Patient and Practitioner Influences on the Placebo Effect in Irritable Bowel Syndrome

John M. Kelley, PhD, Anthony J. Lembo, MD, J. Stuart Ablon, PhD, Joel J. Villanueva, BA, Lisa A. Conboy, DSc, Ray Levy, PhD, Carl D. Marci, MD, Catherine E. Kerr, PhD, Irving Kirsch, PhD, Eric E. Jacobson, PhD, Helen Riess, MD, and Ted J. Kaptchuk

Objective: To determine whether placebo responses can be explained by characteristics of the patient, the practitioner, or their interpersonal interaction. **Methods:** We performed an analysis of videotape and psychometric data from a clinical trial of patients with irritable bowel syndrome who were treated with placebo acupuncture in either a warm empathic interaction (Augmented, n = 96), a neutral interaction (Limited, n = 97), or a waitlist control (Waitlist, n = 96). We examined the relationships between the placebo response and a) patient personality and demographics; b) treating practitioner; and c) the patient-practitioner interaction as captured on videotape and rated by the Psychotherapy Process Q-Set. **Results:** Patient extraversion, agreeableness, openness to experience, and female gender were associated with placebo response, but these effects held only in the augmented group. Regression analyses controlling for all other independent variables suggest that only extraversion is an independent predictor of placebo response. There were significant differences between practitioners in outcomes; this effect was twice as large as the effect attributable to treatment group assignment. Videotape analysis indicated that the augmented group fostered a treatment relationship similar to a prototype of an ideal healthcare interaction. **Conclusions:** Personality and gender influenced the placebo response, but only in the warm, empathic, augmented group. This suggests that, to the degree a placebo effect is evoked by the patient-practitioner relationship, personality characteristics of the patient will be associated with the placebo response. In addition, practitioners differed markedly in effectiveness, despite standardized interactions. We propose that the quality of the patient-practitioner interaction accounts for the significant difference between the groups in placebo response. **Key words:** placebo effect, irritable bowel syndrome, acupuncture, personality, patient-practitioner relationship.

IBS = irritable bowel syndrome; **FFI** = Five Factor Inventory; **PQS** = Psychotherapy Process Q-Set; **M-PQS** = Modified Psychotherapy Process Q-Set.

INTRODUCTION

Patients in the placebo arms of randomized controlled trials in a variety of disorders often experience considerable clinical improvement. However, a well-publicized meta-analysis suggested that this improvement is attributable to natural history and regression to the mean rather than a placebo effect (1). Contrary to this meta-analysis, our team recently completed a trial consisting of patients with irritable bowel syndrome (IBS) that demonstrated a response to placebo beyond regression and natural history (2). The current study uses data from the parent study to determine whether particular characteristics of the patient, the practitioner, or their interpersonal interaction are associated with the placebo effect.

To date, no specific patient characteristics have been shown consistently to affect the placebo response in clinical trials (3–6). There is evidence that practitioners can have differential effects on patient outcomes in clinical trials (7–10); however, to our knowledge, no one has yet investigated practitioner influences on the placebo effect. Likewise, a great deal has been written on the importance of the patient-practitioner relationship for good clinical outcomes (11–13); however, the effect of the patient-practitioner relationship on the placebo response has not been rigorously analyzed.

In the current study, we sought to determine whether specific patient or practitioner characteristics, or the quality of their interpersonal interactions are associated with the placebo effect. To answer these questions, we used data gathered in a large (n = 289), single-center clinical trial of placebo acupuncture for the treatment of patients with IBS. Specifically, in this report, we analyzed the following three sets of variables: 1) patient personality and demographics; 2) practitioner effects; and 3) the nature of the patient-practitioner interaction as captured on videotapes of treatment sessions.

METHODS

Study Design

The parent study was a single-blind clinical trial in which 289 patients were randomized for 3 weeks to: a) Waitlist (n = 96): patient symptoms were monitored periodically but no treatment was delivered; b) Limited (n = 97): placebo acupuncture was delivered twice a week by a neutral practitioner; and c) Augmented (n = 96): placebo acupuncture was delivered twice a week by a warm, empathic practitioner. In the parent study, after the 3-week primary end point, patients were seamlessly re-randomized to either continue on placebo acupuncture or to receive genuine acupuncture. As the current report focuses on placebo effects, we report the results for the 3-week primary end point only. The three treatment groups were designed to add progressively more placebogenic elements at each level. The waitlist group was designed to control for regression to the mean and natural history, but it also provided patients with two potentially placebogenic factors: 1) attention from the study staff who conducted assessments; and 2) the expectation that they would receive genuine treatment at the conclusion of the trial. The limited group included two sessions of placebo acupuncture per week for 3 weeks with only minimal interaction with the practitioner. Finally, the augmented group also included two sessions of placebo acupuncture per week for 3 weeks; however, in contrast to the limited group, the interaction with the practitioner was warm and empathic. We hypothesized that patient improvement in response to our placebo treatments would be ordered as follows: Augmented > Limited > Waitlist. Details of this design and the clinical results have been published

From the Psychology Department (J.M.K.), Endicott College, Beverly, Massachusetts; Psychiatry Department (J.M.K., J.S.A., J.J.V., R.L., C.D.M., H.R.), Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts; Beth Israel Deaconess Medical Center and Harvard Medical School (A.J.L.), Boston, Massachusetts; Osher Research Center (L.A.C., C.K., T.J.K.), Harvard Medical School, Boston, Massachusetts; Psychology Department (I.K.), University of Hull, Hull, UK; and the Department of Global Health and Social Medicine (E.E.J.), Harvard Medical School, Boston, Massachusetts.

Address correspondence and reprint requests to John M. Kelley, Psychology Department, Endicott College, 376 Hale Street, Beverly, MA 01915. E-mail: JohnKelley@Post.Harvard.Edu

Received for publication June 16, 2008; revision received February 12, 2009.

DOI: 10.1097/PSY.0b013e3181acee12

0033-3174/09/7107-0789

elsewhere (2,14). Institutional Review Boards at Beth Israel Deaconess Medical Center and Harvard Medical School approved the project.

Participants

Between January 2004 and August 2006, 350 prospective participants were screened, and 289 were enrolled into the study. Participants were adults (aged ≥18 years) recruited via advertising and referrals from healthcare providers. Participants met the Rome II criteria for IBS (15) and had a score of ≥150 on the IBS Symptom Severity Scale. The diagnosis of IBS was confirmed by a Board-certified gastroenterologist (A.J.L.). Patients were excluded if they had fever, blood in the stools, recent weight loss of >10% of body weight, or a family history of colon cancer or inflammatory bowel disease. Patients were also excluded if they had previously been treated with acupuncture. Participants were allowed to continue on all IBS prescriptions, over-the-counter, or alternative medications and any psychological treatments as long as they had been on a stable regimen for at least 30 days before the start of the study, and they agreed not to change medications or doses during the study.

Interventions

Placebo Acupuncture

Participants were treated with a validated placebo acupuncture device (16). The patient sees and feels the placebo acupuncture needle appear to pierce the skin, but unbeknown to the patient, the needle retracts back up into its hollow handle as it is pressed into the skin. Six to eight placebo needles were placed for 20 minutes over predetermined, nonacupuncture points on the arms, legs, and abdomen. The needle procedure was identical for the augmented and limited groups. Placebo acupuncture treatments lasted for 20 minutes and occurred twice a week for a total of six sessions. All treatment sessions in both the limited and the augmented groups were videotaped.

Limited Group

Participants in the limited group were treated with the placebo acupuncture device by a practitioner whose interaction with the participant was neutral and business-like. Although practitioners were trained to minimize interpersonal interactions with patients, they were explicitly trained not to act in a negative manner.

Augmented Group

Participants in the augmented group received placebo acupuncture from a practitioner whose interaction with the participant was warm and empathic. In contrast to the limited group, the augmented group practitioner explored the psychosocial stressors associated with the patient's symptoms, as well as the patient's understanding of the "meaning" and causes of his or her symptoms. Throughout the interaction, the practitioner used active listening skills and communicated confidence and positive expectations about the treatment. Practitioners were specifically instructed not to use any cognitive or behavioral techniques that have previously been shown to be helpful in IBS (17–19).

Practitioners and Training Methods

The practitioners in this study were four licensed acupuncturists who all had previous experience administering placebo acupuncture in randomized controlled trials. We provided 20 hours of training to ensure that the practitioners were skilled in delivering the treatments in both the augmented and the limited interaction styles. A training manual and a video of model sessions were provided, and role-play was performed on simulated and real patients. The training methods followed those of earlier studies of the patient-physician interaction (20,21).

Adherence to Treatment Protocols

We conducted a fidelity check on the adherence of practitioners to the treatment protocols. All treatment sessions were videotaped, and 10% were selected randomly for evaluation. Two research assistants otherwise unconnected with the trial separately rated each session, using a well-established methodology (22). Reliability between raters was high ($\kappa = 0.92$), and 97% of the sessions were rated as adherent.

Measures

Combined Outcome

For this study, we constructed a single outcome measure by combining the four global outcome measures from the parent study. We standardized the four outcome measures, then averaged across them, and finally converted this variable to a T score with a mean of 50 and a standard deviation (SD) of 10. Thus, across all patients, the mean outcome score was 50. Higher scores indicate better outcomes. Scores of >50 indicate that the patient's outcome was above the mean for the entire sample, and scores of <50 indicate an outcome below the mean for the sample. For our sample, internal consistency for the combined outcome measure was adequate (Cronbach's $\alpha = 0.74$).

The four outcome measures used to construct the combined outcome were: 1) IBS Symptom Severity Scale, a five-item instrument that assesses the frequency and intensity of abdominal pain and distention as well as patient satisfaction with bowel habit (23); 2) IBS Quality of Life scale, a 34-item instrument that measures the impact of IBS symptoms on patient quality of life (24); 3) IBS Global Improvement Scale, which asks participants to rate their overall symptom improvement on a 7-point scale (25,26); and 4) IBS-Adequate Relief, which asks participants: "Over the past week have you had adequate relief of your IBS symptoms?" (27,28). For our sample, internal consistency for the two multiple item scales was adequate to excellent: Cronbach's α was 0.65 for IBS Symptom Severity Scale and 0.94 for IBS Quality of Life. Outcome measures were administered at baseline and at the 3-week end point.

Personality

The Five Factor Inventory (FFI) is a 60-item instrument that measures the "Big Five" dimensions of personality that typically emerge from factor analysis of large data sets of personality descriptors (29). The Big Five are Extraversion, Neuroticism, Agreeableness, Conscientiousness, and Openness to Experience.

Patient-Practitioner Interaction

The Psychotherapy Process Q-Set (PQS) is a 100-item instrument that is rated by independent clinical judges who are kept blind to treatment condition and outcome (30). The PQS is intended to capture the nature of the therapeutic interaction between patient and practitioner, focusing on the process of that interaction as opposed to its content. Content is what is being said or done, whereas process is the manner in which a verbalization or action is carried out. For example, the phrase "that must have been difficult for you" can have radically different meanings (e.g., empathic or sarcastic), depending on the tone and body language of the speaker. Clinical judges view an entire videotape of a treatment session and then sort the 100 PQS items into a set of categories, ranging from 1 = "least characteristic" to 9 = "most characteristic." The middle category, 5, is reserved for items deemed either neutral or irrelevant to the particular session being rated. In addition, judges follow a forced normal distribution when allocating the PQS items to the categories, such that the numbers of items per category are as follows: category 1 (5 items); category 2 (8 items); category 3 (12 items); category 4 (15 items); category 5 (18 items); category 6 (15 items); category 7 (12 items); category 8 (8 items); and category 9 (5 items). For proper interpretation, it is important to note that all PQS items are bipolar. For example, for item 77 (therapist is tactless), a "9" would indicate that the practitioner was very tactless, a "1" would indicate that the practitioner was very tactful, and a "5" would indicate that the practitioner's behavior in the session was neutral with regard to tact, displaying neither tactfulness nor tactlessness.

Five judges with at least a Bachelor's degree in psychology and 1 year of experience working with our clinical research group received 20 hours of training and were required to rate a minimum of five training sessions each. Two judges were used for each treatment session and their ratings were averaged to increase reliability. A third rater was added if interrater reliability (assessed by the intraclass correlation) fell below $r = .50$. Mean interrater reliability for the augmented condition was good ($r = .83$).

Many of the standard PQS items were very difficult to rate for the limited group because of the restricted nature of the interaction. Therefore, for the purposes of this study, we developed a modified version of the PQS (M-PQS). For the limited condition, the judges only used the two most extreme categories at each end of the distribution (i.e., the two most and the two least characteristic categories), putting the remaining items in the neutral category. This method yielded five categories with the following distributions of items: category 1 (5 items); category 2 (8 items); category 5 (74 items); category 8 (8 items); and category 9 (5 items). To establish a common instrument for comparisons between the limited and augmented groups, the augmented PQS ratings were collapsed to convert them to M-PQS ratings. In particular, categories 3, 4, 6, and 7 were all collapsed into the neutral category, category 5. For the augmented group, the correlation between the PQS and the M-PQS was very high ($r = .95$). All comparisons between the limited and augmented groups use the M-PQS. Interrater reliability for the M-PQS as assessed by intraclass correlations was acceptable ($r = .69$ for the limited condition; $r = .76$ for the augmented condition).

Prototype of an Ideal Healthcare Encounter

For this study, we used the PQS to construct a prototype of an ideal healthcare encounter, using a strategy employed successfully in the past to construct prototypes of different forms of psychotherapy (31–34). We asked nine experienced clinicians to use the PQS to describe their understanding of what an ideal patient-practitioner encounter in a general healthcare setting should look like. There was substantial consensus among the raters (average interrater reliability was $r = .70$). We then averaged across all nine raters to produce a single prototype with a reliability of $r = .95$.

To validate this ideal healthcare prototype, we compared it to previously established and validated prototypes of ideal forms of psychotherapy. Previous research by our group has established validated prototypes for four types of psychotherapy: 1) psychodynamic; 2) cognitive-behavioral; 3) interpersonal; and 4) control mastery (31–35). Although these therapies differ substantially in their underlying theories of change and in the particular techniques they prescribe, there is considerable consensus that all forms of psychotherapy share a set of "common factors" that include empathy, unconditional positive regard for the patient, a good working alliance, and the provision of hope and positive expectations for improvement (36–41). These common factors are similar to standard descriptions of "good bedside manner" in general medical settings. We therefore hypothesized that our ideal healthcare prototype should be highly correlated with the average of the four previously established prototypes. Our reasoning was that averaging should tend to remove the elements specific to each form of therapy and reveal the elements common to all psychotherapies. As hypothesized, our ideal healthcare prototype correlated strongly with the average of the four psychotherapy prototypes ($r = .81$, $p < .001$), providing validation that our ideal healthcare prototype is similar to the common factors of psychotherapy.

Finally, we computed correlation coefficients between the M-PQS ratings for each patient-practitioner dyad and the ideal healthcare prototype. These correlations provide an index of the degree to which each patient-practitioner dyad adhered to the prototype of an ideal healthcare interaction.

Statistical Analysis

Missing data were handled by using the multiple imputation algorithms in the Amelia II program (42) to produce a set of five complete imputed data sets. Each data set was then analyzed separately, using SPSS version 15.0. Parameter estimates from each of the five imputed data sets were then combined, using the procedures advocated by Rubin (43). Associations between the independent variables and outcome were assessed both by zero-order correlations as well as by standardized regression coefficients from ordinary least-squares regression models, controlling simultaneously for all other independent variables. To estimate practitioner effects, we dummy-coded for practitioner, then assessed the significance of the change in R^2 when the block of practitioner variables was added to an ordinary least-squares regression model, controlling for treatment condition and for all patient variables. The interrater reliability for the PQS was assessed, using intraclass correlation coefficients.

RESULTS

Participants

Seventy-five percent of the patients were female, and 88% were Caucasian. Their mean ± SD age was 38 ± 14 years (range = 19–76 years). Ninety-seven percent graduated from high school and 80% were employed. IBS subtypes were as follows: 28% IBS-Diarrhea, 23% IBS-Constipation, and 50% Mixed. Most patients (94%) had IBS for >1 year. At baseline, IBS Symptom Severity Scale indicated that: 8% had mild; 59% had moderate; and 33% had severe symptoms.

Outcomes

All observed outcomes for the parent study at the 3-week end point were consistent with our a priori prediction: Augmented > Limited > Waitlist. The results (mean ± SD) for the combined outcome were: Augmented 54.5 ± 8.8 > Limited 50.5 ± 10.0 > Waitlist 45.5 ± 7.7. Figure 1 illustrates the results for the combined outcome. Linear trend contrasts from analysis of variance for each individual outcome measure as well as for the combined outcome were all statistically significant (for all tests, $p < .001$). Details of these results have been reported elsewhere (2).

Using Cohen's method (44), the standardized effect sizes for the differences between the groups on the combined outcome were: limited versus waitlist, d = 0.51; augmented versus limited, d = 0.46; and augmented versus waitlist, d = 0.99. Conventionally, the first two effect sizes are considered medium and the last effect size is large. The results of the combined measure show that our parent trial produced a

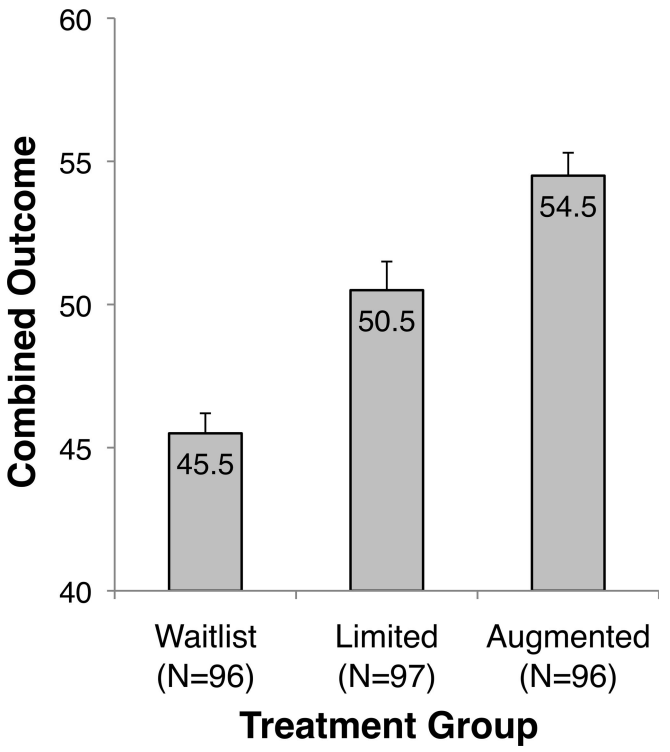

Figure 1. Improvement by treatment group. Error bars represent standard error of the mean.

substantial and statistically significant placebo effect that was beyond natural history and regression to the mean. Moreover, the results show that the effect of placebo acupuncture could be significantly increased by the addition of an augmented patient-practitioner relationship.

Patient Characteristics

Tables 1 and 2 focus on the augmented and limited placebo acupuncture groups, respectively, and show the associations between outcome and patient demographics and personality traits. As can be seen in the tables, the standardized regression coefficients tend to be smaller than the zero-order correlations, indicating that most patient characteristics are not independent predictors of outcome. With the exception of race and gender, the demographic characteristics of the patients were not significantly related to outcome. In the limited group, minority patients (55.4 ± 8.8) had better outcomes than did Caucasians (48.9 ± 9.8), with a standardized effect size of d = 0.70. Minority patients in the limited group had approximately the same mean improvement as did the average patient in the augmented group. In the augmented group, women (55.8 ± 10.3) had significantly better outcomes than men (50.5 ± 8.8) with a standardized effect size of d = 0.55. Men in the augmented group had approximately the same average outcome as the patients in the limited group. After controlling for all other independent variables, the standardized regression coefficients for race and gender were no longer significant.

TABLE 1. Associations Between Patient Characteristics and Outcome for the Augmented Placebo Acupuncture Group

Patient Characteristic	Zero-Order Correlation	Standardized Regression Coefficient
Demographics		
Age	−0.08	−0.03
Female gender	0.22*	0.07
Education	0.11	0.04
Caucasian race	0.07	0.06
Married	−0.08	−0.02
Income	0.05	0.10
Personality		
Extraversion	0.32**	0.25*
Neuroticism	−0.10	0.16
Agreeableness	0.24*	0.16
Conscientiousness	0.08	0.01
Openness to experience	0.25*	0.14

* $p < .05$, ** $p < .01$. Regression coefficients are derived from ordinary least-squares regression, controlling simultaneously for treating practitioner and all other demographic and personality variables.

TABLE 2. Associations Between Patient Characteristics and Outcome for the Limited Placebo Acupuncture Group

Patient Characteristic	Zero-Order Correlation	Standardized Regression Coefficient
Demographics		
Age	−0.08	−0.03
Female gender	0.02	−0.04
Education	0.00	0.02
Caucasian race	−0.25*	−0.20
Married	0.06	0.00
Income	−0.03	0.07
Personality		
Extraversion	0.14	0.14
Neuroticism	0.00	0.15
Agreeableness	0.09	0.10
Conscientiousness	0.10	0.10
Openness to experience	0.03	−0.03

* $p < .05$. Regression coefficients are derived from ordinary least- squares regression controlling simultaneously for treating practitioner and all other demographic and personality variables.

TABLE 3. Correlations Between Patient Characteristics and the Ideal Healthcare Prototype

Patient Characteristic	Limited Group	Augmented Group
Demographics		
Age	−0.01	−0.01
Female gender	0.12	0.14
Education	−0.08	0.25
Caucasian race	−0.12	−0.15
Married	0.15	−0.04
Income	0.23	0.29*
Personality		
Extraversion	0.27	0.29*
Neuroticism	−0.25	−0.50**
Agreeableness	0.19	0.32*
Conscientiousness	0.20	0.13
Openness to experience	0.08	−0.04

* $p < .05$; ** $p < .01$.

Personality traits correlated with patient improvement, but this effect held only in the augmented group. Extraversion, agreeableness, and openness to experience were associated positively with patient improvement ($r = .32$; $r = .24$, and $r = .25$, respectively). Standardized regression coefficients, however, indicated that only extraversion was an independent predictor of outcome.

The correlations between patient characteristics and the ideal healthcare prototype are shown in Table 3. There were very few associations between patient demographic characteristics and the ideal healthcare prototype. For the augmented group, income was associated significantly with ideal healthcare process ($r = .29$), and there was a trend for education ($r = .25$). There were several significant associations between patient personality traits and the ideal healthcare prototype. For the augmented group, extraversion ($r = .29$) and agreeableness ($r = .32$) were associated positively with the ideal healthcare process, whereas neuroticism ($r = -.50$) was negatively associated. A similar pattern held for the limited group, but none of the correlations were significant.

Practitioner Effects

There were dramatic differences between practitioners' in average patient improvement. After controlling for treatment

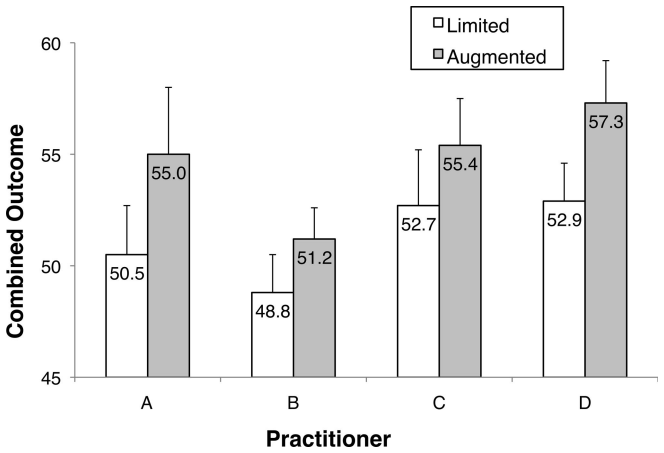

Figure 2. Practitioner effects by treatment group. Error bars represent standard error of the mean.

condition (augmented versus limited), and patient characteristics, practitioners accounted for an additional 6.9% of the variance in outcome ($p = .02$). In contrast, after controlling for practitioner and patient characteristics, treatment condition accounted for only 3.0% of outcome variance ($p = .01$). Thus, the effect attributable to different practitioners was more than twice as large as the effect attributable to treatment condition (i.e., augmented versus limited treatment).

Figure 2 illustrates the practitioner effect. Each practitioner's patients had better outcomes in the augmented group as compared with the limited group. Practitioner B consistently had relatively poor outcomes, whereas Practitioners C and D had more consistently positive outcomes. Practitioner A seemed to be the most successful in altering the therapeutic relationship—this practitioner achieved very good outcomes in the augmented group and relatively poor outcomes in the limited group.

Effects of the Therapeutic Relationship

The M-PQS results indicate that the therapeutic relationships in the augmented and limited groups were strikingly different. The mean correlation with the ideal healthcare prototype was $r = .66$ for the augmented group and $r = .21$ for the limited group. This difference was statistically significant ($t(118) = 13.7$, $p < .001$) with a very large effect size (Cohen's d = 2.5). The mean correlation of the augmented group was >100% of the limited group correlations. These results indicate that the therapeutic process in the augmented group was similar to the prototype of an ideal healthcare interaction. The results for the limited group indicate that the clinical interactions in this group were significantly less ideal, but the positive correlation suggests that the interactions were perceived as being neutral or slightly positive, as opposed to negative or hostile. And indeed, the treatments in the limited group were designed to be neutral, not negative.

Table 4 displays group means for the 20 M-PQS items that most differentiate the two treatment conditions. Independent sample *t* tests indicated that there were significant differences between the augmented and limited groups on all 20 items in the table (for all comparisons, $p < .0001$). Because each therapist conducted treatments in both conditions, the items in Table 4 represent the characteristics that differentiate the treatment conditions, and *not* differences between the therapists. For proper interpretation, it is important to recall that

TABLE 4. Twenty M-PQS Items That Most Differentiate the Augmented From the Limited Treatment Groups (Bold-Faced Items Are More Characteristic of the Limited Group)

Item	Description	Limited	Augmented
3	Therapist's remarks are aimed at facilitating patient speech	1.7	7.0
9	**Therapist is distant, aloof (versus responsive and affectively involved)**	**7.7**	**1.6**
31	Therapist asks for more information or elaboration	3.2	8.0
93	**Therapist is neutral**	**6.9**	**3.6**
45	Therapist adopts supportive stance	4.2	7.0
57	Therapist explains rationale behind his or her technique or approach to treatment	3.9	6.7
16	There is discussion of body functions, physical symptoms, or health	6.3	9.0
6	Therapist is sensitive to the patient's feelings, attuned to the patient; empathic	4.8	7.5
77	**Therapist is tactless**	**4.0**	**1.6**
12	**Silences occur during the hour**	**6.5**	**4.2**
51	**Therapist condescends to, or patronizes the patient**	**3.9**	**1.8**
18	Therapist conveys a sense of nonjudgmental acceptance	5.5	7.6
69	Patient's current or recent life situation is emphasized in discussion	5.2	7.2
65	Therapist clarifies, restates, or rephrases patient's communication	4.4	6.4
37	**Therapist behaves in a teacher-like (didactic) manner**	**7.4**	**5.5**
4	The patient's treatment goals are discussed	3.7	5.6
58	**Patient resists examining thoughts, reactions, or motivations related to problems**	**4.9**	**3.0**
96	**There is discussion of scheduling of hours, or fees.**	**7.0**	**5.2**
17	**Therapist actively exerts control over the interaction (e.g., structuring, introducing new topics)**	**8.6**	**6.9**
99	**Therapist challenges the patient's view (versus validates the patient's perceptions)**	**4.9**	**3.4**

The augmented and limited groups differ significantly on all M-PQS items in the table (for all comparisons, $p < .0001$). M-PQS = Modified Psychotherapy Process Q-Set.

PQS items are bipolar, and therefore, that ratings at or near 5 are neutral with respect to the characteristic being rated.

The following characteristics most differentiated the two treatment conditions. When the practitioners were working with patients in the augmented group, they were more empathic, more sensitive to the patient's feelings, more supportive of the patient, more validating of the patient's perceptions, more nonjudgmental and accepting of the patient, and more responsive and affectively involved. They were more likely to facilitate patient speech by asking for more information or elaboration, and they were more likely to rephrase the patient's speech to clarify its meaning. They were more likely to explain the rationale for the treatment, more likely to discuss the patient's bodily symptoms, and more likely to focus on recent events in the patient's life.

When therapists were working with patients in the limited group, they were more likely to be distant and aloof, more likely to behave didactically, more likely to be neutral, and more likely to control the interaction. They were less likely to be tactful and less likely to validate the patient's perceptions. For their part, the patients in the limited condition were less likely to examine thoughts and motivations related to their problems. This reduction in self-reflection occurred most likely as a result of the practitioners' reduced efforts to encourage patient speech.

DISCUSSION

The parent study showed that patients with IBS treated with placebo acupuncture experienced symptom improvements that were beyond natural history and regression to the mean, and that a positive therapeutic relationship can further increase the effect of placebo acupuncture (2). Using the combined measure computed for this study, the effect sizes for the difference between the two placebo groups and the waitlist were substantial, ranging from $d = 0.46$ to $d = 0.99$. By way of comparison, the mean effect size for all randomized controlled trials of antidepressants that were submitted to the Food and Drug Administration between 1987 and 2004 was $d = 0.31$ (45).

In this study, we have shown that some patient characteristics (extraversion, agreeableness, openness to experience, and female gender) were associated with symptom improvement, but these effects held only in the augmented group. Regression analyses indicate that the most robust patient effect was for extraversion. The other patient effects are no longer significant after controlling for all other independent variables, suggesting that they overlap with one another in predicting outcome variance.

In addition, we have shown that different practitioners can have relatively large differential effects on the placebo response, even when treatment is highly scripted and standardized. Finally, we have shown that practitioners in the augmented treatment group successfully fostered a therapeutic relationship with their patients, which was highly correlated with the prototype of an ideal healthcare encounter. In contrast, practitioners in the limited treatment group fostered a therapeutic interaction that was only modestly correlated with the ideal healthcare prototype. This sharp difference in the nature of the therapeutic relationship is most likely responsible for the significant difference in patient outcomes between the augmented and limited treatment groups in the parent study.

In the augmented and limited treatment groups, patients received identical placebo acupuncture treatments from the same set of four acupuncturists. The only difference between these two conditions was the quality of the interpersonal interaction between practitioner and patient. When working in the augmented condition, practitioners were trained to deliver treatments in a warm, empathic manner; and when working in the limited condition, practitioners were instructed to deliver treatments in a neutral fashion, with minimal interpersonal interaction. The correlations with the ideal healthcare prototype confirm that the treatments delivered in the augmented condition were markedly different from those delivered in the limited condition. Moreover, these correlations show that the augmented treatment was similar to an ideal healthcare interaction, and the limited treatment bore only a slight resemblance to the ideal. The substantial placebo effect observed in this study is most likely the result of the warm, empathic clinical interaction that developed when practitioners worked in the augmented condition.

As powerful as the clinical interaction was, the effect associated with the practitioners themselves was more than twice as large. This study adds to previous work emphasizing the remarkable differences that sometimes occur between practitioners in their effects on patient outcomes (7,46–48). The significant practitioner effect seen in our study is particularly striking, given the fact that the patient-practitioner interactions were manualized, thus minimizing differences between practitioners within each treatment condition. Because patients were not assigned randomly to practitioners, these results must be interpreted cautiously, and future placebo studies that randomly assign patients to practitioners are needed to confirm these results.

Individual differences between patients also seemed to influence the placebo effect in this study; however, most of the patient characteristics that had a significant association with outcome occurred exclusively in the augmented group. In particular, female gender and the personality traits of extraversion, agreeableness, and openness to experience were associated positively with patient improvement. The regression analyses indicate that the most robust personality effect was for extraversion, and that the other patient characteristics were not independent predictors of outcome. Our finding on gender is in line with other studies that have found a role for gender in modulating placebo responses (49). We speculate that female patients and extraverted and agreeable patients who are open to new experiences may have generated a positive feedback cycle with the warm and empathic practitioners in the augmented group, such that practitioner attempts at deepening the therapeutic relationship were rewarded by positive patient responses, thus further increasing practitioner efforts to

deepen the relationship. These same personality traits, however, would be of less benefit in the limited group, in which the practitioners attempted to maintain a neutral, business-like manner.

These speculations are partially supported by the correlations between patient characteristics and the prototype of an ideal healthcare interaction. In the augmented treatment group, neuroticism was negatively correlated with positive therapeutic interactions, whereas extraversion and agreeableness were positively associated. Although there were no significant correlations of personality with the ideal healthcare prototype in the limited condition, there were trends for a positive correlation for extraversion and a negative correlation for neuroticism, mirroring the findings in the augmented group. For demographics, the only significant effect was that income was positively associated with more ideal healthcare interactions in the augmented group (there was a trend for this effect in the limited group). Finally, there was a trend for education to be positively associated with ideal healthcare interactions in the augmented group only.

The negative association between Caucasian race and outcomes in the limited treatment group is intriguing. We have anecdotal evidence from a separate but nested qualitative study (50) that some patients in the limited group interpreted the minimal interaction with their practitioner in a positive way (e.g., "her demeanor is very meditative almost . . . I don't experience the silence as being cold or unfriendly. I experience it more as being a form of meditation . . . quieting the mind; quieting the self."). Thus, some patients seem to have taken the limited interaction as evidence of their practitioner's intense focus on their bodily symptoms and on the acupuncture treatment itself. Many of the non-white patients in this study may have come from cultures in which limited patient-practitioner interactions are more common and more accepted, and this might explain why the racial effect held in the limited treatment group, but not in the augmented group.

Given that this trial was restricted to patients with IBS, it remains unclear whether our results would generalize to other gastrointestinal disorders or more broadly to diseases beyond gastrointestinal disorders. However, because the "treatment" (i.e., the warm, empathic augmented arm of the trial) was not specifically tailored to IBS, we suspect that the results might generalize to other disorders, especially those like IBS for which the outcome measures are primarily subjective in nature. An important question that remains for future research is whether our results would generalize to disorders that have more objective markers. In essence, the question is whether the effects are limited to changes in selective attention to symptoms or to more fundamental changes in the specific biology of the disorder itself.

Similarly, it remains unclear whether our findings would generalize beyond placebo acupuncture treatments. We know from previous work by our team that acupuncture tends to produce a stronger placebo effect than pill placebos (51,52). We speculate that the robust placebo effect of acupuncture is due to the dramatic power associated with placing needles into a patient's body. We presume that the placebo effect of our augmented patient-practitioner interaction must depend, at least in part, on the delivery of a credible treatment that instills positive expectancies in the patient. We therefore speculate that the beneficial effects of our augmented treatment would generalize beyond placebo acupuncture as long as the placebo treatment itself was credible. We plan to test this conjecture in future placebo studies that focus on conventional medical treatments, such as pharmacotherapy, as opposed to alternative treatments, such as acupuncture.

The mechanisms by which the patient-practitioner interaction achieved its effects also remain unclear. Laboratory studies have implicated a number of biochemical, neuroendocrine, and neuroanatomical correlates of the placebo effect, but these studies did not separate out the effect of the treatment interaction from the effect of the placebo treatment itself (53–55). Future clinical and laboratory studies that separate these two components of the placebo effect could begin to identify the mechanisms that underlie each of their effects. At the psychological level, we speculate that the augmented patient-practitioner interaction might have reduced stress and increased expectancies for improvement. In turn, these psychological changes may have improved immune function and reduced pain perception via decreases in stress hormones and increases in the release of endogenous opioids, respectively.

There are a number of limitations to the study. The most important limitation is that, with the exception of the treatment relationship, none of the independent variables we examined were experimentally manipulated. Thus, aside from the findings associated with the treatment relationship, our results must be considered as hypothesis-generating at this point, and should not be considered confirmed until they have been replicated. A second limitation is that this study is the first to use the PQS in a somatic treatment setting; hence, the validity of this form of measurement of the therapeutic relationship may be questioned. A third limitation is that this study focuses on a novel and alternative form of treatment (sham acupuncture) and it is unclear whether our findings would generalize to placebo effects in mainstream medicine. Finally, our study did not include long-term follow-up, and it remains unclear whether the effects we have identified endure beyond 3 weeks.

Overall, our results suggest that patient characteristics are more likely to influence the portion of the placebo effect that is associated with the interpersonal interaction between patient and practitioner (i.e., the augmented group), and *not* the portion that is associated with the placebo device or medication itself (i.e., the limited group). It is possible that the conflicting literature on patient characteristics as predictors of the placebo response may have arisen from differences in the degree to which placebo effects were evoked by the placebo therapy itself as opposed to the treatment relationship. The present study is one of the first to prospectively separate these different components of the placebo effect; and our findings suggest that to the degree that a placebo effect is evoked by the patient-practitioner relationship, personality characteristics of the patient will be associated with placebo response. Con-

versely, to the degree that a placebo effect is evoked primarily through the placebo device or placebo medication with only minimal interaction with a practitioner, then patient characteristics will be less important in predicting placebo response.

A second major finding from this study is that the effect associated with variation in practitioners was even more powerful than the effect attributable to the different treatment groups to which patients were randomized. Because the treatments were highly scripted and variability between practitioners was minimized, this finding was unexpected. Future research that seeks to replicate and extend this finding should be a high priority in placebo studies. In particular, placebo studies that randomly assign patients to practitioners can test prospectively whether these practitioner effects are robust. If practitioner effects can be replicated in future placebo studies, the next logical step is to attempt to elucidate what differentiates successful practitioners from their less successful colleagues.

One anonymous reviewer of this paper suggested an intriguing hypothesis to explain the differences in outcomes between practitioners. The practitioners themselves may have varied in the degree to which their natural inclinations were aligned either with the warm, empathic augmented treatment or with the neutral limited; and this degree of alignment might affected the "genuineness" with which each treatment was delivered. Treatments delivered in a more genuine fashion would presumably produce better outcomes. Although the design of our study precluded a test of this hypothesis, future studies could address this question.

Finally, we have shown that our augmented treatment is similar to the prototype of an ideal healthcare interaction, and we propose that the much greater adherence to this prototype in the augmented group most likely accounts for the significantly better outcomes in this group as compared with the limited group. Additional research is required to determine whether our findings regarding placebo acupuncture effects could generalize to placebo effects in mainstream medicine. In addition, future research should focus on determining whether our positive findings about the power of the therapeutic relationship in placebo effects can be extended to active treatments.

This publication was made possible by Grant 1 R21 AT002564 (J.M.K., J.S.A., J.J.V., R.L., C.D.M., I.K., H.R., T.J.K.), Grant 1 K24 AT004095 (T.J.K.), and Grant 1R01 AT001414-01 (J.M.K., A.JL., L.A.C., C.E.K., E.E.J., H.R., T.J.K.) from the National Center for Complementary and Alternative Medicine (NCCAM). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NCCAM, or the National Institutes of Health.

REFERENCES

1. Hrobjartsson A, Gotzsche PC. Is the placebo powerless? An analysis of clinical trials comparing placebo with no treatment. N Engl J Med 2001;344:1594–602.
2. Kaptchuk TJ, Kelley JM, Conboy LA, Davis RB, Kerr CE, Jacobson EE, Kirsch I, Schyner RN, Nam BH, Nguyen LT, Park M, Rivers AL, McManus C, Kokkoto E, Drossman DA, Goldman P, Lembo AJ. Components of placebo effect: randomized controlled trial in patients with irritable bowel syndrome. BMJ 2008;336:998–1003.
3. Shapiro AK, Shapiro E. The Powerful Placebo: From Ancient Priest to Modern Physician. Baltimore: Johns Hopkins University Press; 1997.
4. Brody H. The Placebo Response. New York: HarperCollins; 2000.
5. Hyland ME, Whalley B, Kirsch I. Consistency of the placebo effect. J Psychosom Res 2008;64:537–41.
6. Kaptchuk TJ, Kelley JM, Deykin A, Wayne PM, Lasagna LC, Epstein IO, Kirsch I, Wechsler ME. Do "placebo responders" exist? Contemp Clin Trials 2008;29:587–95.
7. Kim D, Wampold BE, Bolt DM. Therapist effects in psychotherapy: a random-effects modeling of the National Institute of Mental Health treatment of depression collaborative research program data. Psychotherapy Research 2006;16:161–72.
8. Luborsky L, McLellan AT, Diguer L, Woody G, Seligman DA. The psychotherapist matters: comparison of outcomes across twenty-two therapists and seven patient samples. Clinical Psychology: Science and Practice 1997;4:53–65.
9. Wampold BE, Brown GS. Estimating variability in outcomes attributable to therapists: a naturalistic study of outcomes in managed care. J Consult Clin Psychol 2005;73:914–23.
10. Thomas KB. General practice consultations: is there any point in being positive? Br Med J (Clin Res Ed) 1987;294:1200–2.
11. Hall JA, Harrigan JA, Rosenthal R. Nonverbal behavior in clinician-patient interaction. Appl Prev Psychol 1995;4:21–37.
12. Di Blasi Z, Harkness E, Ernst E, Georgiu A, Kleijnen J. Influence of context effects on health outcomes: a systematic review. Lancet 2001; 357:757–62.
13. Stepien KA, Baernstein A. Educating for empathy. A review. J Gen Intern Med 2006;21.
14. Conboy LA, Wasserman RH, Jacobson EE, Davis RB, Legedza ATR, Park M, Rivers AL, Morey EB, Nam BH, Lasagna L, Kirsch I, Lembo AJ, Kaptchuk TJ, Kerr CE. Investigating placebo effects in irritable bowel syndrome: a novel research design. Contemp Clin Trials 2006;27: 123–34.
15. Thompson WG, Longstreth G, Drossman D, Heaton K, Irvine E, Muller-Lissner S. Functional bowel disorders and functional abdominal pain. Gut 1999;45(Suppl 2):II43–II47.
16. Streitberger K, Kleinhenz J. Introducing a placebo needle into acupuncture research. Lancet 1998;352:354–65.
17. Drossman DA, Toner BB, Whitehead WE, Diamant NE, Dalton CB, Duncan S, Emmott S, Proffitt V, Akman D, Frusciante K, Le T, Meyer K, Bradshaw B, Mikula K, Morris CB, Blackman CJ, Hu Y, Jia H, Li JZ, Koch GG, Bangdiwala SI. Cognitive-behavioral therapy versus education and desipramine versus placebo for moderate to severe functional bowel disorders. Gastroenterology 2003;125:19–31.
18. Keefer L, Blanchard EB. The effects of relaxation response meditation on the symptoms of irritable bowel syndrome: results of a controlled treatment study. Behav Res Ther 2001;39:801–11.
19. Bengtsson M, Ulander K, Borgdal EB, Christensson AC, Ohlsson BA. A course of instruction for women with irritable bowel syndrome. Patient Educ Coun 2006;62:118–25.
20. Lang EV, Benotsch EG, Fick LJ. Adjunctive non-pharmacological analgesia for invasive medical procedures: a randomised trial. Lancet 2000: 1486–90.
21. Morag E, Lieberman G, Volkan K, Shaffer K, Novelline R, Lang EV. Clinical competence assessment in radiology: introduction of an objective structured clinical examination in the medical school curriculum. Acad Radiol 2001;8:74–81.
22. Moncher FJ, Prinz FJ. Treatment fidelity in outcomes studies. Clin Psychol Rev 1991;11:247–66.
23. Francis C, Morris J, Whorwell P. The irritable bowel severity scoring system: a simple method of monitoring irritable bowel syndrome and its progress. Aliment Pharmacol Ther 1997;11:395–402.
24. Drossman DA, Morris CB, Hu Y, Toner BB, Diamant N, Whitehead WE, Dalton CB, Leserman J, Patrick DL, Bangdiwala SI. Characterization of health related quality of life (HRQOL) for patients with functional bowel disorder (FBD) and its response to treatment. Am J Gastroenterol 2007; 102:1442–53.
25. Lembo T, Wright RA, Bagby B, Decker C, Gordon S, Jhingran P, Carter E. Alosetron controls bowel urgency and provides global symptom improvement in women with diarrhea-predominant irritable bowel syndrome. Am J Gastroenterol 2001;96:2662–70.
26. Gordon S, Ameen V, Bagby B, Shahan B, Jhingran P, Carter E. Valida-

tion of irritable bowel syndrome Global Improvement Scale: an integrated symptom end point for assessing treatment efficacy. Dig Dis Sci 2003;48:1317–23.
27. Mangel AW. Personal view: adequate relief as a primary endpoint in irritable bowel syndrome. Aliment Pharmacol Ther2006;23:879–81.
28. Mangel AW, Hahn BA, Heath AT, Northcutt AR, Kong S, Dukes GE, McSorley D. Adequate relief as an endpoint in clinical trials in irritable bowel syndrome. J Int Med Res 1998;26:76–81.
29. Costa PT, McCrae RR. NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI): Professional Manual. Odessa, FL: Psychological Assessment Resources; 1992.
30. Jones EE. Therapeutic Action: A Guide to Psychoanalytic Therapy. Northvale, NJ: Jason Aronson; 2000.
31. Ablon JS, Jones EE. How expert clinicians' prototypes of an ideal treatment correlated with outcome in psychodynamic and cognitive-behavioral therapy. Psychother Res 1998;8:71–83.
32. Ablon JS, Jones EE. Validity of controlled clinical trials of psychotherapy: findings from the NIMH Treatment of Depression Collaborative Research Program. Am J Psychiatry 2002;159:775–83.
33. Pole N, Ablon JS, O'Connor L. Ideal control mastery technique correlates with change in a single case. Psychotherapy: Theory, Research, Practice, Training 2002;39:88–96.
34. Pole N, Ablon JS, O'Connor L. Using psychodynamic, cognitive-behavioral, and control mastery prototypes to predict change: a new look at an old paradigm for long-term single case research. J Counseling Psychol 2008;55:221–32.
35. Ablon JS, Jones EE. Psychotherapy process in the NIMH collaborative study of depression. J Consult Clin Psychol 1999;67:64–75.
36. Wampold BE. The Great Psychotherapy Debate: Models, Methods, and Findings. Mahwah, NJ: Lawrence Erlbaum; 2001.
37. Hubble MA, Duncan BL, Miller SD. The Heart and Soul of Change: What Works in Psychotherapy. Washington, DC: American Psychological Association; 1999.
38. Frank JD. Therapeutic Factors in Psychotherapy. Am J Psychother 1971; 25:350–61.
39. Ackerman SJ, Hilsenroth MJ. A review of therapist characteristics and techniques positively impacting the therapeutic alliance. Clin Psychol Rev 2003;23:1–33.
40. Stiles WB, Shapiro DA, Elliott R. "Are All Psychotherapies Equivalent?" Am Psychol 1986;41:165–80.
41. Horvath AO, Symonds BD. Relation between working alliance and outcome in psychotherapy: a meta-analysis. J Counseling Psychol 1991; 38:139–49.
42. Honaker J, King G, Blackwell M. Amelia II: A Program for Missing Data. Manual. 2006. Available at http://gking.harvard.edu/amelia/docs/amelia.pdf.
43. Rubin DB. Multiple Imputation for Nonresponse in Surveys. New York: John Wiley & Sons; 1987.
44. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
45. Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R. Selective publication of antidepressant trials and its influence on apparent efficacy. N Engl J Med 2008;358:252–60.
46. Blatt SJ, Sanislow CA, Zuroff DC, Pilkonis PA. Characteristics of effective therapists: further analyses of data from the National Institute of Mental Health treatment of depression collaborative research program. J Consult Clin Psychol 1996;64:1276–84.
47. Najavits LM, Weiss RD. Variations in therapist effectiveness in the treatment of patients with substance use disorders: an empirical review. Addiction 1994;89:679–88.
48. Crits-Christoph P, Mintz J. Implications of therapist effects for the design and analysis of comparative studies of psychotherapies. J Consult Clin Psychol 1991;59:20–6.
49. Klosterhalfen S, Enck P. Psychobiology of the placebo response. Autonomic Neuroscience 2006;125:94–9.
50. Kaptchuk TJ, Shaw J, Kerr CE, Conboy LA, Kelley JM, Csordas TJ, Lembo AJ, Jacobson EE. Maybe I made up the whole thing: placebos and patients' experiences in a randomized controlled trial. Cult Med Psychiatry, In Press.
51. Kaptchuk TJ, Stason WB, Davis RB, Legedza ATR, Schnyer RS, Kerr CE, Stone DA, Nam BH, Kirsch I, Goldman P. Sham device versus inert pill: a randomized controlled trial comparing two placebo treatments for arm pain due to repetitive strain injury. BMJ 2006;332:391–7.
52. Kaptchuk TJ, Goldman P, Stone DA, Stason WB. Do medical devices have enhanced placebo effects? J Clin Epidemiol 2000;53:786–92.
53. Benedetti F, Mayberg HS, Wager TD, Stohler CS, Zubieta JK. Neurobiological mechanisms of the placebo effect. J Neurosci 2005;25: 10390–402.
54. Pacheco-López G, Engler H, Niemi MB, Schedlowski M. Expectations and associations that heal: immunomodulatory placebo effects and its neurobiology. Brain Behav Immun 2006:430–46.
55. Enck P, Benedetti F, Schedlowski M. New insights into the placebo and nocebo responses. Neuron 2008;59:195–206.