Anxiety treatment improves physical functioning with oblique scoring of the SF-12 Short Form Health Survey[★]

Andrea N. Niles, M.A.[a], **Cathy D. Sherbourne, Ph.D.**[b], **Peter P. Roy-Byrne, M.D.**[c], **Murray B. Stein, M.D., M.P.H.**[d], **Greer Sullivan, M.D., M.S.P.H.**[e], **Alexander Bystritsky, M.D.**[f], and **Michelle G. Craske, Ph.D.**[a,f,*]

[a]Department of Psychology, University of California, Los Angeles, CA, USA

[b]The RAND Corporation, Santa Monica, CA, USA

[c]Department of Psychiatry and Behavioral Sciences, University of Washington School of Medicine at Harborview Medical Center, Seattle, WA, USA

[d]Departments of Psychiatry and Family and Preventative Medicine, University of California, San Diego, CA, USA

[e]South Central VA Mental Illness Research Education and Clinical Center, North Little Rock, Arkansas and University of Arkansas for Medical Science, Little Rock, AR, USA

[f]Department of Psychiatry and Biobehavioral Sciences, University of California, Los Angeles, CA, USA

Abstract

Objective—No studies have found a positive effect of anxiety treatment on physical functioning, but recent investigations of the 12-item Short Form Health Questionnaire (SF-12), which is frequently used to assess physical functioning, have suggested that orthogonal scoring of the summary measure may distort representations of physical health. The current study reanalyzes whether anxiety treatment improves physical functioning using oblique scoring in the Coordinated Anxiety Learning and Management (CALM) randomized clinical trial for the treatment of anxiety disorders. Replication was tested in reanalysis of data from the earlier Collaborative Care for Anxiety and Panic (CCAP) randomized clinical trial for the treatment of panic disorder.

Method—The CALM study included 1004 primary care patients with panic, social anxiety, generalized anxiety or posttraumatic stress disorders. Patients received usual care (UC) or an evidence-based intervention (cognitive behavioral therapy, psychotropic medication or both; ITV). Physical functioning (SF-12v2) was assessed at baseline and at 6, 12 and 18 months. Oblique and orthogonal scoring methods for the physical functioning aggregate measure from SF-12 scale items were compared.

Results—In CALM, physical functioning improved to a greater degree in ITV than UC for oblique but not orthogonal scoring. Findings were replicated in the CCAP data.

Conclusions—Evidence-based treatment for anxiety disorders in primary care improves physical functioning when measured using oblique scoring of the SF-12. Due to this scoring issue, effects of mental health treatment on physical functioning may have been understated.

[*]Corresponding author. Tel.: +310 825 8403; fax: +310 206 5895. craske@psych.ucla.edu.

Keywords

Short Form Health Survey; Anxiety; Measurement; Physical functioning; Physical health; Health psychology

1. Introduction

Anxiety disorders are the most common psychological disorders and will occur in approximately 29% of individuals at some time in their lives [1]. Not only are symptoms of fear and anxiety debilitating, there is substantial evidence that anxiety disorders are detrimental to physical health and functioning [2–7]. While existing treatments for anxiety disorders, such as psychotropic medication or cognitive behavioral therapy, are effective in reducing symptoms of anxiety, effects on physical functioning, albeit assessed in very few studies, have been negligible. For example, previously published data show that evidence-based care for anxiety disorders significantly reduced anxiety and depressive symptoms, but had no effect on physical functioning [8]. Herein, we reanalyze those data using a different scoring method, given recent questioning of scoring methodology for the 12- and 36-item Short Form Health Questionnaires (SF-12 and SF-36) [9–13].

A number of studies have identified a link between anxiety disorders and poor physical functioning. Sherbourne and colleagues [14] ($N = 875$) found that hypertensive and diabetic patients with comorbid anxiety had significantly worse physical functioning than those without comorbid anxiety. Similarly, using the German Health Study, Sareen and colleagues found that participants with a physical disorder and an anxiety disorder reported worse physical functioning scores and greater disability than participants without an anxiety disorder [6]. In an investigation using the National Comorbidity Survey sample ($N = 5877$), the presence of a comorbid anxiety disorder in the past year was associated with greater role impairment in individuals with one or more medical conditions [5].

Existing treatments such as cognitive behavioral therapy (CBT) and psychotropic medication for anxiety disorders have proven highly effective in reducing anxiety symptoms. In a meta-analytic review, changes from pre- to posttreatment for CBT (effect size = 1.58) were substantially larger for all anxiety disorders than no treatment (effect size = .25) or placebo control treatments (effect size = 1.04) [15]. The effects of psychotropic medication are equal to those of CBT, although relapse rates have been lower in CBT in a number of studies [16]. Recently, the Coordinated Anxiety Learning and Management (CALM) study tested the effectiveness of a collaborative care model of delivery for evidence-based treatment for anxiety disorders that was comprised of medication, CBT or both (ITV) compared to usual care (UC) in 1004 patients with anxiety disorders in primary care [8]. Patients in the ITV group showed significantly greater improvement on primary anxiety-related outcome measures than UC at 6-, 12- and 18-month follow-up [8,17].

In contrast to the robust and substantial effects of evidence-based treatments upon anxiety symptoms, the effects upon physical functioning have been nil. In the Collaborative Care for Anxiety and Panic study (CCAP), an earlier evaluation of treatment for panic disorder in primary care, no differences were found between evidence-based CBT/medication and usual care in physical functioning outcomes as measured by the SF-12 (version 1) [18]. Also, in the CALM study, there was no evidence for group differences on the SF-12 (version 2) between ITV and UC at any follow-up assessment [8]. Findings from depression treatment studies are mixed, with some showing no improvement on SF-12 or SF-36 measures [19,20]; others showing improvement on the subscales, but not on the physical functioning

composite scale [21,22]; and only one showing improvement on the composite scale [23]. These inconsistencies and null findings may be explained by how the scale is scored.

A number of researchers have identified potential problems with the traditional scoring method used for the physical and mental health composite scales of the SF-12 and SF-36. The SF composites are comprised of eight subdimensions, all of which are used to calculate the physical (PCS) and mental (MCS) composite summary scales. In calculating the composite summary scales, the mental items are weighted orthogonally to the physical items under the assumption that mental health and physical health are entirely distinct constructs [9,11]. However, there is a substantial amount of research showing that physical health and mental health are correlated [2,4,24–26]. Therefore, attempts to statistically correct for this correlation could distort the measurement. In addition, as a result of this orthogonal scoring method, summary scores can be inconsistent with subscale scores (which do not use orthogonal scoring) [12]. Finally, such statistical corrections can inhibit the ability to detect changes in physical functioning after mental health treatment [11]. This may explain the decline in physical functioning from baseline to first follow-up observed in intervention studies for anxiety and depression [8,19,20].

Alternative methods of scoring having been proposed that allow the physical and mental health summary scales to be oblique, meaning that the calculation does not correct for the correlation. Fleishman, Selim and Kazis [12] derived new weights for the subscales from the SF-12 version 2, as did Farivar, Cunningham and Hayes [10] for the SF-12 version 1. Because versions 1 and 2 of the SF-12 are scored using different methods, oblique weights were derived separately for each version. In both studies, the authors compared different scoring methods and concluded that the preferred scoring method allowed the physical and mental health items to correlate using oblique scoring. In the current study, we use the weights presented by Fleishman and colleagues (SF-12 version 2) and compare the results to those obtained with the orthogonal scoring using the CALM study data. We hypothesized that, with the oblique scoring of the SF-12, participants assigned to ITV would show greater improvement at follow-up in physical functioning than participants in the UC group. As a replication test, we also reanalyzed data from the CCAP treatment study for panic disorder using weights presented by Farivar and colleagues (SF-12 version 1) and again hypothesized that oblique scoring of the SF-12 would result in greater improvement in physical functioning in the intervention group than the usual care group.

2. Method

2.1. CALM participants, design and procedure

Participants were 1004 primary care patients between the ages of 18 and 75 (see Table 1 for demographics). Between June 2006 and April 2008, 1620 primary care patients consented to complete a study eligibility interview, and 1004 patients with panic disorder (with or without agoraphobia), generalized anxiety disorder, social anxiety disorder or posttraumatic stress disorder were enrolled. Participating research institutions were the University of Washington (Seattle), University of California-Los Angeles, University of California-San Diego, University of Arkansas for Medical Sciences and the RAND Corporation (an assessment site only) (details are presented elsewhere [27]).

2.1.1. Recruitment—Primary care providers and clinic nursing staff directly referred potential participants. At some sites, a five-question anxiety screener, the Overall Anxiety Severity and Impairment Scale (OASIS) [28], was used to identify potential participants. A trained study clinician, the Anxiety Clinical Specialist (ACS), functioned as the main care manager/interventionist, as well as the diagnostician who met with referred patients to

determine eligibility. All participants gave informed, written consent to participate in this study, which was approved by each institution's Institutional Review Board.

2.1.2. Inclusion criteria—An eligible participant had to be a patient at a participating clinic; be 18–75 years old; meet the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition* criteria for one or more of panic disorder, generalized anxiety disorder, social anxiety disorder or posttraumatic stress disorder (based on the Mini International Neuropsychiatric Interview [29]) administered by the ACS (after formal training and diagnostic reliability testing); and score at least 8 (moderate and clinically significant anxiety symptoms on a scale ranging from 0 to 20) on the OASIS.

2.1.3. Exclusion criteria—Patients with unstable or life-threatening medical conditions, marked cognitive impairment, active suicidal intent or plan, psychosis or bipolar I disorder were excluded. Alcohol or marijuana abuse (but not dependence) was permitted, but other drug abuse or dependence was exclusionary. Patients already receiving ongoing CBT ($n = 7$) were excluded, as were patients who could not speak English or Spanish ($n = 2$).

2.2. Randomization

After baseline assessment, participants were randomized using stratified (by clinic and presence of comorbid major depression) permuted block randomization to the intervention (ITV) or usual care (UC) by an automated program at RAND. Block size was masked to all clinical site study members. Of the 1004 patients enrolled in the study, 503 were randomized into the ITV group. Further details are available about patient flow from referral through eligibility screening, consent and randomization in the CALM study [8].

2.3. Intervention

Intervention group patients chose to receive either medication or CBT or both over a 10- to 12-week period. The CBT program (called *CALM Tools for Living*, English and Spanish versions) covered the four anxiety disorders and, for each participant, was targeted at their most distressing and disabling anxiety disorder. CBT was comprised of generic modules (i.e., self monitoring, psychoeducation, fear hierarchy, breathing retraining, relapse prevention) and cognitive restructuring and exposure modules that were tailored to each anxiety disorder through branching mechanisms. The treatment was computer assisted to guide both the therapist and the patient as they sat side by side to view the program on screen (more details are available elsewhere [30]). Occasionally, the ACS used additional strategies, such as behavioral activation and cognitive restructuring for depressed mood, and motivational enhancement strategies to maintain patient engagement. After the first 10–12 weeks, symptomatic participants could receive more of the same modality (CBT or medication) or the alternative modality for up to three more steps (i.e., another 10–12 weeks) of treatment. For patients who chose medication, the ACS monitored adherence, remained in contact with the patient's primary physician and psychiatrist, and provided information about alcohol, sleep hygiene and behavioral activity to promote optimum medication use. All ACSs were supervised weekly by a psychiatrist and psychologist. Further details are available elsewhere [27].

2.4. Usual care

Patients in the UC group were treated by their primary care physicians as they normally would be treated. This consisted of medication, counseling or referral to a mental health professional. After the baseline assessment, patients in the UC group did not have contact with study personnel aside from scheduled assessments by telephone.

2.5. Measures

2.5.1. SF-12 (version 2) – Physical Composite Scale (PCS) and Physical Health Scale (PHS)—The SF-12 [31] is a shortened version of the SF-36, a measure of physical and mental health functioning that has been widely used and validated. The measure is comprised of eight subscales, which are used to calculate physical (PCS) and mental (MCS) composite summary scales. For the PCS, the physical items are weighted more heavily, and for the MCS, the mental items are weighted more heavily. In addition, the mental health items are weighted orthogonally to the physical health items. A second calculation method was based on the oblique item weights derived through confirmatory factor analysis presented in Farivar et al. [10]. Scores calculated using the weights from confirmatory factor analysis are referred to as the PHS.

2.6. Statistical analysis

Analyses on the CALM data were conducted in SAS. To assess the effect of the intervention on change in physical functioning over the 18-month study period, multilevel modeling was used with repeated measures within participants modeled at level 1 and intervention modeled at level 2. Intercepts and slopes were treated as random effects, and time and group were treated as categorical variables. The covariance structure was unstructured, and we fitted the proposed model using a restricted maximum likelihood approach, which produces valid estimates under the missing at random assumption. This approach correctly handles missing data and uses all data to obtain unbiased estimates for model parameters. PCS and PHS were assessed as outcome variables. We corrected for multiple comparisons using Hochberg's correction method. Cohen's *d* effect sizes are reported for between-group differences.

2.7. CCAP replication test

The method used in the CCAP study was similar to that used in the CALM study, but the population was restricted to those with panic disorder, and follow-ups did not extend beyond 12 months. SF-12v1 data from the CCAP study were reanalyzed using the oblique scoring method (PHS). We used the statistical approach described in the manuscript where these data were originally published. For a full description of study methods, see Roy-Byrne et al. [18]. We specified a hierarchical two-level growth curve model with random effects and modeled repeated measures using a piecewise linear growth model, which specifies a linear segment between baseline and 3-month follow-up (segment 1) and another linear segment for 6-, 9- and 12-month follow-ups (segment 2). This model reflects the trend for the strongest effect of treatment to occur in the acute study phase and for effects to level out at later time points. However, previous analyses used a Bayesian approach, whereas the current analyses used xtmixed in Stata 12. The two statistical approaches produced similar mean estimates at each time point and the same pattern of results (i.e., slight reduction in PCS scores over time and no group differences at any time point).

3. Results

Table 1 shows that the CALM sample was 71% women, ethnically diverse (44% nonwhite) and broad in age range. It was a fairly ill group, with more than half having at least two chronic medical conditions and at least two anxiety disorders and with two thirds having comorbid major depression. The intervention and UC groups were comparable on all baseline characteristics.

3.1. CALM: effect of the intervention on physical functioning

3.1.1. PCS—Results for the PCS are depicted in Fig. 1A. PCS scores at baseline were marginally significantly below the population mean in the ITV group [mean difference = −.98; 95% confidence interval (CI) = − 1.98 to .02; *P* = .056] and did not differ from the population mean in the UC group (mean difference = −.67; 95% CI = −1.67 to .33; *P* = .191). Between-group differences at each assessment point were not significant (Table 2). The time by group interaction from baseline to 18 months was marginally significant [β = 1.19, *t*(2486) = 1.88, *P* = .060].

3.1.2. PHS—Results for the PHS are depicted in Fig. 1B. PHS scores at baseline were significantly below the population mean in the ITV group (mean difference = −7.56; 95% CI = −8.44 to −6.68; *P* < .001) and in the UC group (mean difference = −7.36; 95% CI = −8.24 to −6.48; *P* < .001). The groups did not differ significantly at baseline. The ITV group had significantly higher PHS scores than the UC group at 6, 12 and 24 months (Table 2). Cohen's *d* effect sizes ranged from .20 to .21. Also, PHS scores increased significantly from baseline to 6 months in the ITV group (mean difference = 3.40; 95% CI = 2.69 to 4.10; adj *P* = .002) and in the UC group (mean difference = 1.14; CI = .42 to 1.85; adj *P* = .030). The time by group interaction from baseline to 18 months was significant [β = 2.31, *t*(2477) = 4.00, *P* < .001].

3.2. CCAP replication test

3.2.1. PCS—Results for the PCS are depicted in Fig. 2A. Groups did not differ at 3-month (mean difference = 1.48; 95% CI = −1.78 to 4.74; *P* = .374), 6-month (mean difference = 1.56; 95% CI = −1.59 to 4.71; *P* = .332), 9-month (mean difference = 1.64; 95% CI = −1.59 to 4.88; *P* = .320) or 12-month (mean difference = 1.72; 95% CI = −1.77 to 5.22; *P* = .334) follow-up periods. The time by group interaction was not significant for segment 1 (β = .25; 95% CI = −.55 to 1.06; *P* = .533) or for segment 2 (β = .027; 95% CI = −.23 to .29; *P* = .839).

3.2.2. PHS—Results for the PHS are depicted in Fig. 2B. Between-group comparisons were not significant at any time point: at 3 months (mean difference = 2.42; 95% CI = −.74 to 5.58; *P* = .133), 6 months (mean difference = 2.25; 95% CI = −.79 to 5.29; *P* = .146), 9 months (mean difference = 2.08; 95% CI = −1.00 to 5.15; *P* = .186) and 12 months (mean difference = 1.91; 95% CI = −1.37 to 5.18; *P* = .255). The time by group interaction was significant for segment 1 (β = .92; 95% CI = .24 to 1.6; *P* = .008), with participants in the intervention group (slope = 1.48; 95% CI = 1.01 to 1.95) showing significantly greater improvement on the PHS than participants in the UC group (slope = .56; 95% CI = .081 to 1.04). The time by group interaction was not significant for segment 2 (β = −.057; 95% CI = −.30 to .18; *P* = .639).

4. Discussion

Using a revised scoring method for the SF-12, we reexamined whether an evidence-based intervention for anxiety disorders in primary care, comprised of medication, cognitive behavioral therapy or both, significantly improved physical functioning relative to treatment as usual. Patients assigned to the intervention group showed significantly greater improvement than usual care patients in physical functioning, but only when measured using the oblique scoring of the SF-12 (PHS) and not when using the traditional orthogonal scoring (PCS). In fact, the orthogonal scoring results in a decrease in physical functioning (although this was not significant) and no group differences at any time point. In contrast, the oblique scoring showed an increase in physical functioning following baseline in both

groups, with higher functioning levels in the ITV group at each follow-up assessment up to 2 years.

To further validate the findings, we sought to replicate the results with the oblique scoring method using data from an earlier study of primary care treatment for panic disorder. In this data set, the intervention group showed greater improvement in physical functioning from baseline to 3-month follow-up using the oblique scoring but not the traditional orthogonal scoring. Although simple effects analyses of between-group differences were not significant in this panic disorder sample, use of the oblique scoring nearly doubled the mean group differences. The lack of significant between-group differences is likely attributable to the smaller sample size in the CCAP study ($N = 232$) compared to the CALM study ($N = 1008$). In addition, the CALM intervention had a larger impact on SF-12 mental health scores (12-point improvement) than the CCAP intervention (6-point improvement). The traditional orthogonal scoring method is particularly problematic when change in one domain (e.g., mental health functioning) is much larger than change in the other domain. Therefore, the large change in mental health functioning following the CALM intervention concealed changes that occurred in physical functioning when orthogonal scoring was used. These changes were then revealed when the oblique scoring was used. In the CCAP data set, the change in mental health functioning was not as large, and therefore, the orthogonal scoring did not conceal changes in physical functioning to the same degree as in the CALM study data.

These findings have implications for use and interpretation of the summary scores derived from the SF-12v1 and 2. The short form health survey is easy to administer and has made a substantial contribution to the assessment of health-related quality of life. However, reliance on summary scales may produce inconsistent results as demonstrated here. A number of previous studies have addressed the potential impact of different scoring methods in non-treatment-seeking samples [13,32] with little research using the different scoring methods in clinical samples. The decision regarding which scoring method to use is based on one's assumption regarding the association between physical and psychological functioning. If it is assumed that there is no true relationship between these health domains and any observed relationship is based on measurement error, then the original orthogonal scoring should be used. However, if a true association between physical and psychological functioning is assumed, the oblique scoring method should be used. Given that mental health and physical health are strongly related, especially in clinical samples [4,5,2,24], we recommend that the oblique scoring method should be used in studies of patients with physical or psychological disorders.

Our rescoring of the SF-12 measure resulted in evidence-based treatment for anxiety disorders in primary care having significant and positive effects on physical functioning in two separate data sets, which were not found previously with orthogonal scoring of the measure. These findings suggest caution in the interpretation of Physical Composite Summary scores when the condition or treatment of interest has strong effects on the mental health items from the SF-12. In addition, prior null findings [19–22] may prove significant if rescored using the oblique method.

4.1. Limitations and future directions

In the current study, the SF-12 was the only measure of physical functioning that was included at all time points. Replication of the findings across multiple measures of physical functioning would strengthen the conclusions. In addition, our results are limited to health functioning as opposed to physical health–disease status. Furthermore, our results are limited to samples of patients in primary care, without unstable or life-threatening medical conditions, as well as other exclusionary criteria. Thus, the extent to which the findings

replicate in a more severely medically ill population or a sample with psychosis, bipolar disorder, marked cognitive impairment, or drug dependence is unknown.

This is the first study to show that evidence-based treatment for anxiety disorders leads to significantly greater improvement in physical health functioning than usual care. The findings highlight that the benefits of mental health treatments reach beyond psychological symptom outcomes to physical health outcomes. As such, they add to the growing impetus to disseminate evidence-based practices in real-world settings.

Acknowledgments

References

1. Kessler RC, Chiu WT, Demler O, Merikangas KR, Walters EE. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry. 2005; 62(6):617–27. [PubMed: 15939839]

2. Härter MC, Conway KP, Merikangas KR. Associations between anxiety disorders and physical illness. Eur Arch Psychiatry Clin Neurosci. 2003; 253(6):313–20. [PubMed: 14714121]

3. Kessler RC, Ormel J, Demler O, Stang PE. Comorbid mental disorders account for the role impairment of commonly occurring chronic physical disorders: results from the National Comorbidity Survey. J Occup Environ Med. 2003; 45(12):1257–66. [PubMed: 14665811]

4. Roy-Byrne PP, Davidson KW, Kessler RC, Asmundson GJG, Goodwin RD, Kubzansky L, et al. Anxiety disorders and comorbid medical illness. Gen Hosp Psychiatry. 2008; 30(3):208–25. [PubMed: 18433653]

5. Sareen J, Cox BJ, Clara I, Asmundson GJG. The relationship between anxiety disorders and physical disorders in the U.S. National Comorbidity Survey. Depress Anxiety. 2005; 21(4):193–202. [PubMed: 16075453]

6. Sareen J, Jacobi F, Cox BJ, Belik SL, Clara I, Stein MB. Disability and poor quality of life associated with comorbid anxiety disorders and physical conditions. Arch Intern Med. 2006; 166(19):2109–16. [PubMed: 17060541]

7. Stein MB, Roy-Byrne PP, Craske MG, Bystritsky A, Sullivan G, Pyne JM, et al. Functional impact and health utility of anxiety disorders in primary care outpatients. Med Care. 2005; 43(12):1164–70. [PubMed: 16299426]

8. Roy-Byrne PP, Craske MG, Sullivan G, Rose RD, Edlund MJ, Lang AJ, et al. Delivery of evidence-based treatment for multiple anxiety disorders in primary care. J Am Med Assoc. 2010; 303(19):1921–8.

9. Hann M, Reeves D. The SF-36 scales are not accurately summarised by independent physical and mental component scores. Qual Life Res. 2008; 17(3):413–23. [PubMed: 18259888]

10. Farivar SS, Cunningham WE, Hays RD. Correlated physical and mental health summary scores for the SF-36 and SF-12 Health Survey, V.I. Health Qual Life Outcomes. 2007; 5:54. [PubMed: 17825096]

11. Simon GE, Revicki DA, Grothaus L, Vonkorff M. SF-36 summary scores: are physical and mental health truly distinct? Med Care. 1998; 36(4):567–72. [PubMed: 9544596]

12. Fleishman JA, Selim AJ, Kazis LE. Deriving SF-12v2 physical and mental health summary scores: a comparison of different scoring algorithms. Qual Life Res. 2010; 19(2):231–41. [PubMed: 20094805]

13. Taft C, Karlsson J, Sullivan M. Do SF-36 summary component scores accurately summarize subscale scores? Qual Life Res. 2001; 10(5):395–404. [PubMed: 11763202]

14. Sherbourne CD, Wells KB, Meredith LS, Jackson CA, Camp P. Comorbid anxiety disorder and the functioning and well-being of chronically ill patients of general medical providers. Arch Gen Psychiatry. 1996; 53(10):889–95. [PubMed: 8857865]

15. Norton PJ, Price EC. A meta-analytic review of adult cognitive–behavioral treatment outcome across the anxiety disorders. J Nerv Ment Dis. 2007; 195:521–31. [PubMed: 17568301]

16. Baldwin DS, Anderson IM, Nutt DJ, Bandelow B, Bond A, Davidson JRT, et al. Evidence-based guidelines for the pharmacological treatment of anxiety disorders: recommendations from the British Association for Psychopharmacology. J Psychopharmacol. 2005; 19(6):567–96. [PubMed: 16272179]

17. Craske MG, Stein MB, Sullivan G, Sherbourne C, Bystritsky A, Rose RD, et al. Disorder-specific impact of coordinated anxiety learning and management treatment for anxiety disorders in primary care. Arch Gen Psychiatry. 2011; 68(4):378. [PubMed: 21464362]

18. Roy-Byrne PP, Craske MG, Stein MB, Sullivan G, Bystritsky A, Katon W, et al. A randomized effectiveness trial of cognitive–behavioral therapy and medication for primary care panic disorder. Arch Gen Psychiatry. 2005; 62(3):290–8. [PubMed: 15753242]

19. Wells KB, Sherbourne C, Schoenbaum M, Duan N, Meredith L, Unützer J, et al. Impact of disseminating quality improvement programs for depression in managed primary care. JAMA. 2000; 283(2):212–20. [PubMed: 10634337]

20. Hedrick SC, Chaney EF, Felker B, Liu C, Hasenberg N, Heagerty P, et al. Effectiveness of collaborative care depression treatment in veterans' affairs primary care. J Gen Intern Med. 2003; 18(1):9–16. [PubMed: 12534758]

21. Coulehan JL, Schulberg HC, Block MR, Madonia MJ, Rodriguez E. Treating depressed primary care patients improves their physical, mental, and social functioning. Arch Intern Med. 1997; 157(10):1113–20. [PubMed: 9164377]

22. Bair MJ, Robinson RL, Eckert GJ, Stang PE, Croghan TW, Kroenke K. Impact of pain on depression treatment response in primary care. Psychosom Med. 2004; 66(1):17–22. [PubMed: 14747633]

23. Callahan CM, Kroenke K, Counsell SR, Hendrie HC, Perkins AJ, Katon W, et al. Treatment of depression improves physical functioning in older adults. J Am Geriatr Soc. 2005; 53(3):367–73. [PubMed: 15743276]

24. Katon WJ, Lin EHB, Kroenke K. The association of depression and anxiety with medical symptom burden in patients with chronic medical illness. Gen Hosp Psychiatry. 2007; 29(2):147–55. [PubMed: 17336664]

25. Neeleman J, Ormel J, Bijl RV. The distribution of psychiatric and somatic ill health: associations with personality and socioeconomic status. Psychosom Med. 2001; 63(2):239–47. [PubMed: 11292271]

26. Berardi D, Ceroni GB, Leggieri G, Rucci P, Ustun B, Ferrari G. Mental, physical and functional status in primary care attenders. Int'l J Psychiatry Med. 1999; 29(2):133–48.

27. Sullivan G, Craske MG, Sherbourne C, Edlund MJ, Rose RD, Golinelli D, et al. Design of the Coordinated Anxiety Learning and Management (CALM) study: innovations in collaborative care for anxiety disorders. Gen Hosp Psychiatry. 2007; 29(5):379–87. [PubMed: 17888803]

28. Campbell-Sills L, Norman SB, Craske MG, Sullivan G, Lang AJ, Chavira DA, et al. Validation of a brief measure of anxiety-related severity and impairment: the Overall Anxiety Severity and Impairment Scale (OASIS). J Affect Disord. 2009; 112(1–3):92–101. [PubMed: 18486238]

29. Lecrubier Y, Sheehan D, Weiller E, Amorim P, Bonora I, Harnett Sheehan K, et al. The Mini International Neuropsychiatric Interview (MINI). A short diagnostic structured interview: reliability and validity according to the CIDI. Eur Psychiatry. 1997; 12(5):224–31.

30. Craske MG, Roy-Byrne PP, Stein MB, Sullivan G, Sherbourne C, Bystritsky A. Treatment for anxiety disorders: efficacy to effectiveness to implementation. Behav Res Ther. 2009; 47(11):931–7. [PubMed: 19632667]

31. Ware JE, Kosinski M, Keller SD. A 12-item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care. 1996; 34(3):220. [PubMed: 8628042]

32. Ware JE, Kosinski M. Interpreting SF-36 summary health measures: a response. Qual Life Res. 2001; 10(5):405–13. discussion 415–420. [PubMed: 11763203]
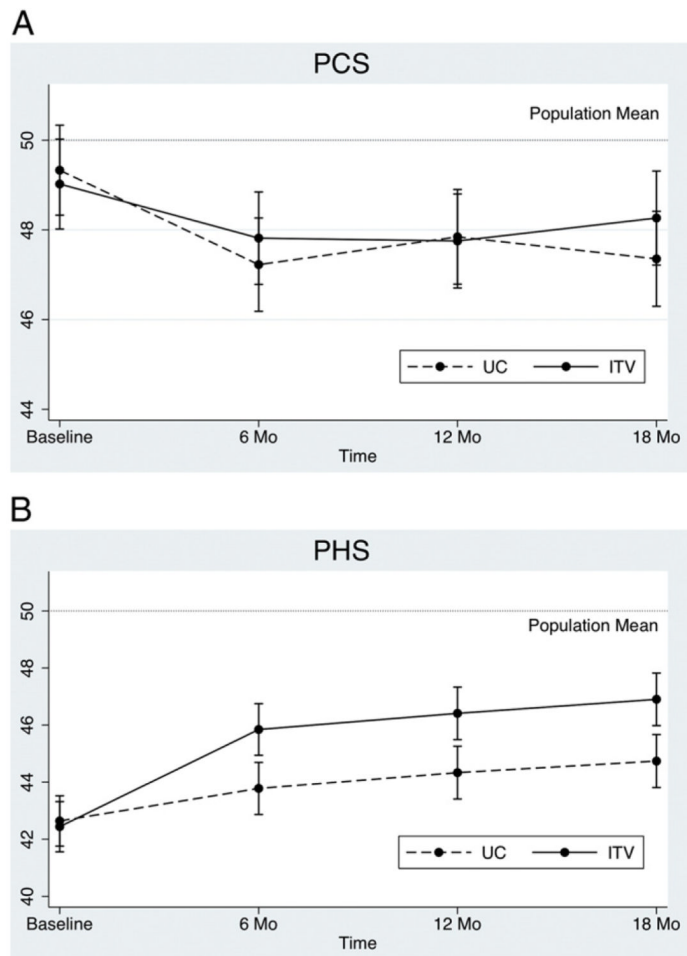
Fig. 1.
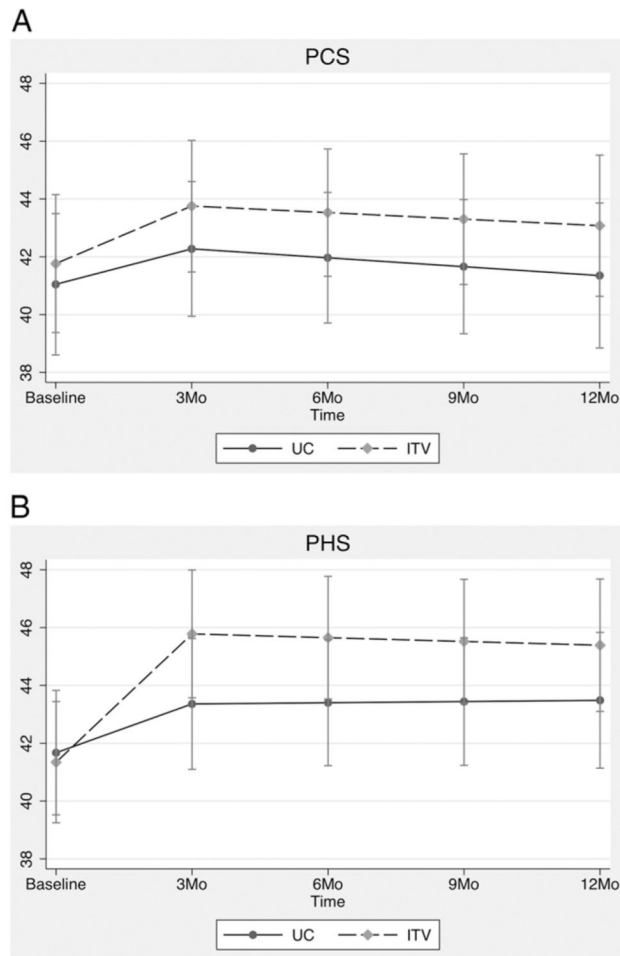(A) PCS and (B) PHS means from the CALM study by time and group.

Fig. 2.
(A) PCS and (B) PHS means from the CCAP study by time and group.

Table 1

Characteristics of patients in the CALM study

	All (*n* = 1004)	Intervention (*n* = 503)	Usual care (*n* = 501)
	No. (%) of patients	No. (%) of patients	No. (%) of patients
Age, mean (S.D.)	43.47 (13.4)	43.3 (13.2)	43.7 (13.7)
Women	714 (71.1)	359 (71.4)	355 (70.9)
Education			
<High school	55 (5.5)	29 (5.8)	26 (5.2)
12 years	166 (16.5)	78 (15.5)	87 (17.4)
>12 years	782 (78.0)	396 (78.7)	388 (77.4)
Race/ethnicity			
Hispanic	196 (19.5)	104 (20.7)	92 (18.4)
Black	116 (11.5)	51 (10.1)	65 (13.0)
White	568 (56.6)	279 (55.5)	289 (57.7)
Other	124 (12.3)	69 (13.7)	55 (11.0)
No. of chronic medical conditions			
0	202 (20.1)	109 (21.7)	93 (18.6)
1	219 (21.8)	108 (21.5)	111 (22.2)
2	582 (58.0)	285 (56.8)	298 (59.3)
Anxiety disorders			
Panic	479 (47.7)	235 (46.7)	244 (48.7)
Generalized anxiety	755 (75.2)	390 (77.5)	365 (72.9)
Social phobia	403 (40.1)	209 (41.6)	194 (38.7)
Posttraumatic stress	181 (18.0)	92 (18.3)	89 (17.8)
Major depressive disorder	647 (64.4)	329 (65.4)	318 (63.5)

Table 2

PCS and PHS means, CIs, Hochberg adjusted significance values and Cohen's *d* effect sizes for patients in intervention compared to usual care across the four assessment points

	Intervention	Usual care	Difference (95% CI)	Adj *P* value	Effect size (*d*)
PCS (SF-12)					
Baseline	49.03 (48.02 to 50.03)	49.33 (48.32 to 50.34)	0.30 (−1.12 to 1.73)	.911	.026
6 months	47.82 (46.80 to 48.85)	47.23 (46.19 to 48.27)	−0.59 (−2.05 to .87)	.911	.051
12 months	47.74 (46.69 to 48.79)	47.82 (46.79 to 48.88)	0.08 (−1.41 to 1.58)	.911	.007
18 months	48.24 (47.18 to 49.29)	47.35 (46.29 to 48.41)	−0.89 (−2.38 to .61)	.911	.077
PHS (SF-12)					
Baseline	42.45 (41.59 to 43.32)	42.64 (41.78 to 43.51)	0.19 (−1.03 to 1.42)	.911	.019
6 months	45.85 (44.95 to 46.75)	43.78 (42.87 to 44.69)	−2.06 (−3.35 to −.78)	.029	.206
12 months	46.37 (45.44 to 47.31)	44.33 (43.39 to 45.27)	−2.05 (−3.37 to −.72)	.036	.200
18 months	46.87 (45.93 to 47.80)	44.75 (43.80 to 45.70)	−2.12 (−3.45 to −.78)	.030	.209