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.

This paper was submitted to a another journal from BMJ but declined for publication following peer review. The authors addressed the reviewers' comments and submitted the revised paper to BMJ Open. The paper was subsequently accepted for publication at BMJ Open.

(This paper received three reviews from its previous journal but only two reviewers agreed to published their review.)

ARTICLE DETAILS

TITLE (PROVISIONAL)	Online decision aids for primary cardiovascular disease prevention: Systematic search, evaluation of quality, and suitability for low health literacy patients
AUTHORS	Bonner, Carissa; Patel, Pinika; Fajardo, Michael; Zhuang, Ruixuan; Trevena, Lyndal

VERSION 1 – REVIEW

REVIEWER	Thomas H. Wieringa Amsterdam UMC (the Netherlands)
REVIEW RETURNED	26-Jul-2018

GENERAL COMMENTS	This is an interesting study and brings to light the shortcommings of online DAs for CVD prevention. In general, I think you can elaborate more in depth about the possible reasons for and meanings of your findings, because I think this is reduced by a minimum while there are multiple remarkable and important findings to discuss and give direction to future research and DA development.
	Furthermore, I would suggest to consider the following in order to improve your manuscript:
	Insert the abreviation in parentheses for IPDAS behind the last sentence in the abstract's objectives. Write down the full name of the PEMAT-P scale ("Patient Education Materials Assessment Tool for Printed Materials") in the first sentence of the abstract's primary outcome measrues section, with the abreviation ("PEMAT-P") behind in parentheses. The second sentence of the results section of the abstract is rather long. I would suggest to insert a dot behind "(mean 61%)" instead of a comma and start a new sentence from "Readability was also…" on. The terms "qualifying criteria" and "certification criteria" are first used in the results section of the abstract. I assume these criteria refer to the IPDAS version 3 and 4. I would suggest to clearly state this, and elaborate a little more on the IPDAS criteria (at least these two) in the abstract as the standards for certification is referred to in the abstract's conclusion as well. Although understandable for practical reasons, I would consider the language restriction (only English DAs) as a limitation when

the aim is to "identify and evaluate all publicly available CVD prevention DAs"

I like the example of the 60 year old female smoker used on page 5, because it clarifies the importance of shared decision making for an individual person.

In line 5 of page 6, the authors give a definition of decision aids. I would suggest to insert a reference on the end of this sentence (It looks like the IPDAS definition)

I think there are some typos in the search methods section ("to collect specific information a given topic", "pertinent to IT internal workings"), please take a closer look.

In general: be consequent in the use of numbers written out and/or numeric symbols

I do not understand the explanation from line 52 on page 8 to line 5 on page 9 ("The first 50 results were considered (not including web advertisements), providing a pool of 1100 results to be title scanned for each searcher (2200 results in total). Scanning the first 100 search results for the first few searches found no additional resources after the first 50 results, so the cut-off of 50 was retained"). What is actually meant with "considered"? Considered for what and how does this result in a pool of 1100 results? Please elaborate a little more on this.

Be consequent in the use of capitals for "decision aid" in the entire document

In line 39 "IPDAS-SF" should be "IPDASi-SF" (I think)

The flow chart is not exactly consistent to the PRISMA flow chart. In some way, I understand why you adapted it a little, but it is for example unclear why there are two times duplicates removed. Please take another look at this flow chart and make it as consistent as possible tot he PRISMA flow chart.

25 DAs are finally included according tot he flow chart, but 26 are included in Table 3 (while the N = 25 according the title of table 3). Can you please explain this?

Please explain the meaning of IPDAS criteria (certifiying, qualifying, and quality) in the method section. Assuming that the criteria of the 3rd and 4th version of IPDAS are the same, but differently named, I think the meaning of the criteria is partly done in the results section of the current version of the article, but seems to be more appropriate fort he methods section. If the third and fourth version of IPDAS are using different criteria, then please indicate this clearly in the methods section, as well as the meaning of every criterion.

I won't regard the first sentence of the strengths and limitations section (using different searchers and raters) as a strength as this is usual in systematic reviews.

GENERAL COMMENTS	Dear authors, This is a well written paper on an important topic. Very good, sound methods. However, I think, there is room for improvements.

	My comments are:
	Major
	1. Title: In my opinion, the term ‚systematic review' should be restricted to the systematic search for primary studies and a qualitative or quantitative synthesis of the results to answer a research question. The best term may be ‚systematic evaluation'.
	2. Many correlations were reported in the methods and results, but without further specifications. Kappa? r? Pearson? …The methods to calculate the correlations should be reported in the methods.
	3. Table 3: same number of decimal numbers: 1 or 2; preferably 1, IDs should be the same like in the appendix table: 24, 26.
	4. It would be nice, if the reader got more information on the IPDAS items/criteria. (e.g.Tables including results on the single items?) Without special knowledge, the reader might have only a vague impression of what these tools measure.
	Minor
	1. The title could also more precisely describe the research including ‚primary' CVD prevention and the appropriateness for low literacy populations. (For example: (Systematic evaluation of the) quality and suitability for low literacy populations of online DA for primary CVD prevention)
	2. I think, the PRISMA checklist for meta-analysis of RCTs is not suitable for the reporting quality. I would recommend STROBE for cross-sectional studies.
	3. Think about more information of the characteristics of the DAs: how long…
	4. In the first paragraph of the discussion, you mentioned the international criteria for certification, you should provide more information about these criteria.
	5. Discussion: some points could be added: correctness of information in the DAs as a measure of quality? Research needed, whether these DA actually support an informed choice? Do low literacy populations actually use/access online DAs?

REVIEWER	Kristen Tecson
	Baylor Heart and Vascular Institute
REVIEW RETURNED	20-Oct-2018

GENERAL COMMENTS	Manuscript ID: bmjopen-2018-025173
	Title: Online decision aids for cardiovascular disease prevention: Systematic review and evaluation of quality
	Summary: The authors conducted a large online search using a combination of google and decision aid (DA) repositories to identify existing DAs for cardiovascular disease prevention. Through an adjudication process, the team identified 25 unique DAs meeting study criteria. The authors conclude that while most of the DAs are fairly well understood, they are neither actionable, nor suitable for reading levels lower than grade 10.
	Major Points:
	The authors conclude that the DA tools for patients are not suitable for those with low literacy; however, some of the tools used in their sample were built for clinicians (mentioned in the abstract, line 16). If tools intended for clinicians were included in this review, intended to capture tools for patients, then the results are biased.
	The topic of patient activation fits nicely with the discussion of health literacy. Consider incorporating.

The correlation between the reviewers was very low. It is unclear as to what conflict would be resolved (page 9, line 41) as these scores are continuous, not categorical.

The authors mention 3 domains and then mention 6 qualifying criteria – it is unclear.

The authors need to explain/introduce icon arrays.

Is the negative correlation (page 12, line 15) a mistake?

The style in which the manuscript is written does not seem to be suited for a medical journal. Having an introduction vastly longer than the discussion (which is only ½ page) is more in line with journals from other disciplines. It is also odd that some results are given in the methods section.

Minor Points:
Abstract- please spell out DA before using the shorthand.
Page 5, line 26: please include units for cholesterol
Page 6, line 4: plurality/singularity disagreement (aids…tool)
Page 6, line 19: please use a semicolon instead of a comma before 'however'
Page 6, line 32: please change 'and' to 'or'
Page 7, lines 25-2: Please remove the sentence regarding the business origins of an environmental scan. Please add the word 'on' between 'information' and 'a'
Page 7, line 30: The second instance of 'it' is a mistake.
Page 7, line 34: Please rephrase without the word 'exploded'
Page 8, line 47: Did you search on 'cholesterol medication and statin' or 'cholesterol medication' and 'statin'?
Page 8, line 48: Please remove the sentence regarding pilot testing.
Page 8 – page 9: Please remove the sentence spanning these pages.
Page 10, line 17: plurality/singularity disagreement (average, were)
Page 10, line 48: 'measures' should be 'measured'
Page 10, line 51: Please remove the text after 'criteria' and before the second 'must'
Page 11, line 2: It is unclear what the numbers are referring to (there is nothing tied to 'respectively')
Figure 1- one of the horizontal lines is overlapping a vertical line
Table 3 – the IDs skip DA_24 and end on A_26

REVIEWER	Ken Redekop Erasmus School of Health Policy and Management, The Netherlands
REVIEW RETURNED	02-Nov-2018

GENERAL COMMENTS	I was asked to examine the statistical methods used in this paper. Therefore, my main comments relate to that part of the paper, although other comments were included to help the readers. MAJOR COMMENTS Methods The authors need to clarify what type of correlation analysis they performed. Moreover, they need to justify their choice. For

	example, if a Pearson correlation analysis was done, this would not detect a rater bias (where one rater on average gives higher scores than the other rater) as long as the two sets of ratings were correlated. The authors probably need to use another way (e.g., an intra-class correlation coefficient) to express the degree of inter-rater agreement.
	Discussion: Regarding study limitations, if another searcher or method would have found additional decision aids, what implications does that have for this study's findings?
	Abstract: more text about the methods, results and discussion is needed.
	MINOR COMMENTS
	Methods: The writeup in some places needs to be improved. One clear example is the subsection on the IPDAS Checklist. The writing style compromises the paper's readability.
	Tables 2 and 3: I'd suggest using the same sequence of parameters in the two tables.
	Table 2: Where are the results for the overall IPDAS score?

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Thomas H. Wieringa
Institution and Country: Amsterdam UMC (the Netherlands)
Please state any competing interests or state 'None declared': None declared
Please leave your comments for the authors below

4. This is an interesting study and brings to light the shortcommings of online DAs for CVD prevention. In general, I think you can elaborate more in depth about the possible reasons for and meanings of your findings, because I think this is reduced by a minimum while there are multiple remarkable and important findings to discuss and give direction to future research and DA development.

Response: We have now provided additional explanation of the implications of our findings throughout the discussion.

5. Insert the abreviation in parentheses for IPDAS behind the last sentence in the abstract's objectives.
Write down the full name of the PEMAT-P scale ("Patient Education Materials Assessment Tool for Printed Materials") in the first sentence of the abstract's primary outcome measrues section, with the abreviation ("PEMAT-P") behind in parentheses.

Response: We have amended based on reviewer's comment.

6. The second sentence of the results section of the abstract is rather long. I would suggest to insert a dot behind "(mean 61%)" instead of a comma and start a new sentence from "Readability was also…" on.

Response. We have amended based on reviewer's comments.

7. The terms "qualifying criteria" and "certification criteria" are first used in the results section of the abstract. I assume these criteria refer to the IPDAS version 3 and 4. I would suggest to clearly state this, and elaborate a little more on the IPDAS criteria (at least these two) in the abstract as the standards for certification is referred to in the abstract's conclusion as well.

Response: We have made additions to the abstract to elaborate on these criteria: "Four decision aids met criteria to be considered a decision aid (i.e. met IPDAS qualifying criteria) and one sufficiently minimised major bias (i.e. met IPDAS certification criteria)."

8. Although understandable for practical reasons, I would consider the language restriction (only English DAs) as a limitation when the aim is to "identify and evaluate all publicly available CVD prevention DAs"

Response: We have amended our aim to address this reviewer's comment: "This study aimed to identify and evaluate all English-language, publicly available online CVD prevention decision aids based on…", and also listed this as a limitation: "Additional decision aids could have been found by a different searcher, search engine or geographical location, and in other languages, which could produce different findings about the overall suitability for low health literate patients."

9. I like the example of the 60 year old female smoker used on page 5, because it clarifies the importance of shared decision making for an individual person.

Response: Thank you for identifying this.

10. In line 5 of page 6, the authors give a definition of decision aids. I would suggest to insert a reference on the end of this sentence (It looks like the IPDAS definition)

Response: We have added the IPDAS references here.

11. I think there are some typos in the search methods section ("to collect specific information a given topic", "pertinent to IT internal workings"), please take a closer look.

Response: We have amended these sentences.

12. In general: be consequent in the use of numbers written out and/or numeric symbols

Response: We have reviewed the use of numbers.

13. I do not understand the explanation from line 52 on page 8 to line 5 on page 9 ("The first 50 results were considered (not including web advertisements), providing a pool of 1100 results to be title scanned for each searcher (2200 results in total). Scanning the first 100 search results for the first few searches found no additional resources after the first 50 results, so the cut-off of 50 was retained"). What is actually meant with "considered"? Considered for what and how does this result in a pool of 1100 results? Please elaborate a little more on this.

Response: We have amended the sentence to aid clarity: "The first 50 results for each unique Google search were exported (not including web advertisements), providing a pool of 1100 results to be title scanned for each searcher (2200 results in total). Scanning the first 100 search results for the first few searches found no additional resources after the first 50 results, so the cut-off of 50 was retained."

14. Be consequent in the use of capitals for "decision aid" in the entire document

Response: Decision Aid is capitalised only when referring to the search term we used (i.e. entering "Decision + Aid" into Google), to delineate from the actual decision aid in lower case.

15. In line 39 "IPDAS-SF" should be "IPDASi-SF" (I think)

Response: We have amended.

16. The flow chart is not exactly consistent to the PRISMA flow chart. In some way, I understand why you adapted it a little, but it is for example unclear why there are two times duplicates removed. Please take another look at this flow chart and make it as consistent as possible to the PRISMA flow chart.

Response: We have amended the chart to be clearer that the first duplicate removal was the removal of duplicate webpages and the second removal of duplicates were the removal of duplicate decision aids that were available on the webpage

17. 25 DAs are finally included according to the flow chart, but 26 are included in Table 3 (while the N = 25 according the title of table 3). Can you please explain this?

Response: Table 3 has 25 Decision Aids. We have amended the ID names to reduce confusion.

18. Please explain the meaning of IPDAS criteria (certifiying, qualifying, and quality) in the method section. Assuming that the criteria of the 3rd and 4th version of IPDAS are the same, but differently named, I think the meaning of the criteria is partly done in the results section of the current version of the article, but seems to be more appropriate fort he methods section. If the third and fourth version of IPDAS are using different criteria, then please indicate this clearly in the methods section, as well as the meaning of every criterion.

Response: We have added to the methods: "IPDASi v3 has three domains: Criteria used to be defined as a patient decision aid (7-items), Criteria to lower risk of making a biased decision (9-items), Other criteria indicating quality (13 items). Criteria used to be defined as a patient decision aid items were rated on a Yes/No scale and the other two domains were rated on a 4-point Likert scale (1=Strongly Disagree to 4=Strongly Agree)."

We have also added:
"Qualifying criteria, if all met, identify the material as a decision aid. Certification criteria are those deemed essential to avoid harmful bias and all six criteria need to be met (i.e. scored 3 or more) for the decision aid to be considered certified. Quality criteria on the other hand were items considered desirable but not essential to avoid harmful bias."

19. I won't regard the first sentence of the strengths and limitations section (using different searchers and raters) as a strength as this is usual in systematic reviews.

Response: Since this is not a standard systematic review of academic literature, we consider adhering to standard protocols as a strength when using an environmental scan methodology that does not require this. We have amended the method in the title to better reflect this and address Review 2's point below.

Reviewer: 2
Reviewer Name: Maren Dreier
Institution and Country: Hannover Medical School, Institute of Epidemiology, Social Medicine and Health Systems Research, Germany
Please state any competing interests or state 'None declared': None declared

This is a well written paper on an important topic. Very good, sound methods. However, I think, there is room for improvements.

Thank you for your positive feedback and helpful suggestions.

Major
1. Title: In my opinion, the term ‚systematic review' should be restricted to the systematic search for primary studies and a qualitative or quantitative synthesis of the results to answer a research question. The best term may be ‚systematic evaluation'.

Response: We have revised the title to be clearer about the methods: "Systematic search and evaluation of quality"

2. Many correlations were reported in the methods and results, but without further specifications. Kappa? r? Pearson? …The methods to calculate the correlations should be reported in the methods.

Response: We have added that these are intraclass correlations and that they were calculated using SPSS v25.

3. Table 3: same number of decimal numbers: 1 or 2; preferably 1, IDs should be the same like in the appendix table: 24, 26.

Response: This has been amended.

4. It would be nice, if the reader got more information on the IPDAS items/criteria. (e.g.Tables including results on the single items?) Without special knowledge, the reader might have only a vague impression of what these tools measure.

Response: We have added more explanation of the certifying and qualifying criteria in the methods. See response to reviewer 1 point 18 above. We have also added tables including results on the single items in the appendix.

Minor
5. The title could also more precisely describe the research including ‚primary' CVD prevention and the appropriateness for low literacy populations. (For example: (Systematic evaluation of the) quality and suitability for low literacy populations of online DA for primary CVD prevention)

Response: We agree and have amended the title as follows "Online decision aids for primary cardiovascular disease prevention: Systematic search, evaluation of quality, and suitability for low health literacy patients"

6. I think, the PRISMA checklist for meta-analysis of RCTs is not suitable for the reporting quality. I would recommend STROBE for cross-sectional studies.

Response: We have used PRISMA because it is the only checklist that covers a systematic search strategy; STROBE would have to be amended for this study because it doesn't cover the systematic search which we regard as an important strength of the method.

7. Think about more information of the characteristics of the DAs: how long…

Response: Due to the dynamic nature of websites, it is difficult to describe characteristics like length or number of pages in an equivalent way across all materials – some were simple PDFs but others were highly interactive with content depending on patient characteristics. We therefore chose standardised scales (PEMAT and IPDAS) as the primary way to describe the decision aids.

8. In the first paragraph of the discussion, you mentioned the international criteria for certification, you should provide more information about these criteria.

Response: We have added into "IPDASi v3 has three domains: Criteria used to be defined as a patient decision aid (7-items), Criteria to lower risk of making a biased decision (9-items), Other criteria indicating quality (13 items). Criteria used to be defined as a patient decision aid items were rated on a Yes/No scale and the other two domains were rated on a 4-point Likert scale (1=Strongly Disagree to 4=Strongly Agree)."

We have also added:
"Qualifying criteria, if all met, identify the material as a decision aid. Certification criteria are those deemed essential to avoid harmful bias and all six criteria need to be met (i.e. scored 3 or more) for the decision aid to be considered certified. Quality criteria on the other hand were items considered desirable but not essential to avoid harmful bias."

9. Discussion: some points could be added: correctness of information in the DAs as a measure of quality? Research needed, whether these DA actually support an informed choice? Do low literacy populations actually use/access online DAs?

We have added further discussion of these issues in the revised manuscript: "In addition, the items did not cover: 1) health literate design issues (e.g. use of white space, images that are consistent with text, and clear direction for next steps); 2) assessment of the accuracy of the information provided (e.g. whether the risks and benefits presented were based on the latest systematic review, if available); 3) ease of access for the intended audience, particularly for low health literacy populations; or 4) how effective the decision aid was, even when an evaluation had been conducted."

Reviewer: 3
Reviewer Name: Kristen Tecson
Institution and Country: Baylor Heart and Vascular Institute
Please state any competing interests or state 'None declared': None declared

Summary: The authors conducted a large online search using a combination of google and decision aid (DA) repositories to identify existing DAs for cardiovascular disease prevention. Through an adjudication process, the team identified 25 unique DAs meeting study criteria. The authors conclude that while most of the DAs are fairly well understood, they are neither actionable, nor suitable for reading levels lower than grade 10.

Major Points:
1. The authors conclude that the DA tools for patients are not suitable for those with low literacy; however, some of the tools used in their sample were built for clinicians (mentioned in the abstract,

line 16). If tools intended for clinicians were included in this review, intended to capture tools for patients, then the results are biased.

Response: decision aids targeting health professionals/clinicians were explicitly excluded; as indicated by the exclusion criteria in the methods: "4) targeted at health professionals". We have added "or clinicians" to this criterion to make explicit that decision aids were all directed to patients.

2. The topic of patient activation fits nicely with the discussion of health literacy. Consider incorporating.

Response: We agree this is an important related topic but due to space limitations have focused on the concept of health literacy only for this paper; both the PEMAT and IPDAS criteria require considerable explanation of their underlying concepts already, and there are alternative measures for patient activation that were not the focus of this study. This could be a good area for future research however.

3. The correlation between the reviewers was very low. It is unclear as to what conflict would be resolved (page 9, line 41) as these scores are continuous, not categorical.

Response: Additions to how the PEMAT-P is used were added to the manuscript to better highlight what conflict was being resolved. "PEMAT-P includes two subscales: 1) understandability, which is a measure of how well a person is able to process and explain the key message of the material; and 2) actionability, which is a measure of how well a person is able to identify what to do based on the information in presented. Items were rated on a binary scale (Yes/No) with some items provided a "Not Applicable" option. Final understandability and actionability scores were calculated as a percentage of "Yes" ratings for all items not including "not applicable" ratings; higher percentages indicate better understandability or actionability. Intraclass correlations were calculated using SPSS v25. For the two independent searchers, the intraclass correlation for final understandability scores was .51 and for actionability scores was .48. Conflicts for individual PEMAT-P items were therefore resolved by the third rater (MF, after discussion with CB) to finalise the PEMAT-P score for each individual decision aid. A threshold of 70% was used to determine whether the decision aid was understandable or actionable."

4. The authors mention 3 domains and then mention 6 qualifying criteria – it is unclear.

Response: An explanation of the IPDAS v3 has been added to the methods to explicate that a Domain of the IPDAS comprises a set amount of criteria.

We have added into "IPDASi v3 has three domains: Criteria used to be defined as a patient decision aid (7-items), Criteria to lower risk of making a biased decision (9-items), Other criteria indicating quality (13 items). Criteria used to be defined as a patient decision aid items were rated on a Yes/No scale and the other two domains were rated on a 4-point Likert scale (1=Strongly Disagree to 4=Strongly Agree)."

We have also added:
"Qualifying criteria, if all met, identify the material as a decision aid. Certification criteria are those deemed essential to avoid harmful bias and all six criteria need to be met (i.e. scored 3 or more) for the decision aid to be considered certified. Quality criteria on the other hand were items considered desirable but not essential to avoid harmful bias."

5. The authors need to explain/introduce icon arrays.

Response: Icon Arrays are graphic representations to show abstract probabilities as more concrete frequencies (e.g. 2% = 2 coloured dots out of 100 black dots), and are considered best practice for risk communication.

6. Is the negative correlation (page 12, line 15) a mistake?

Response: No. Higher percentages of understandability indicate that patients are better able to process and explain the key message of the material. For Gunning Fog: "Scores range from 0 to 20 which corresponds to the US grade level that the text should be easily understood by, for example a score of 6 would indicated the test should be easily understood by those educated to the 6th grade level in the US schooling system". Therefore as GF increases, the text becomes harder to read which helps explain why we are seeing a negative correlation.

For Flesch Kincaid: "The Flesch Kincaid Reading Ease score was also calculated with higher scores indicating greater ease of comprehension. Scores range from 0 to 100 where a score of about 70-80 is the equivalent to school grade 7." Therefore as FK increases, the text becomes easier to read which helps explain why we are seeing a positive correlation

7. The style in which the manuscript is written does not seem to be suited for a medical journal. Having an introduction vastly longer than the discussion (which is only ½ page) is more in line with journals from other disciplines. It is also odd that some results are given in the methods section.

Response: We have expanded on the methods and discussion sections in the revised version of the manuscript to address the comments of all reviewers. Please see tracked changes and/or responses to reviewers 1 and 2 above.

Minor Points:
8. Abstract- please spell out DA before using the shorthand.

Response. We have removed the acronym.

9. Page 5, line 26: please include units for cholesterol

Response: Added in "mmol/L)

10. Page 6, line 4: plurality/singularity disagreement (aids…tool)

Response: Not sure what this is referring to

11. Page 6, line 19: please use a semicolon instead of a comma before 'however'

Response: We have changed it to a semicolon

12. Page 6, line 32: please change 'and' to 'or'

Response: We have changed this.

13. Page 7, lines 25-2: Please remove the sentence regarding the business origins of an environmental scan. Please add the word 'on' between 'information' and 'a'

Response: Have added in "on".

14. Page 7, line 30: The second instance of 'it' is a mistake.

Response: We have amended "it" to "its"

15. Page 7, line 34: Please rephrase without the word 'exploded'

Response: We believe this is appropriate given the OvidSP (a search platform) uses this term.

16. Page 8, line 47: Did you search on 'cholesterol medication and statin' or 'cholesterol medication' and 'statin'?

Response: We have made this list of 11 clearer by adding in a semicolon after cholesterol medication

17. Page 8, line 48: Please remove the sentence regarding pilot testing.

Response: This is an important step for creating the search strategy that we want to communicate therefore we will leave it in.

18. Page 8 – page 9: Please remove the sentence spanning these pages.

Response: This was an important decision we made that should be communicated.

19. Page 10, line 17: plurality/singularity disagreement (average, were)

Response: We have reviewed this

20. Page 10, line 48: 'measures' should be 'measured'

Response: We have changed this

21. Page 10, line 51: Please remove the text after 'criteria' and before the second 'must'

Response: Removing this text after criteria and before the second must changes the meaning of the sentence therefore this text will remain.

22. Page 11, line 2: It is unclear what the numbers are referring to (there is nothing tied to 'respectively')

Response. Respectively is tied to "certification" and "qualifying" items. They have been italicised for clarity.

23. Figure 1- one of the horizontal lines is overlapping a vertical line

Response: This has been amended.

24. Table 3 – the IDs skip DA_24 and end on A_26

Response. We have amended ID numbers to reduce confusion for the reader

Reviewer: 4
Reviewer Name: Ken Redekop
Institution and Country: Erasmus School of Health Policy and Management, The Netherlands
Please state any competing interests or state 'None declared': None declared
Please leave your comments for the authors below
I was asked to examine the statistical methods used in this paper. Therefore, my main comments relate to that part of the paper, although other comments were included to help the readers.

MAJOR COMMENTS

Methods
1. The authors need to clarify what type of correlation analysis they performed. Moreover, they need to justify their choice. For example, if a Pearson correlation analysis was done, this would not detect a rater bias (where one rater on average gives higher scores than the other rater) as long as the two sets of ratings were correlated. The authors probably need to use another way (e.g., an intra-class correlation coefficient) to express the degree of inter-rater agreement.

Response: Thank you for bringing this to our attention. We have recalculated the intra class correlation coefficient. "Intraclass correlations were calculated using SPSS v25. For the two independent searchers, the intraclass correlation for final understandability scores was .51 and for actionability scores was .48." and "the intraclass correlation between two independent ratings was high (Gunning Fog Index was 0.91 and Flesch Kincaid was 0.94) …".

2. Discussion: Regarding study limitations, if another searcher or method would have found additional decision aids, what implications does that have for this study's findings?

Response: We have added further discussion of this issue: "Additional decision aids could have been found by a different searcher, search engine or geographical location, and in other languages, which could produce different findings about the overall suitability for low health literate patients. However, this paper provides a list of known repositories of decision aids, including the primary source of IPDAS-assessed decision aids, to guide future researchers. This may improve the consistency of current and future findings. It also highlights the need for a central reputable location for decision aids that consumers could be referred to rather than search for their own."

3. Abstract: more text about the methods, results and discussion is needed.

Response: Based on other reviewer comments, we have added to the abstract.

MINOR COMMENTS
4. Methods: The writeup in some places needs to be improved. One clear example is the subsection on the IPDAS Checklist. The writing style compromises the paper's readability.

Response: Based on feedback from other reviewers, this section has been revised.

5. Tables 2 and 3: I'd suggest using the same sequence of parameters in the two tables.

Response: We have reordered Table 3 to better reflect the sequence of parameters presented in Table 2

6. Table 2: Where are the results for the overall IPDAS score?

Response: IPDAS-SF scores are in text. IPDAS v3 and v4 do not have an overall score.

REVIEWER	Thomas H. Wieringa
	the Netherlands (currently not working for any institution)
REVIEW RETURNED	18-Dec-2018

GENERAL COMMENTS	Your manuscript is well-using the English language and describes your project in a clear way. The results are clearly presented and the most important limitation (reproducability) is indicated. I think implications for future research and practice are important for development of the decision aid field.

REVIEWER	Maren Dreier
	Hannover Medical School, Institute of Epidemiology, Social Medicine and Health Systems Research. Germany
REVIEW RETURNED	20-Dec-2018

GENERAL COMMENTS	This paper has improved. I have some minor recommendations: 1. I think, the abstract needs a methods section, and the objectives part is too long. 2. P13 last sentence of the results: "The correlation between understandability and readability was -0.60 for Gunning Fog and 0.59 for Flesch Kincaid." In the method section you introduce ICC as a measure for the interrater agreement. What correlation did you present here? Please add in the methods. 3. Methods: "Agreement for the qualifying criteria items ranged from 64% to 100% and the average correlation between certification and qualifying items were .35 and .61 respectively." This is again obviously not the ICC to measure th interrater agreement, please specify and explain. 4. Methods: please add the "0" before the points in the correlations numbers. 5. Discussion --> Comparison to other Research "Other studies using PEMAT-P for patient education materials have found poor results (CVD decision aids in this study: 87% and 61%; CVD risk calculators: 64% and 19%; online heart failure websites: 56% and 35%; printed lifestyle information for chronic kidney disease: 52% and 37%; for understandability and actionability respectively)." Please include the two scales understandibility and actionability, that refer to the percentages. 6. For the authors of the included DA, it would be nice to know the websites of the IDs of the DA. Maybe you give this information in the supplement?

REVIEWER	Kristen Tecson
	Baylor Heart and Vascular Institute, United States
REVIEW RETURNED	13-Dec-2018

GENERAL COMMENTS	The authors greatly strengthened the manuscript as a result of the peer review process. I have no further comments.

REVIEWER	Ken Redekop
	Erasmus School of Health Policy and Management, The Netherlands
REVIEW RETURNED	28-Dec-2018

GENERAL COMMENTS	The authors seem to have done a decent job of addressing the various comments by the editor and reviewer.
	MINOR COMMENTS
	Abstract: Although the abstract contains the phrase "it is unknown how many CVD decision aids are publicly available for clinicians and patients online", the ultimate focus is on patient decision aids, which is made very clear both in study aim and in the final exclusion criterion ("targeted at health professionals or clinicians"). Why not just remove the reference to clinicians in the abstract to avoid misunderstandings about the scope of the review?
	The Discussion section might have benefitted from a couple of examples from the review to highlight 'good practice' approaches.
	Just a remark that the terms eHealth and mHealth are not mentioned in the Discussion, despite their enormous potential in improving decisions and behaviour, perhaps particularly in less literate populations.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Thomas H. Wieringa

Institution and Country: the Netherlands (currently not working for any institution)

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

Please leave your comments for the authors below Your manuscript is well-using the English language and describes your project in a clear way. The results are clearly presented and the most important limitation (reproducability) is indicated. I think implications for future research and practice are important for development of the decision aid field.

Many thanks for your comment.

Reviewer: 2
Reviewer Name: Maren Dreier

Institution and Country: Hannover Medical School, Institute of Epidemiology, Social Medicine and Health Systems Research. Germany

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

Please leave your comments for the authors below This paper has improved. I have some minor recommendations:

1. I think, the abstract needs a methods section, and the objectives part is too long.

The methods are outlined in the design and primary outcome measures sections, as per the journal requirements. We have moved some information from objectives to design to address your comment.

2. P13 last sentence of the results: "The correlation between understandability and readability was -0.60 for Gunning Fog and 0.59 for Flesch Kincaid."

In the method section you introduce ICC as a measure for the interrater agreement. What correlation did you present here? Please add in the methods.

We have added "Pearson's r" to identify the correlation used here.

3. Methods: "Agreement for the qualifying criteria items ranged from 64% to 100% and the average correlation between certification and qualifying items were .35 and .61 respectively." This is again obviously not the ICC to measure th interrater agreement, please specify and explain.

We have calculated the ICC. The sentence now reads: "Agreement for the qualifying items ranged from 64% to 100% and the average intraclass correlation coefficient between independent ratings for qualifying and certifying items were 0.16 and 0.34 respectively."

4. Methods: please add the "0" before the points in the correlations numbers.

We have added "0" before all correlations.

5. Discussion --> Comparison to other Research "Other studies using PEMAT-P for patient education materials have found poor results (CVD decision aids in this study: 87% and 61%; CVD risk calculators: 64% and 19%; online heart failure websites: 56% and 35%; printed lifestyle information for chronic kidney disease: 52% and 37%; for understandability and actionability respectively)." Please include the two scales understandibility and actionability, that refer to the percentages.

We have amended the final sentence to read: "for PEMAT-P understandability and actionability scores respectively."

6. For the authors of the included DA, it would be nice to know the websites of the IDs of the DA. Maybe you give this information in the supplement?

Yes, this information is in the supplement.

Reviewer: 3
Reviewer Name: Kristen Tecson

Institution and Country: Baylor Heart and Vascular Institute, United States

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

Please leave your comments for the authors below The authors greatly strengthened the manuscript as a result of the peer review process. I have no further comments.

Many thanks for your comment.

Reviewer: 4
Reviewer Name: Ken Redekop

Institution and Country: Erasmus School of Health Policy and Management, The Netherlands

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

Please leave your comments for the authors below The authors seem to have done a decent job of addressing the various comments by the editor and reviewer.

Thank you for your comment.

MINOR COMMENTS

Abstract: Although the abstract contains the phrase "it is unknown how many CVD decision aids are publicly available for clinicians and patients online", the ultimate focus is on patient decision aids, which is made very clear both in study aim and in the final exclusion criterion ("targeted at health professionals or clinicians"). Why not just remove the reference to clinicians in the abstract to avoid misunderstandings about the scope of the review?

We have removed clinicians in the abstract as per suggestion.

The Discussion section might have benefitted from a couple of examples from the review to highlight 'good practice' approaches.

The supplement provides more information for readers to access individual decision aids including those with highest evaluation scores; see also point 6 above for reviewer 3.

Just a remark that the terms eHealth and mHealth are not mentioned in the Discussion, despite their enormous potential in improving decisions and behaviour, perhaps particularly in less literate populations.

We agree these fields are important to the broader issue of improving medical decision making. However, the decision aids identified in this review were generally static (e.g. PDF or text on webpages) and do not include any interactive mobile apps, so we have instead focused on discussing more direct implications for the shared decision making field.