PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Communicating personalised statin therapy-effects as 10-year CVD-risk or CVD-free life-expectancy: does it improve decisional conflict? Three-armed, blinded, randomised controlled trial.
AUTHORS	Jaspers, Nicole; Visseren, Frank; van der Graaf, Yolanda; Smulders, Yvo; Damman, Olga; Brouwers, Corline; Rutten, Guy; Dorresteijn, Jannick

VERSION 1 – REVIEW

REVIEWER	Castro, Inar Universidade de Sao Paulo Faculdade de Medicina
REVIEW RETURNED	15-Sep-2020

GENERAL COMMENTS	The authors evaluated the effect of communicating patients with stable CVD under statin prescription about their personalized prognostic, using an algorithm (Smart-Reach model), on their decisional conflict in terms of statin intensification or discontinuation. They concluded that patients who had their prognostic expressed as 10 –years CVD-risk or CVD-free life expectancy showed lower decisional conflict after 1 month than those who receive a standard generalized communication (Control group). It is a challenge to control patients with drugs for a preventive proposal, mainly if the drugs can cause adverse effects. Thus, efforts towards this point are very important. In addition, online tools or “apps” based on mathematical models aiming to help patients and physicians, have increased for several types of chronic diseases, but their efficacy must be evaluated in terms of correlation with clinical end points. This study presents a well carried out randomized controlled trial. However, some parts are not clear and should be revised. Since the begging is important to clarify that “communicating personalized” means an individual prognostic obtained by an algorithm, and which are the practical consequences of a high “decisional conflict”, that is the primary outcome of the study. The authors could exemplify the consequences associated to some DCS classes, since the range is from 0 to 100 (Page 9). The authors justified the lack of more expressive results by the fact that Control group already presented a low DCS. Considering that no difference was observed after 6 months, none of secondary outcomes changed and post-interventional LDL-c concentration was the same for all groups (1.9 mmol), the authors should discuss the efficiency of this type of tool to improve statin adherence and reduce CVD risk. Figure 1: The number of patients is different if the exclusions are discounted. For example, initial n=432 less patients not eligible

	(81) should be 351 and not 384. After, 384 – 81 is 303. In the last group, 101 less 20 is not 90. Please, clarify these numbers. Page 10: Statistical analysis. The authors applied T-test to calculate the sample size. There is a severe statistical restriction about the use of T-test when more than two groups are compared. Next, ANOVA was applied to compare the results but again T-test, instead other post-hoc tests, was used to compare the groups. I suggest the authors comment this procedure. Page 11: The analysis that authors designed as sub-groups should be in fact ANCOVA and not two-way ANOVA, neither sub-groups. Thus, this part should also be checked. Table 7 was cited before Table 1. Table 1: pressure of iLe-Group cannot be 129 mmHg. Revise the legend. Figure 2: Kruskal-Wallis not Wallace. It is more usual to show the p values in the graphic. Page 12 Line 15: It is important to comment that all groups showed DCS values around 25, meaning “following through with a decision”. Legend of Sup. Table 1 must be corrected. Sup.Table 2 : Statistical test column is not necessary. It is not clear what means the “n” of Post-interventional LDL-c considering initial n =101/group. Include the units in the table. Legend of Sup. Table 3 is confused. The authors could explore some association between health literacy and educational level in terms of DCS. Page 14 line20: “one study” [11,32]. The reference seems to be Harmsen et al. [11] Supp. Figures 1 and 2 and Table 2 were not commented in the manuscript. Appeal to consider CONSORT items 1) Outcomes The primary outcome was clearly defined (page 9) and used to calculate the sample size (page 10). It was measured by a validated scale (DCS) that range from 0= no decisional conflict to 100 = extremely high decisional conflict. It was considered scores >37.5 as “feeling unsure about implementation of the decision” and < 25 are “following through with a decision” (page 9) . 2) Sample size Page 10: sample size (n=258) was calculated based on the primary outcome (DCS). The parameters were: alpha risk=0.05; beta risk = 0.20, and a difference of 8.5 in the scale. No information was given about the variance. However, on page 53/54 the authors stated that clinically meaningful effect-sizes ranged between 0.43 and 0.86 on 5-point scale and they based the sample size calculation on mean difference of 0.80* 0.43 =0.344. It is confused because in the manuscript (page 10) the authors stated that T-test or Wilcoxon was applied to compare the groups. There is a severe statistical restriction about the use of T-test when more than two groups are being compared. I suggest that the authors comment this procedure. 3) Sequence generation Page 7: The algorithm R to randomly designed assignment with a 1:1:1 ratio in block-sizes of 12. 4) Allocation concealment

	An independent investigator performed randomization and allocation via an anonymous patient number. 5) Blinding Page 7: Patients were blinded to treatment and allocation. 6) Outcomes and estimation Primary outcome (DCS) was presented (Figure 2) as median (25th-75th percentiles). Secondary outcomes (Sup.Table 2) as median (IQR) 7) Harms No information was provided about harms. 8) Registration Page 7 line 45: The study was registered in the Netherlands Trial Registry (NTR 6227) 9) Protocol NL58608.041.16 10) Funding Partially funded by a Netherlands Heart Foundation grant (2016T026)

REVIEWER	Korenstein, Deborah Memorial Sloan-Kettering Cancer Center, Department of Medicine
REVIEW RETURNED	08-Oct-2020

GENERAL COMMENTS	BMJ Open Review: Communicating personalised statin therapy-effect as 10-year CVD risk or CVD-free life-expectancy Abstract: I don't understand the first 2 sentences of the Results section at all. It doesn't appear related to the main outcomes of the study and I would recommend removing (or if it is related to the main outcomes it needs a lot of clarification). Introduction: The authors make the point that absolute benefit is likely smaller than what patients expect, so being informed of it may discourage medication use. However, their study did not really test disclosure of absolute benefit, it tested disclosure of personalized absolute benefit. Many decision aids do focus on absolute benefits and harms, but they do not personalize them, which is what sets this intervention apart. Page 6; the sentence beginning on line 39 (Moreover,) has a typo and uses the wrong word ("illicit" should be "elicit"). Methods: I'm curious about the choice of study population. Since all patients had known CAD and were already on statins, they are not likely to have much decisional conflict related to statin use. Increasing the dose does not seem likely to lead to much decisional conflict in that context. I'm not sure if the all patients in the SMART study were already on statins, but if not I would have thought the statin-naïve patients would be a better study group for this sub-study. The authors note that the study was hypothesis blinded, which seems to mean that patients were unaware of the hypothesis. Were their treating physicians aware of the hypothesis? Since

	patients were encouraged to discuss statin therapy with their doctor, and since there were physician reported secondary options, blinding of the doctors seems important. Please clarify. It seems that patients enrolled in SMART do not see a clinician, based on the text on page 8. Is this correct? It seemed unlikely to me, so wanted to make sure I am not misunderstanding. Results: In the first paragraph, the authors report the percentages of patients in each group that increased their statin doses. Were all participants recommended to increase the dose? Do we know the number of people who were recommended to increase or decrease their dose compared? This information may not be relevant and I would consider removing. Do you know how many patients watched the video or discussed statins with their doctor? It would be interested in understanding "adherence" to the intervention. At the bottom of page 12, the authors report that physician views about the value of statins did not differ across study arms. How many patients did each physician have in the study? Did they generally have a single patient, or could a physician have cared for several study patients who were allocated to different study arms. Were physicians asked about statin use for a particular patient or more generally? Discussion: In the first paragraph (lines 45-50), the authors state that their results indicate that providing estimated effects "positively impacts patient's opinions on taking these medications". I don't think the findings support this statement- patients in the intervention group had lower decisional conflict, but that decision may have been to continue a lower dose when a higher dose was recommended. I think the statement is particularly misleading since there were no patients in the study who were NOT already on statins. On page 15, the authors state that the algorithm was published after randomization of the last patient- not sure why this is noted, but certainly part of the study took place after the randomization. Was it published after everyone had reported results for the primary outcome? The authors make the point on page 15 about hypothesis blinding. Again I think they need to be clearer about who was blinded to the hypothesis. Under limitations, they discuss the fact that all patients were already on statins. Findings in a group treated for primary prevention might also be very different, which is important to note. It might also be quite different in real practice compared to with patients who are already enrolled in a clinical trial. Tables and Figures: Table 1- there are important apparent differences (rates of diabetes, smoking). Were these statistically significant?

	Figure 1- there appears to be an error- the number 384 appears twice Figure 2- these look quite similar to me. I would recommend considering a review by a statistician to see if the correct tests were done, since it does not look like the differences should be so highly significant. Consort guidelines: 1) Outcomes (Item 6a) Primary outcome is clearly specified, as well as how and when it was measured. The metric was difference across study arms. 2) Sample size (Item 7a) Sample size calculation is described 3) Sequence generation (Item 8a) Block randomization is described 4) Allocation concealment is achieved using centralized block randomization and anonymous patient numbers (Item 9) 5) Blinding (Item 11a) None of the relevant parties (patients, physicians, outcome asssessors) were blinded, and physicians and patients could not be blinded given the nature of the intervention. Patients were blinded to the hypothesis, though it is unclear if physicians were so blinded. Outcome assessors appear not to have been blinded, but given the fact that the study uses surveys I do not think the lack of blinding introduces bias. 6) Outcomes and estimation (Item 17a/b) Relevant outcomes are appropriately presented 7) Harms (Items 19) In this study, the primary outcome could reflect either benefit or harm; there are not other harms to be collected. 8) Registration (Item 23) Yes. Registration was in the Netherlands Trial Registry and the identifier is provided. 9) Protocol (Item 24) Yes. In the discussion a link to the trial protocol is provided. I might favor moving this outside the Discussion and into the Methods section or outside the main text. 10) Funding (Item 25) Funding source is included but role of the funder is not described.

REVIEWER	Wood, Angela

	University of Cambridge
REVIEW RETURNED	23-Nov-2020

GENERAL COMMENTS	This manuscript reports the results of a trial to investigate the impact of communicating 10-year CVD risk or CVD-free life-expectancy on patients' decision-making. There has been much focus on improving CVD risk tools for clinical practice and relatively less effort placed on improving the communication of such tools for both clinicians and patients. Studies like this are important in order to quantify the impact of translating CVD risk tools into accessible estimates for patients, including personalised health profiles incorporating 10-year CVD risk and CVD-free life-expectancy. The trial has been designed and analysed using simple, commonly-used and transparent statistical methods. The results are presented and interpreted in an appropriate manner. My comments are below: 1. The term "personalised treatment-effect" is misleading. I initially assumed the authors were incorporating "treatment effect heterogeneity" (eg, the statin effect differs by age, ethnicity or time since initiation etc…). However, a fixed treatment effect was used for all individuals (ie, HR of 0.80) and the effect of this was assessed on the "personal health profile" including 10-year absolute CVD risk or CVD free life-expectancy. I suggest the text is amended to avoid this misinterpretation. 2. Kruskal-Wallace should be "Kruskal-Wallis". 3. Figure 1 has several typos: (a) "Approached patients screened for eligibility": should be n=303 not n=384 (b) Numbers under the CVD-free life years intervention do not add up. 4. Can the authors provide reasons why patients did not complete the primary endpoint at one-month (the original protocol suggests this was collected)? 5. More description (as supplementary tables or figures) is required to fully understand the nature of the missing outcome data (eg, % missing by trial arm, or table 1 presented for patients with observed outcome). Patients with low literacy were more likely to have missing outcomes – did this vary by trial arm? How does this finding implicate the significant subgroup finding? Are the % of patients with missing outcome significantly differently between arms? Further sensitivity analyses may be required to assess the impact of missing outcomes on the trial results (eg, (i) an analysis adjusting for baseline variables which are no longer randomly distributed between trial arms amongst patients with observed outcome (ii) multiple imputation approach with a delta-adjustment to impute under a missing not at random assumption). 6. Figure 2: typo – "Post-hoc" 7. Page 15, line 51 – remove extra fullstop

	8. Page 62, lines 46-47 (Section 6.5.1 Overview of procedures): Under “additional measurements” no “X” marked for “measurements obtained from SMART work-up

REVIEWER	Chiocchia, Virginia Universität Bern Institut für Sozial- und Präventivmedizin, Institute of Social and Preventive Medicine
REVIEW RETURNED	24-Nov-2020

GENERAL COMMENTS	I would like to thank the authors of this randomised controlled trial (RCT) for submitting their manuscript. I do not have clinical expertise but the study seem to be answering an important question and to be carefully thought-out. However, I think the reporting can be improved substantially and the statistical sections better described, so I am suggesting a major revision. Major comments: • It is not clear whether the authors have followed the CONSORT guidelines for reporting this RCT. They have added in the supplementary material a table listing items to include in an abstract which I do not think it is relevant here. What is needed is the completed CONSORT checklist as required by the journal. I strongly recommend to follow the guidelines to make sure all items/sections are properly described, as also noted in some of my other comments. Also, the SPIRIT guidelines should have been followed for the study protocol but since the trial is now completed, I will just assume they have. • page 7 of 73 - Design, randomization, and follow-up: 1. "hypothesis blinded" is not a commonly used term. I understood why the authors used this term and I think it is explained well in the protocol. I suggest they also explain this rationale in the manuscript and, most importantly, they need to clarify who was blinded (if anyone) and how (see CONSORT). 2. Randomisation: I also think this is better written in the protocol. R is a software not an algorithm. It can simply say computer-generated random allocation sequence. No mention of how allocation concealment was achieved (see CONSORT for all details required). • page 8 of 73 - Intervention arms: when reading this section, it was not very clear to me how the interventions differed; again, I understood it better after reading the protocol (page 18). It does not need a lot of details but simply a clearer and more structured definition. • page 10 of 73 - Statistical analysis: This section can be expanded and should be described in a clearer way (see CONSORT). 1. I would personally separate the sample size calculation in a subsection but, even if not, it should be distinct and described first (here the intention-to-treat is first mentioned, then sample size calculation, then analysis again). It's a three arm trial, among which arms was the 8.5 difference to be detected? This difference need to be justified somewhow and the references in the relevant section of the protocol are not the same as the one cited here. Also, where is 8.5 in the calculation shown in the protocol? It is a bit confusing, it has to be clear so that it is reproducible. The standard deviation used also need to be included. 2. The statistical analysis methods are not clear and, given the lack of statistical analysis plan as a supplement, I was not sure how the analysis was actually carried out. First, the effect size and precision (how the results are reported e.g. OR, RR, mean

	difference, 95% CI, etc.) need to be specified for all outcomes (i.e. both continous and categorical variables). I saw in the results table the test statistic is reported, which is meaningless for the readers as it does not say anything about the effects observed between groups, so I recommend removing. I am unsure about the "direct consecutive comparisons" terminology, "pairwise comparisons" is the term normally used. ANOVA tests if all group means are equal and, if so, then one can compare all pair of means but has to account for multiple testing and it is unclear whether this was done (and accounted in the sample size calculation) and I did not find anything even in the protocol. Since medians are reported instead of means, I suppose it is because of non-normality. However, this is not addressed in the methods. Also, if the assumptions are not met, one could try transformations (e.g. log-normal) before moving on to use non-parametric tests. Also, the Wilcoxon-rank sum test does not test the difference in medians betweeg groups, there are other available methods for that. I suggest discussing with the study statistician the possible options. Also, I would "reorder" the use of tests in the text: ANOVA then t-test (not t-testing; correct throughout the manuscript); if assumptions are not met, Kruskall-Wallis then Wilcoxon-rank. 3. Good to see the subgroup analyses were pre-specified but I would also acknowledge somewhere (methods or results section) that the study was not powered to detect any difference between such subgroups. • Not clear in the Author contributor statement who performed the data analysis. Minor comments: • I did not comment on the results and discussion section as I would like to see the results reported differently (i.e. I do not have a clear idea of the effect and its precision between groups) but otherwise seemed to be presented clearly. • In the Abstracts I do not understand why the changes in CVD risk and CVD-free life-expectancy are mentioned as the first thing in the Results given that they are not even secondary outcomes. The number randomised and analysed should be mentioned first then the result of the primary outcome (see checklist supplement). • Registration: there is a new number on the registry. However, I searched it with this one and it works. I am just not sure which one (old or new) should be included in the article.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

1. Since the begging is important to clarify that “communicating personalized” means an individual prognostic obtained by an algorithm, and which are the practical consequences of a high “decisional conflict”, that is the primary outcome of the study. The authors could exemplify the consequences associated to some DCS classes, since the range is from 0 to 100 (Page 9).

Author response: Thank you for this suggestion. Although the use of a prognostic algorithms was already mentioned in the abstract and aims of this study, we have now better clarified this in the

abstract and introduction. Also, we have now improved our explanation of 'decisional conflict' in the methods section as follows:

"The DCS scale measures the amount of internal conflict a patient feels regarding a medical decision. Summary scores range from 0 (no decisional conflict) to 100 (extremely high decisional conflict). Scores >37.5 are associated with feeling unsure about implementation of the decision, possibly leading to discontinuation of the chosen option or fretting about the chosen option (i.e. using statins as prescribed by the physician), and those <25 are associated with following through with a decision. To limit loss to follow-up, patients who did not initially respond to the follow-up questionnaire were approached telephonically. A reminder was sent by mail if the patient was not reached."

2. The authors justified the lack of more expressive results by the fact that Control group already presented a low DCS. Considering that no difference was observed after 6 months, none of secondary outcomes changed and post-interventional LDL-c concentration was the same for all groups (1.9 mmol), the authors should discuss the efficiency of this type of tool to improve statin adherence and reduce CVD risk

Author response:
We have changed the first paragraph of the discussion to the following:

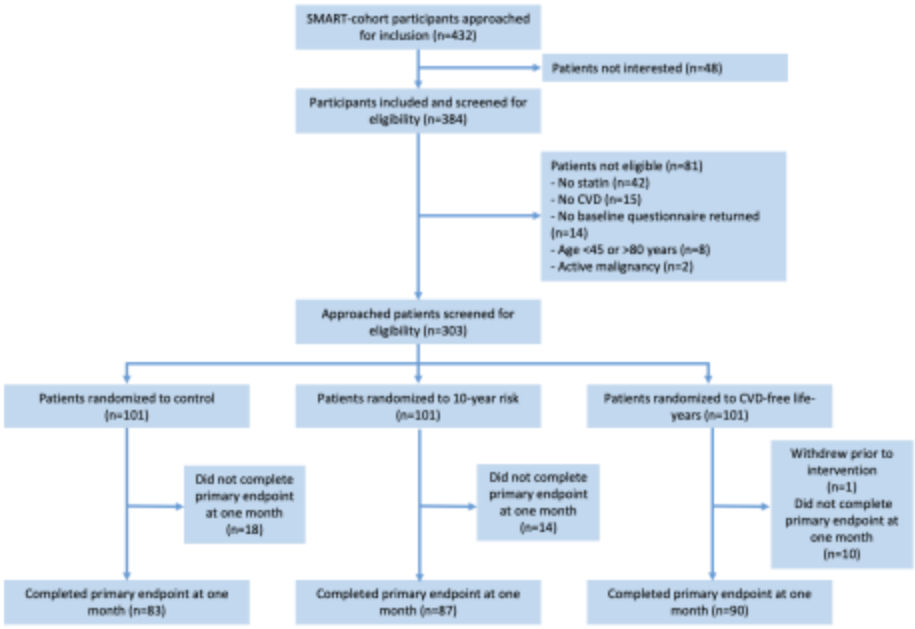
"Providing personalized estimates of the prognostic changes associated with statin use in terms of 10-year CVD risk and CVD-free life-years (compared to a control group) resulted in lower decisional conflict associated with statin use measures after one month. After six months no differences were found. Likewise, no group differences were found in secondary outcomes, which included the degree to which people perceived their CVD to be threatening, how effective patients viewed their statin-medications, or LDL-c levels after six months. The actual benefit from CVD-prevention is smaller than people initially report acceptable. Still, openly communicating the individual estimated of statin use on the prevention of CVD resulted in lower decisional conflict, without may people discontinuing their treatments. However, the effect was small in a population with a low baseline DCS."

We have also altered the paragraph in the discussion relating to the future studies:

"Because the present study showed only a small effect in patients with a low baseline DCS, future studies could focus on populations with higher baseline decisional conflict. For example, patients experiencing adverse effects during statin use or patients considering more intensive preventive treatment options on top of standard treatment, for example intensive blood pressure reduction or combination antithrombotic treatment.

3. Figure 1: The number of patients is different if the exclusions are discounted. For example, initial n=432 less patients not eligible (81) should be 351 and not 384. After, 384 – 81 is 303. In the last group, 101 less 20 is not 90. Please, clarify these numbers.

Author response:
Thank you for noticing this.
In the last group, 1 participant withdrew prior to receiving the intervention. The figure erroneously lists 10 withdrawing. Ten did not complete the questionnaire. This should be 101 – 1 – 10 = 90. And indeed, there was another error in that 384-81 is 303 and not 384. We have corrected this in the following.



4. Page 10: Statistical analysis. The authors applied T-test to calculate the sample size. There is a severe statistical restriction about the use of T-test when more than two groups are compared. Next, ANOVA was applied to compare the results but again T-test, instead other post-hoc tests, was used to compare the groups. I suggest the authors comment this procedure.

Author response:

We used a T-test for the sample size as the primary analysis of this trial was not a three-arm comparison. We were interested in comparing intervention 1 with placebo and also intervention 2 with placebo. The aim was to achieve enough power for both these two-arm T-test comparisons. By conducting an ANOVA test to first detect any difference between groups we aimed to limit chance findings.

We did not plan to account for multiple testing in the protocol. However, as suggested we have applied the (strict) Bonferroni correction to the results here. To protect against a type I error, a Bonferroni correction corresponding to the 22 secondary outcomes was applied. The new p-value was 0.05 / 22 = 0.002 to determine if any of the values were statistically significant. We have added the following to the manuscript:

“The study was not powered to detect any differences between subgroups. A Bonferroni correction corresponding to the 22 secondary outcomes was applied. The new p-value for statistical significance was 0.002.

5. Page 11: The analysis that authors designed as sub-groups should be in fact ANCOVA and not two-way ANOVA, neither sub-groups. Thus, this part should also be checked.

Author response: We did use an ANCOVA test. Our terminology in the original manuscript was confusing for which we apologize. We performed an ANOVA test with the subgroup inserted as an interaction variable. We have altered the text to as follows:

“Pre-specified subgroup analyses via an ANCOVA test investigated whether the effect of the intervention on DCS at one month differed according to the following: ”

6. Table 7 was cited before Table 1.
Author response:
There is no table 7. Did the reviewer perhaps mistake reference number 7 following the word ‘tablet’ for ‘table 7’?

7. Table 1: pressure of iLe-Group cannot be 129 mmHg. Revise the legend.
Author response:
We have corrected the systolic blood-pressure values, the values in the three groups are 131, 131, and 129.

8. Figure 2: Kruskal-Wallis not Wallace. It is more usual to show the p values in the graphic.
Author response:
Thank you, we have corrected this. P-values are shown in the graphic as well as the text.

9. Page 12 Line 15: It is important to comment that all groups showed DCS values around 25, meaning “following through with a decision”.

Author response:
We have added the following to the paper:
“The difference between iAR and iLE arms was not significant (W=3317, p=0.21, Figure 2). All groups showed a DCS of around 25, the value associated with following through with a decision”.

10. Legend of Sup. Table 1 must be corrected.
Author response:
We have altered the table to match the legend and vice versa.

11. Sup.Table 2 : Statistical test column is not necessary. It is not clear what means the “n” of Post-interventional LDL-c considering initial n =101/group. Include the units in the table.

Author response:
We have adjusted the table to be more clear on what the “n” pertained to. For a number of patients, there was no link possible with the general practioner dossier after six months. These patients have been listed a s “unknown.” For all other patients, we know for certain if the LDL-c values were or were not determined.

The statistical test column was applied wo show for which outcomes a parametric (i.e. F-value) and non-parametric (i.e. chi-squared) test was applied. We have altered this to provide the information in the footnote of the table without providing the test column in the table. We have done this with table 2 in the main manuscript as well.

12. Legend of Sup. Table 3 is confused. The authors could explore some association between health literacy and educational level in terms of DCS.

Author response: Thank you for pointing this out. We have made the footnote more clear by discussing the PAM in the main text, and also directly referring the the PAM as “patient activation measure” instead of “patient activation.”

Extra analyses regarding the association between health literacy and educational level in terms of DCS were not pre-specified. Although health-literacy and educational level are related, health-literacy has been shown to have stronger associations with health outcomes than educational level. (3)

13. Page 14 line20: “one study” [11,32]. The reference seems to be Harmsen et al. [11]
Author response:
Thank you, this has been corrected.

14 Supp. Figures 1 and 2 and Table 2 were not commented in the manuscript.

Author response:
Supplemental figure 1 is commented on in the paragraph “predicted therapy effects”. Supplemental figure 2 in the paragraph “subgroup analysis”. Table 2 in patient reported secondary outcomes.

2) Sample size
Page 10: sample size (n=258) was calculated based on the primary outcome (DCS). The parameters were: alpha risk=0.05; beta risk = 0.20, and a difference of 8.5 in the scale. No information was given about the variance.

However, on page 53/54 the authors stated that clinically meaningful effect-sizes ranged between 0.43 and 0.86 on 5-point scale and they based the sample size calculation on mean difference of 0.80* 0.43 =0.344. It is confused because in the manuscript (page 10) the authors stated that T-test or Wilcoxon was applied to compare the groups. There is a severe statistical restriction about the use of T-test when more than two groups are being compared. I suggest that the authors comment this procedure.

Author response: We thank the reviewer for providing us the opportunity to make this section more clear. We have added the following to the manuscript:
“We used a T-test for the sample size as the primary analysis of this trial compared interventions with the control group. Calculations were conducted using G*Power version 3.1. Sample size was based on an effect size (Cohen’s d = mean difference / standard deviation) of 0.43. A standard deviation of 0.80 was used to detect a mean difference of 0.34 on the 5-point scale (ranging from 0-4) which was taken to correspond to 8.6 on the 100-point scale.(7, 8) A power of 80% and a two-tailed alpha of 0.05 was used. A minimum of 86 patients per arm was needed.”

Reviewer: 2

1. Abstract: I don’t understand the first 2 sentences of the Results section at all. It doesn’t appear related to the main outcomes of the study and I would recommend removing (or if it is related to the main outcomes it needs a lot of clarification).

Author response:
We have removed this section from the abstract. These results show that actual personalized effects were smaller than what the majority of patients view as acceptable. Despite this, providing

personalized effects do not negatively affect the outcome. We agree however, that these results are a nuance, and that it has a place in the main text, but not in the abstract. It has been removed from the abstract.

2. Introduction: • The authors make the point that absolute benefit is likely smaller than what patients expect, so being informed of it may discourage medication use. However, their study did not really test disclosure of absolute benefit, it tested disclosure of personalized absolute benefit. Many decision aids do focus on absolute benefits and harms, but they do not personalize them, which is what sets this intervention apart.

Author response:
Thank you for helping us better present the novelty of our study. We have added the following sentence to the introduction, and modified the introduction to be more clear between the difference between population-based and personal benefit.

“In general, decision aids to not provide personalized benefits and harms. (4)”

We have also added the following to the strengths section of the article:

“Strengths of this study involve providing patients with estimations of their actual causal therapy-effects, in contrast to pre-existing decision aids which presented participants with either hypothetical or population-based therapy-effects.”

3. Page 6; the sentence beginning on line 39 (Moreover,) has a typo and uses the wrong word (“illicit” should be “elicit”).

Author response:
Thank you, this has been corrected.

4. Methods: • I’m curious about the choice of study population. Since all patients had known CAD and were already on statins, they are not likely to have much decisional conflict related to statin use. Increasing the dose does not seem likely to lead to much decisional conflict in that context. I’m not sure if the all patients in the SMART study were already on statins, but if not I would have thought the statin-naïve patients would be a better study group for this sub-study.

Author response:
We much understand this question by the reviewer. However, studying statin-naïve patients with clinically manifest vascular disease would not be feasible as statins are usually prescribed per protocol already during hospital admission afor a first cardiovascular event. Therefore, shared decision-making about continued statin use takes place in the outpatient setting in patients already using standard dose statins. Options of choice are continuing current dose, up titration to a higher dose or quitting statin use. In the present study, we aimed to study the effect of individual treatment effect predictions in a setting closely resembling this clinical practice.

Of course, based on the results of this study, we must conclude that the effect of providing personalized treatment effect predictions in this setting is small and that future research may perhaps better focus on decision-making about new add-on therapies.

We have added the following to the discussion:

“Statins are usually prescribed to patients with CVD during hospital admission for the first CVD event. Therefore, outpatient decision-making regarding statins in this population usually pertain to continuing or altering the current statin dose. In the present study, we aimed to examine setting closely resembling the outpatient practice. Therefore future studies could focus on populations with higher baseline decisional conflict such as patients experiencing adverse effects or considering more intensive preventive treatment options on top of standard treatment such as intensive blood pressure reduction or combination antithrombotic treatment.”

We have added the following to the conclusion section:

“Future studies may better focus on decisions associated with higher decisional conflict such as the addition of more intensive preventive treatment options on top of standard treatment including a statin.”

5. The authors note that the study was hypothesis blinded, which seems to mean that patients were unaware of the hypothesis. Were their treating physicians aware of the hypothesis? Since patients were encouraged to discuss statin therapy with their doctor, and since there were physician reported secondary options, blinding of the doctors seems important. Please clarify.

Author response:
Physicians were told that the study involved providing patients with information regarding the effect of statins and were interviewed after having seen a patient’s personal health profile. Physicians were also blinded to the study outcomes and treatment arm differences. Physicians were approached only once, namely after their first patient (but not after consecutive patients) had been included to avoid inadvertently unblinding the physician to the treatment arm differences. We have clarified this by adding the following to the methods:

“General practitioners (GPs) received a copy of the personalized health profile received by their patients. Upon enrolment of the first patient from their practice, GPs were provided a short telephonic explanation of the study and asked to fill in a questionnaire (supplement 6). Questionnaire results and the last known post-intervention LDL-value at 6 months were secondary outcomes. Interviewed general practioners (GP’s) were blinded to the study outcomes and treatment arm differences. GPs were interviewed and questioned after being sent the intervention material of their first included patient. GP’s were not approached if they had subsequent patients included in the study, as receiving material from multiple patients would have unblinded them to treatment arm differences.”

6. • It seems that patients enrolled in SMART do not see a clinician, based on the text on page 8. Is this correct? It seemed unlikely to me, so wanted to make sure I am not misunderstanding.

Author response:
Information provided as part of the SMART-INFORM study was sent by mail. Additionally, patients were contacted by the researcher by telephone calls. As part of the intervention, participants were encouraged to discuss the content of the individual treatment effect predictions with their treating physician, in most cases their general practitioner. Participants were free to decide whether they wanted to follow-up on this advice.

We have modified the following section of the methods as follows:
"Within the SMART-study, patients are encouraged to visit a treating physician to discuss the results and decide whether or not to change their statin prescription. Participants were free to decide whether they wanted to follow-up on this advice or not."

7. Results: • In the first paragraph, the authors report the percentages of patients in each group that increased their statin doses. Were all participants recommended to increase the dose? Do we know the number of people who were recommended to increase or decrease their dose compared? This information may not be relevant and I would consider removing.

Author response:
The intervention itself did not recommend increasing or decreasing the dose of the statin. It simply provided information to the patients regarding the effect of statin medication. According to guideline-based practice, all patients with an LDL-c of >1.8 mmol/L would be recommended intensifying statin therapy. We have added the number of patients with an LDL-c of > 1.8 to the baseline table.

8. Do you know how many patients watched the video or discussed statins with their doctor? It would be interested in understanding "adherence" to the intervention.

Author response:
With the shared-decision making questionnaire in table 2 it is possible to see the number of patients who visited a physician specifically for the purpose of discussing statin use. It was recommended that patients visit a physician. However, the intervention itself did not involve a physician visit, but rather that the patient received a structured telephone consultation to ensure the information regarding personal therapy effects was well-received and understood by the patients. All intervention patients received this phone call.

9. At the bottom of page 12, the authors report that physician views about the value of statins did not differ across study arms. How many patients did each physician have in the study? Did they generally have a single patient, or could a physician have cared for several study patients who were allocated to different study arms. Were physicians asked about statin use for a particular patient or more generally?

Author response:
Physicians were only approached after the first patient had been included to avoid inadvertently unblinding the physician to the treatment arm differences. In response to point (5) this specification has already been added to the manuscript. Supplement 6 lists the questions asked the general practitioner. We have adjusted the legend in the table to be more clear on the interpretation of this outcome. We have also altered the paragraph about physician reported secondary outcomes to as follows:

"General practitioners (GPs) received a copy of the personalized health profile received by their patients. Upon enrolment of the first patient from their practice, GPs were provided a short telephonic explanation of the study and asked to fill in a questionnaire (supplement 6). Questionnaire results and the last known post-intervention LDL-value at 6 months were secondary outcomes. Interviewed general practioners (GP's) were blinded to the study outcomes and treatment arm differences. GPs were interviewed and questioned after being sent the intervention material of their first included patient. GP's were not approached if they had subsequent patients included in the study, as receiving material from multiple patients would have unblended them to treatment arm differences. "

10. Discussion: • In the first paragraph (lines 45-50), the authors state that their results indicate that providing estimated effects "positively impacts patient's opinions on taking these

medications". I don't think the findings support this statement- patients in the intervention group had lower decisional conflict, but that decision may have been to continue a lower dose when a higher dose was recommended. I think the statement is particularly misleading since there were no patients in the study who were NOT already on statins.

Author response:
We have nuanced this paragraph as follows:
"The actual benefit from CVD-prevention is smaller than people initially report acceptable. Still, openly communicating the individual estimates of statin use on the prevention of CVD resulted in lower decisional conflict, without many people discontinuing their treatments."

11. On page 15, the authors state that the algorithm was published after randomization of the last patient- not sure why this is noted, but certainly part of the study took place after the randomization. Was it published after everyone had reported results for the primary outcome?

Author response
The last primary endpoint was reported in September. We had added this sentence to show that the use of personalized treatment effects at the moment the study was conducted was not yet part of standard practice, and that cross-over from the control group was unlikely. However, we agree that this detail may be too much for the manuscript, and have removed the sentence from the manuscript.

12. • The authors make the point on page 15 about hypothesis blinding. Again I think they need to be clearer about who was blinded to the hypothesis.

Author response:
We have addressed this in response to point (5).

13. Under limitations, they discuss the fact that all patients were already on statins. Findings in a group treated for primary prevention might also be very different, which is important to note. It might also be quite different in real practice compared to with patients who are already enrolled in a clinical trial.

We have addressed this in the following paragraphs:

"The effects described here may thus be different in patients who experience adverse effects during statin treatment or consider starting more intensive preventive treatment options on top of standard statin treatment, e.g. intensive blood pressure reduction or antithrombotic treatment. Moreover, the personalized effects were not used directly during a clinical consultation, but provided prior to any potential consultation with a physician, and the effects may be different compared to a population of patients who are involved in a clinical consultation in which statin therapy is discussed."

And:

"Because the present study showed only a small effect in patients with a low baseline DCS, future studies could focus on populations with higher baseline decisional conflict. For example, patients experiencing adverse effects during statin use or patients considering more intensive preventive treatment options on top of standard treatment, for example intensive blood pressure reduction or combination antithrombotic treatment."

14. Tables and Figures: Table 1- there are important apparent differences (rates of diabetes, smoking). Were these statistically significant?

Author response:
We did not test statistical significance of any differences between characteristics presented in table 1. Such statistical tests would investigate whether these differences are likely due to chance or not. However, because this was a randomized trial, any differences between groups are by definition due to chance. Thus, we consider statistical tests in this setting to be uninformative and would introduce issues with multiple testing.

15. Figure 1- there appears to be an error- the number 384 appears twice

Author response:
This has been corrected. Thank you.

16. Figure 2- these look quite similar to me. I would recommend considering a review by a statistician to see if the correct tests were done, since it does not look like the differences should be so highly significant.

17. Consort guidelines:

10) Funding (Item 25) Funding source is included but role of the funder is not described.
Author response:
We have changed this to as follows:
“Partially funded by a Netherlands Heart Foundation grant (2016T026). Funder was not involved in the design or assessment of the study.

Reviewer: 3
Dr. Angela Wood, University of Cambridge
Comments to the Author:

1. The term “personalised treatment-effect” is misleading. I initially assumed the authors were incorporating “treatment effect heterogeneity” (eg, the statin effect differs by age, ethnicity or time since initiation etc…). However, a fixed treatment effect was used for all individuals (ie, HR of 0.80) and the effect of this was assessed on the “personal health profile” including 10-year absolute CVD risk or CVD free life-expectancy. I suggest the text is amended to avoid this misinterpretation.

Author response:
The reviewer rightly points out that there are two types of treatment effect heterogeneity: 1) differences on a relative effect scale, and 2) differences on an absolute effect scale. As far as statins concerned, subgroup analyses in meta-analyses of randomized trials give no reason to suspect that relevant differences on a relative effect scale exist. Therefore, personalized treatment effect estimates provided in this study (based on the previously published SMART-REACH model) are only different on an absolute effect scale. Notably, this study aims to investigate the effect of providing such treatment effect estimates to patients. The methodological background of those predictions may be of lesser relevance to this particular question.

We have added the following to the methods section:
“The model combines hazard ratio’s derived from meta-analyses with a prediction algorithm incorporating individual patient characteristics to derive the personalized therapy effects. A 1 mmol/L

reduction in LDL-c was modelled to correspond to the CVD-specific hazard ratio of 0.80 (5) and the expected LDL-c -reduction from baseline for each statin was derived from a previous meta-analysis.(6) As far as statins concerned, subgroup analyses in meta-analyses of RCT's provide no evidence for relevant differences on a relative effect scale. Therefore, personalized treatment effect estimates based on the SMART-REACH score are only different on an absolute effect scale."

2. Kruskal-Wallace should be "Kruskal-Wallis".
Author response:
This has been corrected

3. Figure 1 has several typos:
(a) "Approached patients screened for eligibility": should be n=303 not n=384
(b) Numbers under the CVD-free life years intervention do not add up.

Author response:
We have corrected this, thank you.

4. Can the authors provide reasons why patients did not complete the primary endpoint at one-month (the original protocol suggests this was collected)?

Author response:
To limit loss to follow-up, patients not initially responding to the follow-up questionnaire were approached telephonically. A reminder was sent by mail if the patient was not reached. Unfortunately we do not have data on why a patient failed to respond. We have added the following to the paper:

"To limit loss to follow-up, patients not initially responding to the follow-up questionnaire were approached telephonically. A reminder was sent by mail if the patient was not reached."

5. More description (as supplementary tables or figures) is required to fully understand the nature of the missing outcome data (eg, % missing by trial arm, or table 1 presented for patients with observed outcome). Patients with low literacy were more likely to have missing outcomes – did this vary by trial arm? How does this finding implicate the significant subgroup finding? Are the % of patients with missing outcome significantly differently between arms? Further sensitivity analyses may be required to assess the impact of missing outcomes on the trial results (eg, (i) an analysis adjusting for baseline variables which are no longer randomly distributed between trial arms amongst patients with observed outcome (ii) multiple imputation approach with a delta-adjustment to impute under a missing not at random assumption).

Author response: We thank the reviewer for these suggestions, which have helped us improve the quality of the manuscript. In the supplement, we have changed supplemental table 1 to show the missing outcomes by trial arm. The table is shown below. Patients with low literacy were indeed shown to have more missing outcomes across all trial arms. We have changed supplemental table 1 to the following:

Supplement table 1: Baseline characteristics per missing and non-missing for primary outcome

	Control		iPOL		iARR	
	Non-missing	Missing	Non-missing	Missing	Non-missing	Missing

Population	n=90	n=11	N=87	N=14	N=83	N = 18
Age	64 (59 - 72)	62 (60 - 67)	66 (59-71)	68 (59-73)	66 (59-72)	63 (59 – 68)
Gender (male)	77 (85%)	9 (82%)	72 (83%)	11 (79%)	71 (86%)	16 (89%)
One CVD location	79 (88%)	10 (91%)	77 (89%)	14 (100%)	75 (90%)	15 (83%)
Years clinically manifest CVD	5 (0 - 11)	0 (0 - 12)	5 (0-10)	0 (0-2)	0 (0-10)	6 (0-12)
Diabetes Mellitus	19 (21%)	4 (36%)	24 (28%)	3 (21%)	13 (16%)	1 (6%)
LDL-c (mmol/L)	2.0 (1.6 - 2.5)	1.8 (1.6 – 2.4)	2.1 (1.7 – 2.4)	1.7 (1.6 – 2.1)	2.0 (1.6 – 2.4)	2.0 (1.7 – 2.3)
LDL-c >1.8mmol/L	56 (62%)	5 (50%)	62 (71%)	6 (43%)	55 (66%)	11 (61%)
Creatinine (umol/L)	85 (75 - 92)	85 (74 - 97)	82 (74 – 96)	91 (80 – 106)	83 (78 – 95)	85 (80 – 90)
Systolic blood pressure (mmHg)	129 (122 - 142)	132 (116 - 150)	132 (121-145)	129 (120-132)	130 (121 - 140)	140 (124-148)
Number of medications per day	5 (4 - 7)	7 (5 - 10)	6 (4-9)	7 (6-8)	5 (4 - 7)	5 (4 - 6)
Months required to offset disutility of daily pill-taking	42 (9 - 97)	97 (35 - 97)	61 (3 – 97)	97 (70 – 97)	61 (9-97)	12 (9-61)
High likelihood limited literacy	6 (7%)	2 (18%)	6 (7%)	2 (14%)	5 (4%)	4 (22%)
Possibility of limited literacy	9 (10%)	2 (18%)	7 (8%)	2 (14%)	7 (9%)	1 (56%)
Adequate literacy	75 (83%)	7 (63%)	74 (85%)	10 (71%)	70 (85%)	13 (72%)

As a sensitivity analysis we have done as suggested and rerun the analyses while corrected for baseline characteristics gender, age, smoking status, diabetes status, LDL-cholesterol (mmol.L), creatinine (umol/L), disutility score, NVS health literacy, and number of medications used per day. After correction none of the outcomes were still significant. We have added the following to the methods:

“Sensitivity analyses A sensitivity analyses was performed to account for baseline characteristics which may differ in missing outcomes between trial arms by conducting an ANCOVA with gender, age, smoking status, diabetes status, LDL-cholesterol (mmol/L), creatinine (umol/L), disutility score, NVS health literacy, and number of medications used per day.

And the following to the results:

“Sensitivity analyses
“Supplemental table 4 shows the sensitivity analyses which corrected for baseline characteristics. After correction none of the outcomes were significant.”

Supplemental table 4: Sensitivity analyses: Patient Reported Secondary Outcomes corrected for baseline characteristics.

	ANCOVA
	p-value
DCS (1)	p=0.49
DCS (6)	p=0.50
Brief-IPQ (1)	p=0.21
Brief-IPQ (6)	p=0.63
PAM (1)	p=0.39
PAM (6)	p=0.29
Perceived Statin Efficacy (1)	p=0.67
Perceived Statin Efficacy (6)	p=0.44
Understanding of therapy-effects (1)	p=0.16
Understanding of therapy-effects (6)	p=0.60
BMQ Adherence Risk Scale (1)	p=0.382
BMQ Adherence Risk Scale (6)	p=0.32
SDMQ9 (1); Reported visiting GP (n)	p=0.21
SDMQ9 (6); reported visiting GP (n)	p=0.35
RAND-36 Quality of life (6)	
Physical functioning	p=0.12
Role limitations due to physical health	p=0.48
Role limitations due to emotional problems	p=0.69
Energy/fatigue	p=0.86
Emotional well-being	p=0.25
Social Functioning	P = 0.71

Pain	p = 0.53
General health	P = 0.86

Legend : Sensitivity analyses for all outcomes at one (1) and six (6) months post intervention corrected for baseline characteristics: gender, age, smoking status, diabetes status, LDL-cholesterol (mmol.L), creatinine (umol/L), disutility score, NVS health literacy, and number of medications used per day

6. Figure 2: typo – “Post-hoc”
This will be adjusted to post hoc.

7. Page 15, line 51 – remove extra fullstop
Author response:
This has been removed.

8. Page 62, lines 46-47 (Section 6.5.1 Overview of procedures): Under “additional measurements” no “X” marked for “measurements obtained from SMART work-up

Author response:
There is indeed an “X” missing for this section of the protocol. We will add this X to the submitted protocol for the revision.

Reviewer: 4

Major comments:

1. It is not clear whether the authors have followed the CONSORT guidelines for reporting this RCT. They have added in the supplementary material a table listing items to include in an abstract which I do not think it is relevant here. What is needed is the completed CONSORT checklist as required by the journal. I strongly recommend to follow the guidelines to make sure all items/sections are properly described, as also noted in some of my other comments. Also, the SPIRIT guidelines should have been followed for the study protocol but since the trial is now completed, I will just assume they have.

Author response: We have added the consort checklist.

2. page 7 of 73 - Design, randomization, and follow-up:
1. "hypothesis blinded" is not a commonly used term. I understood why the authors used this term and I think it is explained well in the protocol. I suggest they also explain this rationale in the manuscript and, most importantly, they need to clarify who was blinded (if anyone) and how (see CONSORT).

Author response:
We have amended the methods to the following:
“The SMART-Inform study was a three-armed, hypothesis-blinded, RCT. Hypothesis blinding means patients and their general practitioners were informed that everyone would receive at least standard SMART-protocol practice and that the study goal was to investigate if information about cholesterol-lowering medications would impact motivation for use, but were unaware whether the content they

received was part of the active or control arms and were unaware what the primary and secondary outcomes were. Researchers and outcome assessors were not blinded."

2. Randomisation: I also think this is better written in the protocol. R is a software not an algorithm. It can simply say computer-generated random allocation sequence. No mention of how allocation concealment was achieved (see CONSORT for all details required).

We have altered the paragraph to the following:
"A computer generated random allocation sequence was used to assign each patient after inclusion, by order of inclusion. The investigator generating the random sequence was not involved other aspects of the study. Investigators enrolling patients and assigning participants to interventions had no access to the sequence."

2. page 8 of 73 - Intervention arms: when reading this section, it was not very clear to me how the interventions differed; again, I understood it better after reading the protocol (page 18). It does not need a lot of details but simply a clearer and more structured definition.

Author response: We have reviewed the wording in the protocol and in the manuscript. We hope it is more clear:

"There were three intervention arms: the control group, the 10-year risk (iAR-group) and CVD-free life-expectancy (iLE-group). The control group received only standard practice. Both intervention arms received standard practice plus: 1) a leaflet entitled personal health profile (supplement 1B and 1C show examples for two fictional patients); 2) a USB-device containing educational videos; 3) a structured telephone consultation enforcing uptake of the information (supplement 2). This 'personal health profile' outlined the individual effect of the following treatment options: 1) continue with the type and dose of statin-therapy ('current prognosis'); 2) discontinue statin therapy ('stop statins'); 3) intensify to maximum statin-therapy, defined as once-daily atorvastatin 80 mg ('increase statins'). The only difference between the intervention arms differed was the measure used to communicate the prognostic change associated with the therapy-effects. Individual treatment effects were estimated in terms of change in 10-year risk (iAR-group) or CVD-free life-expectancy (iLE-group). The USB-device contained intervention-group specific educational videos on how to read and interpret the 'personal health profile' and the effect of statin-medications on CVD. The structured telephone consultation for the intervention arms ensured the information was well-received and understood by the patients."

- page 10 of 73 - Statistical analysis: This section can be expanded and should be described in a clearer way (see CONSORT).

1. I would personally separate the sample size calculation in a subsection but, even if not, it should be distinct and described first (here the intention-to-treat is first mentioned, then sample size calculation, then analysis again). It's a three arm trial, among which arms was the 8.5 difference to be detected? This difference need to be justified somewhow and the references in the relevant section of the protocol are not the same as the one cited here. Also, where is 8.5 in the calculation shown in the protocol? It is a bit confusing, it has to be clear so that it is reproducible. The standard deviation used also need to be included.

Author response:

We have clarified the following in the manuscript:

"We used a T-test for the sample size as the primary analysis of this trial compared interventions with the control group. Calculations were conducted using G*Power version 3.1. Sample size was

based on an effect size (Cohen's d = mean difference / standard deviation) of 0.43. A standard deviation of 0.80 was used to detect a mean difference of 0.34 on the 5-point scale (ranging from 0-4) which was taken to correspond to 8.6 on the 100-point scale.(7, 8) A power of 80% and a two-tailed alpha of 0.05 was used. A minimum of 86 patients per arm was needed."

2. The statistical analysis methods are not clear and, given the lack of statistical analysis plan as a supplement, I was not sure how the analysis was actually carried out. First, the effect size and precision (how the results are reported e.g. OR, RR, mean difference, 95% CI, etc.) need to be specified for all outcomes (i.e. both continous and categorical variables). I saw in the results table the test statistic is reported, which is meaningless for the readers as it does not say anything about the effects observed between groups, so I recommend removing.

I am unsure about the "direct consecutive comparisons" terminology, "pairwise comparisons" is the term normally used. ANOVA tests if all group means are equal and, if so, then one can compare all pair of means but has to account for multiple testing and it is unclear whether this was done (and accounted in the sample size calculation) and I did not find anything even in the protocol.

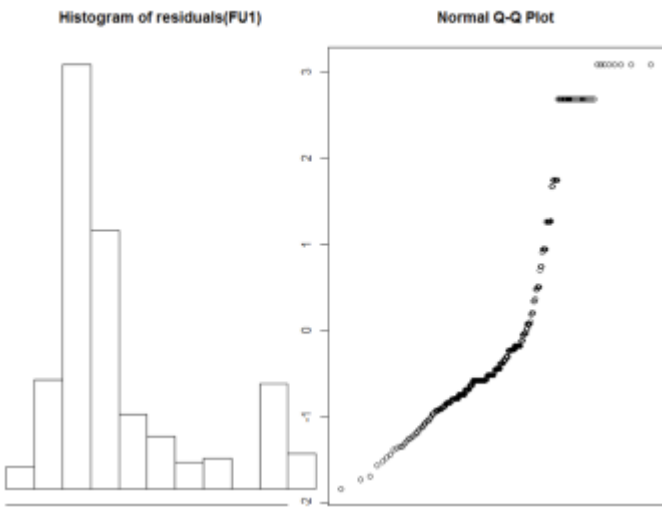
Since medians are reported instead of means, I suppose it is because of non-normality. However, this is not addressed in the methods. Also, if the assumptions are not met, one could try transformations (e.g. log-normal) before moving on to use non-parametric tests. Also, the Wilcoxon-rank sum test does not test the difference in medians between groups, there are other available methods for that. I suggest discussing with the study statistician the possible options. Also, I would "reorder" the use of tests in the text: ANOVA then t-test (not t-testing; correct throughout the manuscript); if assumptions are not met, Kruskall-Wallis then Wilcoxon-rank.

Author response:

We thank the reviewer for the being thorough with how we can improve this section of the manuscript. The statistical analysis plan was cemented in the protocol approved by the medical ethics committee, which we cannot change in retrospect. We have kept the statistical analysis as simple as possible, and believe that an additional supplement repeating the analyses, next to the information in the protocol ad manuscript- would lead to confusion. We hope that with the ammendments made in this revision that it has become more clear.

We did indeed first try to transform the variables. However, the assumptions for the parametric testing were not met. Below is a run of a log-transformed DCS at one month, which shows non-normal residuals. We have also altered the manuscript to be more clear on this part.

"Assumptions of normal (residual) distribution and homoscedasticity were visually inspected. If ANOVA $p<0.05$, pairwise comparisons between arms were determined using a t-test or with the Wilcoxon-rank sum test for the difference in ranked means if ANOVA assumptions were not met after transformation attempts.



We have also changed the wording to remove “t-testing” and reorderd the use of the tests in the text. The statistical test column was reported to show for which outcomes a parametric (i.e. F-value) and non-parametric (i.e. chi-squared) test was applied. We have altered this to provide the information in the footnote of the table without providing the test column in the table. We have done this with table 2 in the main manuscript as well.

We did not plan to account for multiple testing in the protocol. However, as suggested we have applied the (strict) Bonferroni correction to the results here. To protect against a type I error, a Bonferroni correction corresponding to the 22 secondary outcomes was applied (i.e. 23 comparisons). The new p-value was 0.05 / 22 = 0.002 to determine if any of the values were statistically significant. We have added the following to the manuscript:

“A Bonferroni correction corresponding to the 22 secondary outcomes was applied. The new p-value for statistical significance was 0.002.”

3. Good to see the subgroup analyses were pre-specified but I would also acknowledge somewhere (methods or results section) that the study was not powered to detect any difference between such subgroups.
Author response:
We have added the following:
“The study was not powered to detect any differences between subgroups.”

4. Not clear in the Author contributor statement who performed the data analysis.
Author response: We have added the following:
“NEMJ drafted the work and performed the analyses,”

Minor comments:
5. I did not comment on the results and discussion section as I would like to see the results reported differently (i.e. I do not have a clear idea of the effect and its precision between groups) but otherwise seemed to be presented clearly.

Author response: We hope with all the revisions the document has become more clear.

6. In the Abstracts I do not understand why the changes in CVD risk and CVD-free life-expectancy are mentioned as the first thing in the Results given that they are not even secondary outcomes. The

number randomised and analysed should be mentioned first then the result of the primary outcome (see checklist supplement).
Author response:
We have removed the changes in CVD risk and CVD life expectancy from the abstract. These results show that actual personalized effects were smaller than what the majority of patients view as acceptable. Despite this, providing personalized effects do not negatively affect the outcome. We agree however, that these results are a nuance, and that it has a place in the main text, but not in the abstract. It has been removed from the abstract.

7. Registration: there is a new number on the registry. However, I searched it with this one and it works. I am just not sure which one (old or new) should be included in the article.
Author response:
We were unaware the trial had been given two numbers from the NTR. We have added both to be complete.

References:

1. Hibbard JH, Mahoney ER, Stockard J, Tusler M. Development and testing of a short form of the patient activation measure. Health Serv Res. 2005;40(6 Pt 1):1918-30.
2. Fransen MP, Leenaars KE, Rowlands G, Weiss BD, Maat HP, Essink-Bot ML. International application of health literacy measures: adaptation and validation of the newest vital sign in The Netherlands. Patient Educ Couns. 2014;97(3):403-9.
3. Baker DW, Parker, R.M., Williams, M.V., Clark, W.S., Nurss, J. The Relationship of Patient Reading Ability to Self-Reported Health and Use of Health Services. American Public Health Association 87(6):1027-30.
4. Stacey D, Legare F, Lewis K, Barry MJ, Bennett CL, Eden KB, et al. Decision aids for people facing health treatment or screening decisions. Cochrane Database Syst Rev. 2017;4:CD001431.
5. Cholesterol Treatment Trialists C, Baigent C, Blackwell L, Emberson J, Holland LE, Reith C, et al. Efficacy and safety of more intensive lowering of LDL cholesterol: a meta-analysis of data from 170,000 participants in 26 randomised trials. Lancet. 2010;376(9753):1670-81.
6. Law MR, Wald NJ, Rudnicka AR. Quantifying effect of statins on low density lipoprotein cholesterol, ischaemic heart disease, and stroke: systematic review and meta-analysis. BMJ. 2003;326(7404):1423.
7. Schwalm J, Stacey, D., Pericak, D., Natarajan, M. Radial Artery Versus Femoral Artery Access Options in Coronary Angiogram Procedures. Circulation: Cardiovascular Quality and Outcomes. 2012;5:260-6.
8. O'Connor AM. Validation of a Decisional Conflict Scale. Med Decis Making. 1995;15(25).

VERSION 2 – REVIEW

REVIEWER	Castro, Inar Universidade de Sao Paulo Faculdade de Medicina
REVIEW RETURNED	12-Apr-2021

GENERAL COMMENTS	The authors answered the most part of my previous comments and revised the manuscript accordingly. However, the English relative to the new paragraphs included in the revised manuscript must also be improved. I think that the statistical methods applied to determine the sample size and also to treat data could be submitted to a statistician. I

	understand that if you are comparing two interventions with the same control, it means you are comparing three and not two groups. This part was still not clear to me and maybe can be better explained.

REVIEWER	Korenstein, Deborah Memorial Sloan-Kettering Cancer Center, Department of Medicine
REVIEW RETURNED	05-Apr-2021

GENERAL COMMENTS	Thank you for responding to my comments and those of the other reviewers.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 1
Dr. Inar Castro, Universidade de Sao Paulo Faculdade de Medicina
Comments to the Author:
The authors answered the most part of my previous comments and revised the manuscript accordingly. However, the English relative to the new paragraphs included in the revised manuscript must also be improved.
I think that the statistical methods applied to determine the sample size and also to treat data could be submitted to a statistician. I understand that if you are comparing two interventions with the same control, it means you are comparing three and not two groups. This part was still not clear to me and maybe can be better explained.

Author Response:
Thank you for your response. We have had a native speaker read and correct the manuscript to make the sections more clear and reproducible.
Because the aim of this study was a pairwise comparisons between study arms, sample size was calculated based on a t-test comparison. To limit the overall probability of type 1 errors to 0.05, first an ANOVA was used to detect whether there was any difference between the three groups at all before subsequent T-test were performed.
We have altered the paragraph to as follows, and hope it is more clear.
“The aim of this study was a pairwise comparison between study arms. To limit the overall probability of type 1 errors to 0.05, first an ANOVA was used to detect the presence of any differences between the three groups. If the ANOVA detected a difference, a subsequent T-test were performed. Therefore, the sample-size was calculated to detect a difference in two groups using the T-test. Calculations were conducted using G*Power version 3.1.”