Third, only 13% of studies conducted true intention-to-treat analyses. Reliance on inappropriate methods, such as carrying forward the final observation, was common. This practice, coupled with differential attrition between conditions, likely led to biased findings in many cases (42). Finally, very few studies addressed the durability of effects of interventions via follow-up assessment of the majority of randomly assigned participants.

Some but not all of our exploratory hypotheses regarding methodological quality, specific study features, and outcomes were confirmed. For example, there was no clear evidence that the methodological quality of the field improved over time. In fact, some of the more highly rated studies in this sample were published fairly early and in high-impact journals. This suggests a rapidly growing field marked by increasing methodological heterogeneity. Second, while a unique feature of computer-assisted therapies (relative to most behavioral therapies) is that the developers may have a significant financial interest in the approach and hence results of the trial, there was no clear evidence of lower methodological quality in such trials. However, weaker control conditions tended to be used more frequently in those studies where there was a possible conflict of interest. In general, studies that included clearly defined study populations and some clinician involvement were associated with better methodological quality scores. This latter point may be consistent with meta-analytic evidence of larger effect size among studies of computer-assisted treatment that include some clinician involvement (43).

A striking finding was the association of study quality with the reported effectiveness of the intervention, with those studies of lower methodological quality associated with greater likelihood of reporting a significant main effect for the computer-assisted therapy relative to the control condition. Many weaker methodological features (e.g., reliance on self-reported outcomes, differential attrition, mishandling of missing data) are generally expected to bias results toward detecting significant treatment effects, and in this set of studies, use of wait-list rather than active control conditions was particularly likely to be associated with significant effects favoring the computer-assisted treatment. Other features, particularly those associated with reducing power (e.g., small sample size, poor adherence), generally add bias in terms of nonsignificant effects.

While there were no differences in methodological quality scores across the four types of disorders with adequate numbers of studies for review (depression, anxiety, nicotine dependence, drug/alcohol dependence), there were a number of differences across studies associated with use of specific methodological features. For example, the studies evaluating treatments for nicotine dependence were characterized by use of broad, general populations and hence tended to be web-based interventions with less direct contact with participants. Thus, these studies were also characterized by methodological features closely associated with web-based studies (44), such as comparatively little clinician involvement, large sample sizes, and reliance on self-report. In fact, this group of studies consisted of the largest sample sizes among those in our review (approximately 93% reported treatment conditions with greater than 50 participants). The studies involving interventions for anxiety

disorders were some of the earliest to appear in the literature and were characterized by a larger number of replication studies, as well as higher rates of standardized diagnostic interviews to define the study sample in addition to higher levels of clinician involvement and of attention to participant satisfaction and assessment of treatment credibility. These studies had fairly small sample sizes, with 43% reporting fewer than 20 participants per condition. The depression studies were more heterogeneous in terms of focus on clinical populations and use of standardized criteria to define study populations, yet they included relatively large sample sizes (71% contained treatment conditions with more than 50 participants). The drug/alcohol dependence literature contributed the fewest number of studies, a higher proportion of web-based intervention trials, and moderate sample sizes (71% reported treatment conditions with 20–50 participants).

Limitations of this review include a modest number of randomized clinical trials in this emerging area, since only 75 studies met our inclusion criteria. The limited number of studies may have contributed to the modest kappa values for interrater reliability, while percent agreement rates were comparatively good. Although the items on the methodological quality scale were derived by combining several existing systems for evaluating randomized trials and behavioral interventions, lending it reasonable face validity, we did not conduct rigorous psychometric analyses to evaluate its convergent or discriminant validity. Another limitation was failure to verify all information from the relevant study authors. Some of the information used for our methodological ratings may have been eliminated from the specific report prior to publication because of space limitations (e.g., description of evidence base for parent therapy) and therefore may not reflect the true methodological quality of the trial. Finally, we did not conduct a full meta-analysis of effect size, since our focus was evaluation of the methodological quality of research in computer-assisted therapy. Moreover, inclusion of trials of poor methodologic quality and heterogeneity in the rigor of control conditions (and hence likely effect size) would yield meta-analytic effect size estimates of questionable validity (45).

It should be noted that some methodological features were not included in our rating system simply because they were so rarely addressed in the studies reviewed. In particular, only one of the studies included explicitly reported on adverse events occurring during the trial. Most behavioral interventions are generally considered to be fairly low risk (46), as are computerized interventions (17). However, like any potentially effective treatment or novel technology, computer-assisted therapies also carry risks, limitations, and cautions often minimized or overlooked by their proponents (47), and hence there is a pressing need for research on both the safety and efficacy of computer-assisted therapies (24, 34, 47). Relevant adverse events should be appropriately collected and reported in clinical trials of computer-assisted interventions. Another important issue of particular significance for computer-assisted and other e-interventions is their level of security. This issue was not addressed in the majority of studies we reviewed, nor was the type or sensitivity of information collected from participants. Because no computer system is completely secure and many participants in these trials might be considered vulnerable populations, better and more standard reporting of the extent to which program developers and investigators consider and minimize risks of security breaches in this area is needed.

The high level of enthusiasm regarding adoption of computer-assisted therapies conveys an assumption that efficacy readily carries over from clinician-delivered therapies to their computerized versions. However, computer-assisted only conveys that some information or putative intervention is delivered via electronic media, but nothing is indicated about the quality or efficacy of the intervention. We maintain that computer-based interventions should be evaluated using the same rigorous testing of safety and efficacy, with methods that are requirements for establishing empirical validity of behavioral therapies prior to

dissemination. At a minimum, this should include standard features evaluated in the present review (e.g., random assignment; appropriate analyses conducted in the intention-to-treat sample; adequate follow-up assessment; use of standardized, validated, and independent assessment of outcome). In addition, data from this review suggest that particular care is needed with respect to treatment integrity, particularly evaluation and reporting of intervention exposure and participant adherence, as well as assessment of intervention credibility across conditions. Finally, the field has developed to a level where studies that use weak wait-list control conditions and rely solely on participant self-reports are not tenable, since they address few threats to internal validity and convey little regarding the efficacy of this novel strategy of treatment delivery. The poor retention rates in treatment and lack of adequate follow-up assessment provide insufficient evidence of safety and durability. Given the rapidity with which these programs can be developed and marketed to outpatient clinics, healthcare insurance providers, and individual practitioners, our findings should be viewed as a caveat emptor warning for the purchasers and consumers of such products. The vital question for this field is not "Do computer-assisted therapies work?" but "Which specific computer-assisted therapies, delivered under what conditions to which populations, exert effects that approach or exceed those of standard clinician-delivered therapies?"

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Marks IM, Cavanagh K. Computer-aided psychological treatments: evolving issues. Ann Rev Clin Psychol. 2009; 5:121–141. [PubMed: 19327027]

2. Cuijpers P, Marks IM, van Straten A, Cavanagh K, Gega L, Andersson G. Computer-aided psychotherapy for anxiety disorders: a meta-analytic review. Cogn Behav Ther. 2009; 38:66–82. [PubMed: 20183688]

3. Wright JH, Wright AS, Albano AM, Basco MR, Goldsmith LJ, Raffield T, Otto MW. Computer-assisted cognitive therapy for depression: maintaining efficacy while reducing therapist time. Am J Psychiatry. 2005; 162:1158–1164. [PubMed: 15930065]

4. Proudfoot JG. Computer-based treatment for anxiety and depression: is it feasible? is it effective? Neurosci Biobehav Rev. 2004; 28:353–363. [PubMed: 15225977]

5. McCrone P, Knapp M, Proudfoot JG, Ryden C, Cavanagh K, Shapiro DA, Ilson S, Gray JA, Goldberg D, Mann A, Marks I, Everitt B, Tylee A. Cost-effectiveness of computerized cognitive-behavioral therapy for anxiety and depression in primary care: randomized controlled trial. Br J Psychiatry. 2004; 185:55–62. [PubMed: 15231556]

6. Kaltenthaler E, Brazier J, De Nigris E, Tumur I, Ferriter M, Beverley C, Parry G, Rooney G, Sutcliffe P. Computerised cognitive behaviour therapy for depression and anxiety update: a systematic review and economic evaluation. Health Technol Assess. 2006; 10:iii, xi–xiv, 1–168. [PubMed: 16959169]

7. McLellan, AT. What we need is a system: creating a responsive and effective substance abuse treatment system, in Rethinking Substance Abuse: What the Science Shows and What We Should Do About It. Miller, WR.; Carroll, KM., editors. New York: Guilford Press; 2006. p. 275-292.

8. Regier DA, Narrow WE, Rae DS, Manderscheid RW, Locke BZ, Goodwin FK. The de facto US mental health and addictive disorders service system: epidemiological catchment area prospective one-year prevalence rates of disorders and services. Arch Gen Psychiatry. 1993; 50:85–91. [PubMed: 8427558]

9. Institute of Medicine. Broadening the Base of Treatment for Alcohol Problems. Washington, DC: National Academies Press; 1990.

10. Kohn R, Saxena S, Levav I, Saraceno B. The treatment gap in mental health care. Bull World Health Organ. 2004; 82:858–866. [PubMed: 15640922]

11. [Accessed July 2010] www.internetworldstats.com

12. Cuijpers P, van Straten A, Andersson G. Internet-administered cognitive behavior therapy for health problems: a systematic review. J Behav Med. 2008; 31:169–177. [PubMed: 18165893]

13. Spek V, Cuijpers P, Nyklicek I, Riper H, Keyzer J, Pop V. Internet-based cognitive-behaviour therapy for symptoms of depression and anxiety: a meta-analysis. Psychol Med. 2007; 37:319–328. [PubMed: 17112400]

14. Postel MG, de Haan HA, DeJong CA. E-therapy for mental health problems: a systematic review. Telemed J E Health. 2008; 14:707–714. [PubMed: 18817501]

15. Kaltenthaler E, Parry G, Beverley C, Ferriter M. Computerized cognitive-behavioural therapy for depression: systematic review. Br J Psychiatry. 2008; 193:181–184. [PubMed: 18757972]

16. Gibbons, MC.; Wilson, RF.; Samal, L.; Lehmann, CU.; Dickersin, K.; Aboumatar, H.; Finkelstein, J.; Shelton, E.; Sharma, R.; Bass, EB. Impact of consumer health informatics applications, in Evidence Report/Technology Assessment Number 188. Rockville, Md: Agency for Healthcare Research and Quality; 2009.

17. Portnoy DB, Scott-Sheldon LA, Johnson BT, Carey MP. Computer-delivered interventions for health promotion and behavioral risk reduction: a meta-analysis of 75 randomized controlled trials, 1988–2007. Prev Med. 2008; 47:3–16. [PubMed: 18403003]

18. Higgins, JPT.; Green, S., editors. Cochrane Handbook for Systematic Reviews of Interventions, Version 4.2.6., in the Cochrane Library, Issue 4. Chichester, United Kingdom: John Wiley and Sons; 2006.

19. Williams, JBW.; Spitzer, RL. Psychotherapy Research: Where Are We and Where Should We Go?. New York: Guilford Press; 1984.

20. Chambless DL, Hollon SD. Defining empirically supported therapies. J Consult Clin Psychol. 1998; 66:7–18. [PubMed: 9489259]

21. Chambless DL, Ollendick TH. Empirically supported psychological interventions: controversies and evidence. Ann Rev Psychol. 2001; 52:685–716. [PubMed: 11148322]

22. Lilienfeld SO. Psychological treatments that cause harm. Perspect Psychol Sci. 2007; 2:53–70.

23. Dimidjian S, Hollon SD. How would we know if psychotherapy were harmful? Am Psychologist. 2010; 65:21–33.

24. Barlow DH. Negative effects from psychological treatments: a perspective. Am Psychol. 2010; 65:13–20. [PubMed: 20063906]

25. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. J Epidemiol Community Health. 1998; 52:377–384. [PubMed: 9764259]

26. Marks, IM.; Cavanagh, K.; Gega, L. Hands-On Help: Computer-Aided Psychotherapy. Hove, United Kingdom: Psychology Press; 2007.

27. Tumur I, Kaltenthaler E, Ferriter M, Beverley C, Parry G. Computerized cognitive behaviour therapy for obsessive-compulsive disorder: a systematic review. Psychother Psychosom. 2007; 76:196–202. [PubMed: 17570957]

28. Nathan, PE.; Gorman, JM. A Guide to Treatments That Work. 2nd ed. New York, Oxford: University Press; 2002.

29. Miller WR, Wilbourne PL. Mesa Grande: a methodological analysis of clinical trials of treatments for alcohol use disorders. Addiction. 2002; 97:265–277. [PubMed: 11964100]

30. Cuijpers P, van Straten A, Bohlmeijer E, Hollon SD, Andersson G. The effects of psychotherapy for adult depression are overestimated: a meta-analysis of study quality and effect size. Psychol Med. 2010; 40:211–223. [PubMed: 19490745]

31. Yates SL, Morley S, Eccleston C, de C Williams AC. A scale for rating the quality of psychological trials for pain. Pain. 2005; 117:314–325. [PubMed: 16154704]

32. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. J Pharmacol Pharmacother. 2010; 1:100–107. [PubMed: 21350618]

33. Chalmers TC, Smith HT Jr, Blackburn B, Silverman B, Schroeder B, Reitman D, Ambroz A. A method for assessing the quality of a randomized control trial. Control Clin Trials. 1981; 2:31–49. [PubMed: 7261638]

34. Carroll KM, Rounsaville BJ. Computer-assisted therapy in psychiatry: be brave—it's a new world. Curr Psychiatry Rep. 2010; 12:426–432. [PubMed: 20683681]

35. Postel MG, de Jong CA, de Haan HA. Does e-therapy for problem drinking reach hidden populations? Am J Psychiatry. 2005; 162:2393. [PubMed: 16330613]

36. Devilly GJ, Borkovec TD. Psychometric properties of the credibility/expectancy questionnaire. J Behav Ther Exp Psychiatry. 2000; 31:73–86. [PubMed: 11132119]

37. Carroll KM, Ball SA, Martino S, Nich C, Babuscio TA, Nuro KF, Gordon MA, Portnoy GA, Rounsaville BJ. Computer-assisted delivery of cognitive-behavioral therapy for addiction: a randomized trial of CBT4CBT. Am J Psychiatry. 2008; 165:881–888. [PubMed: 18450927]

38. Gwet, KL. Handbook of Inter-Rater Reliability. 2nd ed. Gaithersburg, Md: Advanced Analytics; 2010.

39. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977; 33:159–174. [PubMed: 843571]

40. Feinstein AR, Cicchetti DV. High agreement but low kappa, I: the problems of two paradoxes. J Clin Epidemiol. 1990; 43:543–549. [PubMed: 2348207]

41. Kaltenthaler E, Sutcliffe P, Parry G, Beverley C, Rees A, Ferriter M. The acceptability to patients of computerized cognitive behaviour therapy for depression: a systematic review. Psychol Med. 2008; 38:1521–1530. [PubMed: 18205964]

42. Lavori PW. Clinical trials in psychiatry: should protocol deviation censor patient data? Neuropsychopharmacology. 1992; 6:39–48. [PubMed: 1571068]

43. Cuijpers P, Marks IM, van Straten A, Cavanagh K, Gega L, Andersson G. Computer-aided psychotherapy for anxiety disorders: a meta-analytic review. Cogn Behav Ther. 2009; 38:66–82. [PubMed: 20183688]

44. Cunningham JA. Access and interest: two important issues in considering the feasibility of web-assisted tobacco interventions. J Med Internet Res. 2008; 10:e37. [PubMed: 18984558]

45. Kraemer HC, Kuchler T, Spiegel D. Use and misuse of the Consolidated Standards of Reporting Trials (CONSORT) guidelines to assess research findings: comment on Coyne, Stefanek, and Palmer (2007). Psychol Bull. 2009; 135:173–178. discussion 179–182. [PubMed: 19254073]

46. Petry NM, Roll JM, Rounsaville BJ, Ball SA, Stitzer M, Peirce JM, Blaine J, Kirby KC, McCarty D, Carroll KM. Serious adverse events in randomized psychosocial treatment studies: safety or arbitrary edicts? J Consult Clin Psychol. 2008; 76:1076–1082. [PubMed: 19045975]

47. Dimidjian S, Hollon SD. How would we know if psychotherapy were harmful? Am Psychol. 2010; 65:21–33. [PubMed: 20063907]
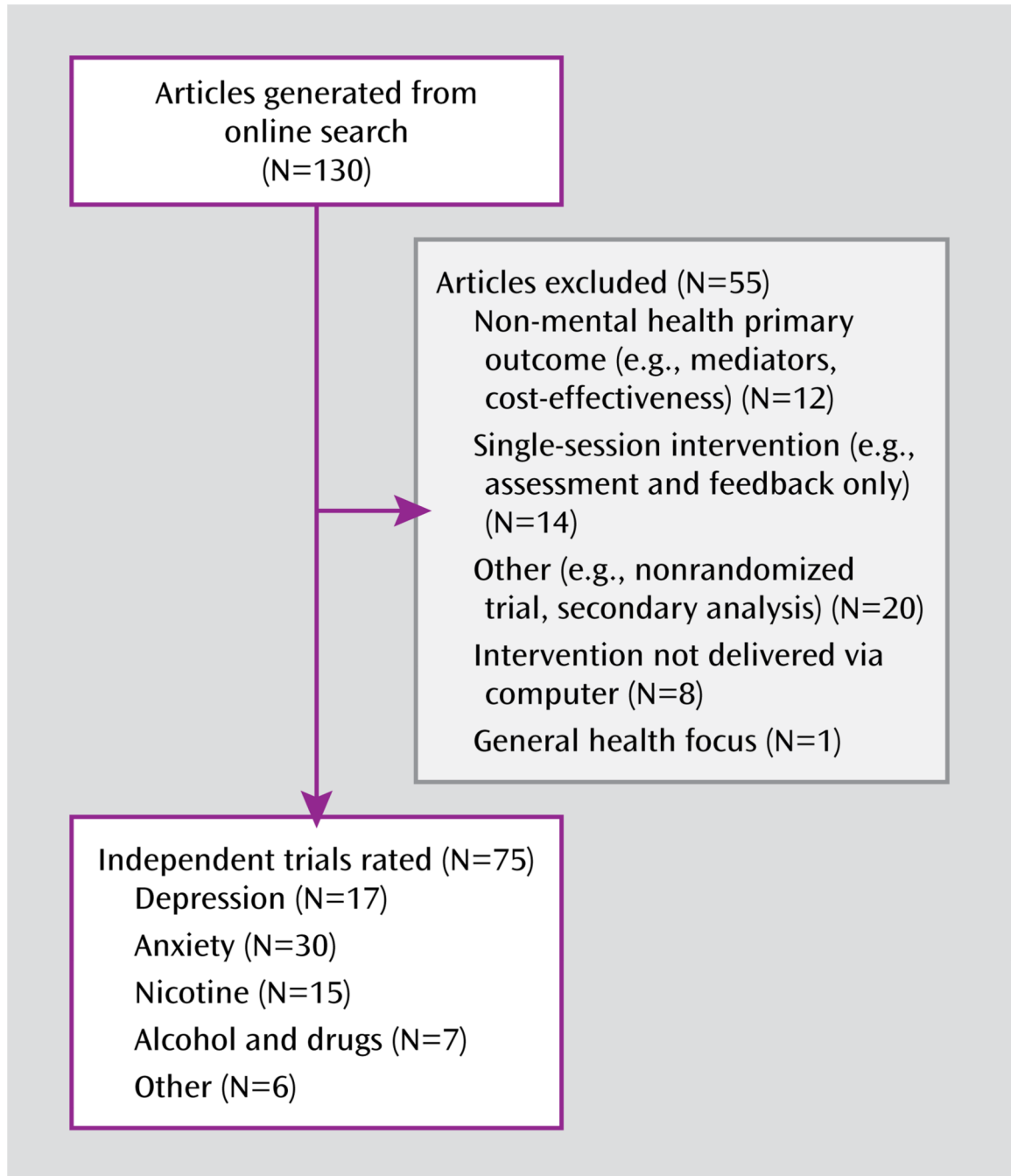
FIGURE 1.
Computer-Assisted Therapies for Adult Psychiatric Disorders Reviewed for This
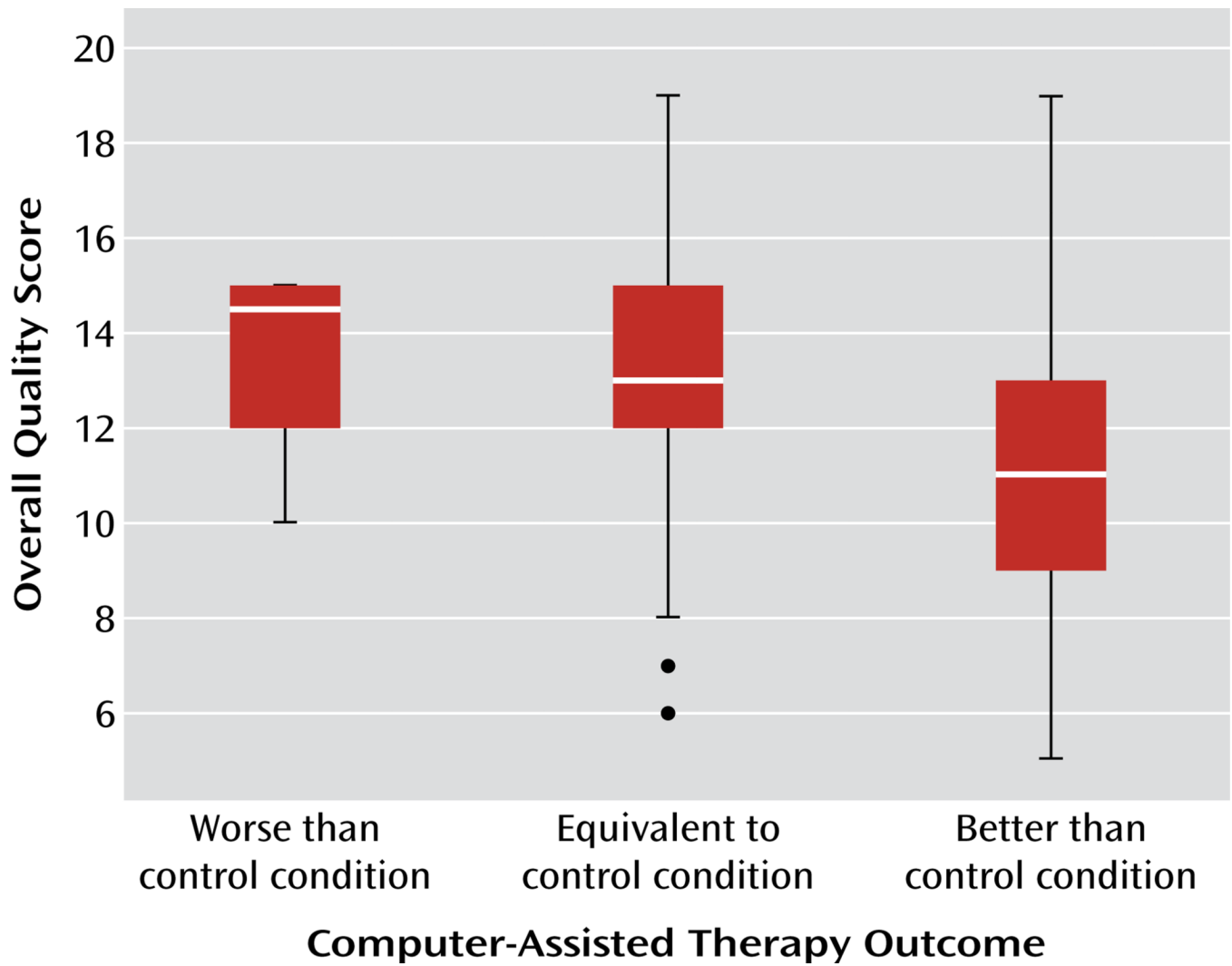Methodological Analysis

FIGURE 2.
Overall Quality Scores According to Reported Effectiveness of Intervention in Randomized
Controlled Trials of Computer-Assisted Therapies

TABLE 1

Comparison of Clinical Methodological Quality Rating Systems

Item	Computer-Assisted Therapy Methodology Rating Index	Chalmers et al. (33)	Chambless and Hollon (20)	Downs and Black (25)	Miller and Wilbourne (29)	Yates et al. (31)	Cuijpers et al. (30)	CONSORT[a] (32)
Criteria relevant to all randomized controlled trials								
Randomization method described (groups balanced)	×	×		×	×	×	×	×
Follow-up assessment for >80% of intention-to-treat sample	×		×	×	×	×	×	
Outcome assessment by blind assessor (validated assessments)	×		×	×	×	×		×
Adequate sample size/power	×		×	×		×	×	×
Appropriate, clearly described statistical analysis	×	×	×	×	×	×		×
Appropriate management of missing data, conformity to intention-to-treat principles	×		×	×	×	×	×	×
Replication in independent population	×		×		×			
Criteria relevant to trials of behavioral and computer-assisted therapies								
Clinical condition defined by standard measure (diagnosis), clear inclusion/exclusion criteria	×		×	×		×	×	×
Computer-assisted approach based on manualized evidence-based therapy	×							
Level of stringency for control condition	×		×			×		
Independent, biological assessment of outcome	×		×	×	×	×	×	×
Quality control, measure of compliance or adherence	×	×	×	×	×	×	×	
Equivalence of time, exposure across conditions	×		×	×		×		
Credibility measure	×					×		

[a] Consolidated Standards of Reporting Trials guidelines.

TABLE 2

Item Score Frequency Among 75 Trials of Computer-Assisted Therapies for Psychiatric Disorders[a]

Item	Rating	N	%	Mean	SD
Randomized control trial quality					
Randomization				1.73	0.44
No method	0				
Unspecified method	1	20	27		
Randomized method (equivalent groups)	2	55	73		
Follow-up evaluation				0.80	0.68
None	0	26	35		
Conducted posttreatment	1	38	51		
Data for >80% of total sample	2	11	15		
Assessment				0.45	0.70
Self-report only	0	50	67		
Outcome interview	1	16	21		
Blind interview	2	9	12		
Sample size				1.20	0.79
Small (<20 per group)	0	17	23		
Moderate (20–50 per group)	1	26	35		
Large (>50 per group)	2	32	43		
Statistics				1.05	0.49
Weak approach	0	7	9		
Reasonable method	1	57	76		
Strong analytic method	2	11	15		
Missing data				0.73	0.68
Not mentioned/inappropriate	0	30	40		
Handled with imputations	1	35	47		
Intention to treat (all subjects included)	2	10	13		
Replication				0.57	0.76
None	0	44	59		
One	1	19	25		
Multiple populations	2	12	16		

Item	Rating	N	%	Mean	SD
Computer-assisted therapy-specific quality					
Defined problem				1.45	0.62
No clear criteria	0	5	7		
Self-report or cutoff score	1	31	41		
Diagnostic interview	2	39	52		
Evidence-based				1.43	0.60
None	0	4	5		
Some (no manual)	1	35	47		
Strong	2	36	48		
Control condition				1.13	0.76
Wait-list	0	17	23		
Placebo/attention/education	1	31	41		
Active (equal time)	2	27	36		
Outcome measure				1.16	0.57
Unvalidated	0	7	9		
Self-report only	1	49	65		
Independent assessment	2	19	25		
Quality control/adherence				0.97	0.72
No measure	0	20	27		
Measure access	1	37	49		
Considered in analysis	2	18	24		
Credibility				0.33	0.64
No measure	0	57	76		
Measure included	1	11	15		
Comparability across treatments	2	7	9		
Time exposure				0.61	0.93
Not equal	0	52	69		
Equivalent	2	23	31		

[a] The totals with regard to score range for randomized control quality items and computer-assisted therapy-specific quality items are 2–11 and 2–13, respectively; the total mean score values are 6.55 (SD=1.90) and 7.09 (SD=2.67), respectively; and the totals for overall quality score range and mean value are 6–21 and 13.64 (SD=3.62).

TABLE 3

Methodological Features by Disorder/Problem Area in Trials of Computer-Assisted Therapies for Psychiatric Disorders[a]

Feature	Depression		Anxiety		Nicotine Dependence		Alcohol/Drug Dependence		Other		Total	
	N	%	N	%	N	%	N	%	N	%	N	%
Web-based (yes)	12	71	22	73	12	80	5	71	3	50	54	72
Sample type (clinical)	6	35	24	80	0		2	29	5	83	37	49
Stand-alone (yes)	11	65	27	90	9	60	4	57	3	50	54	72
Clarity regarding recruitment (yes)	17	100	28	93	15	100	7	100	6	100	73	97
Clarity regarding handling of missing data and dropouts (yes)	15	88	22	73	14	93	6	86	3	50	60	80
For-profit sponsor (yes)	4	24	4	13	6	40	4	57	0		18	24
Declaration of conflict of interest (yes)	12	71	5	17	7	47	4	57	1	17	29	39
Clinician involvement (none)	9	53	6	20	12	80	6	86	2	33	35	47
Peer support (yes)	1	6	9	30	7	47	4	57	0		21	28
Patient satisfaction measured (yes)	2	12	18	60	6	40	4	57	3	50	33	44
Conclusions supported by data (yes)	15	88	25	83	14	93	6	86	5	83	65	87
Consolidated Standards of Reporting Trials diagram included (yes)	14	82	15	50	7	47	3	43	3	50	42	56
Defined length of treatment and reported treatment exposure (yes)	3	17	6	20	5	33	1	14	1	17	16	21
Effect size reported (yes)	10	59	17	57	9	60	4	57	3	50	43	57
Intervention type												
Cognitive-behavioral therapy	15	88	27	90	4	27	4	57	4	67	54	72
Motivational	0	0	0	0	1	7	2	29	0		3	4
Stage-based	0	0	0	0	4	27	0		0		4	5
Problem-solving	1	7	0	0	1	7	0		0		2	3
Other	1	7	3	10	5	33	1	14	2	33	12	16

[a] Analyses showed significant differences (p<0.05) for the following study features: sample type, stand-alone, declaration of conflict of interest, clinician involvement, peer support, and patient satisfaction.