PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Effectiveness and implementation of SHared decision-making supported by OUTcome information among patients with breast cancer, stroke and advanced kidney disease: SHOUT study protocol of multiple Interrupted Time Series
AUTHORS	Hackert, Mariska; Ankersmid, Jet; Engels, Noel; Prick, Janine; Teerenstra, Steven; Siesling, Sabine; Drossaert, Stans; Strobbe, Luc; Van Riet, Yvonne; van den Dorpel, Marinus; Bos, Willem Jan; van der Nat, Paul; Van den Berg-Vos, RM; van Schaik, Sander; Garvelink, Mirjam; Van der Wees, Philip; Van Uden-Kraan, Cornelia

VERSION 1 – REVIEW

REVIEWER	Lindig, Anja University Medical Center Hamburg-Eppendorf, Department of Medical Psychology
REVIEW RETURNED	23-Nov-2021

GENERAL COMMENTS	PEER REVIEW Effectiveness and implementation of SHared decision-making supported by OUTcome information among patients with breast cancer, stroke and advanced kidney disease: SHOUT study protocol of multiple Interrupted Time Series bmjopen-2021-055324, BMJ Open Many thanks for the opportunity to review this study protocol. The authors intend to foster shared decision-making by implementing patients' outcome data in discussions about treatment options. Therefore, they developed personalized decision aids and a training for healthcare professionals. Especially the decision aids are well designed and are an interesting new tool to provide relevant information to patients. This is a very interesting and well-designed study with a clear objective and comprehensible described methodology. The authors plan to assess a range of measures, which will help to understand effects of the implementation strategies. All in all, this study has the potential to contribute to existing literature and I am looking forward to read about your results. Still, I have some minor comments and suggestions to further strengthen the manuscript. Introduction

	1. For readers, who are not familiar with the concept of SDM, please insert a short definition after the first paragraph. Methods and analysis 2. Page 4, line 41, "… trends on outcomes will be established …" – I am not sure if you use the correct phrase here. Is it rather "evaluated"? Intervention 3. Page 6, line 26, "A multicomponent intervention …" – I would not define the implementation strategy as one component of the intervention. I would suggest to explain that you developed a strategy to implement the interventions, PtDAs and trainings. 4. Page 6, line 34, "A literature review and needs assessment studies …" – Here, references are missing. 5. Page 6, line 43, "Usablility testing …" – How did you perform the usability testing? Observations in a real-world setting, interviews, focus groups? 6. I just wondered how you plan to enter personalized data into the PtDAs. Is this task of the physicians or are those data entered into the PtDA by physicians and patients together during their consultation? 7. Page 7, line 16, "INFLUENCE" – Is this an abbreviation? What do you mean here? 8. Page 7, last paragraph – Is the described content of the trainings also part of the e-learnings or are the e-learnings shorter with less content? 9. Page 7, line 49, "… reflection on audio-taped consultations …" – Are these consultations of physicians taking part in the trainings? Study design and procedure 10. Page 8, last paragraph – Why are hospitals not randomized? How do you decide for the order, hospitals start with the intervention? Data collection and methods

	11. Why are there different dates for follow up for patients with stroke compared to patients with breast cancer and AKD? Outcomes Effectiveness 12. Page 10, line 7 to 33 – This is the introduction for this section but it is quite redundant to the following paragraphs. I would suggest to shorten this introduction. Primary outcome measure 13. Page 10, line 41 – Here is a mistake with the references. I would also suggest to include a reference for the original German SDM-Q-9 scale by Scholl et al. Secondary outcome measures 14. I would suggest to structure this paragraph with numbers: "Secondary outcomes will be (1) patient reported SDM, […], (2) decisional conflict, measured …". This would increase readability. 15. Page 10, line 50 – What is meant by knowledge? Patients' knowledge about SDM or their own disease? 16. You do not plan to assess each measure for each disease or during each assessment phase. For example, the SF-12 will not be assessed for patients with stroke and quality of life measures for stroke will only be assessed for T1. Can you shortly explain why you choose specific measures for specific time points and diseases? 17. According to your overview on measures, patients have to fill out several measures with many items. This seem to result in quite long surveys. Do you offer some kind of incentives for study participation?

REVIEWER	Ubbink, Dirk Amsterdam University Medical Centres, Surgery
REVIEW RETURNED	06-Jan-2022

GENERAL COMMENTS	Many thanks for the opportunity to review this manuscript. It is a laudable attempt to improve shared decision-making (SDM) among three important and large groups of patients. Although I realise it is late to change the protocol at this stage, I do have the following remarks: Implementation of SDM is not only a matter of providing outcome information (patients may not make rational choices based on data, but on other outcomes clinicians tend to overlook), nor is it achieved by merely introducing decision support tools. The main

	implementation issues are awareness, willingness and behaviour among clinicians and patients themselves. This study focuses on the former two interventions. These are doomed to fail if there is no focus on the latter aspects: are patients and clinicians willing and capable to start SDM and to use these tools? I do not agree with the statement in the 'strengths and limitations of this study' that "randomization is not feasible". Implementation studies can surely be conducted in a randomised fashion, for example by means of a stepped-wedge cluster-RCT. For example: Scholl I, et al. Implement Sci 2021; Scalia P, et al. Implement Sci 2019, and also in the Netherlands: Thunnissen FM, et al. BMC Med Inform Decis Mak. 2021; de Mik SML, et al. BMC Med Inform Decis Mak. 2020; Al-Itejawi HHM, et al. BMJ Open. 2017. The same study strengths and limitations section focuses on the study design only. There are no remarks on expected outcome. Another limitation is the small effect size the authors want to achieve, while the implementation effort is large, involving over 1700 patients. Moreover, the purported effect is based on a subjectively scored outcome measure that is known to suffer from a ceiling effect. So, what will be the clinical relevance, even if such a small effect is found? Hence, why was the objective measure of patient involvement not chosen as primary outcome? It seems likely that the effects of the implementation differ due to the differences in interventions (treatment and post-treatment options) and disorders involved. The authors may want to consider studying and reporting on the effects of implementation for each of the three disorders separately. As the interventions focus on different aspects of the care process for each disorder, the effect size may vary as well. This would also have implications for the sample size calculation if conducted per disorder. The authors describe they will start the study as of November 2019. Given the time frame of 20 months pre- and post-implementation as shown in figure 2, the study has already been completed by now. As far as I understand, this is an exclusion criterion for publication in BMJ Open ("If data collection is complete, we will not consider the manuscript"). The authors state that "The moment by which hospitals switch from standard care to use of the intervention will not be randomized." This may lead to contamination/information bias as the clinicians know beforehand (when) they will have to change their approach. How will the authors correct for intercurrent activities or studies promoting SDM in the mean time? What will happen if patients do not consent to the study? Will and can clinicians switch to their previous habit of consulting? The authors describe many outcomes to be measured. How will they correct for multiple testing and the risk of type-1 errors? The rating by means of the OPTION-5 instrument is highly dependent on the experience of, and the agreement among, the raters. How will this be assured? The sample size calculation and statistics section are murky, as it uses terms clinicians are unfamiliar with (e.g., "linear exponent

	autoregressive correlation structure") and does not show what the clinical relevance will be of an increase in SDM-Q-9 as primary outcome. For example, if SDM-Q-9 score indeed shows a 'moderate effect size' and would increase, let's say, 5 out of 45 points, what does it mean for the enhancement of SDM/VBHC in clinical practice?

VERSION 1 – AUTHOR RESPONSE

Reviewer 1

Reviewer comment 1.1: Many thanks for the opportunity to review this study protocol. The authors intend to foster shared decision-making by implementing patients' outcome data in discussions about treatment options. Therefore, they developed personalized decision aids and a training for healthcare professionals. Especially the decision aids are well designed and are an interesting new tool to provide relevant information to patients. This is a very interesting and well-designed study with a clear objective and comprehensible described methodology. The authors plan to assess a range of measures, which will help to understand effects of the implementation strategies. All in all, this study has the potential to contribute to existing literature and I am looking forward to read about your results.

Authors' response to reviewer comment 1.1: Thank you for your positive feedback.

Introduction

Reviewer comment 1.2: For readers, who are not familiar with the concept of SDM, please insert a short definition after the first paragraph.

Authors' response to reviewer comment 1.2: We thank the reviewer for this suggestion. We have added the following short definition of SDM after the first paragraph of the introduction on page 0 on lines 13 and 14:

SDM is the process in which patients and health care professionals make well-informed, collaborative choices by combining the best available evidence and the patient's values and preferences. [8, 11]

8. Stiggelbout AM, Van der Weijden T, De Wit MP, Frosch D, Legare F, Montori VM, et al. Shared decision making: really putting patients at the centre of healthcare. BMJ. 2012;344:e256.
11. Elwyn G, Frosch D, Thomson R, Joseph-Williams N, Lloyd A, Kinnersley P, et al. Shared decision making: a model for clinical practice. J Gen Intern Med. 2012;27(10):1361-7.

Methods and analysis

Reviewer comment 1.3: Page 4, line 41, "… trends on outcomes will be established …" – I am not sure if you use the correct phrase here. Is it rather "evaluated"?

Authors' response to reviewer comment 1.3: We agree with the reviewer and changed the wording to 'evaluated' instead of 'established' on page 1 on line 50.

Intervention

Reviewer comment 1.4: Page 6, line 26, "A multicomponent intervention …" – I would not define the implementation strategy as one component of the intervention. I would suggest to explain that you developed a strategy to implement the interventions, PtDAs and trainings.

Authors' response to reviewer comment 1.4:

We agree with the reviewer that the implementation strategy should not be considered as a component of the intervention. We adjusted the manuscript by replacing the description of the

implementation strategy from “intervention” to a new subheading “implementation strategy for multicomponent intervention” on page 5 on lines 155 to 159.

Reviewer comment 1.5: Page 6, line 34, “A literature review and needs assessment studies …” – Here, references are missing.
Authors’ response to reviewer comment 1.5:
Detailed results of the developmental process of the PtDA’s will be published. Currently, only a manuscript on the needs assessment among breast cancer patients is published. We have added this reference on page 3 on line 100. The other manuscripts are under review. To inform future readers we added the following line to the manuscript in the paragraph “Interactive patient decision aids containing personalized outcome data”: Detailed results of the developmental process of the PtDA’s will be published.

Reviewer comment 1.6: Page 6, line 43, “Usability testing …” – How did you perform the usability testing? Observations in a real-world setting, interviews, focus groups?
Authors’ response to reviewer comment 1.6:
Usability testing consisted of going through the PtDA, combined with think-aloud sessions with patients, an online survey (stroke) and/or interviews by telephone (breast cancer, stroke and advanced kidney failure) among health care professionals. We have added this information on pages 3 and 4 on lines 103 to 106. All participants in the usability test did not participate in the needs assessment or co-creation sessions. The development process of all PtDAs, including a detailed description of the needs assessments, the co-creation sessions, the acceptability testing, and usability testing will be published.

Reviewer comment 1.7: I just wondered how you plan to enter personalized data into the PtDAs. Is this task of the physicians or are those data entered into the PtDA by physicians and patients together during their consultation?
Authors’ response to reviewer comment 1.7:
The personalized data is entered into the PtDA by the both health care professionals and patients. Health care professionals provide personalized clinical data (e.g., for patients with stroke: type of stroke, NIHSS score) when introducing the PtDA. Patients enter patient-reported data, by means of PROMs, into the PtDA while using the PtDA (e.g., for patients with advanced kidney disease: physical condition, treatment goals). On the summary sheet, an overview of these patient-reported data can be found, as a base for final decision-making in a consultation with the health care professional. We have added this information on page 4 on lines 108 and 109, 111 and 112, 115 to 117, and 118 to 121.

Reviewer comment 1.8: Page 7, line 16, “INFLUENCE” – Is this an abbreviation? What do you mean here?
Authors’ response to reviewer comment 1.8:
In this part of the manuscript, we refer to the INFLUENCE-nomogram. INFLUENCE is an acronym of **in**dividualized **f**o**ll**ow **u**p for br**e**ast ca**nce**r. The INFLUENCE-nomogram is a validated prediction model with which the five-year risk for locoregional recurrences after breast cancer can be estimated. We have clarified this in the manuscript on page 4 on lines 124 to 128.

Reviewer comment 1.9: Page 7, last paragraph – Is the described content of the trainings also part of the e-learnings or are the e-learnings shorter with less content?
Authors’ response to reviewer comment 1.9:
The e-learning and the group training are two separate training components. Both are aimed at increasing the knowledge, skills and willingness of healthcare professional regarding SDM supported by (personalized) outcome data. The e-learning is focused on providing theoretical background and practical tips and tricks on applying outcome information in the four steps of SDM in clinical consultations (including text, video’s and self-assessment tests). Completion of the

e-learning takes approximately one hour. The group training is aimed at practicing with applying outcome information and conversation skills in SDM with training actors. By offering the e-learning before the group training sessions, we reduce the time spent on theoretical background in the training, leaving more time to practice on SDM conversational skills. We have expanded the information on the training components in the manuscript on page 5 on lines 141 to 149.

Reviewer comment 1.10: Page 7, line 49, "… reflection on audio-taped consultations …" – Are these consultations of physicians taking part in the trainings?
Authors' response to reviewer comment 1.10:
These are indeed consultations performed by physicians taking part in the training sessions. During the training, physicians will be receiving feedback on audio-taped consultations that they have audio-taped as part of the data collection in the pre-implementation phase of this study. We have added this information to the manuscript on page 5 on lines 145 and 146.

Study design and procedure
Reviewer comment 1.11: Page 8, last paragraph – Why are hospitals not randomized? How do you decide for the order, hospitals start with the intervention?
Authors' response to reviewer comment 1.11: The moment by which hospitals switch from standard care to use of the intervention are not randomized because of our focus on sustainable implementation of SDM in clinical practice. Effectiveness studies have demonstrated that PtDAs are beneficial, yet structural implementation falls short of expectations. To promote that PtDAs become successfully implemented ito routine clinical settings, we focused on removing barriers to implementation on e.g. the organizational level. One of these barriers considers "unsettled organization": degree to which there are other changes in progress (organizational or otherwise) that represent obstacles to the process of implementing the innovation, such as reorganizations, mergers, cuts, staffing changes or the simultaneous implementation of different innovations (Fleuren et al., 2014). To overcome this barrier, we asked HCP's in the hospitals when it was most convenient for them to proceed with implementation, within the framework of the design of the study. We have added this information to the manuscript on page 6 in table 2 (bullet 3) and on lines 169 to 171.

Data collection and methods
Reviewer comment 1.12: Why are there different dates for follow up for patients with stroke compared to patients with breast cancer and AKD?
Authors' response to reviewer comment 1.12:
The differences in timing of the follow-up questions are due to the course and nature of the three conditions in combination with the phase of the treatment in which the intervention is implemented per condition. For example, for stroke the decision for rehabilitation takes place in an acute setting, whereas the decision for breast cancer follow-up takes place at a less acute moment in the care pathway. These differences in care pathways led to differences in the time points at which the three questionnaires are sent for the three conditions (see also reviewer comment 1.17). We have added the argumentation for the differences in timing of follow-up questionnaires to the manuscript on page 7 on lines 189 to 192.

Outcomes
Effectiveness
Reviewer comment 1.13: Page 10, line 7 to 33 – This is the introduction for this section but it is quite redundant to the following paragraphs. I would suggest to shorten this introduction.
Authors' response to reviewer comment 1.13:
Thank you for your relevant comment. We have decided to remove the introduction paragraph (on page 8), as well as the introduction to the section on implementation (on page 12).

Primary outcome measure

Reviewer comment 1.14: Page 10, line 41 – Here is a mistake with the references. I would also suggest to include a reference for the original German SDM-Q-9 scale by Scholl et al.
Authors' response to reviewer comment 1.14: We agree with your comment and have included a reference to the original German SDM-Q-9 scale on page 8 on line 211, added the correct reference and adjusted table 3.

Secondary outcome measures
Reviewer comment 1.15: I would suggest to structure this paragraph with numbers: "Secondary outcomes will be (1) patient reported SDM, […], (2) decisional conflict, measured …". This would increase readability.
Authors' response to reviewer comment 1.15: Thank you for your feedback. We have restructured this paragraph accordingly on pages 8 and 9 on lines 217 to 233.

Reviewer comment 1.16: Page 10, line 50 – What is meant by knowledge? Patients' knowledge about SDM or their own disease?
Authors' response to reviewer comment 1.16: Here knowledge refers to patients' knowledge regarding their disease and treatment options. This will be evaluated with a self-constructed survey designed to explore patients' knowledge on topics discussed during standard educational consultations in the participating hospitals. We have clarified this in the manuscript on page 8 on lines 220 and 221 and in table 3.

Reviewer comment 1.17: You do not plan to assess each measure for each disease or during each assessment phase. For example, the SF-12 will not be assessed for patients with stroke and quality of life measures for stroke will only be assessed for T1. Can you shortly explain why you choose specific measures for specific time points and diseases?
Authors' response to reviewer comment 1.17:
Thank you for your question. Some measures, such as the Cancer Worry Scale (CWS), the Revised Illness Perception Questionnaire for Breast Cancer Survivors (IPQ-BCS), the modified Ranking Scale (mRS), or the Utrecht Scale for Evaluation of Rehabilitation-Participation (USER-P) are disease specific and will therefore only be assessed in the appropriate patient groups. Furthermore, not all measures are suitable for each time point due to differences in the nature and the course of the condition (see also reviewer comment 1.12). For example, we felt that evaluating the quality of life of patients with stroke during the acute phase of the condition would be inappropriate because measurements may be biased by overwhelming experiences of the acute phase and hospital admission. When considering the measures in combination with the time points, we accounted to take the care pathway for the three conditions. This led to the differences in the time points in which the T0, T1 and T2 questionnaires will be send to patients for the three conditions. We have added the argumentation for the differences in timing of follow-up questionnaires and the differences with regards to disease specific outcome measures to the manuscript on page 7 on lines 189 to 192.

Reviewer comment 1.18: According to your overview on measures, patients have to fill out several measures with many items. This seems to result in quite long surveys. Do you offer some kind of incentives for study participation?
Authors' response to reviewer comment 1.18:
Thank you for your comment. Participants do not receive incentives for study participation. However, it is important to note that participants only fill out the questionnaires for one of the three conditions and that the time that it takes to fill out the questionnaires differs per measurement moment. For example, the first questionnaire takes about 30 to 45 minutes to complete and the second and third questionnaires take 15 to 20 minutes to complete. We have added the information about the time it takes to complete the questionnaires to the manuscript on page 7 on lines 187 to 189.

Reviewer 2

Reviewer comment 2.1: Many thanks for the opportunity to review this manuscript. It is a laudable attempt to improve shared decision-making (SDM) among three important and large groups of patients.
Authors' response to reviewer comment 2.1: Thank you for your positive feedback.

Although I realise it is late to change the protocol at this stage, I do have the following remarks:

Reviewer comment 2.2: Implementation of SDM is not only a matter of providing outcome information (patients may not make rational choices based on data, but on other outcomes clinicians tend to overlook), nor is it achieved by merely introducing decision support tools. The main implementation issues are awareness, willingness and behaviour among clinicians and patients themselves. This study focuses on the former two interventions. These are doomed to fail if there is no focus on the latter aspects: are patients and clinicians willing and capable to start SDM and to use these tools?
Authors' response to reviewer comment 2.2:
We agree with the reviewer that implementation of SDM is not only a matter of providing outcome information, nor that it is achieved by merely introducing decision support tools. Therefore besides providing outcome information (including not only clinical outcomes, but also PROMs) and by introducing PtDA's (including value clarification), we designed an implementation strategy focusing on awareness, willingness and behaviour (see Table 2). In addition, the SHOUT study is part of the Santeon Experiment Outcome Indicators. At the start of the Experiment the multidisciplinary improvement teams of all conditions active in the Santeon VBHC program had the opportunity to pitch for participation in the Experiment focusing on SDM, meaning that focusing on willingness even started before the SHOUT study. Amongst others, willingness for SDM was also considered when deciding on the key moments for SDM supported by outcome information with clinicians, patient associations and patients. Clinicians capability was cared for by training clinicians in SDM, including skills such as the teach back method. For patients, informative video's on SDM with outcome information were created. PtDA's were tested on patients being capable of usinge intervention. We added information on the implementation strategy in the manuscript on page 3 on lines 93 to 95, on page 5 on lines 157 and 158 and on page 6 in table 2 (bullet 1).

Reviewer comment 2.3: I do not agree with the statement in the 'strengths and limitations of this study' that "randomization is not feasible". Implementation studies can surely be conducted in a randomised fashion, for example by means of a stepped-wedge cluster-RCT. For example: Scholl I, et al. Implement Sci 2021; Scalia P, et al. Implement Sci 2019, and also in the Netherlands: Thunnissen FM, et al. BMC Med Inform Decis Mak. 2021; de Mik SML, et al. BMC Med Inform Decis Mak. 2020; Al-Itejawi HHM, et al. BMJ Open. 2017.
Authors' response to reviewer comment 2.3: We agree with the reviewer that randomization is indeed feasible in implementation studies. Therefore we removed the statement accordingly.

Reviewer comment 2.4: The same study strengths and limitations section focuses on the study design only. There are no remarks on expected outcome. Another limitation is the small effect size the authors want to achieve, while the implementation effort is large, involving over 1700 patients. Moreover, the purported effect is based on a subjectively scored outcome measure that is known to suffer from a ceiling effect. So, what will be the clinical relevance, even if such a small effect is found? Hence, why was the objective measure of patient involvement not chosen as primary outcome?
Authors' response to reviewer comment 2.4: We have added a remark on the expected outcome as well as on the small effect size in the strengths and limitation section: "It is unclear whether the effect size aimed to achieve, constitutes a clinically meaningful difference."
Our primary outcome measure, the SDM-Q-9, aims to measure the process of SDM in the medical consultation from the patients' perspective. As recommended by Ubbink et al. (2022) patient-reported SDM measures, should be combined with objective scores of SDM, as these may differ from the patients' subjective interpretation. Therefore we included the OPTION in our study as well (see

page 90, lines 235 to 238). We weighed the benefits and limitations of both instruments and decided to choose the SDM-Q-9 as our primary outcome, as the focus of our study is to assess patients' view on how they are helped by SDM supported by outcome data. In addition, the SDM-Q-9 is a commonly applied tool to assess patient-reported levels of SDM and has been validated in the Netherlands.

Reviewer comment 2.5: It seems likely that the effects of the implementation differ due to the differences in interventions (treatment and post-treatment options) and disorders involved. The authors may want to consider studying and reporting on the effects of implementation for each of the three disorders separately. As the interventions focus on different aspects of the care process for each disorder, the effect size may vary as well. This would also have implications for the sample size calculation if conducted per disorder.
Authors' response to reviewer comment 2.5: As the effect of the implementation may differ due to the differences in interventions and conditions involved, we will report the effects of implementation for each of the three disorders separately. Therefore, we conducted sample size calculations per condition (see the sample size calculations on page 13, line numbers 263 to 286). In addition, a strength of the current study is that we will be able to assess an overall effect by means of a meta-analysis across all patient groups. Finally, we will be able to investigate implementation across all patient groups by using several the same outcome measures at a similar points in time. We have added this information on page 15 on lines 317 and 318.

Reviewer comment 2.6: The authors describe they will start the study as of November 2019. Given the time frame of 20 months pre- and post-implementation as shown in figure 2, the study has already been completed by now. As far as I understand, this is an exclusion criterion for publication in BMJ Open ("If data collection is complete, we will not consider the manuscript").
Authors' response to reviewer comment 2.6: We do meet the BMJ Open criterion for protocol manuscripts to report on planned or ongoing research studies. Data collection is still ongoing (see page 6, line 168): i.e. T2 follow-up questionnaires will be send out at 12 months following the baseline questionnaire.

Reviewer comment 2.7: The authors state that "The moment by which hospitals switch from standard care to use of the intervention will not be randomized." This may lead to contamination/information bias as the clinicians know beforehand (when) they will have to change their approach. How will the authors correct for intercurrent activities or studies promoting SDM in the mean time?
Authors' response to reviewer comment 2.7: Clinicians indeed know beforehand when they will have to change their approach. However, only at that moment clinicians will be trained and the PtDA will be introduced. Internal validity will be increased, as each hospital will act as its own historical control group and the hospitals will not switch at the same time (see page 7, lines 172 and 173). For example, national publicity on SDM would affect all hospitals. Thus, if one hospital changes when the intervention is introduced while those remaining in the pr-implementation phase do not, one can be confident that the change is not due to the publicity. Moreover, segmented regression will enable us to control for other variables, that can cause a change in level or trend of the outcomes of interest (see page 14, lines 308 and 309).
In addition, we would like to refer to the publication by Biglan at all (2000) *The Value of Interrupted Time-Series Experiments for Community Intervention Research* to explain how the interrupted Time-Series design controls for threats to internal validity, such as intercurrent activities or studies promoting SDM in the meantime (see page 1, line 47 and reference 13).

Reviewer comment 2.8: What will happen if patients do not consent to the study? Will and can clinicians switch to their previous habit of consulting?
Authors' response to reviewer comment 2.8:

The aim of the Experiment Outcome indicators of which the SHOUT study is part, is to gain experience with the use of outcome information to support SDM in clinical practice. The study design is focused at implementing the multicomponent intervention in all seven Santeon hospitals for three conditions. Therefore, in the post-implementation phase the multicomponent intervention will be offered as standard care. Patients are free to decline participation in the SHOUT-study but will still be offered the PtDA as the standard form of care. We have added this information to the manuscript on page 7 on lines 179 to 181.

Reviewer comment 2.9: The authors describe many outcomes to be measured. How will they correct for multiple testing and the risk of type-1 errors?
Authors' response to reviewer comment 2.9: We will apply a Bonferroni-Holm procedure across the set of primary and secondary endpoints as they were prespecified in the protocol before start of the trial. We have added this information to the manuscript on page 15 on lines 313 and 314.

Reviewer comment 2.10: The rating by means of the OPTION-5 instrument is highly dependent on the experience of, and the agreement among, the raters. How will this be assured?
Authors' response to reviewer comment 2.10:
The raters have been educated on the OPTION-5 and jointly participated in multiple practice ratings of video and audio consultations under the guidance of an experienced advisor and trainer in shared decision-making. One of the raters has ample experience in rating consultations with the OPTION-5. Agreement among the raters will be assured because all consultations will be double-coded by 2 raters. In case of disagreement, a third rater will be consulted. We have added this information on page 9 on lines 238 and 239.

Reviewer comment 2.11: The sample size calculation and statistics section are murky, as it uses terms clinicians are unfamiliar with (e.g., "linear exponent autoregressive correlation structure") and does not show what the clinical relevance will be of an increase in SDM-Q-9 as primary outcome. For example, if SDM-Q-9 score indeed shows a 'moderate effect size' and would increase, let's say, 5 out of 45 points, what does it mean for the enhancement of SDM/VBHC in clinical practice?
Authors' response to reviewer comment 2.11: We agree with the reviewer that the terminology used is new and unfamiliar. We have provided an explanation and motivation in words on page 13 on lines 273 to 280. Unfortunately there is no agreement on what constitutes a clinically meaningful difference on the SDM-Q-9. Hence, we estimated the size of the expected effect on previous studies using the SDM-Q-9. We have added this information to the manuscript on page 13 on lines 264 to 266.

VERSION 2 – REVIEW

REVIEWER	Lindig, Anja University Medical Center Hamburg-Eppendorf, Department of Medical Psychology
REVIEW RETURNED	08-Apr-2022

GENERAL COMMENTS	Dear authors, I had the pleasure to again read your revised version of the manuscript. Thanks a lot for the detailed and comprehensive answers to my comments. This is a very interesting and well-written manuscript using a clear and precise language and containing the relevant information. I had the impression that the manuscript improved a lot and you answered all comments very satisfying. Since I have nothing else to comment or add, I would like to recommend to accept your manuscript for submission. Best regards and good luck with your future work.

VERSION 2 – AUTHOR RESPONSE

Reviewer 1

Reviewer comment 1.1: I had the pleasure to again read your revised version of the manuscript. Thanks a lot for the detailed and comprehensive answers to my comments. This is a very interesting and well-written manuscript using a clear and precise language and containing the relevant information. I had the impression that the manuscript improved a lot and you answered all comments very satisfying. Since I have nothing else to comment or add, I would like to recommend to accept your manuscript for submission. Best regards and good luck with your future work.

Authors' response to reviewer comment 1.1: Thank you for your positive feedback.