Author’s response to reviews

Title: Effectiveness of a decision aid for promoting colorectal cancer screening in Spain: a randomized trial

Authors:

Lilisbeth Perestelo-Perez (lilisbethp@gmail.com)

Amado Rivero-Santana (amado.riversantana@sescs.es)

Alezandra Torres-Castaño (atorrcas@sescs.es)

Vanesa Ramos-García (vanesa.ramosgarcia@sescs.es)

Yolanda Alvarez-Perez (yolanda.alvarezperez@sescs.es)

Nerea Gonzalez-Hernandez (NEREA.GONZALEZHERNANDEZ@osakidetza.eus)

Andrea Buron (ABuron@parcdesalutmar.cat)

Michael Pignone (Michael_Pignone@med.unc.edu)

Pedro Serrano-Aguilar (pserafu@gobiernodecanarias.org)

Version: 1 Date: 22 May 2018

Author’s response to reviews:

Santa Cruz de Tenerife, May 21, 2018

Dear Editor Dr. Dirk Krüger:

Please find enclosed our revised manuscript entitled “Effectiveness of a decision aid for promoting colorectal cancer screening in Spain: a randomized trial” (MIDM-D-17-00233) by Lilisbeth Perestelo-Perez, Amado Rivero-Santana, Alezandra Torres-Castaño, Vanesa Ramos-García, Yolanda Alvarez-Perez, Nerea Gonzalez-Hernandez, Andrea Buron, Michael Pignone, and Pedro Serrano-Aguilar, for its consideration as an original article in BMC Medical Informatics and Decision Making.

We would like to thank the reviewers and you for your careful revision and thoughtful comments on our paper. All the comments have been carefully taken into consideration for the preparation
of our revised. We appreciate the constructive feedback that without doubt has improved our original.

Yours sincerely,

Lilisbeth Perestelo-Perez, MPsysch, Ph.D.
Servicio de Evaluación y Planificación
Dirección del Servicio Canario de la Salud
Camino Candelaria, 44. C.S. El Chorrillo
lperperr@gobiernodecanarias.org
lilisbethp@gmail.com
Tel. +34 922 684019 (Ext 241)

We have identified and corrected two errata: one in the text (page 9, line 14): the difference between the DA and control groups in knowledge was 17.8%, not 28%. The other in table 2: F and p-value for the interaction term in knowledge are 2.96 and 0.088, respectively, and not -7.85 (0.360)

Adam Dunn (Reviewer 1)
Accept after minor essential revisions

Thank you for the opportunity to review this manuscript. The authors present the evaluation of a web-based decision support tool for colorectal cancer (CRC). The strengths of the research include the clarity of the presentation, the clear manner in which the limitations and constraints of the study design are presented, and the survey of the specific concerns of patients, which was nice to see. I was unable to assess the actual content of the decision tool, and I did not see whether the patients were presented with the rates of false positives, false negatives, and the potential risks associated with population level screening as part of the decision tool. I only have general suggestions and very minor suggestions for corrections in the presentation of the results.
Major comments/suggestions:

1. While it would have been interesting to examine the patients' intentions pre and post intervention to see the proportion that changed their intentions, I understand that it is not possible to go back and do this post hoc.

Certainly, that is a limitation of the study, even more relevant after obtaining such “inflated” results in intention.

2. I note that the limitations are very clear and well described. Concerns I had about the issues with no inclusion of a "sham" comparator or an alternative to the decision support tool were addressed directly.

Thanks to the reviewer for this comment. Relating to this and in response to other reviewer, we have now mentioned in the discussion that, except for the application of the DA, we tried to offer the same experience to intervention and control participants (i.e., appointment in their primary care center, standardized instructions by the researcher, who did not mediate in participants’ responses to questionnaires, which were delivered in the same web interface).

3. The analysis of the differences between feeling uninformed and answering knowledge-based questions directly is interesting - but I wanted to know more - can the authors speculate about the reasons in more detail? Was the amount of information inadequate, or is this related to typical issues in risk communication where people find it hard to judge risks based on how they are presented? It might be worth investigating the risk communication literature to examine this in more detail.

We agree with the reviewer that it is a very interesting issue. Although the mentioned difficulty that people have when dealing with risk probabilities, and the way they are presented, likely influence this result, we think that it has also to do with a more generalized effect observed in psychology research about the low associations between objective and self-perceived abilities. For instance, actual intelligence measured with intelligence tests correlates little with self-perceptions of intelligence (Chamorro-Premuzic et al., 2004), and the same applies with measures of the so-called emotional intelligence objectively assessed (e.g., recognizing emotions) versus self-perception of emotional abilities assessed by self-report (Brackett & Mayer, 2003). We have not included this reflection in the discussion because we think that it is not strictly related to the aim of the study, and there are restrictions in the number of words in the article.
4. Overall, the authors should be clearer about the null results in the study. When reporting the significance, the results clearly indicate that there was no clear difference between the groups (and the limitations suggest why it might have been inflated in the DA group because of "novelty") but there is absolutely nothing wrong with concluding that there was very limited evidence to suggest that the tool improved the intentions, likely because patients intended to undergo the screening regardless of the intervention and because the sample size may have been relatively small.

Thanks for the advice. We now have tried to be more explicit about null results in the intention to be screened, although we have maintained the comments about the p-value of 0.062 in the intention to undergo FOBT for the whole sample.

Minor comments:

1. Results: I think it should be "fewer women" rather than "less women".

We have changed the word, according to the reviewer suggestion.

2. Results: Avoid using "at the limit of significance" and saying "obtained a higher knowledge score". It is better to state that there was no clear evidence that the knowledge score was higher.

We have eliminated that sentence. We thanks Professor Dunn for his comments.


Pedro Pereira Rodrigues (Reviewer 2)

Reject
The manuscript presents the results of a clinical trial to estimate the impact of a decision aid in colorectal cancer screening in Spain.

Although the methodology seems to have been correctly followed according to the registered protocol, there are several issues that, in my opinion, make the results unsound:

a) First and foremost, since the outcome evaluation is done immediately after intervention, this study at most evaluates short-term memory efficacy, not effectiveness of the intervention;

As the reviewer can confirm in the Cochrane Collaboration systematic review on DAs [1], including more than 100 studies in a wide number of medical conditions, the immediate assessment of psychological outcomes (e.g., correct knowledge acquisition, decisional conflict, self-efficacy, readiness to decide, interest/intentions) after a DA application is an usual practice in DA research, even in situations like CRC screening in which the decision has not be made in the short-term [2-5]. The effect of a DA on these basic psychological processes related to decision-making is immediate, and in fact, if we do not find these immediate effects it is not reasonable to expect them later in time. Therefore, there is no reason to not consider the immediate effects of a DA on psychological processes as an effectiveness result.

This fact does not preclude the convenience of investigating the temporal evolution of these decisional processes (as well as, of course, actual decisions), in situations where an immediate decision is not mandatory (we could not carry out a follow up due to funding limitations). Nonetheless, the reviewer has to take into account that nothing prevent the patients from reviewing again the DA when thinking about the decision. On the contrary, the aim is that they had an available resource to be consulted whenever they need it.

Finally, perhaps the reviewer has in mind that in this context of CRC screening the concept of effectiveness is restricted to an increase in screening rates; that is a valuable objective from a public health perspective, but the primary aim of DA is to help patients to make informed decisions, considering their own values and preferences and respecting their autonomy, whatever the final decision they made. Correct knowledge acquisition and reduced decisional conflict are desirable outcomes from the first moments of the decisional process.

b) Given the specific population (islands), the authors should discuss if the population has particular characteristics that could have biased the study results;

Canary Islands are one of the 17 autonomous regions of Spain. It has one of the lowest socioeconomic levels in the country (e.g., gdp per capita, unemployment), but there are not any characteristics that reasonably could bias the results.
c) Selection criteria are somewhat strict (e.g. no family history of CRC, center B with individuals who have refused screening already);

Individuals with family history of CRC present a higher risk of developing the disease, and therefore they could be more prone to undergo screening. Previous trials of decision aids in CRC screening also have applied this exclusion criteria.

DAs for CRC screening are mainly aimed to people who never have undergone the procedure, and therefore this was an inclusion criteria. Center B was actually an additional sample, included to assess the DA effect in people who already had been invited to screening (but not exposed to a DA) and refused it, and to compare it to that observed in participants who never had been invited. We have now tried to make it clearer in the text (last paragraph of the introduction).

d) Evidence coming from this study is limited due to the several potential bias (some well discussed by the authors; other not so much) including hawthorn effect, selection, desirability and measurement bias, and for using proxy outcomes (intention to be screened);

The use of a proxy outcome (which was not the primary outcome of the study) is certainly a limitation, but as we mention in the text, we consider that the used of an intention measure is well justified both theoretically and empirically (previous findings have shown moderate to strong relationships between intentions and actual decisions). Please, see the response to comment h) for further discussion about this measure.

Research on educational or decision support interventions in health care is more subjected to potential biases than clinical studies based on clinical parameters. The difficulty of blinding participants and educators, or potential investigator’s effects certainly introduce threats to internal validity. We tried to minimize these bias as much as possible by keeping the researcher who recruited patients (by phone) blinded to allocation. Except for the application of the DA, we tried to offer the same experience to intervention and control participants (i.e., appointment in their usual primary care center, standardized instructions provided by the researcher, who did not mediated in participants’ responses to questionnaires, which were delivered in the same web interface). We have modified the manuscript to explicitly reflect these aspects (discussion of limitations).

e) Statistical analysis was not clear; why did the authors use ANOVA to compare two groups (A vs B; DA vs control)?

The aim of the study was to assess the effect of the DA and its interaction with the fact of having been invited previously to be screened. Therefore, ANOVA is an appropriate technique to test this. Using several t-tests would have incremented the probability of type I error.
f) Why was intention to be screened dichotomized using cut-off 50% of the predictive model? Would it be better to fine-tune the cut-off value using ROC analysis?

Currently, there are not a “gold standard” statistical technique to assess concordance or congruency between people’s values/goals/concerns about medical procedures and their actual decisions (or intentions, as it was our case). The technique described has been proposed by Sepucha et al. [reference 29 in the article], one of the leading authors in the field of decision support interventions. Other proposals are welcomed, and the suggestion of using ROC analysis is certainly more rigorous. However, regarding our study the issue becomes irrelevant since there were not significant associations between goals/concerns and intentions (probably due to the low variability in both measures), and therefore a measure of concordance was not constructed.

g) Evaluating the outcomes immediately after the intervention is a bias too strong in this field of application; I cannot well value the results using such strategy;

Thanks for your comment, please see the response to comment “a”.

h) Moreover, limitations in evaluating instrument (discussed in the manuscript) raise even higher doubts regarding the evidence exposed by the study; it seems several issues represent potential biases towards and overestimation of the effect, although the results are in line with the previously published meta-analysis (Volk et al., 2016);

The primary outcome of the study was decisional conflict, defined as uncertainty about the course of action to take when facing a health decision (this definition is now included in the text). The scale used is a widely validated instrument and one of the most used measure in research on decision aids. We assume that the reviewer refers to the measure of intention to be screened. Certainly, based on previous findings and the actual rate of screening in our region, we did not expected the observed so high rate of participants stating a positive intention. In the discussion we recognized the likely presence of a (desirability?) bias, and it would has been better to also measure the “intensity” of intention, or to constrain the hypothetical implementation of the decision to a short-term period. However, we want to highlight that even in a context of a ceiling effect for the whole sample, in which is more difficult to obtain significant differences, the DA yielded a significant (uncorrected) or near-significant (corrected) increase in the rate of positive intention to undergo FOBT.
i) The decision aid should be better described and discussed, to allow for a better assessment of the potential benefits and drawbacks.

Thanks for your suggestion. In response to this comment and to other reviewers, we have changed the text to explain better the DA developing and content:

“The DA was a Spanish translation and adaptation of the one developed by Pignone et al. [21]. The adaptation process was performed with the support of an advisory group of health professionals related to CRC care. We carried out several focus groups with professionals and citizens eligible for screening who had undergone the procedure or not, before and during the adaptation process. The final DA is presented in a web format that presents information based on scientific evidence about CRC (available in: www.pydesalud.com/toma-de-decisiones-en-cancer-colorrectal/). It is organized in three sections; the first one explains the usefulness of DAs and a brief summary of their content. The second section explains CRC causes, symptoms and available treatments. Third, information on FOBT and colonoscopy is presented, including quantitative data about incidence and mortality risk reduction, as well as potential adverse effects. Finally, a summary table is presented with the two tests and their characteristics (i.e., description, preparation, recovery, frequency, need for additional tests, incidence and mortality risk reduction, adverse effects) compared to not being screened. Outcomes variables are then assessed in the same interface. Once the questionnaires have been completed, a summary document including the content explored together with the answers given by the patients is automatically generated and available for them, in a printed format or via e-mail.”

Overall, a collection of potential biases and inespecific effects prevent me from assessing the evidence of this study, which would go beyond the existing meta-analysis of Volk et al. (2016) and would yield publication of the manuscript.

We thanks Mr. Pereira for his comments.


Georgina Kennedy (Reviewer 3)

Accept after discretionary revisions

Overall this is a helpful study, with a solid rationale and methods appropriate to answer the research questions posed.

We thanks the reviewer for her comments.

A number of sentences require re-writing for parsability and excessive length e.g.

- method pg of abstract

- results 1st sentence, results pg2 last sentence, results pg3 last 2 sentences

Multiple places: population based --> population-based ...

nor did we [make] explicit a time frame...

We have made the changes recommended by the reviewer.

Volk et al is mentioned, however it should be called-out that it covers the effectiveness of CRC screening DAs, not just DAs in general, and also provide rationale as to why none of the reviewed tools are appropriate (population? language?)

In the third paragraph of the introduction we mention the Cochrane review of Stacey et al. [17], that includes studies evaluating DAs for a wide range of medical conditions, and the one of Volk et al. [18], focused on CRC screening.
In response to the second issue, we now have mentioned in the text that our DA was a translation/adaptation of a previous one developed in the USA (Pignone et al. [21]), and explained its development and content a little more in detail (see Decision Aid section).

Method p1 - '...reminders to undergo the screening [in the past].' We have changed it accordingly.

Decision aid design - this is described, but not rationalized - perhaps ground in the Volk et al results?

As commented above, we have now included two sentences explaining the process development.

Why was one patient not analyzed in each group for intention to undergo screening?

Those two participants did not answer to the “intention” question.

Peter Loewen (Reviewer 4)

Major revisions required

This is a fairly well-written report of an evaluation of a DA aimed at increasing awareness of and screening rates for CRC on the island of Tenerife.

Methods

"Computer-based simple randomization was performed by an independent researcher." RCT reporting best-practices (CONSORT) require additional disclosures: "Type of randomisation; details of any restriction (such as blocking and block size)", "Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned", "Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions", "who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how".

We have modified the text to better explain these aspects (method, second paragraph):

“Computer-based simple randomization was performed by a statistician not involved in the study, and the researcher who recruited participants and established an appointment by phone was blinded to allocation (we used a centralized off-site computer allocation process). All
participants were received in a room of their primary care center by a different researcher, who had the allocation list; all participants were assigned to the group in which they had been randomized.”

The two centers were described as different in terms of access to CRC screening services. Was randomization blocked by center?

We used simple randomization, without blocking nor stratifying by center.

Disclose re: allocation concealment and blinding or lack thereof.

"reviewed the DA accompanied by a researcher" - reviewed it where? in the clinic? sitting beside them as they used the computer? on the phone while they completed it on their own computer?

We have now explained this in the manuscript (method, second paragraph):

“All participants were received in a room of their primary care center by other researcher […]. Intervention participants signed informed consent and reviewed the DA in the computer; the researcher was sitting beside them and briefly explained them the functioning of the DA and gave support in navigation when necessary.”

"...in a web format that gathers information based on scientific evidence about CRC..." Since I infer that the DA does not collect any personal health information from users, "presents" might be a more appropriate verb than "gathers" here.

The reviewer is right, and we have changed the word.

The URL, " www.pydesalud.com/toma-de-decisiones-encancer-colorrectal/ ", es "No Encontrado" so I was unable to review its content.

We have checked the link and it works correctly (www.pydesalud.com/toma-de-decisiones-en-cancer-colorrectal/). In response to other reviewers, we have changed the text to explain better the DA developing and content:

“The DA was a Spanish translation and adaptation of the one developed by Pignone et al. [21]. The adaptation process was performed with the support of an advisory group of health professionals related to CRC care. We carried out several focus groups with professionals and citizens eligible for screening who had undergone the procedure or not, before and during the adaptation process. The final DA is presented in a web format that presents information based on scientific evidence about CRC (available in: www.pydesalud.com/toma-de-decisiones-en-cancer-colorrectal/). It is organized in three sections; the first one explains the usefulness of DAs and a brief summary of their content. The second section explains CRC causes, symptoms and available treatments. Third, information on FOBT and colonoscopy is presented, including quantitative data about incidence and mortality risk reduction, as well as potential adverse effects. Finally, a summary table is presented with the two tests and their characteristics (i.e.,
description, preparation, recovery, frequency, need for additional tests, incidence and mortality risk reduction, adverse effects) compared to not being screened. Outcomes variables are then assessed in the same interface. Once the questionnaires have been completed, a summary document including the content explored together with the answers given by the patients is automatically generated and available for them, in a printed format or via e-mail.”

"Therefore, we instead used intention to be screened as a proxy instead"... recurring "instead".

Thanks for the correction.

The Methods do not disclose the differences in the intervention between arms to which participants were randomized. What exactly were the two groups exposed to for the study, what was identical between the arms, and what was different?

Except for the application of the DA, we tried to offer the same experience for intervention and control participants: appointment in their usual primary care center, standardized instructions provided by the researcher, who did not mediated in participants’ responses to questionnaires, which were delivered in the same web interface. We have now discussed this issue in the discussion (last paragraph, limitations).

The primary outcome variable is appropriately specified. The secondary outcome variables should be specifically designated as such.

We have followed the reviewer suggestion.

Results:

There are valid statistical arguments for and against using Yates' correction for continuity in this circumstance. Why was it used?

We used it because several of the expected frequencies in the “No screening” cells were lower than 5 (this is now explicitly stated in the text). Given that in some analyses uncorrected and corrected values differed in their significance with a 95% confidence level, we decided to report both.

"When the same analyses were performed separately by subsample, in center A the difference favoring the DA was significant for both procedures" - Table 3 does not appear to show this. It shows non-significant results for both procedures for Center A with Yates’ correction.

The reviewer is right. We have now commented the corrected non-significant results in the text.

Table 4: re-check the means. On inspection, it seems improbable that the mean for row 1 is below 4 given the distribution of responses.
We have checked that calculation and it is correct. Twelve subjects (11.2%) scored 1, 10 (9.3%) scored 2, 15 (14%) scored 3, 25 (23.4%) scored 4 and 45 (42.1%) scored 5.

\[
(12+20+45+100+225)/107= 3.76
\]

Discussion:

There are serious limitations to the "intention" outcome in this study, but the authors appropriately acknowledge them in this section.

We thanks Professor Loewen for his valuable comments.

Lucia Sacchi (Reviewer 5)

Reject

This paper describes the results of a randomized control trial to assess the effects of the use of a decision aid for promoting colorectal cancer screening. The study considered two medical centers in two different areas with different characteristics in terms of screening policies.

Although the final goal of the paper is very interesting, unfortunately the study has several limitations, the most important being the low sample size and the impossibility of evaluating the actual screening uptake assessment.

The sample size calculation is described in page 7, line 1, yielding 126 participants to detect an effect size of 0.50 (Cohen’s d) in the primary outcome (decisional conflict). Because of time constraints we only could recruit 107 participants, but nonetheless results were significant in center A (n=83) due to the intense effect observed.

The use of the intention to be screened as a proxy of the screening uptake seemed not to be appropriate in this specific population.

The use of intention as a proxy outcome (which was not the primary outcome of the study) is certainly a limitation, but as we mention in the text, we consider that the used of an intention measure is well justified both theoretically and empirically (previous findings has shown moderate to strong relationships between intentions and actual decisions, as we commented in the article), and in fact, it has been used in many previous trials assessing DA in colorectal cancer screening (please, see the review of Volk et al. (2016), indexed in the references).
In addition, the presentation of the methodology and of the results should be improved and clarify in several points.

In the Method section, it is important to describe all the parameters that are collected in the study, and the moment of the study when these are collected for the two patients' populations (control vs intervention). In particular:

* It is not clear what are the "questionnaires" that are given to assess the outcome variables, mentioned in the first part of the Method section. At this point of the paper, the authors have not yet mentioned what will be the outcomes of the study, and they still have not introduced the DA

We have modified the text to better explain the procedure (method, second paragraph), but we have maintained the description of the procedure before the description of the intervention and measures:

“All participants were received in a room of their primary care center by other researcher, who had the allocation list; all participants were assigned to the group in which they had been randomized. Intervention participants signed informed consent and reviewed the DA in the computer; the researcher was sitting beside them and briefly explained them the functioning of the DA and gave support in navigation when necessary. After that, participants filled the questionnaires assessing the outcome variables, in the same web-based interface, with no mediation of the researcher. Control participants only signed informed consent and completed the questionnaires, in a different interface of the web site. Therefore, all measures were assessed only once (after reviewing the DA in the intervention group), except knowledge in the DA group, which was assessed both before and after the DA application.”

* The DCS questionnaire should be better described, first of all giving a definition of Decisional Conflict, and explaining why this measure is appropriate as the primary outcome of the study. In addition, the paper should specify when this questionnaire is administered to the study population, and specifically to the intervention group: is it administered both before and after the delivery of the DA? Finally, the DCS includes several sub-sections, which are then used separately for results evaluation. The paper should include the number of questions in each of the subsections and the possible range of the scores for each one (is it normalized as the global score? Is it the raw score given by the sum of the single questions scores?).

We have modified the text to better describe the DCS and its subscales (measures, first paragraph):

“The primary outcome measure was decisional conflict, defined as uncertainty about the course of action to take when choosing among several medical procedures [21]. It was measured with the Spanish version of the Decisional Conflict Scale (DCS) [22]. It includes 16 items and 5 subscales (with 3 items each, except the latter, with 4): feeling informed, having clear values
about benefits and risks, support to take the decision, uncertainty, and perceived effectiveness of the decision. Items are scored from 0 (strongly agree) to 4 points (strongly disagree), with higher scores indicating higher decisional conflict. Scores on the total scale and subscales are transformed to a 0-100 scale.”

* The DA should be described in more detail, especially because the available web version is only in Spanish and it could be difficult to understand for the readers.

We have changed the text to explain better the DA developing and content:

“The DA was a Spanish translation and adaptation of the one developed by Pignone et al. [21]. The adaptation process was performed with the support of an advisory group of health professionals related to CRC care. We carried out several focus groups with professionals and citizens eligible for screening (who had undergone the procedure or not), before and during the adaptation process. The final DA is presented in a web format that presents information based on scientific evidence about CRC (available in: www.pydesalud.com/toma-de-decisiones-en-cancer-colorrectal/). It is organized in three sections; the first one explains the usefulness of DAs and a brief summary of their content. The second section explains CRC causes, symptoms and available treatments. Third, information on FOBT and colonoscopy is presented, including quantitative data about incidence and mortality risk reduction, as well as potential adverse effects. Finally, a summary table is presented with the two tests and their characteristics (i.e., description, preparation, recovery, frequency, need for additional tests, incidence and mortality risk reduction, adverse effects) compared to not being screened. Outcomes variables are then assessed in the same interface. Once the questionnaires have been completed, a summary document including the content explored together with the answers given by the patients is automatically generated and available for them, in a printed format or via e-mail.”

It is important to underline that the questions about knowledge on CRC and its screening options are included in the DA (if I understood correctly by looking at it) for the intervention group, and they are repeated twice, before and after the patient is informed. As regards the knowledge questionnaire, it should also be made clear how this is delivered to the control group (on paper? Together with the DCS?)

We have modified the text to state that knowledge questions were assessed before and after the DA information. For the control group, they were delivered via web, as commented above, along with remaining dependent variables.

* The characteristics of screening that are assessed for importance are not detailed at all. They must be listed and, also in this case, it should be specified if these questions are included in the DA for the intervention group, or how they are administered to the two groups.

The characteristics of screening assessed for importance are listed in table 4. In the DA group their evaluation by the participants was made after the DA information, along with the remaining
dependent variables. For the control group they were presented in the web-based questionnaires, along with the remaining dependent variables.

The statistical analysis section should be better detailed as well. Stating that "we performed several analysis of variance" is too vague, especially because the study has two populations and two centers that are considered, so different effects need to be considered.

We performed one 2(DA/control) x 2(center A/B) ANOVA for each dependent variable (knowledge, DCS total score and subscales). We have now stated it more clearly in the manuscript.

When talking about the comparison between the intervention and the control group, it is important to have previously clarified when the measurements are collected. For example, comparing the knowledge score between the two groups considering the score for the intervention group collected before the DA was administered can help to assess the baseline, while using the knowledge score of the intervention group after the DA might help clarifying how the knowledge has improved thanks to the information received.

Thanks for the suggestion. We have now reported that there was not a significant difference in knowledge between the DA group before the intervention and the control group (baseline assessment). We also report the pre-post difference in the DA group, and the between group difference (after reviewing the DA).

As regards logistic regression, there exist several techniques able to automatically select the variables during the learning phase (stepwise regression, lasso). It is not clear why the authors have decided to manually select the features instead of using an automatic method.

We used the strategy proposed by Sepucha et al. [28] to construct a measure of concordance between goals/concerns and decisions (intention in our case), although other methods could be used. Since there was not significant associations at the univariate level, automatic selection would not change the results.

The fact that in center B patients had already been invited and reminded for screening (and thus probably already informed) affects their knowledge and information level, as well as their intention to be treated (as it is shown by the results). This is a selection bias that might influence the obtained results, as the patients had already been exposed to the intervention, and I think this affects the overall results of the study. It would be important to understand how much information they were provided with during the first contacts, and if they had personally provided to collect information about the disease and the screening program.

Including center B was not a selection bias but a secondary aim of the study, that is, to compare the effect of the DA in the two populations (invited and not invited to the public program). Invitations to participate in the program do not include specific information about CRC
screening, and therefore it cannot be considered that they had been exposed to the intervention, as much they only received general information about screening (the near-significant higher score in knowledge for center B was due exclusively to the high score obtained by the DA group; the control group obtained the same score than control group in center A). We have modified the text to better explain this issue (introduction, last paragraph; method, first paragraph).

As regards the evaluation of the concordance among the expectations of the patients and their intention to be treated, it could be useful to include the center as a correction variable in the logistic regression.

Thanks to the reviewer for this suggestion. We have re-analyzed the data controlling for center and results have not changed. We now mention this in the text.

The fact that the intention to be treated for patients belonging to center B seems not to be different from that of patients in center A and that the number of patients who stated their intention not to be screened is very low, makes me wonder how appropriate is this measure as a proxy of the screening uptake measure in this specific sample. The authors should better discuss this point.

Based on previous findings and the actual rate of screening in our region, we did not expected the observed so high rate of participants stating a positive intention. In the discussion we recognized the likely presence of a (desirability?) bias, and that it would has been better to also measure the “intensity” of intention, or to constrain the hypothetical implementation of the decision to a short-term period. We have now added that the fact of being physically in a health context (a primary health care center) also could introduced a desirability bias. However, we want to highlight that even in a context of a ceiling effect for the whole sample, in which is more difficult to obtain significant differences, the DA yielded a significant (uncorrected) or near-significant (corrected) increase in the rate of positive intention to undergo FOBT.

Jiang Bian (Reviewer 6)

Major revisions required

(1) The paper needs some proofreading. Some of the sentences are awkward and with grammatical errors. A professional editing service might be helpful.

We have tried to improve language expression.

(2) Why the need for "no family history of CRC" as a inclusion criteria? Please justify.
Individuals with family history of CRC present a higher risk of developing the disease, and therefore they could be more prone to undergo screening. Previous trials of decision aids (DA) in CRC screening also have applied this exclusion criteria.

(3) How were participants identified? Through reviewing of the EHR system and then sent out invitation mails? Please explain.

Participants were identified by collaborating primary care physicians, who assessed eligibility criteria (page 5, line 3), preliminarily informed them about the study and invited them to be contacted by our team. We have modified the text to reflect explicitly this issue.

(4) "www.pydesalud.com/toma-de-decisiones-encancer-colorrectal/" is no longer available. It’s hard to assess the adequacy of the decision aids. How was the DA developed? e.g., went through focus groups, etc?

We have checked the link and it works correctly (www.pydesalud.com/toma-de-decisiones-encancer-colorrectal/). We have changed the text to explain better the DA developing and content:

“The DA was a Spanish translation and adaptation of the one developed by Pignone et al. [21]. The adaptation process was performed with the support of an advisory group of health professionals related to CRC care. We carried out several focus groups with professionals and citizens eligible for screening who had undergone the procedure or not, before and during the adaptation process. The final DA is presented in a web format that presents information based on scientific evidence about CRC (available in: www.pydesalud.com/toma-de-decisiones-en-cancer-colorrectal/). It is organized in three sections; the first one explains the usefulness of DAs and a brief summary of their content. The second section explains CRC causes, symptoms and available treatments. Third, information on FOBT and colonoscopy is presented, including quantitative data about incidence and mortality risk reduction, as well as potential adverse effects. Finally, a summary table is presented with the two tests and their characteristics (i.e., description, preparation, recovery, frequency, need for additional tests, incidence and mortality risk reduction, adverse effects) compared to not being screened. Outcomes variables are then assessed in the same interface. Once the questionnaires have been completed, a summary document including the content explored together with the answers given by the patients is automatically generated and available for them, in a printed format or via e-mail.”

(5) "We could not assess actual screening uptake in center B, and participants of center A did not have access to the population based screening program and therefore could only undergo screening by means of private services, but this was considered improbable given the out of
pocket costs." seems to simple of an explanation. First, if cost of the procedures is the most prohibitive factor, high intention to screening does not mean much. Second, center B does have a program. Why not wait 3 month, do another survey and assess the actual uptake rate?

In center A, the public screening program was not available at that moment, but it was planned to progressively cover all the population in the following years. Therefore, the assessment of intention to screening was justified. Funding limitations precluded us to recruit a larger sample in center B and to carry out a follow up assessment in both centers.

Joaquim Cezar Felipe (Reviewer 7)

Major revisions required

The article presents a procedure to assess the influence of the use of decision aid (DA) material to promote colorectal cancer screening, showing statistical results, applied to a group of primary care patients in Spain.

The text is clear and well written. The results are presented in a synthetic and clear way. However, some key issues need to be resolved:

1) What is the original contribution of this study? The authors cite other papers (eg, references [17], [18] and [19]) that have already performed the same study, on the same disease or not. It was not clear what this work adds, comparing it to what has already been published.

In the field of colorectal cancer screening, most studies have assessed actual screening (or at least intention/interest to be screened) since it is the most relevant outcome from a public health perspective, but only a few have assessed decisional conflict, a basic psychological process involved in the decision-making about screening. Therefore, we choose this construct as our primary measure.

Second, to our knowledge this is the first study in colorectal cancer screening that compare the effect of a DA in patients who had been invited (although not exposed to a DA or otherwise informed about the procedure) or not to participate in a public screening program. Although limited by the small sample size in center B, results indicate that in this population the DA is effective at increasing knowledge but not at decreasing decisional conflict, probably because conflict is already low in these participants.

Third, some previous studies have analyzed the effect of a DA on the concordance between the stated preferred screening option and actual decision, but not between goals/concerns regarding screening and the final decision (or, as in our case, intention to be screened). This kind of concordance has been proposed as an indicator of good decision quality by Sepucha et al. [28].
Nonetheless, since we did not find significant associations, we could not construct that measure of concordance.

We have modified the text to reflect these aspects (discussion, last paragraph).

2) How was defined the number of patients (107) who participated in the study? If the goal is to represent a specific population, there are statistical formulas to determine that number, taking into account the desired degree of confidence and margin of error, as well as the standard deviation or a proportion of belonging to category within the population. This number of patients seems to me not to be enough to generate reliable statistical results. In addition, the title of the article induces an expectation of being a study carried out nationally in Spain.

Sample size calculation is described in page 7, line 1. We needed 126 patients to detect a small-to-moderate effect size \( (d = 0.5) \) in the primary outcome measure (Decisional Conflict Scale). Because of time constraints we only could recruit 107 participants, but results were significant due to the intense effect observed.

3) Also, the characteristics of the patients should be defined and presented in such a way that we had a clear definition of the scope of the study. Patients could have been stratified in, for example, age range, socioeconomic conditions, demographic conditions, clinical conditions, and so forth. With this, the study may get conclusions such as: "in a population with these characteristics…, the influence of the application of DA is this…," regardless of nationality. This allows further comparison of analyzes, thus giving a degree of importance to the study.

As the reviewer comments, in an ideal design participants should be stratified by all the relevant characteristics that could influence the results (for instance, socioeconomic status and clinical comorbidities are two relevant confounders). However, it requires large sample sizes to obtain enough statistical power to assess the intervention effect in the different subgroups. The aim of this study regarding subgroups was more modest, limited to compare the effect of the DA in participants who previously have been invited to participate in the public screening program to those who did not.