

## 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

<b>TITLE (PROVISIONAL)</b>	The ability of observer and self-report measures to capture shared decision making in clinical practice in the United Kingdom: a mixed-methods study.
<b>AUTHORS</b>	Williams, Denitza; Edwards, Adrian; Wood, Fiona; Lloyd, Amy; Brain, Kate; Thomas, Nerys; Prichard, Alison; Goodland, Annwen; Sweetland, Helen; McGarrigle, Helen; Hill, Gwenllian; Joseph-Williams, Natalie

### VERSION 1 – REVIEW

<b>REVIEWER</b>	Ian Hargraves Mayo Clinic Rochester, Minnesota, USA
<b>REVIEW RETURNED</b>	11-Feb-2019

<b>GENERAL COMMENTS</b>	<p>The authors present a very interesting study on a significant issue—the extent to which common measures of the process and behaviors of SDM align with the way that decisions are made in a shared manner in real practice. Any misalignment here compromises the effectiveness and purpose of SDM evaluation when used at policy, practice quality, and training levels. The study uses a novel mixed methods approach to contrast qualitative analysis of decision making encounters with established quantitative observer and participant measures of SDM, and found factors in decision making in real practice that are not adequately accounted for in the quantitative scales.</p> <p>This paper has important implications for the promotion and implementation of SDM, it could however benefit from some tightening of the argument that it builds and presents—specifically the linking of a problem to methods. The argument presented in the background would be more compelling and drive the paper if it spoke more to any signs that SDM in practice may be significantly different from what is evaluated for policy, quality, and training purposes (e.g. literature on the distributed nature of decision making) and the consequences of this issue should it exist. This argument is made somewhat in the current background section but could be effectively strengthened.</p> <p>Out of the significance of this plausible issue, the case could be made that research is needed to understand if the discordance of SDM in practice and SDM in evaluation exists, the nature, and extent of any discrepancy. Methodologically, this requires comparing evaluations using current measures with analysis of how decisions made together by patients and clinicians happen in vivo...Hence this study.</p>
-------------------------	--

	<p>I question the objective of the study as listed in the abstract. The current objective: To examine measures of shared decision-making (SDM), in two clinical teams that had embedded SDM in practice.; appears to be the objective of the methods But it seems that the objective of the study is to develop a preliminary assessment of if and how observer and participant measures of SDM encounters evaluate the decision making activities that patients and clinicians undertake together</p> <p>There were a few minor issues that the authors may wish to consider</p> <ul style="list-style-type: none"> <li>• The sentence beginning on line 31 of page 13 was unclear to me</li> <li>• The authors use the formulation of “most patients” without providing what that means in numerical terms</li> <li>• Consider moving the quote on page 11 to between the SDM was distributed and preference elicitation paragraphs i.e. between line 39 &amp; 40, as the quote seems to refer to distribution rather than preference elicitation</li> </ul>
--	--

<b>REVIEWER</b>	margaret schwarze University of Wisconsin USA
<b>REVIEW RETURNED</b>	23-Feb-2019

<b>GENERAL COMMENTS</b>	<p>I have been asked to review “The challenges of measuring shared decision-making in the NHS: rarely one clinician, one consultation, one decision” for BMJ Open. This is an observational study using mixed methods to describe shared decision making for patients with breast cancer and late-stage kidney disease. The investigators audio recorded conversations between clinicians and patients facing decisions about breast cancer treatment or type of dialysis in clinics where interventions to improve shared decision making have been employed. The authors use patient and clinician reported impressions of shared decision making, observer measured shared decision making and qualitative content analysis to evaluate SDM in 43 conversations. The authors conclude that there are some real challenges of measurement of SDM and, in part, this stems from the longitudinal nature of SDM.</p> <p>This is an interesting paper, yet as presented is not acceptable for publication. The major problem is that the aim of the study is to “examine the process of SDM in two routine clinical settings” which is not a particularly specific research aim or research question. The manuscript is thus completely disoriented based on lack of a clear goal driving data collection and presentation and seems to jump back and forth between efforts to characterize SDM and efforts to evaluate measures of SDM. It is very difficult to follow and hard to support the authors conclusions with this lack of clear direction and organization. I think this could be reconciled but it would need a fair amount of restructuring to clarify why the study design was appropriate and to present the data in a way that answers a specific (or 2 specific) research question.</p> <p>There are several other minor concerns</p> <ol style="list-style-type: none"> <li>1. In the abstract, the first conclusion is about thematic analysis, when thematic analysis is not even mentioned in the methods section.</li> <li>2. It is completely unclear how mean scores for OPTION 5 domains (rated on a scale of 0 – 4 ( not 0 – 5 as noted by the authors) or 0 – 20) could be in the range of 92 with an SD of 62-</li> </ol>
-------------------------	---

	<p>100...the total score is out of 100, so how is 1/5 of the score modified to a scale of 100?</p> <p>3. There are several items in the introduction that are confusing: “as well as recommendations” is a challenging way to start a sentence and is very unclear about who made such recommendations, in paragraph 2 it is unclear what it means to “identify measurements of successful SDM” – does this mean to identify good strategies to measure SDM or identify episodes of SDM that have been done well?</p> <p>4. In the methods section: what does it mean to have undergone a process of SDM? Passive voice is used in many sections of the methods when active voice is far more appropriate, I am sure the research team did these things, they should say that.</p> <p>5. Who enrolled these patients? The clinicians? How do we know this was done in a sequential and unbiased way, rather than when the clinicians thought they had done a good job?</p> <p>6. Who scored OPTION 5 and how do you know that another person scoring with OPTION 5 would have made the same assessment? Was there an attempt to measure IRR?</p> <p>7. NVIVO can catalog data, it cannot support or do qualitative analysis. There is no mention of second order, or higher level analysis and how this was done. Qualitative analysis is not complete when the coding is done.</p> <p>8. The results were presented in a very non-specific way when quantitative data was presented e.g., “a high proportion of respondents” “surescore showed that patients and clinicians thought SDM was happening” some numbers here are required to present quantitative data appropriately</p> <p>9. “item 2 consistently scored lowest” – I don’t think item 2 actually did any scoring.</p> <p>10. Discussion of “implicit and broader SDM tasks” that were not accounted for by the OPTION scoring needs a lot more specifics with examples from the data to support the statement that these items were not measured/assessed, what were these things that were unaccounted for?</p> <p>11. It is unclear to me why a deferred decision was not counted as a recommendation or scored well in the OPTION 5 scoring system...this does help to account for the longitudinal nature of decision making within the OPTION 5 construct.</p> <p>12. The authors use the word “distributed” repeatedly to describe decision making over time. It is a very unclear word and quite misleading.</p>
--	--

<b>REVIEWER</b>	haske van veenendaah ESHMP, Erasmus University, The Netherlands
<b>REVIEW RETURNED</b>	12-Mar-2019

<b>GENERAL COMMENTS</b>	<p>An interesting article as one of the major challenges in the field of SDM is measurement in relation to implementation in clinical/daily practice. The study is of explorative nature but gives a good insight into how two SDM measurement-tools function in practice. I have the following comments/suggestions:</p> <ul style="list-style-type: none"> <li>- Title: I somehow miss in the title that this research specifically focusses on daily practice and its relation to measurement.</li> <li>- The teams 'have embedded SDM (by participation in MAGIC). I assume that applying SDM is more of a continuous than dichotome variable. This may also have consequences for the results find in the study.</li> </ul>
-------------------------	--

	<ul style="list-style-type: none"> <li>- The study is explorative and specifically looks at teams that are successful in applying SDM. It has also several limitations about which little information is provided.</li> <li>- Three measurement challenges emerge from the thematic analysis (table 4). These are interesting and relevant challenges, but I miss some insight in how/why these 3 have emerged.</li> <li>- In my opinion the added value of the article is primarily how trying to measure SDM is most relevant for SDM in routine clinical practice (I assume for implementation and learning). It is especially interesting how to solve the challenges, i.e. by making choices about when and whom to observe measure (as stated at page 17, top). This can be solved not only by the modifying instruments but also about choosing when and by whom to use (all consultations of involved clinicians of 1 decision i.e.).</li> <li>- I like the mixed methods approach in the study. This may also be applied on the challenges raised in the study. The OPTION is an observation so it will only capture what can be observed. But developing new instruments or adapt the OPTION might not be the most logical next step as it might be one of the instruments that works at an acceptable level. Instead, it might be interesting to look at an appropriate combination of instruments. i.e. qualitative ones to cover contexts/settings? The statement that caution is needed when using instruments that only capture a 'snapshot' of the entire process seems to suggest that it is possible to have 1 instrument covering the clinical practice that is very diverse in its nature?</li> <li>- I would suggest to (better) align the reported challenges (table 4) in the discussion and conclusion section. For example, the conclusion that "We urgently need to develop tools that are valid and responsive to the non-standardised process of achieving SDM", does not seem to fully cover the findings of the study?</li> </ul>
--	---

### VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

This paper has important implications for the promotion and implementation of SDM, it could however benefit from some tightening of the argument that it builds and presents—specifically the linking of a problem to methods. The argument presented in the background would be more compelling and drive the paper if it spoke more to any signs that SDM in practice may be significantly different from what is evaluated for policy, quality, and training purposes (e.g. literature on the distributed nature of decision making) and the consequences of this issue should it exist. This argument is made somewhat in the current background section but could be effectively strengthened.

Out of the significance of this plausible issue, the case could be made that research is needed to understand if the discordance of SDM in practice and SDM in evaluation exists, the nature, and extent of any discrepancy. Methodologically, this requires comparing evaluations using current measures with analysis of how decisions made together by patients and clinicians happen in vivo...Hence this study.

I question the objective of the study as listed in the abstract. The current objective: To examine measures of shared decision-making (SDM), in two clinical teams that had embedded SDM in practice.; appears to be the objective of the methods. But it seems that the objective of the study is to develop a preliminary assessment of if and how observer and participant measures of SDM encounters evaluate the decision making activities that patients and clinicians undertake together

We agree with the reviewer's comment regarding the clarity of the objective listed in the abstract. This has now been amended to reflect the key objective of the study more closely. The objective stated in the abstract section has been modified to read: "To examine if and how observer measures of SDM encounters evaluate the decision making activities that patients and clinicians undertake in consultation."

There were a few minor issues that the authors may wish to consider:

1. The sentence beginning on line 31 of page 13 was unclear to me

The sentence on line 31 has been altered for further clarity to read:

"The amount of information about options that was presented by clinicians varied based on the patient's previous knowledge about the options, as well as their emotional response. For example, some patients were understandably emotionally distressed following their diagnosis of breast cancer. In such cases, clinicians would sometimes limit the depth of information offered at that appointment and would defer in-depth explanations to the follow-up appointment."

2. The authors use the formulation of "most patients" without providing what that means in numerical terms

As well as presenting numerical data in Table 1, we have added numerical representations of terms such as "most patients" in the results section.

3. Consider moving the quote on page 11 to between the SDM was distributed and preference elicitation paragraphs i.e. between line 39 & 40, as the quote seems to refer to distribution rather than preference elicitation

We have moved the quote so that it is placed within the distribution section.

Reviewer: 2

I have been asked to review "The challenges of measuring shared decision-making in the NHS: rarely one clinician, one consultation, one decision" for BMJ Open. This is an observational study using mixed methods to describe shared decision making for patients with breast cancer and late-stage kidney disease. The investigators audio recorded conversations between clinicians and patients facing decisions about breast cancer treatment or type of dialysis in clinics where interventions to improve shared decision making have been employed. The authors use patient and clinician reported impressions of shared decision making, observer measured shared decision making and qualitative content analysis to evaluate SDM in 43 conversations. The authors conclude that there are some real challenges of measurement of SDM and, in part, this stems from the longitudinal nature of SDM. This is an interesting paper, yet as presented is not acceptable for publication. The major problem is that the aim of the study is to "examine the process of SDM in two routine clinical settings" which is not a particularly specific research aim or research question. The manuscript is thus completely disoriented based on lack of a clear goal driving data collection and presentation and seems to jump back and forth between efforts to characterize SDM and efforts to evaluate measures of SDM.

We acknowledge the reviewer's comments regarding lack of clarity between the stated aim of the study and the reported results. We agree that the stated objective in the abstract section is not clear and it has now been amended to reflect the study more closely.

The objective stated in the abstract section has been modified to read: “To examine if and how observer measures of SDM encounters evaluate the decision making activities that patients and clinicians undertake consultation.”

It is very difficult to follow and hard to support the authors conclusions with this lack of clear direction and organization. I think this could be reconciled but it would need a fair amount of restructuring to clarify why the study design was appropriate and to present the data in a way that answers a specific (or 2 specific) research question.

We acknowledge the reviewer’s feedback. We have modified the manuscript to improve the focus and to ensure the objectives are clearer.

There are several other minor concerns:

1. In the abstract, the first conclusion is about thematic analysis, when thematic analysis is not even mentioned in the methods section.

Thematic analysis was described on page 8 of the original manuscript, in the Methods section under the ‘Qualitative analysis’ sub-heading. We have added further description of thematic analysis in this section.

2. It is completely unclear how mean scores for OPTION 5 domains (rated on a scale of 0 – 4 (not 0 – 5 as noted by the authors) or 0 – 20) could be in the range of 92 with an SD of 62- 100...the total score is out of 100, so how is 1/5 of the score modified to a scale of 100?

We thank the reviewer for highlighting our typographical error, which has led to the confusion. The OPTION 5 domains were indeed rated on a 5-point Likert scale, but ranging from 0-4 not 0-5 as stated in the manuscript. This typo has been amended.

3. There are several items in the introduction that are confusing: “as well as recommendations” is a challenging way to start a sentence and is very unclear about who made such recommendations, in paragraph 2 it is unclear what it means to “identify measurements of successful SDM” – does this mean to identify good strategies to measure SDM or identify episodes of SDM that have been done well?

The sentence “As well as recommendations, incentives for doing SDM with patients are being considered to help overcome some of the barriers associated with implementation” has been modified to read “Incentives for doing SDM with patients are being considered to help overcome some of the barriers associated with implementation”.

We have clarified the meaning of the following sentence “This includes identifying measurements of successful SDM”; it has been re-worded it to read “This includes identifying strategies to successfully measure SDM in practice.”

4. In the methods section: what does it mean to have undergone a process of SDM? Passive voice is used in many sections of the methods when active voice is far more appropriate, I am sure the research team did these things, they should say that.

We have ensured that the language in our manuscript reflects the activities of the research team.

5. Who enrolled these patients? The clinicians? How do we know this was done in a sequential and unbiased way, rather than when the clinicians thought they had done a good job?

Patient recruitment was described in the 'Methods' section under 'Participant Recruitment'. Members of the clinical team (nurse specialists in renal and research nurse in breast) sequentially recruited patients who they deemed appropriate to the study, prior to their consultation (based on the inclusion / exclusion criteria and judgement about emotional / personal circumstances). Consultation recordings were collected by the clinicians, and then sent directly to clinicians by the research team. As this was not a typical 'research study', but an 'implementation study', the research team could not guarantee that this was done in a completely unbiased way. However, we are confident that we were sent all consultations that were recorded, due to the range of scores that we did observe with the OPTION 5 instrument (e.g. we did not only receive very high scoring consultations). Further, the objective of this study was to explore the ability of shared decision-making tools to capture shared decision-making in clinical practice, not a controlled trial setting; we have now further discussed this limitation in the 'Discussion' section of the paper.

6. Who scored OPTION 5 and how do you know that another person scoring with OPTION 5 would have made the same assessment? Was there an attempt to measure IRR?

We thank the reviewer for their comment. This process was described in the 'Methods' section under the 'Quantitative Analysis' sub-section of the manuscript. We have added some additional information to aid clarity.

7. NVIVO can catalog data, it cannot support or do qualitative analysis. There is no mention of second order, or higher level analysis and how this was done. Qualitative analysis is not complete when the coding is done.

We thank the reviewer for their comment. To clarify, we did not use NViVo to 'conduct' or 'support' the analysis per se, however it was used to order and manage the data, and to support the overall process. We have amended the term used in the manuscript and ensured that it is clear that we used NViVo to manage the data. We have provided further description of the thematic analysis in the 'Methods' section of the 'qualitative analysis' sub-section of the manuscript.

8. The results were presented in a very non-specific way when quantitative data was presented e.g., "a high proportion of respondents" "surescore showed that patients and clinicians thought SDM was happening" some numbers here are required to present quantitative data appropriately

We acknowledge the reviewer's comment. As well as presenting numerical data in Table 1, we have also added numerical representations of terms such as "most patients" in the results section.

9. "item 2 consistently scored lowest" – I don't think item 2 actually did any scoring.

The phrasing has been changed to "item 2 was consistently given a low score".

10. Discussion of "implicit and broader SDM tasks" that were not accounted for by the OPTION scoring needs a lot more specifics with examples from the data to support the statement that these items were not measured/assessed, what were these things that were unaccounted for?

The discussion of implicit and broader SDM tasks has been expanded for further clarity:

“By listening to the consultation audio-recordings, it was clear that the implicit and unspoken understandings that occur across consultations are difficult for an assessor to score. For example, when clinicians had previously spoken to patients over the telephone when making an appointment and had outlined the rationale for offering a choice. Supportive SDM tasks were often implicit and broader than the instrument assessed. Lower scores on item 2 were likely to have resulted from the instrument's inability to capture embedded and unspoken understandings, and did not reflect the actual support offered. The fact that the clinician was spending a long time outlining treatment options, discussing pros and cons as well as eliciting preferences implied that they were supportive of the patient. This is implicit support. Item 2 specifically states that a high score can only be given if support is offered explicitly “The clinician reassures the patient or re-affirms that the clinician will support the patient to become informed or deliberate about the options”.

11. It is unclear to me why a deferred decision was not counted as a recommendation or scored well in the OPTION 5 scoring system...this does help to account for the longitudinal nature of decision making within the OPTION 5 construct.

We observed that during some consultations, decisions were deferred to a follow-up consultation. This was typically not because the clinician and patient had reached a decision about deferring the decision to a later date, whereby the clinician had integrated the patient's preference to come back to the decision. Within the renal settings, it was typically because the patient was not yet at the point where they are being required to make a decision (thus further illustrating the complexity of 'when' SDM should be measured).

Item 5 which assesses decision making in the Observer OPTION 5 measure states: “The clinician makes an effort to integrate the patient's elicited preferences as decisions are made. If the patient indicates how best to integrate their preferences as decisions are made, the clinician makes an effort to do so”. Therefore, if a decision is deferred and not made (and not based on a patient's preference to come back to the decision at a later date, but on the structure of the clinical care pathway and the time of these discussions), it is difficult to score high on Item 5 on the OPTION 5 measure.

12. The authors use the word “distributed” repeatedly to describe decision making over time. It is a very unclear word and quite misleading.

We have used the term 'distributed' to refer to the process of shared decision-making which can be distributed over multiple consultations and with multiple clinicians. The term has been used in previous papers focusing on shared decision-making and notably is the focus of the widely cited paper by Rapley (2008) “Distributed decision making: the anatomy of decisions-in-action”. Based on what we observed, whereby many patients made a 'shared decision' over many consultations (rather than one consultation), sometimes over a long time period (e.g. 'choice' and 'option' talk could take place, and then the final 'decision' talk would take place up to 12 months later), and with a number of different people (consultants, surgeons, nurses, carers, family members etc), we felt that the term 'distributed' accurately reflected this process.

Reviewer: 3

An interesting article as one of the major challenges in the field of SDM is measurement in relation to implementation in clinical/daily practice. The study is of explorative nature but gives a good insight

into how two SDM measurement-tools function in practice. I have the following comments/suggestions:

1. Title: I somehow miss in the title that this research specifically focusses on daily practice and its relation to measurement.

The title has been amended to better reflect the study: The ability of observer and self-report measures to capture Shared Decision- Making in clinical practice.

2. The teams 'have embedded SDM (by participation in MAGIC). I assume that applying SDM is more of a continuous than dichotome variable. This may also have consequences for the results find in the study.

We have used the term 'embedded' throughout the manuscript because the teams involved had not only participated in MAGIC, but they had adopted an ethos of SDM. SDM had become part of their routine care and had therefore become 'embedded'. We acknowledge that SDM is indeed a continuous variable and not dichotomous and therefore the nature of an SDM discussion and its components at one time can change.

3. The study is explorative and specifically looks at teams that are successful in applying SDM. It has also several limitations about which little information is provided.

Limitations of the measures used in the study have been presented in the discussion section. However, we have also highlighted further limitations of the study in the discussion section.

4. Three measurement challenges emerge from the thematic analysis (table 4). These are interesting and relevant challenges, but I miss some insight in how/why these 3 have emerged.

The challenges described emerged from the data that we analysed. By reflecting on the data provided by the different types of measurement, we determined that these were key issues. We have added some clarity to this discussion in the manuscript.

5. I like the mixed methods approach in the study. This may also be applied on the challenges raised in the study. The OPTION is an observation so it will only capture what can be observed. But developing new instruments or adapt the OPTION might not be the most logical next step as it might be one of the instruments that works at an acceptable level. Instead, it might be interesting to look at an appropriate combination of instruments. i.e. qualitative ones to cover contexts/settings? The statement that caution is needed when using instruments that only capture a 'snapshot' of the entire process seems to suggest that s possible to have 1 instrument covering the clinical practice that is very diverse in its nature?

We thank the reviewer for highlighting this issue, and our lack of clarity. We have amended the discussion to reflect the complexity of measurement in clinical practice and recommend that it would be unfeasible to have one instrument for measuring SDM. We have made it explicitly clear that different instruments or a combination of different instruments are needed based on the context of the SDM discussion.

6. I would suggest to (better) align the reported challenges (table 4) in the discussion and conclusion section. For example, the conclusion that ‘We urgently need to develop tools that are valid and responsive to the non-standardised process of achieving SDM’, does not seem to fully cover the findings of the study?

We have modified sections of the discussion and conclusion to ensure that the findings of the study are represented.

**VERSION 2 – REVIEW**

<b>REVIEWER</b>	Ian Hargraves Mayo Clinic, USA
<b>REVIEW RETURNED</b>	26-Apr-2019

<b>GENERAL COMMENTS</b>	<p>This is a well written paper addressing an important concern—do current SDM measures evaluate if and how patients and clinicians make decisions together in real life? This question has significant implications for quality measures, SDM incentivizing, and implementation.</p> <p>A few minor changes may make the paper clearer for the reader.</p> <p>Throughout the paper there is a tension between four senses of SDM: 1.) the a-priori/accepted definition of SDM stated clearly for example in line 36 of page 25; 2.) SDM consisting of theoretically derived elements/behaviors (three talk/option) 3.) the SDM encounter which is the time when mandated SDM (as a policy) is supposed to occur; and 4.) SDM as what patients and clinicians are actually doing (inside or outside of the encounter) in coming to decisions. This tension is central to the argument of the paper but because the term SDM is used for all senses it can be a little confusing for the reader. This might be resolved by calling attention to these distinctions, or creating terms for them and using them within the text. E.g. a-priori SDM/the accepted definition of SDM; SDM in-vivo; theoretical SDM elements/tasks/behaviors etc. You could consider acronyms for each.</p> <p>A clear example of where this confusion arises is line 34 of page 14 where it was not immediately clear to me which sense of SDM “to capture SDM” refers to. Another example, where “SDM” is used in line 27 of page 5, I think it means the theoretically described process of SDM, but it could also mean what patients and clinicians actually do in real life in coming to decisions.</p> <p>I want to note however, there are several times within the paper where these distinctions are made clearly and usefully.</p> <p>Typos:</p> <p>Remove “been” from line 16 page 6          Add “the” before “MAGIC” in line 8 page 10          Change “The” to “To” at the end of line 31 page 10          There may be an additional space between “measure” and “were” line 53 page 10</p>
-------------------------	---

<b>REVIEWER</b>	margaret schwarze University of Wisconsin
<b>REVIEW RETURNED</b>	17-Apr-2019

<p><b>GENERAL COMMENTS</b></p>	<p>I have been asked to re-review “The ability of observer and self-report measures to capture SDM in clinical practice.” This manuscript is much improved since the previous iteration with a stronger focus on precise aims and tightening of the research question and presentation of results. I think the topic is interesting and adds to the literature, their focus on contextual factors and longitudinal nature of decision making is particularly important. There are still a few points that need adjustment as I believe the authors have stated a conclusion about current measurement practices that are incorrect.</p> <p>Remaining MAJOR concern: There are 2 items related to OPTION5 scoring that I believe need to be rectified. The authors still do not seem to be calculating the OPTION 5 score correctly. They have scaled EACH DOMAIN up to a 100 point scale. So they report, for example, “eliciting patient preferences” has a mean score of 80.76. The MAX score for each domain is 20 (once it is scaled), this is calculated by scoring each domain on a scale of 0 – 4 and then multiplying the score by 5, or as instructed by the OPTION5 manual, adding all domains (max score 20) and then scaling that resultant to 100. So, multiplying the final score by 5. Have they averaged each domain score to get a final OPTION5 score? That is not how it is calculated, it’s a summative score of each domain. Here are the instructions verbatim: Items should be scored independently. Scoring should be summed, so that the potential total score is between 0 and 20. For ease of interpretation, we advocate rescaling this score to be between 0 and 100.</p> <p>Next, the authors note that “the observational and self-report measures we used were unable to fully capture the SDM process in clinical practice.” This needs to be qualified – noting that if the measurement measures only one clinic visit, that is the only way their statement is correct. OPTION5 scoring is designed for multiple visits with clear instructions about how to score over time. The problem is not the measure, the problem is a problem of data collection. (Agree, that is potentially a feasibility problem).</p> <p>Remaining MINOR concerns: At beginning of methods.... “all consultations were audio recorded and transcribed verbatim.” This needs to reflect what was done, not suggest that over a period of time all transcripts were recorded by an investigational team. If this statement is a summary of the study design then it needs to reflect that more accurately – e.g., “consultations that clinicians audio recorded over a period of time were reviewed by the research team” would not mislead readers to believe that over a period of time all the visits were recorded.</p> <p>In patient recruitment section how did they know when to stop enrollment? The authors note they used a premise of information power, but do not describe how this was translated to sample size or an instruction to researchers or clinicians to stop recruitment. It would be helpful to think more in depth about how clinicians should respond to patients who present with pre-determined notions of treatment plans. While this might suggest a need to not repeat information, for a decision to be shared, the clinician needs to understand the patient’s reasoning behind the treatment choice, as simple statements of choice or experience with treatment provides minimal, if any, insight as to why this treatment is best for the patient or reassurance that the final plan is consistent with patient preferences. Efforts to explore a treatment preference, show how a treatment preference reflects patient’s goals and values could be used to satisfy requirements for eliciting</p>
--------------------------------	---

	preferences or confirming understanding of pros/cons, which would still allow for appropriate scoring of shared decision making even when the patient arrives with previous information and/or preferences.
<b>REVIEWER</b>	haske van veenendaal ESHPM, Erasmus University
<b>REVIEW RETURNED</b>	15-Apr-2019
<b>GENERAL COMMENTS</b>	Thank you for addressing the feedback in a sufficient way.

## VERSION 2 – AUTHOR RESPONSE

### Reviewer 1

Throughout the paper there is a tension between four senses of SDM: 1.) the a-priori/accepted definition of SDM stated clearly for example in line 36 of page 25; 2.) SDM consisting of theoretically derived elements/behaviors (three talk/options) 3.) the SDM encounter which is the time when mandated SDM (as a policy) is supposed to occur; and 4.) SDM as what patients and clinicians are actually doing (inside or outside of the encounter) in coming to decisions. This tension is central to the argument of the paper but because the term SDM is used for all senses it can be a little confusing for the reader. This might be resolved by calling attention to these distinctions, or creating terms for them and using them within the text. E.g. a-priori SDM/the accepted definition of SDM; SDM in-vivo; theoretical SDM elements/tasks/behaviors etc. You could consider acronyms for each.

A clear example of where this confusion arises is line 34 of page 14 where it was not immediately clear to me which sense of SDM “to capture SDM” refers to. Another example, where “SDM” is used in line 27 of page 5, I think it means the theoretically described process of SDM, but it could also mean what patients and clinicians actually do in real life in coming to decisions.

I want to note however, there are several times within the paper where these distinctions are made clearly and usefully.

We would like to thank the reviewer for drawing our attention for the need to clarify the use of the term SDM in relation to definitions of SDM, theoretical conceptions of SDM, policy related SDM and SDM in clinical practice. We have added relative descriptions to the use of the term ‘SDM’ to reflect which sense of SDM is being discussed at that time. We have also clarified the use of SDM on the two occasions noted here (P34 and P27).

Typos:

Remove “been” from line 16 page 6

Add “the” before “MAGIC” in line 8 page 10 Change “The” to “To” at the end of line 31 page 10 There may be an additional space between “measure” and “were” line 53 page 10

We would like to thank the reviewer for highlighting the typos. They have now been amended.

Reviewer: 2

1. There are 2 items related to OPTION5 scoring that I believe need to be rectified. The authors still do not seem to be calculating the OPTION 5 score correctly. They have scaled EACH DOMAIN up to a 100 point scale. So they report, for example, “eliciting patient preferences” has a mean score of 80.76. The MAX score for each domain is 20 (once it is scaled), this is calculated by scoring each domain on a scale of 0 – 4 and then multiplying the score by 5, or as instructed by the OPTION5 manual, adding all domains (max score 20) and then scaling that resultant to 100. So, multiplying the final score by 5. Have they averaged each domain score to get a final OPTION5 score? That is not how it is calculated, it’s a summative score of each domain. Here are the instructions verbatim: Items should be scored independently. Scoring should be summed, so that the potential total score is between 0 and 20. For ease of interpretation, we advocate rescaling this score to be between 0 and 100.

We would like to thank the reviewer for highlighting the need for further clarity. Two raters scored 25 breast cancer and 26 CKD consultations using the observer OPTION5 measure and the mean of the two global scores was taken for each consultation (out of a possible score of 20, which was then rescaled to 100). Therefore, standardised OPTION5 global scores ranged from 65-95 with a mean score of 82.82 (S.D. 8.54) for breast consultations and 47.50-80.0 with a mean score of 66.53 for pre-dialysis consultations. We have also presented the mean of each domain score to show the extent to which the different domains were present within the consultations e.g. outlining alternative options, eliciting preferences etc. However, we do note that for the presentation of data in Table 2, we had rescaled each domain from 0-100 instead of 0-20. We have therefore rescaled the results presented in table 2 to range from 0-20. We would like to thank the reviewer for noting this discrepancy, which has now been rectified. As a result, the table has been amended as follows:

OPTION-5 item	Mean score, S.D.		Minimum score		Maximum score	
	Breast	Renal	Breast	Renal	Breast	Renal
1 Alternative options	15	3.60	11.05	2.66	10	7.5
2 Support deliberation	20	15				

12.5,3.06

8.65, 2.79

7.5

5.0

20

15

### 3 Information about options

17.6, 2.22

16.05,2.14

12.5

10

20

20

### 4 Eliciting preferences

17.7, 2.87

16.15,3.48

10

10

20

20

### 5 Integrating preferences

18.5, 2.39

15.38, 3.65

12.5

7.5

20

20

Table 2 Standardised (0-20) observer OPTION-5 domain mean scores, standard deviation, minimum and maximum scores for each item (n= 25 breast cancer and n=26 pre-dialysis consultations)

2. Next, the authors note that “the observational and self-report measures we used were unable to fully capture the SDM process in clinical practice.” This needs to be qualified – noting that if the measurement measures only one clinic visit, that is the only way their statement is correct. OPTION5 scoring is designed for multiple visits with clear instructions about how to score over time. The problem is not the measure, the problem is a problem of data collection. (Agree, that is potentially a feasibility problem).

We would like to thank the reviewer for highlighting this, and we agree that the observer OPTION5 measure could be used repeatedly over a number of consultations (and we did this in the breast care team, where two consultations per participant were recorded). However, as noted by the reviewer the feasibility of doing so in practice can be a significant barrier. In response to the reviewers feedback we have clarified in the discussion section that the OPTION5 measure could be used to capture SDM over time, but that feasibility to do this in clinical practice might be problematic. We clarified further in the text:

“We observed that the process of SDM is distributed across consultations and clinicians and this poses important questions about when we should be measuring whether SDM took place, and whom we are assessing. Although the OPTION5 measure could be used to capture SDM over time, by recording multiple visits, and scores, collecting this data might be unfeasible within the context of clinical practice. “

3. At beginning of methods.... “all consultations were audio recorded and transcribed verbatim.” This needs to reflect what was done, not suggest that over a period of time all transcripts were recorded by an investigational team. If this statement is a summary of the study design then it needs to reflect that more accurately – e.g., “consultations that clinicians audio recorded over a period of time were reviewed by the research team” would not mislead readers to believe that over a period of time all the visits were recorded.

We agree that this sentence can be misleading and thank the reviewer for highlighting this. We have subsequently amended the sentence to read:

“Consultations audio-recorded by clinicians over a period of time, were passed to the research team and were subsequently transcribed verbatim.”

4. In patient recruitment section how did they know when to stop enrollment? The authors note they used a premise of information power, but do not describe how this was translated to sample size or an instruction to researchers or clinicians to stop recruitment.

Recruitment was based on the judgement of information power by the researchers, who advised the clinicians to stop recording consultations once this was achieved. We have provided further clarity in the Qualitative analysis section of the manuscript, and have added the following sentence:

“Sample size was iteratively determined based on the themes that were emerging through the data, and continued until no major new themes were being identified. Clinicians were informed when to stop collecting data based on the researchers’ assessment of information power.”

5. It would be helpful to think more in depth about how clinicians should respond to patients who present with pre-determined notions of treatment plans. While this might suggest a need to not repeat information, for a decision to be shared, the clinician needs to understand the patient’s reasoning behind the treatment choice, as simple statements of choice or experience with treatment provides minimal, if any, insight as to why this treatment is best for the patient or reassurance that the

final plan is consistent with patient preferences. Efforts to explore a treatment preference, show how a treatment preference reflects patient's goals and values could be used to satisfy requirements for eliciting preferences or confirming understanding of pros/cons, which would still allow for appropriate scoring of shared decision making even when the patient arrives with previous information and/or preferences.

We would like to thank the reviewer for this reflection, and agree that efforts to explore treatment preference and confirming understanding of pros and cons can result in appropriate scoring of shared decision making even if the patient arrives with pre-determined preferences. We have added further reflection of this in the discussion section:

“Detailed thematic analysis showed that not only was the SDM encounter distributed, it also identified that it was contextually adapted and that individuals would often need to make multiple and interacting decisions (as explained in Table 4). We identified that some patients already had pre-determined treatment preferences. In such situations it is important to acknowledge that for SDM to occur, the clinician needs to understand the reasoning behind the treatment preference. Clinicians should be encouraged to explore how the treatment preference meets with patient goals, as well as their understanding of associated pros and cons. If the treatment preference meets the goals of the patient and if their understanding of the pros and cons is accurate, then the consultation should be seen as a process of SDM.”

We hope that the above responses and the changes to the manuscript address the editors' and reviewers' comments. We hope that we have been able to improve the manuscript further and that it can now proceed to publication.

### VERSION 3 – REVIEW

<b>REVIEWER</b>	Ian Hargraves Mayo Clinic
<b>REVIEW RETURNED</b>	29-May-2019

<b>GENERAL COMMENTS</b>	<p>This is an excellent paper that has usefully incorporated previous reviewer comments. I highly recommend it for publication.</p> <p>I have 1 very minor suggestion and observed one typo.</p> <p>I suggest adding the word "theoretical" in front of "elements of SDM" in column 2, row 2 of Table 3</p> <p>In the first sentence of the conclusion correct "ptactice" to "practice"</p>
-------------------------	---

<b>REVIEWER</b>	margaret schwarze University of Wisconsin
<b>REVIEW RETURNED</b>	22-May-2019

<b>GENERAL COMMENTS</b>	The authors have addressed the reviewers comments well, except that they have not changed the results presented in the abstract with respect to OPTION5 scoring, which still seems to be on a 100 point scale for each domain.
-------------------------	--

### VERSION 3 – AUTHOR RESPONSE

Reviewer: 2

The authors have addressed the reviewers comments well, except that they have not changed the results presented in the abstract with respect to OPTION5 scoring, which still seems to be on a 100 point scale for each domain.

We would like to thank the reviewer for highlighting this inconsistency. We have now amended the results presented in the abstract to reflect the scoring in the paper.

Reviewer: 1

This is an excellent paper that has usefully incorporated previous reviewer comments. I highly recommend it for publication.

I have 1 very minor suggestion and observed one typo.

I suggest adding the word "theoretical" in front of "elements of SDM" in column 2, row 2 of Table 3

In the first sentence of the conclusion correct "ptactice" to "practice"

We would like to thank the reviewer for the suggestion to add 'theoretical' to the " in front of "elements of SDM" in column 2, row 2 of Table 3. We have done this.

We would also like to thank the reviewer for highlighting the typo which we have now corrected.