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)	Routine patient-reported experience measurement of shared
	decision-making in the US: a qualitative study of the current state
	according to frontrunners
AUTHORS	Forcino, Rachel; Meinders, Marjan; Engel, Jaclyn; O'Malley, A.
	James; Elwyn, Glyn

VERSION 1 – REVIEW

REVIEWER	Shannon Fuller
	University of California, San Francisco (USA)
REVIEW RETURNED	14-Feb-2020

GENERAL COMMENTS	Thank you for the opportunity to review this manuscript. This paper provides a comprehensive description of the SDM measurement practices in healthcare facilities across the US and uses qualitative methods to understand the barriers and facilitators to routine SDM measurement.
	The methods are described thoroughly in accordance with the SRQR checklist, and an interview guide has been included as a supplemental file. The authors effectively use diffusion of innovation theory to organize to frame and organize the results.
	My suggestions are relatively minor. Mainly, I wanted to see more detail if the word count allows.
	For example, in the Innovation paragraph under barriers, when the authors note that the relative advantage of patient-reported measurement is not yet sufficient in one of the sites, it would be interested to hear more about why that is the case. Then, in either the background or the discussion section, the authors could add further commentary on the proxy measures for decision aid use. I was left wondering what sites were missing by only having the proxy measures, especially if the relative advantage was perceived as low.
	Other comments for the results section:
	Table 4 – there are a couple of boxes where illustrative quotes are not included. Is there anything that could be added there?
	Barriers The text says "several key barrierswere identified," but only four are described.
	Quote under "Use of SDM Data" section – This quote in the second paragraph is really interesting. Could you expand more on this fear? Are there specific metrics that providers are worried

about? And when the quote says that they "haven't quite gone there yet," what does that mean?
For the same quote above, is it possible to pull out one longer excerpt of the quote, rather than to piece it together? That will help retain the participant's voice, and the authors can add further paraphrasing before or after the quote to help contextualize and clarify the information.

REVIEWER	Stig Bjønness Centre for Resilience in Healthcare (SHARE), Department for Public Health, Faculty of Health Sciences, University of Stavanger, Stavanger, Norway
REVIEW RETURNED	26-Mar-2020

GENERAL COMMENTS

I have reviewed the manuscript entitled "Routine patient-reported experience measurement of shared decision-making in the US: a qualitative study of the current state according to frontrunners". I find this article interesting. It is important as it explores patient-reported experiences with SDM. As a part of quality improvement, this field has received limited attention in the research literature. However, there are a few points that should be taken into consideration to improve the paper further. When reviewing the manuscript, I struggled with some of the details and use of concepts. Given the importance and relevance of the study for a broader audience, it should be revised for easier readability (examples of sentences containing about 50 words: line 34-39 p.3, line 37-42 p.4, line 24-29 p.5, line 52-57 p.13). Below I will briefly account for my issues of concern, and I hope you will find them helpful.

Introduction:

In the last paragraph, you describe the study aim. I suggest you present the aim and then objectives or research questions. Methods:

The design is described as a descriptive multiple case study. I believe some adjustments or further description would be clarifying. You made contact with individuals from 23 US centers. I suggest changing the term qualitative survey. A survey is rarely qualitative, even though it has open-ended questions. The orientation of the study to explore how and why. In this context, I perceive the survey to be a tool to identify sites to recruit participants to qualitative interviews. If so, the design is somehow misleading. However, if the survey is a method to answer objective 1, as reflected in the result, it should be described.

Recruitment:

Please add information about the sample in the method section after the description of the recruitment (survey and interviews, n=?). Description of the participants is missing in this section. I would also like to know how many participants were included from each unit, not just the total number of participants in the study. Line 29-30: "participate in a survey and/or telephone interview". information about how the in-depth interviews were conducted and if it were the same participants in the survey and the interviews would be useful.

A discussion of sample size and considerations regarding saturation or information power would strengthen the study. Data:

I suggest changing this topic to Data collection

The 12-item survey is well presented. However, I would prefer to see the questionnaire as an appendix.

The interviews are also well described. It would be nice if you could add the duration of the interviews.

Analysis

The survey is used to identify instances of routine SDM measurement, and no further description is needed unless this is considered as a method to gather data to answer objective 1. Regarding the qualitative thematic analysis, some more information would be helpful: What themes and how many emerged from the analysis? Can you give examples on codes? Results:

The first part of the result (page 5 and table 1) is a description of the recruitment and sample. Therefore, I suggest reorganizing the placement of text sections.

The numbers are a bit puzzling. As earlier described, 32 individuals from 23 centers were contacted. However, 42 were referred to the research team by your sampling approach, and 36 of those (from 26 organizations) contacted. If some were excluded and some additional participants and/or sites included, it should be described how. According to table 1, the study consists of 8 sites. Out of those 36 invited participants, 15 did not respond or provide measurement details. As far as I understand, 21 conducted the survey, and 10 participated in interviews (cf. the abstract)? Anyhow, please provide a description of this and place it in the method section.

Tables are a nice way to present results. However, the use of tables in the article is somewhat exaggerated.

Table 2 contains the questionnaires in three SDM measures. It would have been sufficient to give a brief description of them, and possibly add it as an appendix.

Regarding table 4: Is this reflecting the themes and codes, if so please specify.

The results are reported as quantitative data. Note that qualitative studies are not aiming for quantifying data to generalize them, rather to give in-depth descriptions of findings providing insight to a certain population and phenomenon. When using tables to report results, the content of the tables should not be the same as the text, but rather complement each other. The text in the result section following table 4 should, therefore, be revised. Quotations are used to illustrate. However, the same quotations are used in both table 4 and the text. I would recommend using quotes that are more distributed among the participants (nine out of 12 quotes are from two participants). The quotes are not results but rather used to illuminate the main findings in the different categories. The rest of the quotes should be incorporated in the analytical text. Discussion:

It is nice to start the discussion with a short summary of the results, then compare and discuss results with other research. The key findings should form one paragraph and can be shortened. It would be nice if you could include methodology and how it influences the results in the discussion.

Limitations

The content of this paragraph depends on the changes made in the manuscript. The sample size, number of sites/cases, and data collection should be addressed.

I wish the authors good luck with the revision.

VERSION 1 – AUTHOR RESPONSE

Reviewers' Comments to Author:

Reviewer: 1

Reviewer Name: Shannon Fuller

Institution and Country: University of California, San Francisco (USA)

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

Thank you for the opportunity to review this manuscript. This paper provides a comprehensive description of the SDM measurement practices in healthcare facilities across the US and uses qualitative methods to understand the barriers and facilitators to routine SDM measurement.

The methods are described thoroughly in accordance with the SRQR checklist, and an interview guide has been included as a supplemental file. The authors effectively use diffusion of innovation theory to organize to frame and organize the results.

My suggestions are relatively minor. Mainly, I wanted to see more detail if the word count allows.

Reviewer comment 1.1: For example, in the Innovation paragraph under barriers, when the authors note that the relative advantage of patient-reported measurement is not yet sufficient in one of the sites, it would be interested to hear more about why that is the case.

Response 1.1: Thank you for the opportunity to elaborate on this interesting finding. The impact of financial incentives we observed at sites 1 and 2 suggests that relative advantage has much to do with what activities are rewarded by payers. For the most part, the SDM process between patients and clinicians is not yet rewarded by payers in the US; instead, some emerging payer initiatives focus on distribution of decision aids, understandably leading to uptake of this kind of proxy measure as observed at sites 7 and 8. We've now elaborated on these points in the results and discussion section, as follows:

"The success of financial incentives for patient-reported SDM measurement at sites 1 and 2 suggests that relative advantage is associated with those activities that are rewarded by payers." (Page 12)

"Underlying the sparse routine use of patient-reported SDM measurement is a US context in which the SDM process is not yet widely rewarded by healthcare payers. There are a few emerging exceptions, including the Centers for Medicare and Medicaid Services requiring documentation of SDM for lung cancer screening.[26] However, such initiatives tend not to differentiate distribution of patient decision aids from an SDM process in which patients and clinicians share information about potential benefits and harms, engage in dialogue about preferences and values, and jointly decide on next steps. The relative advantage of a valid and reliable SDM measure, inclusive of potential data collection costs, over low-cost proxy measures such as extent of decision aid distribution, is therefore currently absent in two of the participating sites (sites 7 and 8)." (Page 13)

Reviewer comment 1.2: Then, in either the background or the discussion section, the authors could add further commentary on the proxy measures for decision aid use. I was left wondering what sites were missing by only having the proxy measures, especially if the relative advantage was perceived as low

Response 1.2: We have now further elaborated on this point in the discussion section: "Underlying the sparse routine use of patient-reported SDM measurement is a US context in which the SDM process is not yet widely rewarded by healthcare payers. There are a few emerging exceptions, including the Centers for Medicare and Medicaid Services requiring documentation of SDM for lung cancer screening.[26] However, such initiatives tend not to differentiate distribution of

patient decision aids from an SDM process in which patients and clinicians share information about potential benefits and harms, engage in dialogue about preferences and values, and jointly decide on next steps. The relative advantage of a valid and reliable SDM measure, inclusive of potential data collection costs, over low-cost proxy measures such as extent of decision aid distribution, is therefore currently absent in sites 7 and 8. In settings where the SDM process is already routine, monitoring decision aid distribution can be a helpful proxy; however, measures of the SDM process itself are needed for patient-centered culture change and SDM skill-building." (Page 13)

Reviewer comment 1.3: Table 4 – there are a couple of boxes where illustrative quotes are not included. Is there anything that could be added there?

Response 1.3: We have now added quotes to the result summary table (now named Table 3).

Reviewer comment 1.4: Barriers -- The text says "several key barriers...were identified," but only four are described.

Response 1.4: We have now removed that introductory text. (Page 11)

Reviewer comment 1.5: Quote under "Use of SDM Data" section – This quote in the second paragraph is really interesting. Could you expand more on this fear? Are there specific metrics that providers are worried about? And when the quote says that they "haven't quite gone there yet," what does that mean?

Response 1.5: We've now added a longer quote in which the participant explains that regardless of metric (e.g. SDM, efficiency, outcomes), having a health insurance company collect performance data causes providers anxiety about the insurance company's motives, e.g. the possibility of the payer steering patients to higher-scoring providers. For that reason, site 1 has so far declined to pursue using their SDM measurement results to find and train low-scoring providers. We updated this section as follows:

"One site, however, struggles to find a use for its extensive SDM data that is deemed acceptable by its community of clinicians (site 1). As a payer organization, site 1 finds that its collection of SDM data has "created a little bit of trepidation" within the clinician community due to a perception that they could "weaponize this information" (P01). The participant explains:

[Low SDM scores] make the physician look bad and we, as a health plan, could frankly use that information to steer patients away from those kinds of doctors and towards the doctors that get better scores. That's part of the problem with anything when you're collecting data, any type of data. Whether it's shared decision-making data or efficacy data around quality scores or even around outcomes, the perception is that health plans can use that data against them to steer patients away and send them to higher performers. That's the concern from providers and so we have this data. We don't intend on doing that. We don't intend on using the scores in a way to punish or, right now, even provide benefit to those high scorers. We just want to collect the data to better understand shared decision-making. Is the process occurring? How the patients – how are they responding to it? (P01, site 1)

Site 1 aspires to "use the information to try to educate" and offer training to lower-performing clinicians (P01). However, despite a desire to "use it as a mechanism to help educate maybe the lower-scored folks versus the higher-scored folks...[site 1 hasn't] quite gone there yet" (P01) with regard to training low-scoring providers in SDM." (Pages 12-13)

Reviewer comment 1.6: For the same quote above, is it possible to pull out one longer excerpt of the quote, rather than to piece it together? That will help retain the participant's voice, and the authors can add further paraphrasing before or after the quote to help contextualize and clarify the information.

Response 1.6: Thank you for this suggestion! As described in response 1.5 above, we've now incorporated a longer excerpt. (Pages 12-13)

Reviewer: 2

Reviewer Name: Stig Bjønness

Institution and Country: Centre for Resilience in Healthcare (SHARE), Department for Public Health,

Faculty of Health Sciences, University of Stavanger, Stavanger, Norway Please state any competing interests or state 'None declared': None declared

I have reviewed the manuscript entitled "Routine patient-reported experience measurement of shared decision-making in the US: a qualitative study of the current state according to frontrunners". I find this article interesting. It is important as it explores patient-reported experiences with SDM. As a part of quality improvement, this field has received limited attention in the research literature.

However, there are a few points that should be taken into consideration to improve the paper further. When reviewing the manuscript, I struggled with some of the details and use of concepts.

Reviewer comment 2.1: Given the importance and relevance of the study for å broader audience, it should be revised for easier readability (examples of sentences containing about 50 words: line 34-39 p.3, line 37-42 p.4, line 24-29 p.5, line 52-57 p.13).

Response 2.1: Thank you for pointing this out. We have revised the manuscript for easier readability with particular attention to the areas you specified.

Reviewer comment 2.2: Introduction: In the last paragraph, you describe the study aim. I suggest you present the aim and then objectives or research questions.

Response 2.2: We have now listed the research question, as follows:

"In this study, we aim to 1) identify and describe instances of routine patient-reported SDM measurement in the US; and 2) explore barriers and facilitators of routine patient-reported SDM measurement for quality improvement using the Greenhalgh et al. diffusion of innovations theoretical framework.[13] Our primary research question was: what are the barriers and facilitators of routine patient-reported SDM measurement in the US?" (Page 3)

Reviewer comment 2.3: The design is described as a descriptive multiple case study. I believe some adjustments or further description would be clarifying. You made contact with individuals from 23 US centers. I suggest changing the term qualitative survey. A survey is rarely qualitative, even though it has open-ended questions.

Response 2.3: We have removed the term 'qualitative survey' and reworded as follows:

"To describe examples of patient-reported SDM measurement, we employed a multi-pronged data collection approach, including a survey of representatives from leading SDM centers, and, as available, in-depth interviews of representatives from relevant sites." (Page 3)

Reviewer comment 2.4: The orientation of the study to explore how and why. In this context, I perceive the survey to be a tool to identify sites to recruit participants to qualitative interviews. If so, the design is somehow misleading. However, if the survey is a method to answer objective 1, as reflected in the result, it should be described.

Response 2.4: Thank you for this opportunity to clarify. The survey was a method to gather data specific to the study aim for those participants who were unable to participate in a semi-structured interview. We have now added the questionnaire as an appendix and clarified the multi-pronged data collection approach in the methods section, as follows:

"To describe examples of patient-reported SDM measurement, we employed a multi-pronged data collection approach, including a survey of representatives from leading SDM centers, and, as available, in-depth interviews of representatives from relevant sites." (Page 3)

"Where we were unable to conduct semi-structured interviews with relevant contacts, we conducted a 12-item open-ended survey hosted by Qualtrics online survey software to gain insight into routine SDM measurement efforts. Participants were asked to provide information on which SDM measures were in routine use at their organizations, how the measures were selected, details on measurement volume, what concerns are voiced in their organizations about SDM measurement, and how the organizations use the SDM data they collect for quality improvement (see Appendix 2)." (Page 4)

Reviewer comment 2.5: Please add information about the sample in the method section after the description of the recruitment (survey and interviews, n=?). Description of the participants is missing in this section. I would also like to know how many participants were included from each unit, not just the total number of participants in the study.

Response 2.5: We sought to thoroughly describe recruitment procedures in the methods section, but we think that a participant profile is best suited for the results section. We therefore included a description of participants, including specification of how many participants enrolled from each site with active SDM measurement programs, in the far-right column of Table 1. (Pages 5-6)

Reviewer comment 2.6: Line 29-30: "participate in a survey and/or telephone interview". information about how the in-depth interviews were conducted and if it were the same participants in the survey and the interviews would be useful.

Response 2.6: We have now elaborated on these important recruitment and data collection procedures, as follows:

"The research team made initial contact by email, followed by either an emailed link to the survey or an interview invitation, depending on participant availability and preference." (Page 4) "In-depth interviews were conducted by Zoom teleconference (audio only)." (Page 4)

Reviewer comment 2.7: A discussion of sample size and considerations regarding saturation or information power would strengthen the study.

Response 2.7: Through our broad snowball sampling approach, we sought to contact all active SDM researchers and leading SDM practitioners in the US. Data derived from this small but heterogeneous group of sampled institutions did not reach thematic saturation, though we did observe several commonalities as described in the key findings. We have now made this explicit in the strengths and limitations section of the discussion:

"Through our broad snowball sampling approach, we sought to conduct a thorough search of active SDM researchers and leading SDM practitioners in the US. Data derived from this small but heterogeneous group of institutions did not reach thematic saturation, though we observed several key commonalities as described in the key findings. As this study is an early exploration into routine SDM measurement, we found that the landscape is diverse and currently without consensus. This study therefore presents views of early adopters, relevant even without thematic saturation." (Page 14)

Reviewer comment 2.8: Data: I suggest changing this topic to Data collection Response 2.8: We have now renamed this subsection "Data collection." (Page 4)

Reviewer comment 2.9: The 12-item survey is well presented. However, I would prefer to see the questionnaire as an appendix.

Response 2.9: We have now added the questionnaire as Appendix 2.

Reviewer comment 2.10: The interviews are also well described. It would be nice if you could add the duration of the interviews.

Response 2.10: We have now added this information as follows:

"Six participants completed semi-structured interviews, with an average interview duration of 40 minutes." (Page 5)

Reviewer comment 2.11: The survey is used to identify instances of routine SDM measurement, and no further description is needed unless this is considered as a method to gather data to answer objective 1.

Response 2.11: The survey was a method to gather data specific to the study aim for those participants with whom we were unable to conduct a semi-structured interview. We have now added the questionnaire as an appendix and clarified the multi-pronged data collection approach in the methods section, as follows:

"To describe examples of patient-reported SDM measurement, we employed a multi-pronged data collection approach, including a survey of representatives from leading SDM centers, and, as available, in-depth interviews of representatives from relevant sites." (Page 3)

"Where we were unable to conduct semi-structured interviews with relevant contacts, we conducted a 12-item open-ended survey hosted by Qualtrics online survey software to gain insight into routine SDM measurement efforts. Participants were asked to provide information on which SDM measures were in routine use at their organizations, how the measures were selected, details on measurement volume, what concerns are voiced in their organizations about SDM measurement, and how the organizations use the SDM data they collect for quality improvement (see Appendix 2)." (Page 4)

Reviewer comment 2.12: Regarding the qualitative thematic analysis, some more information would be helpful: What themes and how many emerged from the analysis? Can you give examples on codes?

Response 2.12: We have now relabeled the second and third columns in Table 3 to specify themes and relevant codes. (Page 8)

Reviewer comment 2.13: The first part of the result (page 5 and table 1) is a description of the recruitment and sample. Therefore, I suggest reorganizing the placement of text sections. Response 2.13: We have now moved summary text about the sample and Table 1 up to the first results paragraph:

"The recruitment process and full snowball sample referral network is depicted in Figure 1 and Figure 2. Table 1 summarizes SDM measurement at each included site with active SDM measurement initiatives. One health insurance company (site 1) and two provider organizations (sites 2 and 3) routinely measure SDM from patients' perspectives." (Page 5)

Reviewer comment 2.14: The numbers are a bit puzzling. As earlier described, 32 individuals from 23 centers were contacted. However, 42 were referred to the research team by your sampling approach, and 36 of those (from 26 organizations) contacted. If some were excluded and some additional participants and/or sites included, it should be described how. According to table 1, the study consists of 8 sites. Out of those 36 invited participants, 15 did not respond or provide measurement details. As far as I understand, 21 conducted the survey, and 10 participated in interviews (cf. the abstract)? Anyhow, please provide a description of this and place it in the method section.

Response 2.14: Thank you for this clarification. We have now corrected the abstract and clarified the results as follows:

"Participants: Current or former adult employees of healthcare organizations with prior SDM activity and that may be conducting routine SDM measurement." (Page 2)

"Of 42 people referred to the research team through our initial sample (32 people) and our snowball sampling approach (10 people), 36 people from 26 organizations were contacted for survey and interview recruitment." (Page 5)

"Six participants completed semi-structured interviews, with an average duration of 40 minutes." (Page 5)

Reviewer comment 2.15: Tables are a nice way to present results. However, the use of tables in the article is somewhat exaggerated. Table 2 contains the questionnaires in three SDM measures. It would have been sufficient to give a brief description of them, and possibly add it as an appendix. Response 2.15: Table 2 has now been converted to an appendix (Appendix 3).

Reviewer comment 2.16: Regarding table 4: Is this reflecting the themes and codes, if so please specify.

Response 2.16: We have now changed the second column heading of the former Table 4 (now labeled Table 3) to "Attributes/Themes" to maintain reference to the diffusion of innovations model, which refers to those specific attributes, while clarifying that coded data mapped to those themes. (Page 9)

Reviewer comment 2.17: The results are reported as quantitative data. Note that qualitative studies are not aiming for quantifying data to generalize them, rather to give in-depth descriptions of findings providing insight to a certain population and phenomenon. When using tables to report results, the content of the tables should not be the same as the text, but rather complement each other. The text in the result section following table 4 should, therefore, be revised. Quotations are used to illustrate. However, the same quotations are used in both table 4 and the text. I would recommend using quotes that are more distributed among the participants (nine out of 12 quotes are from two participants). The quotes are not results but rather used to illuminate the main findings in the different categories. The rest of the quotes should be incorporated in the analytical text.

Response 2.17: We have now updated the table with additional illustrative quotes to further vary the quoted participants and differentiate the table's content from the results described in-text. (Table 3, pages 8-10)

Reviewer comment 2.18: It is nice to start the discussion with a short summary of the results, then compare and discuss results with other research. The key findings should form one paragraph and can be shortened.

Response 2.18: We have now shortened the key findings sub-section as follows:

"In organizations where patient-reported SDM measurement is routine, facilitators include: compatibility of SDM measurement with core organizational values; brevity of the collaboRATE patient-reported SDM measure; trialability (and potential for subsequent expansion) of patient-reported SDM measurement within the organization; flexibility in how measures can be implemented; involvement of both clinical champions and rank-and-file clinicians in the decision to measure SDM performance; an environment in which payers (e.g., health insurance companies) have begun to require provider organizations to measure patients' experiences of SDM; and dedicated resources (i.e. personnel) within the organizations to design and maintain their SDM measurement programs. Barriers include inadequate perceived relative advantage of patient-reported SDM measurement over proxy measures, a paucity of patient-reported SDM measures that are sufficiently pragmatic for routine and widespread use, and the existence of competing priorities for organizational leadership

when it comes to patient experience. The few organizations we identified with routine patient-reported SDM measurement tend to use the resulting information for internal benchmarking and quality improvement initiatives. However, site 1, due to constraints unique to payer-only organizations, is still in the process of developing a tenable use of the extensive patient-reported SDM data it collects." (Page 12)

Reviewer comment 2.19: It would be nice if you could include methodology and how it influences the results in the discussion.

Response 2.19: We have included the following discussion of methodology and its influence on results: "Through the authors' professional networks and a snowball sampling approach, recruitment efforts for this study involved a near-census of major SDM initiatives in the United States. Through our broad snowball sampling approach, we sought to conduct a thorough search of active SDM researchers and leading SDM practitioners in the US. Data derived from this small but heterogeneous group of institutions did not reach thematic saturation, though we observed several key commonalities as described in the key findings. As this study is an early exploration into routine SDM measurement, we found that the landscape is diverse and currently without consensus. This study therefore presents views of early adopters, relevant even without thematic saturation. However, the multi-modal data collection approach led to varying levels of detail available across included participants and sites, which is a limitation." (Page 15)

Reviewer comment 2.20: Limitations: The content of this paragraph depends on the changes made in the manuscript. The sample size, number of sites/cases, and data collection should be addressed. Response 2.20: We have added to the limitations section as follows:

"Through our broad snowball sampling approach, we sought to conduct a thorough search of active SDM researchers and leading SDM practitioners in the US. Data derived from this small but heterogeneous group of institutions did not reach thematic saturation, though we observed several key commonalities as described in the key findings. As this study is an early exploration into routine SDM measurement, we found that the landscape is diverse and currently without consensus. This study therefore presents views of early adopters, relevant even without thematic saturation. However, the multi-modal data collection approach led to varying levels of detail available across included participants and sites, which is a limitation." (Page 14)

VERSION 2 - REVIEW

REVIEWER	Shannon Fuller
	University of California, San Francisco
REVIEW RETURNED	05-May-2020
GENERAL COMMENTS	The authors have responded sufficiently to prior reviews. No
	further comments.