PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Assessment of patients' expectations: development and validation of the Expectation for Treatment Scale (ETS)
AUTHORS	Barth, Jürgen; Kern, Alexandra; Lüthi, Sebastian; Witt, Claudia M.

VERSION 1 - REVIEW

REVIEWER	Johannes Laferton Philipps-University of Marburg Germany
REVIEW RETURNED	25-Oct-2018

GENERAL COMMENTS	Peer Review – BMJ Open – 2018-024290 Title: Assessment of patients' expectations: development and validation of the Expectation for Treatment Scale (ETS) In their manuscript Barth et al. present a three or two step approach in developing a new measure for treatment expectations that tries to overcome the problem of ceiling effects and treatment specificity. They use a strong methodological approach (three independent samples including exploratory and confirmatory evaluation of the scale in pain and acupuncture patients). The initial results are convincing that this scale is promising to be evaluated further as treatment expectation measure that might be valuable to an important and broad field of research. The strong methodological approach and the promising results, however, might benefit from more precision in the introduction section and the presentation of methods and results. For details please see below. Note: Given that the reviewer is a non-native English speaker, the judgment of sufficient written English (Reviewer item #15) is referred to a native reviewer or editor. Abstract P2L6-12: It might help to outline one of the strengths of the study (exploratory and confirmatory psychometric evaluation in independent samples), if that would be reflected in the methods section of the abstract. P2L13-21: The correlation of the ETS with the existing expectation measure for acupuncture is an important finding regarding convergent validity. It should be mentioned in the abstract. P2L24f: Consider revising the sentence: 'As a next step, it should be psychometrically evaluated among other disorders and treatments…' (as is stated by the authors in the strengths and limitations section) P4L29-31: The authors mention patient involvement as a strength of their approach. It would be helpful, if they could describe the

way patients were involved in the development of the scale (patient focus groups? Or patient advocacy organization consulting? …)

P5L54: 'used a specific measure'. Do the authors refer to a disease/or treatment specific measure (i.e. only applicable for acupuncture) or to various scales among the different studies?

P5L54-58: '… differences between measures might hide or exaggerate associations between expectations and intervention outcomes.' It might be helpful to clarify how different expectation measures would exaggerate/inflate associations. I would agree that unreliable scales, or scales with lacking validity would produce inconsistent findings. Do the authors refer to that?

P5L59f: The reviewer strongly agrees with the authors that clear definition is key, especially when researching expectations. Therefore, it might improve the manuscript further, if the definition of expectations applied by the authors could be more clear, hence explicitly stated. The sentence (P5L61f)'…''patient expectations'' cover treatment-related outcome expectations' appears to suggest a focus on treatment outcome expectations in the realm of patients' expectations. It would be helpful to the reader, if a definition of expectations (i.e. a persons subjective probability of the incidence or non-incidence of an event) would be stated in a previous sentence. Further, without knowing the authors exact definition of expectations it is not entirely clear why they would disregard 'self-efficacy' (personal expectation to be capable to perform/endure performing certain actions) or optimism (generalized outcome expectations) of being an expectation construct. Knowing the authors' clear definition of expectations might further help to understand the sentence (P5L64f) 'Related constructs are sometimes included in expectation measures, and in other cases, only cognitions about treatment outcome relationships are included`. It would appear to the reviewer, that treatment outcome expectations could also be categorized as future oriented subjective cognitions about a treatment outcome relationship. In this crucial abstract it might be helpful to explicitly describe a conceptualization of expectations (definition, different sub-constructs) to the reader. The manuscript already refers to valuable manuscripts offering such conceptualizations i.e. the Maddux chapter in reference #2 or the conceptualization of expectations in Ref#16, which integrates expectation concepts of reference #12,#13,#14, #21 and the Maddux chapter in Ref#2.

In the end, the reviewer agrees that a focus on treatment outcome expectations is a valid argument, especially when opting for a brief and general measure. However, especially a reader not expert to the field of expectations, might benefit from a more explicit introduction into the concept(s).

P6L72f: The reviewer agrees with the problem of ceiling effects. It might be even more striking if the authors would add the reason why this is a problem (i.e. insufficient variance in the predictor variable).

P6L69-81: This para lists shortcomings in the field of expectation measures regarding acupuncture and alternative medicine treatments. Since the authors opt for a measure that is generally applicable to different conditions, it might also strengthen their cause, if they would list shortcomings in the field of expectation measurement in general. See reference #12,16 or (Haanstra et al., 2012; van Hartingsveld et al., 2010; Zywiel, Mahomed, Gandhi, Perruccio, & Mahomed, 2013)

Haanstra, T. M., van den Berg, T., Ostelo, R. W., Poolman, R. W., Jansma, E. P., Jansma, I. P., … de Vet, H. C. (2012). Systematic review: do patient expectations influence treatment outcomes in

total knee and total hip arthroplasty? Health and Quality of Life Outcomes, 10(August), 152. http://doi.org/10.1186/1477-7525-10-152

van Hartingsveld, F., Ostelo, R. W. J. G., Cuijpers, P., de Vos, R., Riphagen, I. I., & de Vet, H. C. W. (2010). Treatment-related and patient-related expectations of patients with musculoskeletal disorders: a systematic review of published measurement tools. The Clinical Journal of Pain, 26(6), 470–88. http://doi.org/10.1097/AJP.0b013e3181e0ffd3

Zywiel, M. G., Mahomed, A., Gandhi, R., Perruccio, A. V, & Mahomed, N. N. (2013). Measuring Expectations in Orthopaedic Surgery: A Systematic Review. Clinical Orthopaedics and Related Research, 471, 3446–56. http://doi.org/10.1007/s11999-013-3013-8

P7L90-96: The reviewer is not exactly clear whether the development of the ETS (reduction from 17 to 9 items) is part of this study (since the reduction from 17 to 9 is mentioned in the abstract, or an (unpublished?) pilot study. If possible, it would be interesting to include the first step (17 to 9 items) as an appendix, including a brief description of the sample and the statistics of the items used for item selection. I think using three different samples and the selection of items based on empirical data on variance to reduce ceiling effects is a strong methodological approach, which deserves more credit within this manuscript.

Moreover, as mentioned regarding the abstract, it might be helpful to guide the reader through the development: i.e. ' first study 1: excluding items with ceiling effects in sample XY; second study two, explorative investigation of psychometric merit in sample YZ; step 3: confirming psychometric merit in sample YY.

P7L107-108: patients were excluded due to insufficient pain…. Please explicitly state inclusion and exclusion criteria for you samples used.

Discussion: Initial data on the psychometric properties of the ETS are promising. However, so far there is no data on predictive validity (the RCT could be used for this, now or in the future?), and construct validity is not entirely clear. I.e. correlations of the ETI with personal control expectations/self-efficacy would be interesting for future investigations to determine the specific aspect of expectations the ETS is measuring. While most items of the ETS have face validity regarding treatment outcome expectations, the coping item might be considered a personal control expectation. Moreover, psychometric evaluation in different samples (other disease/treatment) are necessary before advocating the (routine) use of the scale in different samples. Additionally, investigating sensitivity to change would be valuable. Hence, to advance the rigorous development of this initially promising scale and to motivate other researchers to conduct investigations, the discussion might benefit of balancing the promising results with open questions regarding the scale development process.

P20L398-400: 'The ETS is a short and validated measure that can be used in the German and English languages and can contribute to the understanding of patient expectations for treatment outcomes.'

I believe this statement is premature. The authors do not mention any process of translation into English (something that might enhance the manuscript) and there is no evidence for the application in of the ETS in other patient populations.

	Table 4: I can only guess what the three different values refer to (r, p-value, and n?). Please revise the table.
	Appendix 2: This table is (at least the way I downloaded it), completely incomprehensible! When submitting the proof pdf, please make sure that the manuscript is on order.
	Minor: P2L2: including 'self-report' instrument might add valuable information P6L72-75: please check this sentence for typos and missing words. P11L191: Do you mean the distinctiveness of the ETS regarding neuroticism?

REVIEWER	Joel Bialosky University of Florida USA
REVIEW RETURNED	07-Dec-2018

GENERAL COMMENTS	Line 66-68: I appreciate the authors attempting to operationally define expectations for their questionnaire; however, I believe this would better fit within the methods.
	Second to last paragraph of the introduction (lines 69-81) are confusingly worded and require clarity. The intention seems to be to indicate limitations in prior expectation measurement tools; however, the AES and EXPECT are singled out in a seemingly arbitrary way
	The development of the ETS requires greater detail in the Methods section lines 91-96. The authors indicate the tool was based on existing questionnaires; however, what was the process for identifying and including the indicated questionnaires. Additionally, the authors indicate creating a list of 17 items; however, was there a systematic process in how these were selected?
	More information is necessary regarding the criteria for inclusion in the study. it appears this was intentionally broad with the only requirements being having sought treatment for a musculoskeletal condition and fluent in German. Can the authors be any more specific? For example, how was "a musculoskeletal condition" defined? Line 108 indicates participants excluded for insufficient pain suggesting a required minimal pain rating. How was pain measured and what was considered "insufficient"?
	Line 129 (NRS for acupuncture success) and 132 (NRS for pain) require greater clarification. How were each anchored? High scores indicate more/ less success/ pain?
	line 158: what is "a high peculiarity for optimism or pessimism"?
	Greater justification is required for the comparative constructs. The authors briefly suggest this in the introduction line 82-88 in stating the tool should correlate highly to prior measures of expectation, moderately to related constructs, and minimally with personality. Justification for this should be better supported. Additionally, depression, resilience, body efficacy expectation, perceived

	sensitivity to medicine, neuroticicm and openness to experience were all included. What was the basis and justification for considering the association of the ETS to these?
	line 276-277: How was "pre- study feedback" obtained?
	I found the discussion to be excessively speculative and beyond the scope of the findings. I believe this section requires extensive edits and should align more with the stated purposes and findings of the study. Specific examples:
	Line 339-348: No basis for suggesting this instrument could be used for "mental disorders" or "lower levels of health literacy"
	Line 349-356 does not add anything to the discussion. This could perhaps be included in the introduction to make the case for limitations in how expectation has traditionally been measured and reported justifying the need for this study.
	line 375-382: Not clear what you are getting at here or how this relates to your study or findings
	The ETS is presented in both and English and German version. Was a formal process undergone to translate one version to the other?
	line 220 to 221 indicates using a total sum score of the ETS in the analysis; however, page 29 (immediately following the German version of the ETS) suggests scoring is done through percentage i.e. "Add up the score of the individual items and use the number of filled in questions as denominator"

VERSION 1 – AUTHOR RESPONSE

Reviewer(s)' Comments to Author:

Reviewer: 1

Reviewer Name: Johannes Laferton

Institution and Country: Philipps-University of Marburg

Germany

Please state any competing interests or state 'None declared': None declared

Please leave your comments for the authors below

Peer Review – BMJ Open – 2018-024290

Title: Assessment of patients' expectations: development and validation of the Expectation for Treatment Scale (ETS)

Comment

In their manuscript Barth et al. present a three or two step approach in developing a new measure for treatment expectations that tries to overcome the problem of ceiling effects and treatment specificity.

They use a strong methodological approach (three independent samples including exploratory and confirmatory evaluation of the scale in pain and acupuncture patients). The initial results are convincing that this scale is promising to be evaluated further as treatment expectation measure that might be valuable to an important and broad field of research. The strong methodological approach and the promising results, however, might benefit from more precision in the introduction section and the presentation of methods and results. For details please see below.

Response

We thank the reviewer for this positive evaluation of the material, and we are happy to provide more details in the revisions. Specifically, we considered many of the conceptual clarifications.

Comment

Note: Given that the reviewer is a non-native English speaker, the judgment of sufficient written English (Reviewer item #15) is referred to a native reviewer or editor.

Response

We sent out the manuscript again for an external professional editing service to improve language and grammar.

Abstract

Comment

P2L6-12: It might help to outline one of the strengths of the study (exploratory and confirmatory psychometric evaluation in independent samples), if that would be reflected in the methods section of the abstract.

Response

This suggestion is an excellent one. We mention the confirmatory factor analysis in the revised manuscript as a strength.

Comment

P2L13-21: The correlation of the ETS with the existing expectation measure for acupuncture is an important finding regarding convergent validity. It should be mentioned in the abstract.

Response

We now also report on convergent validity in the abstract. We added one sentence on this topic:

"Convergent validity was confirmed with a high correlation (r > .90) between ETS and a treatment-specific measure of expectations."

Comment

P2L24f: Consider revising the sentence: 'As a next step, it should be psychometrically evaluated among other disorders and treatments…' (as is stated by the authors in the strengths and limitations section)

Response

We revised the concluding sentence and now provide similar conclusions as already stated in the strengths and limitations section. It now reads,

"As a next step, the ETS might be implemented in different clinical conditions and settings to investigate psychometrics and its predictive power for treatment outcomes."

Comment

P4L29-31: The authors mention patient involvement as a strength of their approach. It would be helpful, if they could describe the way patients were involved in the development of the scale (patient focus groups? Or patient advocacy organization consulting? …)

Response

We see the point that our description in the methods part is not sufficiently comprehensive to obtain insights in our efforts to pre-select proper items for the first version of the scale (9 items) from the initial list of 17 items. We are happy to describe our effort in developing this measure in more detail. It now reads,

"Twenty patients completed the questionnaires, provided sociodemographic information and were asked for written comments about the accessibility of the questions. In addition, two patients were interviewed by a qualitative researcher. Two health professionals (one acupuncturist and a doctor assistant) also verbally commented about the appropriateness of the questions. Based on these data, we selected items with low skewness and a large range of responses (i.e., variation). High correlations between items and the findings from the qualitative feedback were also considered."

Additionally, we prepared more material for the reviewer. We leave it to the Editor whether this material should be presented as an appendix.

Comment

P5L54: 'used a specific measure'. Do the authors refer to a disease/or treatment specific measure (i.e. only applicable for acupuncture) or to various scales among the different studies?

Response

The wording was ambiguous. In fact, the studies most often used specifically designed measures for their own purposes. Therefore, replication is difficult. Some measures are specified for the disease; others refer to a specific intervention, and some mention a disease and an intervention in the items. We have rephrased it as follows:

"These differences in the association of expectations and outcomes might be partly explained by the fact that each study used a newly invented measure, and differences between measures might hide or exaggerate associations between expectations and intervention outcomes."

Comment

P5L54-58: '… differences between measures might hide or exaggerate associations between expectations and intervention outcomes.' It might be helpful to clarify how different expectation measures would exaggerate/inflate associations. I would agree that unreliable scales, or scales with lacking validity would produce inconsistent findings. Do the authors refer to that?

Response

The reviewer mentioned two important quality dimensions (validity and reliability), but our argument refers to the fact that items differ from study to study. In consequence, the label "expectation" might be used, but one study assesses the general belief about treatment effectiveness, and another study assesses the expectation to have less pain at the end of the treatment. According to our understanding, such differences might also be responsible for differences between studies.

We rephrased the sentences (see earlier response), and we hope that the reviewer agrees with this change.

Comment

P5L59f: The reviewer strongly agrees with the authors that clear definition is key, especially when researching expectations. Therefore, it might improve the manuscript further, if the definition of expectations applied by the authors could be more clear, hence explicitly stated. The sentence (P5L61f)'…"patient expectations" cover treatment-related outcome expectations' appears to suggest a focus on treatment outcome expectations in the realm of patients' expectations. It would be helpful to the reader, if a definition of expectations (i.e. a persons subjective probability of the incidence or non-incidence of an event) would be stated in a previous sentence. Further, without knowing the authors exact definition of expectations it is not entirely clear why they would disregard 'self-efficacy' (personal expectation to be capable to perform/endure performing certain actions) or optimism (generalized outcome expectations) of being an expectation construct. Knowing the authors' clear definition of expectations might further help to understand the sentence (P5L64f) 'Related constructs are sometimes included in expectation measures, and in other cases, only cognitions about treatment outcome relationships are included`. It would appear to the reviewer, that treatment outcome expectations could also be categorized as future oriented subjective cognitions about a treatment outcome relationship. In this crucial abstract it might be helpful to explicitly describe a conceptualization of expectations (definition, different sub-constructs) to the reader. The manuscript already refers to valuable manuscripts offering such conceptualizations i.e. the Maddux chapter in

reference #2 or the conceptualization of expectations in Ref#16, which integrates expectation concepts of reference #12,#13,#14, #21 and the Maddux chapter in Ref#2.

In the end, the reviewer agrees that a focus on treatment outcome expectations is a valid argument, especially when opting for a brief and general measure. However, especially a reader not expert to the field of expectations, might benefit from a more explicit introduction into the concept(s).

Response

We now incorporate some definitions of related constructs. By doing so, we want to ensure that the difference from treatment expectations becomes apparent. In addition, we defined expectations (in our case, treatment-related outcome expectations). The paragraph now reads,

"Expectation is a well-known and oft-used term. A clear definition and a sharp distinction from associated constructs is important for the development of a measure[11]. In the context of medical treatments, the term "expectations" describes cognitions about treatment-related health outcomes in the future after a specific intervention[12][13]. Patients can consider a treatment more or less beneficial for their complaints or disease at a specific time-point (i.e., outcome expectations)[13]. Role expectations also capture the role of a patient and the therapist during the treatment. In other words, a patient might consider himself rather inactive during treatment in defining treatment goals and expects an active therapist to achieve a good treatment outcome. However, our purpose was to develop a scale on ''patient expectations'' that covers treatment-related outcome expectations. Bowling et al. (2012) provide an insightful summary about the theoretical underpinning of expectations[14]. Related constructs such as optimism, self-efficacy, and hope share some facets with expectations but differ on the level of the construct[15][16]. Optimism can be viewed as a trait characteristic of a person with high stability over time and situations. Optimism is defined as "the extent to which people hold generalized factorable expectancies for their future" (Carver et al, 2010, p. 879)[16]. Self-efficacy is also a construct at a general level (i.e., "Perceived self-represents an optimistic sense of personal competence […]"; Scholz et al., 2002; p. 342)[17]. If self-efficacy is related to a specific behaviour or problem, it captures the strength of a belief to cope in a situation successfully (for example Pain Self-Efficacy Questionnaire; PSEQ)[18]. Hope should also be considered conceptually different: "Expectations and hopes are very different concepts. Hopes tend to be based more upon emotions or wishes, things that individuals want reality to be, whereas expectations tend to rely more heavily upon rational thought and logical reasoning" Woolhead et al., 2003 p. 1656)[19]. Related constructs are sometimes included in expectation measures, whereas in other cases, only cognitions about treatment outcome relationships are included[9][20][21]. Following the definition by Bowling et al. (2012), we therefore designed our measure to assess expectations related to a clinical intervention with a clinically relevant outcome from a patient's perspective."

Comment

P6L72f: The reviewer agrees with the problem of ceiling effects. It might be even more striking if the authors would add the reason why this is a problem (i.e. insufficient variance in the predictor variable).

Response

We agree with the reviewer that this problem deserves more attention. We therefore added one sentence about the fact that expectations are commonly used as predictors, but in cases of low variation, this research question cannot be answered properly. This sentence was added:

"Second, ceiling effects were a common problem in the measurement of expectations, because patients who are seeking help from a specific treatment often expect large benefits; otherwise, they would not be attracted by this treatment. This problem was apparent in an established expectation measure for acupuncture treatment that served as a benchmark measure for our scale (Acupuncture Expectancy Scale (AES))25. Ceiling effects are particularly problematic because the predictive power of such skewed variables is low. In the case of expectations, many research questions address the prediction of treatment outcomes; therefore, a measure with sufficient variation between patients is needed."

Comment

P6L69-81: This para lists shortcomings in the field of expectation measures regarding acupuncture and alternative medicine treatments. Since the authors opt for a measure that is generally applicable to different conditions, it might also strengthen their cause, if they would list shortcomings in the field of expectation measurement in general. See reference #12,16 or (Haanstra et al., 2012; van Hartingsveld et al., 2010; Zywiel, Mahomed, Gandhi, Perruccio, & Mahomed, 2013)

Haanstra, T. M., van den Berg, T., Ostelo, R. W., Poolman, R. W., Jansma, E. P., Jansma, I. P., … de Vet, H. C. (2012). Systematic review: do patient expectations influence treatment outcomes in total knee and total hip arthroplasty? Health and Quality of Life Outcomes, 10(August), 152. http://doi.org/10.1186/1477-7525-10-152

van Hartingsveld, F., Ostelo, R. W. J. G., Cuijpers, P., de Vos, R., Riphagen, I. I., & de Vet, H. C. W. (2010). Treatment-related and patient-related expectations of patients with musculoskeletal disorders: a systematic review of published measurement tools. The Clinical Journal of Pain, 26(6), 470–88. http://doi.org/10.1097/AJP.0b013e3181e0ffd3

Zywiel, M. G., Mahomed, A., Gandhi, R., Perruccio, A. V, & Mahomed, N. N. (2013). Measuring Expectations in Orthopaedic Surgery: A Systematic Review. Clinical Orthopaedics and Related Research, 471, 3446–56. http://doi.org/10.1007/s11999-013-3013-8

Response

We included the suggested sources in the initial statement of shortcomings. This introduction clarifies that the assessment problem is not only given in the field of CAM but also present in other domains. It now reads,

"Earlier findings about the expectation outcome association in clinical studies have been limited by the diversity of measures. Several authors claimed diversity in covered concepts, time-point of assessment and problems to evaluate the validity of the measures22–24. A strong measure is a prerequisite to accurately predict treatment responses based on pretreatment expectations. A closer investigation of the results from a systematic review about acupuncture expectation measures by Prady and colleagues20 showed that of ten trials, only five provided their exact item wording for measuring expectations."

Comment

P7L90-96: The reviewer is not exactly clear whether the development of the ETS (reduction from 17 to 9 items) is part of this study (since the reduction from 17 to 9 is mentioned in the abstract, or an (unpublished?) pilot study. If possible, it would be interesting to include the first step (17 to 9 items) as

an appendix, including a brief description of the sample and the statistics of the items used for item selection. I think using three different samples and the selection of items based on empirical data on variance to reduce ceiling effects is a strong methodological approach, which deserves more credit within this manuscript.

Moreover, as mentioned regarding the abstract, it might be helpful to guide the reader through the development: i.e. ' first study 1: excluding items with ceiling effects in sample XY; second study two, explorative investigation of psychometric merit in sample YZ; step 3: confirming psychometric merit in sample YY.

Response

We agree that we could mention all of the steps more extensively in the manuscript; however, we would like to adhere to the word limit of the usual paper in this journal. We also would like to mention our multi-method approach and highlight the relevant steps in the pilot study (17 items). In a short paragraph, we describe what we have done. We also uploaded an appendix (items in German; Decision (include / exclude in reduced scale) and rationale for the decision were documented there (for the reviewer)).

"Twenty patients completed the questionnaires, provided sociodemographic information and were asked for written comments about the accessibility of the questions. In addition, two patients were interviewed by a qualitative researcher. Two health professionals (one acupuncturist and a doctor assistant) also verbally commented about the appropriateness of the questions. Based on these data, we selected items with low skewness and a large range of responses (i.e., variation). High correlations between items and the findings from the qualitative feedback were also considered. "

Comment

P7L107-108: patients were excluded due to insufficient pain…. Please explicitly state inclusion and exclusion criteria for you samples used.

Response

We added a short statement about the inclusion criteria. It now reads,

"The included pain patients had to suffer from at least some pain at the day of the assessment (> 0 on a numeric rating scale from 0 (no pain) to 10 (worst pain))."

Comment

Discussion: Initial data on the psychometric properties of the ETS are promising. However, so far there is no data on predictive validity (the RCT could be used for this, now or in the future?), and construct validity is not entirely clear. I.e. correlations of the ETI with personal control expectations/self-efficacy would be interesting for future investigations to determine the specific aspect of expectations the ETS is measuring. While most items of the ETS have face validity regarding treatment outcome expectations, the coping item might be considered a personal control expectation. Moreover, psychometric evaluation in different samples (other disease/treatment) are necessary before advocating the (routine) use of the scale in different samples. Additionally,

investigating sensitivity to change would be valuable. Hence, to advance the rigorous development of this initially promising scale and to motivate other researchers to conduct investigations, the discussion might benefit of balancing the promising results with open questions regarding the scale development process.

Response

Together with some suggestions from reviewer 2, we rephrased parts of the discussion to provide a more balanced interpretation of the strengths of our scale development but mention also open questions after this initial step. We removed parts in which we speculated about applicability, generalizability and predictive power. We rephrased the section as suggested by the reviewer as open questions that deserve attention in the application of the ETS.

Comment

P20L398-400: 'The ETS is a short and validated measure that can be used in the German and English languages and can contribute to the understanding of patient expectations for treatment outcomes.'

I believe this statement is premature. The authors do not mention any process of translation into English (something that might enhance the manuscript) and there is no evidence for the application in of the ETS in other patient populations.

Response

We now describe in the manuscript how we translated the items into English, but we agree that a validation of this English scale must be done.

"These five items were translated into English by two bilingual researchers and translated back into German by two other bilingual researchers. The wording was improved based on feedback from Dr. George Lewith. The final English version is presented in Table 1."

We now do not emphasize the English version in the conclusion. The conclusion now reads,

"The ETS is a short and validated measure that can contribute to the understanding of patient expectations for treatment outcomes."

Comment

Table 4: I can only guess what the three different values refer to (r, p-value, and n?). Please revise the table.

Response

We had already indicated in the heading of the table what the three values indicate. "Table 4. Convergent and divergent validity of the 5-item version of the ETS for subjects with current pain (N = 102). Pearson Correlation, Significance level (2-tailed), N of patients". However, we agree with the

reviewer that the order in which the rows are presented is not 100% clear. We now add in the first column and first row (left upper corner) descriptions of the values.

Comment

Appendix 2: This table is (at least the way I downloaded it), completely incomprehensible! When submitting the proof pdf, please make sure that the manuscript is on order.

Response

We checked the upload. Sorry for this technical problem.

Minor:

Comment

P2L2: including 'self-report' instrument might add valuable information

Response

This suggestion is an excellent one to clarify our approach.

Comment

P6L72-75: please check this sentence for typos and missing words.

Response

We rephrased the sentence. It now reads,

"Second, ceiling effects were a common problem in the measurement of expectations, since patients who are seeking help from a specific treatment often expect large benefits, otherwise they would not go to this treatment. This problem was apparent in one established expectation measure for the case of acupuncture (Acupuncture Expectancy Scale (AES)) 17."

Comment

P11L191: Do you mean the distinctiveness of the ETS regarding neuroticism?

Response

Thank you for making us aware of this misspelled word.

Reviewer: 2

Reviewer Name: Joel Bialosky

Institution and Country: University of Florida

USA

Please state any competing interests or state 'None declared': None declared

Please leave your comments for the authors below

Comment

Line 66-68: I appreciate the authors attempting to operationally define expectations for their questionnaire; however, I believe this would better fit within the methods.

Response

We understand this point of the reviewer, that is, that such conceptual discussions can also be provided in the methods section. However, we would like to take the approach of reviewer 1, who asked for even more conceptual clarity in the background section. We hope that reviewer 2 agrees with this decision.

Comment

Second to last paragraph of the introduction (lines 69-81) are confusingly worded and require clarity. The intention seems to be to indicate limitations in prior expectation measurement tools; however, the AES and EXPECT are singled out in a seemingly arbitrary way

Response:

We choose the AES and EXPECT as reference points for the context of acupuncture, which served us as the treatment we used throughout the study. We fully agree that there are many other measures, but they were developed for many other clinical situations and contexts. We streamlined the paragraph to make it more accessible for readers.

The aim of our study was twofold: First, we wanted to develop a measure that connects to other measures in the field of acupuncture treatment. Second, we wanted to develop a measure that can also be introduced in other clinical domains. We hope the reviewer agrees that this approach reflects a rational choice to highlight these two scales.

Comment

The development of the ETS requires greater detail in the Methods section lines 91-96. The authors indicate the tool was based on existing questionnaires; however, what was the process for identifying and including the indicated questionnaires. Additionally, the authors indicate creating a list of 17 items; however, was there a systematic process in how these were selected?

Response

We now explain in more detail which items were selected from the existing questionnaires in this domain. It now reads,

"Based on existing questionnaires on patient expectations[9 25 27–30], we created a list of 17 items covering different facets of expectations that fit with our aim to develop a measure for treatment-related outcome expectations."

Additionally, we show the 17 items in an appendix for the reviewer (German language only).

Comment

More information is necessary regarding the criteria for inclusion in the study. it appears this was intentionally broad with the only requirements being having sought treatment for a musculoskeletal condition and fluent in German. Can the authors be any more specific? For example, how was "a musculoskeletal condition" defined? Line 108 indicates participants excluded for insufficient pain suggesting a required minimal pain rating. How was pain measured and what was considered "insufficient"?

Response

We agree that this paragraph was not written clearly. Our web-based survey recruited patients, former patients and subjects from the public. With this approach, we were unable to have a formal diagnosis of these patients. Therefore, we asked in this survey only for actual pain intensity measured on a numeric rating scale (NRS from 0 to 10), and we also asked for the location of the pain. If a participant responded with "0" on the NRS, we excluded this person from the analysis. In turn, an NRS from 1 to 10 was the main inclusion criterion. Our sample had a mean pain intensity rating of 6.5, which can be considered clinically relevant.

Intention to seek a treatment was not a requirement. However, the participants had to indicate that they suffered pain irrespective of the localization. Table 2 therefore includes a broad range of different body regions with (musculoskeletal) pain. However, "other" regions also were included (such as headache).

We slightly adapted the respective paragraph to clarify the details of the population in this study.

Comment

Line 129 (NRS for acupuncture success) and 132 (NRS for pain) require greater clarification. How were each anchored? High scores indicate more/ less success/ pain?

Response

We added this information. High values reflect more pain or more success.

Comment

line 158: what is "a high peculiarity for optimism or pessimism"?

Response

We rephrased this sentence. It now reads,

"Each score can range from 0 to 12, with higher values indicating either higher optimism or pessimism."

Comment

Greater justification is required for the comparative constructs. The authors briefly suggest this in the introduction line 82-88 in stating the tool should correlate highly to prior measures of expectation, moderately to related constructs, and minimally with personality. Justification for this should be better supported. Additionally, depression, resilience, body efficacy expectation, perceived sensitivity to medicine, neuroticicm and openness to experience were all included. What was the basis and justification for considering the association of the ETS to these?

Response

It is difficult to have a strong rationale for these assumptions because they are largely based on theoretical work (such as Bowling et al.). In the case of optimism, we could argue that general cognitions about the future (with a positive outcome) should be related to specific cognitions. In turn, however, extraversion (which is a very general and stable construct) should not be that strongly related because expectations can vary over time, whereas personality traits are considered stable. We believe the cited work of Bowling provides a good introduction to these overlaps and differences. We therefore would prefer not to repeat this discussion in our background section. However, we highlighted the importance of this source as follows:

"Bowling et al. (2012) provide an insightful summary about the theoretical underpinning of expectations14."

Additionally, we incorporated some important references mentioned by reviewer 1 in this paragraph, which might provide a better rationale for the choice of measures.

Expectation is a well-known and oft-used term. A clear definition and a sharp distinction from associated constructs is important for the development of a measure[11]. In the context of medical treatments, the term "expectations" describes cognitions about treatment-related health outcomes in the future after a specific intervention[12][13]. Patients can consider a treatment more or less beneficial for their complaints or disease at a specific time-point (i.e., outcome expectations)[13]. Role expectations also capture the role of a patient and the therapist during the treatment. In other words, a patient might consider himself rather inactive during treatment in defining treatment goals and expects an active therapist to achieve a good treatment outcome. However, our purpose was to develop a scale on "patient expectations" that covers treatment-related outcome expectations. Bowling et al. (2012) provide an insightful summary about the theoretical underpinning of expectations[14]. Related constructs such as optimism, self-efficacy, and hope share some facets with expectations but differ on the level of the construct[15][16]. Optimism can be viewed as a trait characteristic of a person with high stability over time and situations. Optimism is defined as "the extent to which people hold generalized factorable expectancies for their future" (Carver et al, 2010, p. 879)[16]. Self-efficacy is also a construct at a general level (i.e., "Perceived self-represents an optimistic sense of personal competence [...]"; Scholz et al., 2002; p. 342)[17]. If self-efficacy is related to a specific behaviour or problem, it captures the strength of a belief to cope in a situation successfully (for example Pain Self-Efficacy Questionnaire; PSEQ)[18]. Hope should also be considered conceptually different: "Expectations and hopes are very different concepts. Hopes tend to be based more upon emotions or wishes, things that individuals want reality to be, whereas expectations tend to rely more heavily upon rational thought and logical reasoning" Woolhead et al., 2003 p. 1656)[19]. Related constructs are sometimes included in expectation measures, whereas in other cases, only cognitions about treatment outcome relationships are included[9][20][21]. Following the definition by Bowling et al. (2012), we therefore designed our measure to assess expectations related to a clinical intervention with a clinically relevant outcome from a patient's perspective.

The reviewer asked for reasons to include other measures. Depression, resilience and self-efficacy might also alter treatment outcomes. Because we aim at predictive models with the ETS, we wanted to ensure that ETS does not measure similar aspects, such as already established predictors. The choice to include sensitivity to medicine is founded on the idea to also use ETS in the context of placebo and nocebo research. We wanted to explore whether expected positive consequences are not only the opposite of expected negative consequences of a medical treatment.

We remain happy with this choice of measures and hope that the reviewer agrees that our decisions are sound.

Comment

line 276-277: How was "pre- study feedback" obtained?

Response

We streamlined the terminology. Pre-study feedback refers to the pilot testing of the 17 items. Therefore, we added some information:

"Twenty patients completed the questionnaires, provided sociodemographic information and were asked for written comments about the accessibility of the questions. In addition, two patients were interviewed by a qualitative researcher. Two health professionals (one acupuncturist and a doctor assistant) also verbally commented about the appropriateness of the questions. Based on these data,

we selected items with low skewness and a large range of responses (i.e., variation). High correlations between items and the findings from the qualitative feedback were also considered. "

Comment

I found the discussion to be excessively speculative and beyond the scope of the findings. I believe this section requires extensive edits and should align more with the stated purposes and findings of the study. Specific examples:

Response

We agree that some parts are beyond the empirical basis of our study. Therefore, we edited the entire section considering both reviewers' comments.

Comment

Line 339-348: No basis for suggesting this instrument could be used for "mental disorders" or "lower levels of health literacy"

Response

We agree that such a statement is premature. We did not specifically examine such a study population. We therefore removed these parts.

Comment

Line 349-356 does not add anything to the discussion. This could perhaps be included in the introduction to make the case for limitations in how expectation has traditionally been measured and reported justifying the need for this study.

Response:

As suggested by the reviewer, we moved this part to the background section in the paragraph about limitations of earlier research. We did not highlight the moved section because it remains unchanged.

Comment

line 375-382: Not clear what you are getting at here or how this relates to your study or findings

Response

We removed this part.

Comment

The ETS is presented in both and English and German version. Was a formal process undergone to translate one version to the other?

Response

We conducted a formal translation of the items and the instructions. A researcher (already mentioned in the initial submission in the acknowledgement) helped us to make a final version, which remains to be validated in English. We added in the manuscript a short description about the translation process.

"These five items were translated into English by two bilingual researchers and translated back into German by two other bilingual researchers. The wording was improved based on feedback from Dr. George Lewith. The final English version is presented in Table 1."

Comment

line 220 to 221 indicates using a total sum score of the ETS in the analysis; however, page 29 (immediately following the German version of the ETS) suggests scoring is done through percentage i.e. "Add up the score of the individual items and use the number of filled in questions as denominator"

Response

Thank you for detecting this wrong formula. The term "score" was used with two meanings (item score and sum score). We have rephrased this instruction. It now reads,

"The values from the 5 single items are summed to build the ETS sum score (min 5, max 20). If one item has a missing value, the sum score can be calculated. Therefore, the values of the remaining 4 items are summed, divided by 4 and multiplied by 5. However, multiple imputation procedures to impute the missing value should be preferred over this re-calculation. In the case of more than one missing value, imputation procedures are needed, and no manual recalculation should be considered.
"

VERSION 2 – REVIEW

REVIEWER	Johannes Laferton Philipps University of Marburg, Germany
REVIEW RETURNED	12-Feb-2019

GENERAL COMMENTS	The authors have responded to all my concerns and the manuscript has further improved.

REVIEWER	Joel Bialosky University of Florida USA
REVIEW RETURNED	05-Mar-2019

GENERAL COMMENTS	Introduction: Line 61-88: The authors have attempted to operationally define expectation and identify "outcome expectations" as the focus of the new measurement tool. Bowling et al is provided as the theoretical model for their approach; however, in my opinion, this paragraph lacks development and distracts from the primary message. Specifically,line 72-85 become a list of related constructs i.e. optimism, pessimism, self efficacy, hope all of which are defined without a direct explanation for their relevance to "outcome expectations" or the design of the measurement tool. This could then be tied into the remainder of the manuscript to support the comparisons supporting convergent validity.
	The derivation of the questionnaire is very superficially described. I am sympathetic to the need to keep the length of the manuscript manageable as well as keeping focus; however, I continue to struggle with how the initial 17 items were chosen (line 128-131). The authors indicate these were based on "existing questionnaires"; however, was there a systematic process to identifying questionnaires and then selecting appropriate items? If so, the process should be indicated. If not, a more descriptive explanation of the process should be provided. Greater detail could also be provided re: the skewness, variation, and correlations observed in the chosen 9 items
	Line 146-147 specifies "patients from the previous year seeking treatment for a musculoskeletal condition"; however, the inclusion criteria was only that the participant experienced experienced pain on the day of the assessment. Are you certain the participants were all experiencing musculoskeletal pain complaints OR would it be more accurate to simply indicate they were seeking treatment for pain?
	Similar to my comment in the initial review, I continue to think the authors must provide a better rationale for the chosen comparative measures. Within the methods, the authors should specify why each measure was considered in validating the scale i.e. optimism, pessimism, depression, resilience, body-efficacy expectation, perceived sensitivity to medication, neuroticism and openness to experience.
	Similarly, in the results, the authors indicate the findings for convergent and divergent validity and specify "we expected" or "which is unexpected". The basis for why these findings were expected or unexpected must be clearly made for the reader earlier in the manuscript.
	I found the discussion improved; however, believe it still requires focus. Specifically:
	line 388-396: Not clear that this adds anything. You found a high correlation between the 2 instruments; however, the relevance of the differences between the instruments is not clear. You indicate in the introduction the relatively low association between the EXPECT and AES in a prior study. Perhaps commenting on why the ETS may have been more highly correlated would add to the discussion
	Line 397-405: I think the lack of association between optimism and the ETS as well as the moderate association between pessimism

	and the ETS is interesting and requires consideration. I found your explanation (line 403-405) to be overly speculative. You may wish to consider this in the theoretical context you establish for expectation. Perhaps considering how this supports or contradicts the Bowling model.

VERSION 2 – AUTHOR RESPONSE

Reviewer(s)' Comments to Author:

Reviewer: 1

Reviewer Name: Johannes Laferton

Institution and Country: Philipps University of Marburg, Germany

Please state any competing interests or state 'None declared': none declared

Please leave your comments for the authors below

Comment

The authors have responded to all my concerns and the manuscript has further improved.

Response

We thank the reviewer for this positive evaluation.

Reviewer: 2

Reviewer Name: Joel Bialosky

Institution and Country: University of Florida, USA

Please state any competing interests or state 'None declared': None declared

Please leave your comments for the authors below

Comment

Introduction: Line 61-88: The authors have attempted to operationally define expectation and identify "outcome expectations" as the focus of the new measurement tool. Bowling et al is provided as the theoretical model for their approach; however, in my opinion, this paragraph lacks development and distracts from the primary message. Specifically,line 72-85 become a list of related constructs i.e.

optimism, pessimism, self efficacy, hope all of which are defined without a direct explanation for their relevance to "outcome expectations" or the design of the measurement tool. This could then be tied into the remainder of the manuscript to support the comparisons supporting convergent validity.

Response

The list of related constructs has been included according to the suggestion of the other reviewer in the last revision. Reviewer 2 already mentioned in the first round that a theoretical introduction of related constructs would be needed in the measurement section and not in the theoretical background. Finally in this revision, we shifted some parts of the background section to the measures and hope that this helps to understand 1) the rationale for the choice of additional measures and 2) the assumptions about the magnitude of the correlation with ETS and the other measures. It now reads as shown in the next comment.

Comment

The derivation of the questionnaire is very superficially described. I am sympathetic to the need to keep the length of the manuscript manageable as well as keeping focus; however, I continue to struggle with how the initial 17 items were chosen (line 128-131). The authors indicate these were based on "existing questionnaires"; however, was there a systematic process to identifying questionnaires and then selecting appropriate items? If so, the process should be indicated. If not, a more descriptive explanation of the process should be provided. Greater detail could also be provided re: the skewness, variation, and correlations observed in the chosen 9 items

Response

The reviewer asked for the selection process of the initial 17 items. We now have included this information.

We especially had been interested in questionnaires from the field of acupuncture research and other non-pharmacological intervention, and the questions should be applicable in an applied context. The first and senior author were responsible for the selection of these items.

Concerning the second point of this comment, we submitted in the last revision additional material for the reviewers including all this information. In the meantime this material was received by the both reviewers and we think that no changes in the Ms are needed for this part.

Comment

Line 146-147 specifies "patients from the previous year seeking treatment for a musculoskeletal condition"; however, the inclusion criteria was only that the participant experienced experienced pain on the day of the assessment. Are you certain the participants were all experiencing musculoskeletal

pain complaints OR would it be more accurate to simply indicate they were seeking treatment for pain?

Response

Since we invited patients from the outpatient clinic who were treated for musculoskeletal pain we would like to keep this information as it is relevant for the recruitment. One inclusion criteria, however, was that patients had to suffer from pain. The first part of this paragraph describes the recruitment, the second part refers to the inclusion criteria.

Comment

Similar to my comment in the initial review, I continue to think the authors must provide a better rationale for the chosen comparative measures. Within the methods, the authors should specify why each measure was considered in validating the scale i.e. optimism, pessimism, depression, resilience, body-efficacy expectation, perceived sensitivity to medication, neuroticism and openness to experience.

Response

The suggestion was contradicting the one from the other reviewer in the last review round, however, we have now revised the Ms according to this suggestion. It now reads:

In a second step, we used the reduced version of the scale to generate a total sum score of the five items (ranging from 5 to 20). To examine divergent and convergent validity, correlations between the ETS sum score and the other measures were calculated. The selection of measures was based on theoretical assumptions: With another measure of expectations (AES) we hypothesised very high correlations, since three of the five items of the ETS cover similar topics as the AES (coping, disappearance of complaints, energy) even though instruction and response options differ. We assumed a very high correlation between the ETS and the most strongly related construct (r about .70; AES) and a moderate correlation with strongly related constructs (r about .30; LOT-R optimism, inverse with LOT-R pessimism). Optimism can be viewed as a trait characteristic of a person with high stability over time and situations. Optimism is defined as "the extent to which people hold generalized favourable expectancies for their future" (Carver et al, 2010, p. 879)34. We included optimisms and the counterpart pessimism to assess the overlap between expectation and this personality trait.

Small correlations with less related constructs (r about .20; PHQ-9; RS-11, BEE) were assumed. Explanatory styles (i.e. expectation about future events) are associated with depressed mood with similar correlations 35. Self-efficacy is also a construct at a general level (i.e., "Perceived self-represents an optimistic sense of personal competence […]"; Scholz et al., 2002; p. 342)36. If self-efficacy is related to a specific behaviour or problem, it captures the strength of a belief to cope in a situation successfully (for example Pain Self-Efficacy Questionnaire; PSEQ)37. In our study we used the Body Efficacy Expectation (BEE) since this scale assess the "conviction that one's body is able to heal and take care of itself by dealing with pathogens and other health-threatening demands on its own." (Schützler & Witt, 2013; p. 2). Resources for health also capture positive beliefs on how to deal with a difficult situation in life. We hypothesized for both dimension only low correlations since such resources are rather general and neither related to a specific disorder nor to a specific time-point in

life. Finally we assumed no correlation with unrelated constructs (PSM, NEO-FFI neuroticism, NEO-FFI openness to experience), however we wanted to assess these dimension since the ETS might be used in upcoming placebo / nocebo research and several studies have shown that these dimensions are possibly related to placebo / nocebo responses 38 39.

Comment

Similarly, in the results, the authors indicate the findings for convergent and divergent validity and specify "we expected" or "which is unexpected". The basis for why these findings were expected or unexpected must be clearly made for the reader earlier in the manuscript.

I found the discussion improved; however, believe it still requires focus. Specifically:

line 388-396: Not clear that this adds anything. You found a high correlation between the 2 instruments; however, the relevance of the differences between the instruments is not clear. You indicate in the introduction the relatively low association between the EXPECT and AES in a prior study. Perhaps commenting on why the ETS may have been more highly correlated would add to the discussion

Response

We have now removed this paragraph. The explanation why EXPECT and AES did not correlate in another study was not the focus of our work. We mention in the background section the similarities of ETS and AES. This might give the reader an impression why ETS and AES are correlated.

Comment

Line 397-405: I think the lack of association between optimism and the ETS as well as the moderate association between pessimism and the ETS is interesting and requires consideration. I found your explanation (line 403-405) to be overly speculative. You may wish to consider this in the theoretical context you establish for expectation. Perhaps considering how this supports or contradicts the Bowling model.

Response

The finding was contrary to the Bowling model and we added one sentence on that and removed the earlier explanation.

It now reads:

Furthermore, we found no correlation between optimism and the ETS in our study, which is contrary to the underlying theoretical framework of Bowling et al. in this study.

REVIEWER	Joel Bialosky University of Florida USA
REVIEW RETURNED	15-Apr-2019

GENERAL COMMENTS	Thank you to the authors for their response. My concerns have been adequately addressed and I have no further comments.