

PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (<http://bmjopen.bmj.com/site/about/resources/checklist.pdf>) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Measuring Engagement in Advance Care Planning: A Cross-sectional Multi-centre Feasibility Study
AUTHORS	Howard, Michelle; Bonham, Aaron; Heyland, Daren; Sudore, Rebecca; Fassbender, Konrad; Robinson, Carole; McKenzie, Michael; Elston, Dawn; You, John

VERSION 1 - REVIEW

REVIEWER	Karen Detering Austin Health Melbourne, Australia
REVIEW RETURNED	13-Dec-2015

GENERAL COMMENTS	<p>Thank you for the opportunity to review this paper, which researches a very important and new area in the ACP arena. I have the following minor comments to make to assist you further</p> <p>Re abstract - I believe the abstract needs to include the number of items in the tool, and a more balanced statement related to the duration required to complete the tool (not just stating 34minutes to > 1 hour). I also believe more detail in the outcomes (i.e. more on how you assessed feasibility, acceptability and clinical sensibility). This is especially important given your conclusion. I also do not think your conclusion necessarily makes sense given the outcomes section.</p> <p>Re limitations - you do not discuss or mention why only 20 dialysis patients were recruited, despite the plan to recruit 50. It is unclear how (if at all) this may influence any of your conclusions, especially given that these patients were the quickest at completing the tool. I think it is important to at least mention this in the body of the paper. A further limitation may be the addition of the extra constructs, and therefore comparing to previous studies is more difficult. You do sort of cover this in your discussion.</p> <p>In terms of your conclusion I am not sure a survey with a mean time to administer of 49 minutes is necessarily feasible (I agree the language etc is fine, and no emotional burden) It is also not likely to be a feasible tool for repetitive measurements as you suggest in your final sentence. I suggest these issues are clarified.</p> <p>A final comment is that the discussion is a bit hard to follow - especially paragraph 3 and 6.</p>
-------------------------	---

REVIEWER	Jared Chiarchiaro MD, MS and Robert Arnold MD University of Pittsburgh, USA
REVIEW RETURNED	09-Jan-2016

GENERAL COMMENTS	The authors have completed a cross sectional study assessing the
-------------------------	--

feasibility of administering a survey to measure engagement in advance care planning. This is an important area and a well written manuscript. There are some major points I recommend the authors address to hopefully strengthen the message. First, I think the main weakness is a lack of clarity regarding what this survey is supposed to do and why it is important. It's also unclear in what context this survey would be used, for example in clinical trials, in real life, before and/or after an ACP intervention. This speaks to what audience you are trying to reach and why people should care. Second, why are family members/surrogates of the patients not included? This is a major weakness in my view. Third, the expansion of the original survey is a major issue for me. This potentially invalidates the survey and adds additional burden to the participants in the form of increased time to an already lengthy interview process. A particular concern is whether a survey that takes this much time is really feasible in larger studies. For a study about feasibility, it is curious to me why the authors would add 33 new items to an already validated survey. The authors may consider presenting the results with and without the additional items to see if they really add anything to the data.

Specific points:

INTRODUCTION

1. Could the authors expand on why this measure of ACP engagement is important? They make the statement in the second to last paragraph of the Introduction, "To measure engagement in ACP and to develop and evaluate the effectiveness of tools to facilitate all steps in the ACP process..." Why is measuring engagement in advance care planning important? Is this what we really care about in advance care planning or is it a surrogate for something else? Do the authors envision using it only in the context of clinical trials for advance care planning interventions or do they envision it also being used in the real world? Furthermore, I'm not clear on how measuring advance care planning engagement will help "evaluate the effectiveness of tools to facilitate all steps in the ACP process." Are the authors arguing that advance care planning engagement should be the outcome of advance care planning interventions or that it should be measured prior to intervention deployment, or both?

METHODS

2. Under recruitment and eligibility. Could the authors discuss why they chose these ages and disease states? I ask because people of different ages and with different underlying medical conditions likely have different propensities to be engaged in advance care planning. For example, a 19 year old with a hematologic malignancy is probably different than an 80 year old in the hospital with a broken hip. Would the authors consider presenting the data stratified by age and/or disease?

3. Why were family members/surrogates not included? This is a major weakness in my view.

4. The authors make a point that the advance care planning engagement survey has been validated. However, the advance care planning engagement survey used in this study has been

significantly changed from the initial version of the study. Is this survey still valid? To me, it seems that the internal consistency and other measures for the original survey reported in the Methods (page 9) are not relevant to the survey used in this study. The authors describe the changes made to this survey; however, it's unclear whether these 33 new questions add anything meaningful to the survey. Also, addition of the 33 questions bring the total number of survey items to 116 items which seems quite burdensome. A particular concern is whether a survey that takes this much time is really feasible in larger studies. Could the authors consider presenting the data with and without these 33 additional questions? This may help to discern whether these additional items add anything of value.

RESULTS

5. Could the authors discuss why feasibility and acceptability scores > 4 on this 6 point Likert scale are good enough? To me, the feasibility and acceptability scores in Table 2 seem quite modest and not convincing. Is there any data on what is a "good" score? Are their plans to further refine the ACP engagement survey based on these data (beyond what the authors mention about shortening the survey)?

6. This is simply a question: do the authors think that presenting the results in two groups: inpatient and outpatient may make the results clearer and a bit easier to interpret? To me the primary care, dialysis, and cancer groups could all go together as one "outpatient" group. Just a suggestion. There are pros and cons to keeping them separate vs combining, particularly when considering the statistical analysis.

7. In the Discussion, it is mentioned that some participants required more than one session to complete the survey. I would include the number who required more than one session in the Results. It seems like an important piece of feasibility data.

DISCUSSION

8. The authors make the point in the Discussion that prior literature indicates an interview time of 53-60 minutes was too long. This speaks to the point above about the survey length of 116 items. Given this prior data, could the authors justify why they feel the benefit of lengthening the survey by 33 questions outweighs the added burden?

9. The authors state, "Further study is needed to determine benchmarks or thresholds for what is considered clinically meaningful ACP engagement..." Could the authors expand on the implications of measuring ACP engagement? Is it a surrogate outcome? If so, what is it a surrogate for? Is it meant to be a measure of readiness to engage in ACP or is it a measure of engagement with an actual ACP intervention? Is it important only in the context of clinical trials or does it have real world applicability?

10. On page 21, "hospitalized patients in our study were also older than patients from other settings which could explain their higher engagement in ACP." If this is the hypothesis, then I would stratify the results by age.

	<p>11. The authors note prior data that illness severity is not associated with advance directive completion. Then why is it “surprising that ACP engagement among patients in the cancer and dialysis care settings was similar to patients in primary care?”</p> <p>12. I don’t think I understand the point of paragraph 6 of the Discussion. The authors seem to be conflating “engagement” with “readiness.” Are they the same or different constructs? (To me they are different, but I could be wrong.) For example, one could be at different stages of readiness for a behavior but still engage in that behavior. The engagement may not be effective, but that’s a different matter. Smoking cessation is one example of this.</p> <p>13. In the Conclusions, how does this survey both “track progress and evaluate the impact of potential interventions.” I guess I’m still just unclear on what this survey is supposed to do. More clarity about this would strengthen the manuscript and key readers in to the importance of this work.</p>
--	--

VERSION 1 – AUTHOR RESPONSE

Reviewer #1

1) Re abstract - I believe the abstract needs to include the number of items in the tool, and a more balanced statement related to the duration required to complete the tool (not just stating 34minutes to > 1 hour). I also believe more detail in the outcomes (i.e. more on how you assessed feasibility, acceptability and clinical sensibility). This is especially important given your conclusion. I also do not think your conclusion necessarily makes sense given the outcomes section.

Response: These suggestions are helpful and we have revised the abstract.

2) Re limitations - you do not discuss or mention why only 20 dialysis patients were recruited, despite the plan to recruit 50. It is unclear how (if at all) this may influence any of your conclusions, especially given that these patients were the quickest at completing the tool. I think it is important to at least mention this in the body of the paper.

Response: As suggested, we have noted the difference in sample size from this setting in the body of the paper.

3) A further limitation may be the addition of the extra constructs, and therefore comparing to previous studies is more difficult. You do sort of cover this in your discussion.

Response: We have provided further description of the stage of survey development in this paper and rationale for the added constructs and items (see also response to Reviewer #2, comment 4). We do not see it as a limitation of the methodology that these items were added in the version.

4) In terms of your conclusion I am not sure a survey with a mean time to administer of 49 minutes is necessarily feasible (I agree the language etc. is fine, and no emotional burden) It is also not likely to be a feasible tool for repetitive measurements as you suggest in your final sentence. I suggest these issues are clarified.

Response: We have revised the conclusion as suggested and have provided more information in the paper to describe that this survey continues to be revised to address the length.

5) A final comment is that the discussion is a bit hard to follow - especially paragraph 3 and 6.

Response: We have revised the discussion to help clarify the points. These paragraphs have also been revised in addressing the second reviewer’s comments.

Reviewer #2

The authors have completed a cross sectional study assessing the feasibility of administering a survey to measure engagement in advance care planning. This is an important area and a well written manuscript.

There are some major points I recommend the authors address to hopefully strengthen the message.

Specific points:

INTRODUCTION

1. Could the authors expand on why this measure of ACP engagement is important? They make the statement in the second to last paragraph of the Introduction, "To measure engagement in ACP and to develop and evaluate the effectiveness of tools to facilitate all steps in the ACP process..." Why is measuring engagement in advance care planning important? Is this what we really care about in advance care planning or is it a surrogate for something else? Do the authors envision using it only in the context of clinical trials for advance care planning interventions or do they envision it also being used in the real world? Furthermore, I'm not clear on how measuring advance care planning engagement will help "evaluate the effectiveness of tools to facilitate all steps in the ACP process." Are the authors arguing that advance care planning engagement should be the outcome of advance care planning interventions or that it should be measured prior to intervention deployment, or both?

Response: As suggested, we have clarified the reasoning for developing this survey instrument. We have re-worded the introduction to better describe the rationale for this survey, highlighting that advance care planning is a process comprised of several behaviors in addition to advance directive completion. Currently, the survey is being used predominantly as an outcome measure of advance care planning interventions.

METHODS

2. Under recruitment and eligibility. Could the authors discuss why they chose these ages and disease states? I ask because people of different ages and with different underlying medical conditions likely have different propensities to be engaged in advance care planning. For example, a 19 year old with a hematologic malignancy is probably different than an 80 year old in the hospital with a broken hip. Would the authors consider presenting the data stratified by age and/or disease?

Response: The main objective of the study was to assess the feasibility and acceptability of the survey in a wide variety of clinical settings/populations in which ACP would be important. We chose the clinical settings based on evidence from the literature that ACP is lacking in all health care settings even though it could be beneficial. Since it was not the objective of the paper to make inferences about socio-demographic or health predictors of ACP engagement such as age or health status and we would have limited sample size to examine these associations, we do not feel it would be appropriate to stratify data.

3. Why were family members/surrogates not included? This is a major weakness in my view.

Response: We thank the reviewer for pointing out this issue. We agree that the role of family members/surrogates is key in advance care planning. This version of the survey includes domains asking about the engagement of and communication with both surrogates and other family/friends. The survey was developed to be administered to individuals as it is asking about ACP engagement of the respondent, as a measure of the effectiveness of ACP interventions. Such interventions would include engagement of family members and surrogates as a key element of ACP, and this

engagement is captured from the patient's perspective in this survey. It would be an important outcome of ACP interventions to assess the surrogate's perceptions also, requiring a different measurement instrument.

4. The authors make a point that the advance care planning engagement survey has been validated. However, the advance care planning engagement survey used in this study has been significantly changed from the initial version of the study. Is this survey still valid? To me, it seems that the internal consistency and other measures for the original survey reported in the Methods (page 9) are not relevant to the survey used in this study. The authors describe the changes made to this survey; however, it's unclear whether these 33 new questions add anything meaningful to the survey. Also, addition of the 33 questions bring the total number of survey items to 116 items which seems quite burdensome. A particular concern is whether a survey that takes this much time is really feasible in larger studies. Could the authors consider presenting the data with and without these 33 additional questions? This may help to discern whether these additional items add anything of value.

Response: The survey reported in this paper is one iteration that continues to be developed and validated. To clarify for the reader, we have changed the wording of the survey description to indicate that it was created from a version that was validated, with the addition of some questions that have not been validated from a psychometric perspective. We agree with the reviewer that the changes to the survey need to be very clear for the reader.

Most of the suggested additions to the survey were based on recurring feedback from stakeholders / experts in ACP, in particular, the items about "future wishes for treatment" and the survey developer (RS) did not feel it would be prudent to dismiss these concerns during the continual process of trying to refine and improve the survey. We have revised the wording in the methods (pg 9 paragraph 2, pg 10 paragraph 1) to help clarify how the decisions to add questions arose.

We feel that despite the new items in this version, the previously established high internal consistency within the sub-scales, suggests that the survey is robust. We have highlighted the publication of the original survey describing that the internal consistency holds even for the individual domains of the survey (pg 9 paragraph 1)

"Overall internal consistency (Cronbach's alpha) for the ACP Process Measures was 0.94. Internal consistency is also high within each sub-scale ranging between 0.84-0.94". We agree that further work needs to be done on item reduction.

RESULTS

5. Could the authors discuss why feasibility and acceptability scores > 4 on this 6 point Likert scale are good enough? To me, the feasibility and acceptability scores in Table 2 seem quite modest and not convincing. Is there any data on what is a "good" score? Are their plans to further refine the ACP engagement survey based on these data (beyond what the authors mention about shortening the survey)?

Response: We agree that the reviewer's comment is important to address for interpretation of the results. The score of 4 is 'good' and '5' is very good and most mean scores in most settings (except hospital) were in the range of 4 to 5. We have added information comparing the feasibility/acceptability results to those of other published, validated, and widely used surveys asking about ACP and end of life communication, which were assessed with the same measures.

6. This is simply a question: do the authors think that presenting the results in two groups: inpatient and outpatient may make the results clearer and a bit easier to interpret? To me the primary care, dialysis, and cancer groups could all go together as one "outpatient" group. Just a suggestion. There are pros and cons to keeping them separate vs combining, particularly when considering the statistical analysis.

Response: We take the point about the potential benefits of comparing only 2 groups but since we

were primarily interested in description rather than doing hypothesis testing/comparisons between groups, we think the results may be helpful to a greater audience by presenting all of the health care settings separately. Also, collapsing 3 groups into one would result in a loss of information.

7. In the Discussion, it is mentioned that some participants required more than one session to complete the survey. I would include the number who required more than one session in the Results. It seems like an important piece of feasibility data.

Response: Unfortunately we were not initially aware during the data collection process that one interviewer in one hospital was completing the survey over more than one occasion, and the frequency of this was not recorded. Since we are not clear to what extent this was driven by the interviewer we have removed this anecdote from the discussion.

DISCUSSION

8. The authors make the point in the Discussion that prior literature indicates an interview time of 53-60 minutes was too long. This speaks to the point above about the survey length of 116 items. Given this prior data, could the authors justify why they feel the benefit of lengthening the survey by 33 questions outweighs the added burden?

Response: We agree that further work is needed on the survey. By using this iteration that includes the additional domains, we can perform further psychometric analysis while including all important ACP domains. We are currently undertaking factor analysis to create a shorter version of the survey.

9. The authors state, "Further study is needed to determine benchmarks or thresholds for what is considered clinically meaningful ACP engagement..." Could the authors expand on the implications of measuring ACP engagement? Is it a surrogate outcome? If so, what is it a surrogate for? Is it meant to be a measure of readiness to engage in ACP or is it a measure of engagement with an actual ACP intervention? Is it important only in the context of clinical trials or does it have real world applicability?

Response: ACP engagement is the term used to describe a person's behavior with respect to ACP according to stage of behavior change (knowledge, contemplation, self-efficacy, readiness). The survey is intended to measure the impact of interventions, by detecting changes in behavior along the spectrum. It allows measurement at a more granular level to provide information on where specifically the intervention is having an impact or not, and how much of an impact. The ultimate goal of ACP is to improve concordance between a patient's wishes and actual health care received but measurement of this outcome is fraught with challenges including that it is a very "distal" outcome and so can be subject to many other influences besides an initial/"upstream" ACP intervention. Thus, a comprehensive measure is needed to capture in a more "proximal" fashion whether ACP is actually occurring (beyond completion of advance directives) in response to an ACP intervention.

10. On page 21, "hospitalized patients in our study were also older than patients from other settings which could explain their higher engagement in ACP." If this is the hypothesis, then I would stratify the results by age.

Response: Due to the convenience sample and small numbers, we did not set out to test specific hypotheses a priori. The sentence notes a potential interpretation of the result.

11. The authors note prior data that illness severity is not associated with advance directive completion. Then why is it "surprising that ACP engagement among patients in the cancer and dialysis care settings was similar to patients in primary care?"

Response: We have changed the wording of this paragraph to discuss the similarity in ACP engagement between these three settings, noting also that it may seem counter-intuitive for healthy patients and patients with a serious illness to have similar engagement and that this phenomenon has been found in other studies.

12. I don't think I understand the point of paragraph 6 of the Discussion. The authors seem to be conflating "engagement" with "readiness." Are they the same or different constructs? (To me they are different, but I could be wrong.) For example, one could be at different stages of readiness for a behavior but still engage in that behavior. The engagement may not be effective, but that's a different matter. Smoking cessation is one example of this.

Response: This paragraph has been revised to clarify how we have used 'engagement' to mean the stages of behaviour change including readiness.

13. In the Conclusions, how does this survey both "track progress and evaluate the impact of potential interventions." I guess I'm still just unclear on what this survey is supposed to do. More clarity about this would strengthen the manuscript and key readers in to the importance of this work.

Response: The conclusion paragraph has been revised to reflect the importance of the survey measure for evaluating ACP improvement interventions (see also Reviewer #2, comment 9).

VERSION 2 – REVIEW

REVIEWER	Jared Chiarchiaro MD, MS and Robert Arnold MD University of Pittsburgh, USA
REVIEW RETURNED	22-Feb-2016

GENERAL COMMENTS	<p>I thank the authors for their response. I think the manuscript could still benefit from some clarity, especially in the Discussion section.</p> <ol style="list-style-type: none"> 1. I would include some explanation in the text about why surrogates were not included. I think this is a limitation that many people will question. 2. I applaud the authors for adding the additional items to the survey in response to stakeholder feedback. This is a core tenant of creating a patient-centered outcome measure. I agree with noting in the text that the original survey has robust validity data. I would consider including in the limitations that since new items were added to the survey, it may not have the same validity as the original outcome measure. I believe adding this as a potential limitation was a comment from Reviewer 1 as well. 3. In the third paragraph of the discussion, I like the changes that the authors have made. Could the authors discuss in this paragraph why they think there is such a discrepancy in the time to survey completion for inpatients vs outpatients? This may guide the future study they indicate is necessary. (Although as we discuss below, it might be clearer to cut all the p values and just focus on the feasibility and descriptive data. At most a paragraph at the end that focuses on differences would be enough!) 4. The authors talk in their response letter several times that their goal is description and not hypothesis testing. However, they performed a lot of hypothesis testing. All their tables have p-values. Multiple times in the text they talk about differences between inpatients and outpatients. For example, all of the p-values in Table 2 are highly significant. But the actual scores are not that different. What should the reader take away from this? Also, in Table 3, the readiness p-value is significant while the others are not. Two questions – first if this just descriptive why give p values. And second, if they take the comparison seriously how do they interpret
-------------------------	--

	<p>these changes?</p> <p>5. I would consider adding some of the authors' response to my previous comment #9 to the text.</p> <p>6. Paragraph 5 of the Discussion is still unclear. For example you say, "It was interesting that we found differences between settings for action measures that would suggest sicker and older hospitalized patients are more engaged in ACP." This seems to contradict the prior paragraph where you say, "ACP engagement in patients from the cancer and dialysis care settings, who have a serious illness, was similar to that of patients in primary care, who would be expected to be healthier." It might be simpler to just focus on descriptive findings and cut all of this.</p> <p>7. The conclusions paragraph also seems contradictory. The authors state that the survey is feasible and then say that development of a more feasible version is needed. This is confusing.</p>
--	---

VERSION 2 – AUTHOR RESPONSE

Reviewer's Comment:

1. I would include some explanation in the text about why surrogates were not included. I think this is a limitation that many people will question.

Response: We have added the explanation that the design of the survey did not include questioning of surrogates in the Methods and have noted this as a limitation in the Discussion.

While the role of family members/surrogates is key in ACP, the Survey was designed to be answered by an individual and does not attempt to assess surrogates' perspectives. However, the Survey does include domains relating to the respondents' engagement of and communication with both surrogates and other family/friends. **(page 12, paragraph 2)**

The Survey asks the patient, not the surrogate, about the surrogate's involvement in ACP. This may result in social desirability bias as it is known that agreement between individuals and their surrogates on perceived extent of ACP involvement is often poor.¹⁵ **(page 26, paragraph 1)**

2. I applaud the authors for adding the additional items to the survey in response to stakeholder feedback. This is a core tenant of creating a patient-centered outcome measure. I agree with noting in the text that the original survey has robust validity data. I would consider including in the limitations that since new items were added to the survey, it may not have the same validity as the original outcome measure. I believe adding this as a potential limitation was a comment from Reviewer 1 as well.

Response: As suggested, we have added a statement to the Limitations section concerning the use of a version of the Advance Care Planning Engagement Survey that was modified from the original validated version.

The Survey version in used this study contained new items that were added to the previous validated version based on feedback from stakeholders. Therefore, the version of the Survey used in this study may not have the same validity as the original version. **(page 26, paragraph 1)**

3. In the third paragraph of the discussion, I like the changes that the authors have made. Could the authors discuss in this paragraph why they think there is such a discrepancy in the time to survey completion for inpatients vs outpatients? This may guide the future study they indicate is necessary. (Although as we discuss below, it might be clearer to cut all the p values and just focus on the feasibility and descriptive data. At most a paragraph at the end that focuses on differences would be enough!)

Response: Please see our response to comment # 4 below regarding p values. We have added statements in the Discussion paragraph mentioned, to suggest why the survey took longer among the inpatients and also a point about the implications of the survey length among inpatients (new text in bold):

The inclusion criteria of the hospitalized patients led to a sample more than 10 years older on average than the patients in the other settings. Hospitalized patients were likely experiencing an acute health event, which may explain the greater time to complete the survey. In previous studies on similar topics and in similar settings and populations that administered surveys taking on average 40 to 60 minutes, patients and family members reported that the duration was too long. **These findings have implications for administering research surveys in different settings and patient populations. For example, it may not be feasible or prudent to administer the Survey in the inpatient setting. (page 22, paragraph 3 to page 23 paragraph 1)**

4. The authors talk in their response letter several times that their goal is description and not hypothesis testing. However, they performed a lot of hypothesis testing. All their tables have p-values. Multiple times in the text they talk about differences between inpatients and outpatients. For example, all of the p-values in Table 2 are highly significant. But the actual scores are not that different. What should the reader take away from this? Also, in Table 3, the readiness p-value is significant while the others are not. Two questions – first if this just descriptive why give p values. And second, if they take the comparison seriously how do they interpret these changes?

Response: We agree that we did not address this issue adequately in the first revision. We have taken the reviewers' advice and removed the p values for the results of the actual ACP engagement survey across settings (and the demographics Table 1), since it was difficult to make interpretations without a priori hypotheses. We have made changes to the Statistical Analysis section and Results tables and text accordingly.

To describe the samples, we presented demographic characteristics and self-reported health variables by health care setting. We compared feasibility across the four different health care settings using the Kruskal-Wallis non-parametric test. Where a statistically significant difference was found between settings, pairwise post-hoc tests were performed using the Bonferroni correction for multiple comparisons (two-sided). The criterion of statistical significance was set at $\alpha=0.05$. We also summarized Survey scores for the overall study sample and descriptively examined the scores on the Survey by health care setting. **(page 13, paragraph 2)**

New results text:

Mean process scales varied little across settings and ranged from 3.0 in cancer, renal and primary care settings to 3.2 in hospitalized patients. Scores were lowest for contemplation and highest for self-efficacy (Table 3). Hospitalized patients reported the highest knowledge and readiness. **(page 19, paragraph 1)**

The total mean action score was 11.2 ± 5.6 , on a scale of 0 to 25. The lowest mean score was reported for patients in primary care settings (9.6 ± 4.9) and the highest was reported by hospitalized patients (13.6 ± 5.5) (Table 4). **(page 20, paragraph 2)**

We also clarified the primary objective of the study as follows:

The objective of this study was to evaluate the feasibility, acceptability and clinical sensibility of using the ACP Engagement Survey in primary care, hospital care, cancer care and dialysis care. **(page 8 paragraph 3)**

We did not remove the p values from the comparisons of feasibility and acceptability between the settings as this was the primary objective of the paper. However we have simplified Table 2 by removing the column for p values (all were highly significant) and added a footnote regarding statistical significance. We also removed the reference to the appendix of pairwise comparisons since it was possible to briefly summarize them in the text. We summarized findings in the Results text, including the pairwise comparisons results, as follows:

There was a statistically significant difference between settings for all the feasibility and acceptability items. Mean scores for acceptability and feasibility items were greater than 4 on 6-point scales for all items in all settings except hospital where scores were lower (Table 2). In the pairwise comparisons, for all items hospitalized patients were significantly different from patients in each of the other settings, and patients in primary care, cancer care and dialysis care did not differ significantly from each other on any of the items. **(page 17, paragraph 2)**

5. I would consider adding some of the authors' response to my previous comment #9 to the text.

Response: Thank you for the suggestion. We have incorporated language from the previous response to the Introduction as follows:

To measure **the full range** of engagement in ACP **behaviors** and to develop and evaluate the effectiveness of tools to facilitate ACP beyond advance directive documentation, we need a feasible and valid self-report measure of ACP behaviours. **ACP engagement, in this paper, is the term used to describe a person's behavior with respect to the various aspects of ACP according to stage of behavior change (knowledge, contemplation, self-efficacy, readiness).** **(page 7, paragraph 2)**

The ultimate goal of ACP is to improve concordance between a patient's wishes and actual health care received but measurement of this outcome is fraught with challenges including that it is a very "distal" outcome and so can be subject to many other influences besides an initial "upstream" ACP intervention. Thus, a comprehensive measure is needed to capture in a more "proximal" fashion whether ACP is actually occurring (beyond completion of advance directives) in response to an ACP intervention. Based on Social Cognitive Theory and a stage-of-change conceptual model, the ACP Engagement Survey has been developed and its reliability and validity examined among older adults in community and ambulatory care settings in the U.S.⁹ The ACP Engagement Survey is intended to measure the impact of interventions, by detecting changes in behavior along the spectrum. It allows measurement at a more granular level to provide information on where specifically the intervention is having an impact or not, and how much of an impact. To date the ACP Engagement survey has not been tested for feasibility of use among varying age groups, disease processes and various health care settings. **(page 8 paragraph 2)**

6. Paragraph 5 of the Discussion is still unclear. For example you say, "It was interesting that we found differences between settings for action measures that would suggest sicker and older hospitalized patients are more engaged in ACP." This seems to contradict the prior paragraph where you say, "ACP engagement in patients from the cancer and dialysis care settings, who have a serious illness, was similar to that of patients in primary care, who would be expected to be healthier." It might be simpler to just focus on descriptive findings and cut all of this.

Response: Thank you for noting that this paragraph in the Discussion was inconsistent. We have revised the paragraph as follows to distinguish the patterns among inpatients and outpatients and noting that the study we compare to was conducted in outpatient settings:

Although our analysis of ACP engagement was exploratory, hospitalized patients appeared to be slightly more engaged in ACP than patients in other settings. Hospitalized patients reported the highest knowledge and readiness, and two to four more ACP actions on average, than patients from the other settings. We found that ACP engagement was similar across all outpatient settings including cancer care and dialysis care where we would expect greater illness severity, and primary care settings where we would expect lesser illness severity. A previous study also found that, among outpatients, illness severity was not associated with stage of readiness for ACP. **(page 24, paragraph 2 to page 25 paragraph 1)**

7. The conclusions paragraph also seems contradictory. The authors state that the survey is feasible and then say that development of a more feasible version is needed. This is confusing.

Response: We agree the paragraph remained contradictory. It has been revised as follows:

We have shown that the Survey demonstrated feasibility acceptability and clinical sensibility in outpatient settings, but to a lesser extent among hospitalized patients. The main concern for feasibility was length. To facilitate the use of this measurement to evaluate interventions to improve engagement in ACP, further work is needed to reduce the length, especially if the Survey is to be used in hospitalized patients. **(page 27, paragraph 1)**

VERSION 3 - REVIEW

REVIEWER	Robert Arnold and Jared Chiarchiaro University of Pittsburgh, USA
REVIEW RETURNED	27-Apr-2016
GENERAL COMMENTS	We thank the authors for a very thoughtful and thorough response. All of our concerns have been addressed.