PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Cost-effectiveness of multicomponent interventions in type 2 diabetes mellitus in a cluster randomized controlled trial: the INDICA Study
AUTHORS	García-Pérez, Lidia; Ramallo Farina, Yolanda; Vallejo Torres, Laura; Rodríguez-Rodríguez, Leticia; Gonzalez-Pacheco, Himar; Santos-Hernández, Beatriz; García-Bello, Miguel Angel; Wägner, Ana Maria; Carmona, Montserrat; Serrano-Aguilar, Pedro

VERSION 1 – REVIEW

REVIEWER	Li, Jinshuo University of York, Department of Health Sciences
REVIEW RETURNED	22-Oct-2021

GENERAL COMMENTS	General comments This is an economic evaluation alongside a trial with a large sample size and long follow-up period. The data sources included electronic medical records, which improved the accuracy of the results. However, the methods need more details regarding what and how analyses were performed, while the introduction of the trial could be simplified. I am slightly concerned with the post-hoc subgroup analysis which left an impression of reporting bias. Although it could be done, I would suggest not to draw conclusion from it or stress on it too much. Another general issue is that statistical significance has been emphasized on a lot in this article. I would suggest the authors re-consider its use in a cost-effectiveness analysis. The effect size for the significance is not clear for either QALYs or costs. Even if one could do reverse calculation based on sample size and difference, that effect size might not be clinically or practically meaningful. Specific comments Abstract 1. The perspective of the CEA should be stated in abstract. 2. I suspect the term ‘multiple imputation’ is misused here. By protocol, the missing data were handled with multiple methods of imputation (mean, LVCF), which is different from multiple imputation. The term ‘multiple imputation’ is itself a method for dealing with missing data. If there was no deviation from protocol, please do not use ‘multiple imputation’ as it might cause confusion. 3. Results part needs uncertainty measures. 4. As mentioned in general comments above. It may not be appropriate to put subgroup analysis here, especially in conclusions.

	Introduction 5. The first sentence of the second paragraph needs clarification: prevalence is a point estimate, how could it be 'over 15 years'? Or did it mean 'average prevalence over the last 15 years'? And why would prevalence show a higher mortality and a higher incidence? Methods 6. Trial design: QALYs is the outcome measure of effectiveness, not the main outcome of CEA. The main outcome of CEA should be ICER. 7. Interventions: Please briefly describe what 'the standard practice' consists of. 8. Setting, recruitment and randomization: PHCP, FCU and patients were all randomly selected. Please clarify how was the random selection achieved? 9. Healthcare utilization and costs: As the content of usual care was not described, it is not clear if the costs of usual care were included. 10. P10 Line 5: AUC approach needs to cite a reference. 11. Sample size calculation: I suggest remove this part. Simply state the number of patients per arm in trial design would be sufficient. 12. P10 Line 21: By protocol, 'multilevel models' might be more appropriate than 'mixed models'. 13. P10 Line 24-26: By protocol, the covariates were selected after assessing their significance, which needs to be briefly explained if that were the case. 14. P10 Line31: As point 2 above, please reconsider the use of the term 'multiple imputation'. If it means 'multiple methods of imputation', please list the methods used and which data they applied to. If it does mean the method multiple imputation, then it needs to be explained why it deviated from protocol and the details of the imputation model. In any case, it may be necessary to have a sensitivity analysis to assess the impact of missing data imputation. 15. P10 Line 33-36: Uncertainty surrounding the point estimate of ICER needs to be examined and presented. 16. P10 Line 36-38: What parameters were re-estimated? How were they selected? How alternative values were selected for each of them and what values were they? Were the alternative values used together or the values were replaced one at a time? Results 17. Before reporting the results, please briefly describe the follow-up rates and if they were different by arms, and the missing data patterns. 18. Table 1: please specify statistical test used and add P-values as it is stated in the text that none of these were statistically significantly different. 19. When stating statistical significance, associated numbers should be included in the text, even if they were presented in tables. 20. Where mean values were stated, uncertainty measures should also be included in the text. 21. Table 2: please specify which statistical test was used. 22. Cost-effectiveness analysis base case: The uncertainty surrounding the ICERs should be examined and presented using CEACs. 23. Cost-effectiveness analysis sensitivity analysis: As mentioned in point 16, it is not clear how the sensitivity analysis was performed. And the results of sensitivity analysis should at least be

	included in Appendix. It is not sufficient to simply say they were similar without data presented. Discussion 24. The strengths are more of the trial as a whole than of the current analysis. As this is the discussion of cost-effectiveness analysis, I suggest the discussion focusing more on the economic evaluation side. 25. It also needs comparison with similar studies.

REVIEWER	Oyagüez, Itziar Pharmacoeconomics & Outcomes Research Iberia
REVIEW RETURNED	04-Nov-2021

GENERAL COMMENTS	It is a very interesting work. The paper is well written and easily understable for readers. I was quite surprised by the results, so as it is mentioned in the discussion, I would have expected better outcomes for CBI, and PFI than for UC. My main concern is regarding the result shown about ICER. It seems it is calculated for PTI vs UC, but according the authors themselves in the methods section (page 10 of 49, line 32 to 35), "Incremental cost-effectiveness ratio (ICER), that is, the differences between costs divided by the differences in QALYs, were calculated when one alternative was more effective and more costly than another." This part and results sections (page 12 of 49) should be aligned to be coherent. After doing quick calculations, ICER would result €17896/QALY with UC vs PTI. I guess this is not the expected outcome to report but that's it. Minor comments: - In suggest to avoid the term more cost-effective in the conclusion (page 15 of 49, line 5) Methodologically I think an intervention is cost-effective or not cost-effective, but not more cost-effective. ICER could be lower or higher but it doesn't mean "more cost-effectiveness"

REVIEWER	Pollard, Daniel University of Sheffield, School of Health and Related Research
REVIEW RETURNED	12-Nov-2021

GENERAL COMMENTS	The paper presents an interesting analysis of information and communication technologies to support self-management of type 2 diabetes in the Canary Islands in Spain. The exact statistical model fitted should be specified and a reference to the method should be provided. Mixed models are an inadequate description. Furthermore, it is unclear why the control variates were chosen as they are different for costs and QALYs. Why are QALYs not adjusted for age and sex. Why are costs not adjusted for time since diagnosis? The multiple imputation should be specified, with details on the imputation model, imputation covariates and the number of imputation iterations. Some diagnositics should be shown to

	demonstrate that enough imputation iterations have been conducted. Sensitivity analyses are conducted; however it is unclear what sensitivity analyses have been conducted as they are not defined the main text. There is extensive use of p values to present results. It would be clearer to a reader to present mean costs and 95% confidence intervals. The threshold ICER for decisions makers in the Canary Islands is not mentioned. Therefore, it is unclear whether the conclusions are supported or not. The results that some of the interventions produce fewer QALYs than the comparator arm. In diabetes, HbA1c itself has at best a minimal effect on utility (see Beaudet et al for systematic review of utility values for health states in diabetes simulation modelling). The big drivers of changes in utility are diabetes related complications. HBA1c itself, does influence the progression to these health states. However, as the trial was only powered on HbA1c and not the incidence of long term complications I would not expect that the intervention would produce more QALYs than a control arm. The intervention costings precented in the appendix are very unclear: Why have development / reviewing costs (for materials and computed systems) been included? Surely these are one-off costs in the intervention development Why have the full cost of the laptops been included in the costs? Surely the costs of laptops should be spread out over multiple years if they were to be used in routine practice. Why is the website 10 times cheaper in the both column of Table A2 compared to the Patients column? Table A3 appears to show that the numbers of: Visits to a general practitioner (PTI arm and PFI arm) Visits to a nurse at primary care (PTI arm and PFI arm) Lab test by general practitioner (PTI arm) Are the differences between multiple different arms (e.g. PTI v UC and PTI v CBI and PFI v PTI) of the study. How? Furthermore, it is stated that in a footnote that there was a negative binomial regression model, adjusted by time since diagnosis and baseline resource use. It would be useful to specify exactly what the negative binomial models were predicting. For example, are they used for each subcategory of resource use or were they only used for some categories. The baseline resource use variables in these regressions should be more clearly defined, as the specific varaibles that were controled for. It is unclear whether all baseline resource use was controlled for or only a particular category of resoruce use (e.g. Hospital Stays). References

	Beaudet A, Clegg J, Thuresson PO, Lloyd A, McEwan P. Review of utility values for economic modeling in type 2 diabetes. Value in Health. 2014 Jun 1;17(4):462-70.

REVIEWER	Sarmiento, Samuel Johns Hopkins University School of Medicine, Plastic and Reconstructive Surgery
REVIEW RETURNED	20-Nov-2021

GENERAL COMMENTS	Cost-effectiveness of multicomponent interventions in type 2 diabetes mellitus in a cluster randomized controlled trial: the INDICA Study General **1.** I like that you designed your trial with a cost-effectiveness (CE) analysis in mind, as well as a potential qualitative study in the future. I think this makes your results easier to interpret, even for those without a background in health economics. **2.** Please proofread the document again and correct minor grammatical errors throughout the text. I'll point out some of them in the appropriate sections. Attention to such details will give more credibility to your study. **3.** All references to lines and pages I make here are based on the proof you submitted to BMJ, not on the original manuscript. Abstract 4. Minor grammatical adjustments: • Line 7 & elsewhere: In primary care. • Line 11: Within-trial period. • Line 14 & elsewhere: In the Canary Islands. • Line 37: Aimed at professionals or usual care. • Line 46: Keywords: primary care. • Don't forget to leave a space before symbols such as "greater than" or "less than" (>, <) throughout the document. For instance, "p value <0.05; HbA1c >7%." Please be consistent with this and punctuation in general. Strengths & Limitations **5.** Line 10: From a healthcare perspective. **6.** Line 23: The main limitation is the relatively short duration of the trial, two years. Introduction **7.** Line 9: Comma after "…patients who have already developed T2DM". Methods **8.** Line 7: Controlled clinical trial. **9.** Line 13: Was the main outcome. **10.** Line 13: If quality-adjusted life years is in the plural, you can abbreviate as "QALYs." **11.** Patient involvement: • Line 38: Canary Islands. • Line 43: In the preparation of the protocol. • Line 48: "…interviews that will be the subject of another study."

12. Health care utilization and costs:

- Line 14: "…and medications."
- Line 29: Did you consider adjusting for inflation and reporting costs for the year you analyzed the data?

13. Outcomes:

- Line 53: It's been a while since I reviewed the EQ instrument, but can a health state really be worse than death? The answer is obviously no. Negative scores simply represent *decreases* in life quality, and not health states worse than death.

Results

14. CE analysis: base case:

- Line 15: No space after €.
- Line 17: This is the most important line in the document and you must elaborate on this in the discussion. In most health systems, an ICER of €30,000 can be considered costeffective. However, in this case, this number is affected by the intervention being cheaper than the standard of care, even though it was also slightly less effective.

However, in the subgroup analysis among those with an HbA1c >7%, this was clearly the dominant strategy. You have to drive this point home in the discussion.

Discussion

15. Line 27: "The differences are explained by the differences…" Rephrase to improve clarity.

16. Line 56: "Non-significant."

17. Page 14, line 3: Change "moreover" to "however." You are about to introduce a statement that contrasts with the previous one, that is, despite a lower effectiveness, PTI was dominant in patients with poorly controlled diabetes.

18. Page 14, line 9: Change to "patients with poorly controlled blood glucose levels."

19. Please create subheadings for "strengths" and "limitations." This will improve readability.

20. In the section "Strengths & Limitations," you present the duration of the trial as a limitation. Therefore, you cannot also include it as a strength in the discussion. This is either a strength or a limitation, but not both. In my opinion, given the long-term effect of certain interventions on quality of life, two years is a relatively short period of time to draw any definitive conclusions.

21. Page 14, line 27: "Potential efficacy," not efficiency.

22. Page 14, line 32: "Some degree of missing data."

23. Page 15, line 9: "With the highest needs…" "…its impact on the…"

Graphs & Tables

24. Table 2: Please add an asterisk or another symbol next to the statistically significant results so that readers can more easily find them or put them in bold.

25. Table 3: I don't have access to your raw numbers, but regarding the ICER between PTI and UC, going by the information in this table:

Arm	Cost	QALYs	ICER

	PTI	2571.48	1.71	
	UC	2750.44	1.72	
	Incremental	-178.96	-0.01	17896

Am I missing something?

REVIEWER	Chen, Yiwei Stanford University, Department of Economics
REVIEW RETURNED	21-Nov-2021

GENERAL COMMENTS	Overall, this is a well-written paper and a well-conducted study. The importance and interpretation of the finding however need to be strengthened. Background and Setting: 1. The authors mentioned that the usual care arm had an intensive trial follow-up that may explain the good health outcomes. Can you elaborate more on this? Do the other arms have this feature too or it's a unique feature for the usual care arm? Is the cost reflected? 2. What's the health care and patient profile setting for Canary Islands? Is it similar to Spain? Shall the readers interpret the results just for Canary Islands or for a broader region? Cost 1. The cost of INDICA interventions include a lot of fixed costs, e.g. material development time for patients; material development time for GP; development of computer system, etc. Such costs will dimmish very quickly when we include more patients. The current cost calculation is appropriate for the experiment population, but it's not appropriate if readers want to extend it to the whole population of the Canary Islands. I suspect this will change the cost calculation and is more appropriate for the goal of this paper. Statistical methods: 1. Where is the sensitivity analysis? I can't find it the in the paper. 2. Attrition rate isn't clear. How many participants did the study lose over the 24 period? What are the baseline stats for the population who remained at the end? Is the analysis result sensitive to imputation method?

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Ms. Jinshuo Li, University of York

General comments:
This is an economic evaluation alongside a trial with a large sample size and long follow-up period. The data sources included electronic medical records, which improved the accuracy of the results. However, the methods need more details regarding what and how analyses were performed, while the introduction of the trial could be simplified. I am slightly concerned with the post-hoc subgroup analysis which left an impression of reporting bias. Although it could be done, I would suggest not to draw conclusion from it or stress on it too much. Another general issue is that statistical significance

has been emphasized on a lot in this article. I would suggest the authors re-consider its use in a cost-effectiveness analysis. The effect size for the significance is not clear for either QALYs or costs. Even if one could do reverse calculation based on sample size and difference, that effect size might not be clinically or practically meaningful.

Specific comments
Abstract
R1.1 The perspective of the CEA should be stated in abstract.
Thank you. We have added the perspective of the analysis in the abstract.

R1.2. I suspect the term 'multiple imputation' is misused here. By protocol, the missing data were handled with multiple methods of imputation (mean, LVCF), which is different from multiple imputation. The term 'multiple imputation' is itself a method for dealing with missing data. If there was no deviation from protocol, please do not use 'multiple imputation' as it might cause confusion.

Thank you for your comment. The phrase in the abstract and Statistical methods section were confusing. To process the missing data the technique Multiple imputation using chained equations [1] was used. Both paragraphs have been modified to improve their wording.
This is actually a deviation from the protocol because at the time of design of the study we did not know that the Multiple imputation technique was better than routine methods. At the time of the analysis we believed it was correct to use the best imputation technique available for our data [2] and this led to the change in regard to the protocol. A phrase has been added in the limitations in which the change of imputation method with regard to the protocol has been recognized. Moreover, at the request of another reviewer an appendix that details the imputation process has been included (supplemental Appendix 2).
[1] White IR et al. Multiple imputation using chained equations: issues and guidance for practice. Stat Med 2011 Feb 20;30(4):377-399.
[2] Faria R, Gomes M, Epstein D, White IR. A guide to handling missing data in cost-effectiveness analysis conducted within randomised controlled trials. *Pharmacoeconomics*. 2014;32(12):1157-1170. doi:10.1007/s40273-014-0193-3

R1.3. Results part needs uncertainty measures.

The results section of the abstract has been extended and modified according to your and other reviewers' indications.

R1.4. As mentioned in general comments above. It may not be appropriate to put subgroup analysis here, especially in conclusions.

We have removed from the abstract conclusions the phrase making explicit reference to the analysis by subgroups

Introduction
R1.5. The first sentence of the second paragraph needs clarification: prevalence is a point estimate, how could it be 'over 15 years'? Or did it mean 'average prevalence over the last 15 years'? And why would prevalence show a higher mortality and a higher incidence?

We are grateful for your comment because we realize it is insufficiently clear in its wording. This paragraph in the introduction has been modified.

Methods

R1.6. Trial design: QALYs is the outcome measure of effectiveness, not the main outcome of CEA. The main outcome of CEA should be ICER.

We have changed the trial design section. We now state “incremental cost per quality-adjusted life years (QALYs) was the main outcome” instead of “quality-adjusted life years (QALYs) was the main outcome”. Moreover, we have changed the subheadings so there is no ‘Outcomes section’ now, and the ICER is mentioned in the ‘Analysis’ section.

R1.7. Interventions: Please briefly describe what ‘the standard practice’ consists of.
Thank you for your comment. The description of “the standard practice” group has been improved.

R1.8. Setting, recruitment and randomization: PHCP, FCU and patients were all randomly selected. Please clarify how was the random selection achieved?

Thank you for your comment. We have tried to summarize this section in the methods because they are published in detail in the clinical trial protocol [1]. Despite this we have improved the wording of the paragraph to give further details in this paper that facilitate understanding the study’s design.

[1] . Y. Ramallo-Fariña, L. García- Pérez et al. Effectiveness and cost-effectiveness of knowledge transfer and behavior modification interventions in type 2 diabetes mellitus patients—the INDICA study: a cluster randomized controlled trial, Implement. Sci. 10 (2015) 47. https://doi.org/10.1186/s13012-015-0233-1.

R1.9. Healthcare utilization and costs: As the content of usual care was not described, it is not clear if the costs of usual care were included.

As we set out in your prior comment we have better reported the meaning of usual care. We have also clarified in this section the costs included in each arm, also for usual care.

R1.10. P10 Line 5: AUC approach needs to cite a reference.

Thank you. We have decided to cite Glick et al. here as a reference textbook.

R1.11. Sample size calculation: I suggest remove this part. Simply state the number of patients per arm in trial design would be sufficient.

The section has been simplified according to your suggestion. The full description on the calculation of sample size can be found in the paper published on the study protocol [1].

[1] . Y. Ramallo-Fariña, L. García- Pérez et al. Effectiveness and cost-effectiveness of knowledge transfer and behavior modification interventions in type 2 diabetes mellitus patients—the INDICA study: a cluster randomized controlled trial, Implement. Sci. 10 (2015) 47. https://doi.org/10.1186/s13012-015-0233-1.

R1.12. P10 Line 21: By protocol, ‘multilevel models’ might be more appropriate than ‘mixed models’.
This has been modified in the text and mixed model has been modified to multilevel model, which is the correct term. Thank you very much for your comment.

R1.13. P10 Line 24-26: By protocol, the covariates were selected after assessing their significance, which needs to be briefly explained if that were the case.

The best model was adjusted for each dimension (QALYs and Costs) and models were evaluated independently. The criteria to include some covariates or others is set out in the study protocol: “To identify the covariates to be included in the model, we will first fit separate models including each covariate, one at a time. The final model will include those covariates such that their inclusion changes the estimates’ treatment effect by at least 10%. As suggested in the CONSORT statement, decisions about covariates will not be based on p value” [1].
A clarification has been added on the procedure in the Statistical Methods section.

[1] Ramallo-Fariña, Y., García-Pérez, L., Castilla-Rodríguez, I. et al. Effectiveness and cost-effectiveness of knowledge transfer and behavior modification interventions in type 2 diabetes mellitus patients—the INDICA study: a cluster randomized controlled trial. Implementation Sci 10, 47 (2015). https://doi.org/10.1186/s13012-015-0233-1

R1.14. P10 Line31: As point 2 above, please reconsider the use of the term ‘multiple imputation’. If it means ‘multiple methods of imputation’, please list the methods used and which data they applied to. If it does mean the method multiple imputation, then it needs to be explained why it deviated from protocol and the details of the imputation model. In any case, it may be necessary to have a sensitivity analysis to assess the impact of missing data imputation.
The multiple imputation method has been used for imputation of lost data as explained for comment R1.2.
In the papers already published on clinical effectiveness [1] and self-reported by patients [2] in this study Multiple Imputation procedure was also performed.
The estimate of costs and QALYs for imputed and non-imputed data are similar and lead us to draw the same conclusions. Imputation improves the accuracy of the estimate and reduces possible bias associated with loss of data. This was also verified for efficacy measures, both clinical and self-reported by patients. A phrase on this has been included in the results section.
The imputation model for variables included in the cost-effectiveness evaluation has been included as Supplemental Appendix 2 in this publication.
As discussed in comment R1.2 the change of imputation method has been included in the limitations in the discussion section.

[1] Y. Ramallo-Fariña, M.A. García-Bello, L. García-Pérez, et al. Effectiveness of internet- based multicomponent interventions, for patients and health care professionals, to improve clinical outcomes in type 2 diabetes. The INDICA study: a multiarm cluster randomized controlled trial, JMIR MHealth UHealth. (2020). https://doi.org/10.2196/18922.
[2] Yolanda Ramallo Fariña, Amado Rivero Santana, Lidia García Pérez et al. Patient-reported outcome measures for knowledge transfer and behaviour modification interventions in type 2 diabetes—the INDICA study: a multiarm cluster randomised controlled trial. BMJ Open 2021;11:e050804. doi: 10.1136/bmjopen-2021-050804

R1.15. P10 Line 33-36: Uncertainty surrounding the point estimate of ICER needs to be examined and presented.

Due to the complexity of the models performed, which Include a multilevel analysis and imputation of lost data, it was not possible to perform a bootstrapping analysis that would enable us to characterize uncertainty around the value of the ICER. We have added this study limitation in the manuscript. In any case we believe it is appropriate not to highlight the value of the ICER of the base case given that the differences in terms of effectiveness (QALYs) and costs are not significant. This has all been added to the limitations in the discussion.

R1.16. P10 Line 36-38: What parameters were re-estimated? How were they selected? How alternative values were selected for each of them and what values were they? Were the alternative values used together or the values were replaced one at a time?

We are very sorry we did not include the results of the one-way sensitivity analysis in the Supplementary material. We varied the costs by +-20%. We have added this detail in the Methods and a table with the results in online supplemental Appendix 1. Regardless of the analysis, PFI and CBI are always dominated.

Results

R1.17. Before reporting the results, please briefly describe the follow-up rates and if they were different by arms, and the missing data patterns.

A phrase has been added to the results section referring to the paper on evaluation of clinical results; in which the flowchart can be found that explains in detail the follow-up rates by arms. This also shows that at 24 months the percentage of data lost in the intervention and control group arms is approximately 30% and 23%, respectively.

The missing data patterns is also published in the paper of clinical trial results [1]. Reference has been made to this in the Statistical Methods.

[1] Y. Ramallo-Fariña, M.A. García-Bello, L. García-Pérez, et al. Effectiveness of internet- based multicomponent interventions, for patients and health care professionals, to improve clinical outcomes in type 2 diabetes. The INDICA study: a multiarm cluster randomized controlled trial, JMIR MHealth UHealth. (2020). https://doi.org/10.2196/18922.

R1.18. Table 1: please specify statistical test used and add P-values as it is stated in the text that none of these were statistically significantly different.

A column has been added to Table 1 that includes the p-values associated with baseline comparison between groups for each variable. We have corrected the errata in results in regard to this variable as there were differences in the percentage of men in the PTI and PFI groups.

The statistical analysis used for each variable has been added in the Statistical Methods section.

R1.19. When stating statistical significance, associated numbers should be included in the text, even if they were presented in tables.

Thank you for the recommendation. The meaning has been included in the text when relevant.

R1.20. Where mean values were stated, uncertainty measures should also be included in the text.

Thank you for the recommendation. Standard deviation or confidence interval has been included in the text when relevant (for total values mainly).

R1.21. Table 2: please specify which statistical test was used.

The explanation of the models used has been improved in the Statistical methods section. The title of the tables has been improved to make this more explanatory. The model used and the variables for which it has been adjusted has also been added to the table footnote.

R1.22. Cost-effectiveness analysis base case: The uncertainty surrounding the ICERs should be examined and presented using CEACs.

As we said above it was not possible to complete the bootstrapping or depict the CEAC as we would have liked.

R1.23. Cost-effectiveness analysis sensitivity analysis: As mentioned in point 16, it is not clear how the sensitivity analysis was performed. And the results of sensitivity analysis should at least be included in Appendix. It is not sufficient to simply say they were similar without data presented.

We are very sorry we did not include the results of the one-way sensitivity analysis in the Supplementary material. We varied the costs by +-20%. We have added this detail in Methods and a table with the results in the online supplemental Appendix 1.

Discussion

R1.24. The strengths are more of the trial as a whole than of the current analysis. As this is the discussion of cost-effectiveness analysis, I suggest the discussion focusing more on the economic evaluation side.

Although we could not overcome the limitations of the trial (they are also limitations in the economic evaluation), we agree with the reviewer. Hence, we have elaborated this section to focus both limitations and strengths on the economic evaluation.

R1.25. It also needs comparison with similar studies.

We have added some references of similar studies to compare with our study and discuss the generalizability in the Discussion section.

Reviewer: 2
Dr. Itziar Oyagüez, Pharmacoeconomics & Outcomes Research Iberia

Comments to the Author:
It is a very interesting work. The paper is well written and easily understable for readers.
R2.1 I was quite surprised by the results, so as it is mentioned in the discussion, I would have expected better outcomes for CBI, and PFI than for UC.
Yes, we were surprised as well. We have tried to comment on this in the results. We are now in a position to cite the paper with the medical results in the Discussion so we expect the reader could find more answers there.

R2.2 My main concern is regarding the result shown about ICER. It seems it is calculated for PTI vs UC, but according the authors themselves in the methods section (page 10 of 49, line 32 to 35), "Incremental cost-effectiveness ratio (ICER), that is, the differences between costs divided by the differences in QALYs, were calculated when one alternative was more effective and more costly than another."

We have amended this last sentence to not exclude the southwest quadrant from the results and interpretation. Now we say that the ICER “was calculated when one alternative was more (less) effective and more (less) costly than another”, once the dominated alternatives were excluded”.

R2.3 This part and results sections (page 12 of 49) should be aligned to be coherent.
After doing quick calculations, ICER would result €17896/QALY with UC vs PTI. I guess this is not the expected outcome to report but that's it.

We have added all the decimals in the table as they are determinant in the estimation of the ICER (especially the decimals of the QALYs: the difference between arms is not 0.1 but 0.00465). We have also added the incremental costs and QALYs as the reviewer suggested; the previous values were those obtained by means of the adjusted model. Nevertheless, the ICER remains above 25,000 Euros per QALY when we use all the decimals.

Minor comments:
R2.4 - In suggest to avoid the term more cost-effective in the conclusion (page 15 of 49, line 5)

R4.3 Methodologically I think an intervention is cost-effective or not cost-effective, but not more cost-effective. ICER could be lower or higher but it doesn't mean "more cost-effectiveness"

Thank you, we have rewritten this sentence.

Reviewer: 3
Mr. Daniel Pollard, University of Sheffield

Comments to the Author:
The paper presents an interesting analysis of information and communication technologies to support self-management of type 2 diabetes in the Canary Islands in Spain.

R3.1 The exact statistical model fitted should be specified and a reference to the method should be provided. Mixed models are an inadequate description.

Thank you very much for your comment. The explanation of the model used has been improved and referenced. As another reviewer has said the best denomination for the model used is Multilevel Model. We have modified all references to the model in the text and tables.

R3.2 Furthermore, it is unclear why the control variates were chosen as they are different for costs and QALYs. Why are QALYs not adjusted for age and sex. Why are costs not adjusted for time since diagnosis?

The best model was adjusted for each dimension (QALYs and Costs), the models are evaluated independently. The criteria to include some covariates or others is specified in the study protocol: “To identify the covariates to be included in the model, we will first fit separate models including each covariate, one at a time. The final model will include those covariates such that their inclusion changes the estimates’ treatment effect by at least 10%. As suggested in the CONSORT statement, decisions about covariates will not be based on p-value” [1].
A clarification on the procedure has been added in the section Statistical Methods.
After your comment we have also recovered the models performed adjusting QALYs for age and sex; estimates are similar with minor changes as of the third decimal. We have therefore left the estimates as we had to stay faithful to what was set out in the protocol.
[1] Ramallo-Fariña, Y., García-Pérez, L., Castilla-Rodríguez, I. *et al.* Effectiveness and cost-effectiveness of knowledge transfer and behavior modification interventions in type 2 diabetes mellitus patients—the INDICA study: a cluster randomized controlled trial. *Implementation Sci* 10, 47 (2015). https://doi.org/10.1186/s13012-015-0233-1

R3.3 The multiple imputation should be specified, with details on the imputation model, imputation covariates and the number of imputation iterations. Some diagnositics should be shown to demonstrate that enough imputation iterations have been conducted.
Thank you for your comment, other reviewers have also shown interest in this topic.
As explained in the answer to comment R1.2 the description of Multiple Imputation procedure has been improved. The model of imputation for variables included in the cost-effectiveness evaluation has been included as Supplemental Appendix 2 in this publication. Reference has been made in the Statistical Methods section to the analysis on the losses pattern already published in the paper for evaluation of the clinical efficacy of the ECA [1].
In regard to the number of imputation iterations, it was calculated for the variables included in the model of imputation the number of imputations necessary with the module "how_many_imputations" from STATA [2]. A total of 90 iterations was performed, this amount was greater than those necessary for each variable. For example, for the EQ-5D-5L Index, the necessary number of imputations was 35, and we performed 90.

```
. mi estimate: mean TARIFA_EQ_M0 TARIFA_EQ_M6 TARIFA_EQ_M12 TARIFA_EQ_M18 TARIFA_EQ_M24

Multiple-imputation estimates     Imputations     =         90
Mean estimation                   Number of obs   =      2,334
                                  Average RVI     =     0.4511
                                  Largest FMI     =     0.3432
                                  Complete DF     =       2333
DF adjustment:   Small sample     DF:     min     =     510.45
                                          avg     =   1,055.19
Within VCE type:     Analytic             max     =   2,295.14
```

	Mean	Std. Err.	[95% Conf.	Interval]
TARIFA_EQ_M0	.8622507	.0038131	.8547732	.8697281
TARIFA_EQ_M6	.8489342	.0042466	.8406011	.8572673
TARIFA_EQ_M12	.8496559	.0046112	.8405985	.8587133
TARIFA_EQ_M18	.8249558	.0051637	.8148112	.8351005
TARIFA_EQ_M24	.8344975	.0044732	.8257184	.8432766

```
. how_many_imputations
Fraction of missing information (95% CI):  0.34 (0.28, 0.41)
Imputations in pilot:                      90
Imputations needed:                        35
Imputations to add:                        0
```

[1] Y. Ramallo-Fariña, M.A. García-Bello, L. García-Pérez, et al. Effectiveness of internet- based multicomponent interventions, for patients and health care professionals, to improve clinical outcomes in type 2 diabetes. The INDICA study: a multiarm cluster randomized controlled trial, JMIR MHealth UHealth. (2020). https://doi.org/10.2196/18922.
[2] Paul T. von Hippel, 2018. "HOW_MANY_IMPUTATIONS: Stata module to determine required number of imputations," Statistical Software Components S458452, Boston College Department of Economics.

R3.4 Sensitivity analyses are conducted; however it is unclear what sensitivity analyses have been conducted as they are not defined the main text.

We are very sorry we did not include the results of the one-way sensitivity analysis in the Supplementary material. We varied the costs by +-20%. We have added this detail in Methods and a table with the results in the online supplemental Appendix 1.

R3.5 There is extensive use of p values to present results. It would be clearer to a reader to present mean costs and 95% confidence intervals.

Thank you very much for your comment. The 95%CI for costs and QALYs have been included in the tables.

R3.6 The threshold ICER for decisions makers in the Canary Islands is not mentioned. Therefore, it is unclear whether the conclusions are supported or not.

We are very sorry. We mention the Spanish threshold now at the end of analysis. This would certainly facilitate interpretation.

R3.7 The results that some of the interventions produce fewer QALYs than the comparator arm. In diabetes, HbA1c itself has at best a minimal effect on utility (see Beaudet et al for systematic review of utility values for health states in diabetes simulation modelling). The big drivers of changes in utility are diabetes related complications. HBA1c itself, does influence the progression to these health states. However, as the trial was only powered on HbA1c and not the incidence of long term complications I would not expect that the intervention would produce more QALYs than a control arm.

Thank you for the reference. We use it now in our paper. Here we present the short term economic evaluation analysis conducted alongside the clinical trial. For the long term we have scheduled a model in which the QALYs will hopefully vary according to whether or not more or less complications appear. We have improved the discussion to collate these ideas.

R3.8 The intervention costings presented in the appendix are very unclear:
Why have development / reviewing costs (for materials and computed systems) been included? Surely these are one-off costs in the intervention development
Why have the full cost of the laptops been included in the costs? Surely the costs of laptops should be spread out over multiple years if they were to be used in routine practice.

We conducted a budget impact analysis of the implementation of any of these interventions and included their results in a report for the public health authorities. Certainly we could have reported this to complement the within-trial cost-effective analysis. However, we believe this would require much more material than that presented here and it was not our original intention. For example, the resources and costs needed to implement the intervention for patients in the whole region will not be the multiplication of resources just based on the number of people with diabetes in the region, but will require having in mind the number of people in each island, the adherence in each area, and some scales and access equity consideration, etc. That is, the number of nurses and laptops will not be proportional to the population as there are many scarcely inhabited areas that should be serviced anyway. So laptops will last more than two years but also its cost should be assigned to more patients. In summary, we decided to report here the within-trial costs and explore these issues by means of the sensitivity analysis, that now we report in the appendix 1. We discuss all this in the Discussion section.

R3.9 Why is the website 10 times cheaper in the both column of Table A2 compared to the Patients column?

Thank you for checking the additional material so carefully. In this case it was a typographical error: a zero was missing one of the columns. It has been corrected. The cost of the website is the same regardless of the alternative.

R3.10 Table A3 appears to show that the numbers of:
Visits to a general practitioner (PTI arm and PFI arm)
Visits to a nurse at primary care (PTI arm and PFI arm)
Lab test by general practitioner (PTI arm)
Are the differences between multiple different arms (e.g. PTI v UC and PTI v CBI and PFI v PTI) of the study. How?

We have modified the table's title so that it is clearer that this involves use of resources during entire follow up. Super indices denote those variables in which statistically significant differences were identified between groups.

R3.11 Furthermore, it is stated that in a footnote that there was a negative binomial regression model, adjusted by time since diagnosis and baseline resource use.
It would be useful to specify exactly what the negative binomial models were predicting. For example, are they used for each subcategory of resource use or were they only used for some categories.
The baseline resource use variables in these regressions should be more clearly defined, as the specific variables that were controlled for. It is unclear whether all baseline resource use was controlled for or only a particular category of resource use (e.g. Hospital Stays).

All categories on use of resources were estimated with models with the same structure: Negative Binomial Regression Model adjusted by time since diagnosis and baseline resource use.
To estimate QALYs a Multilevel model adjusted by time elapsed since diagnosis and baseline utility was used for each period (0-6 months, 0-12 months, 0-18 months and 0-24 months).
For costs on use of resources the same models structure for each resources was maintained (Hospital stays, Laboratory tests, etc): Multilevel model, adjusted by age, sex and baseline utility.
We have included footnotes for each table clarifying the models used.

References
Beaudet A, Clegg J, Thuresson PO, Lloyd A, McEwan P. Review of utility values for economic modeling in type 2 diabetes. Value in Health. 2014 Jun 1;17(4):462-70.

Reviewer: 5
Dr. Yiwei Chen, Stanford University

Comments to the Author:
Overall, this is a well-written paper and a well-conducted study. The importance and interpretation of the finding however need to be strengthened.

Background and Setting:
R5.1 The authors mentioned that the usual care arm had an intensive trial follow-up that may explain the good health outcomes. Can you elaborate more on this? Do the other arms have this feature too or it's a unique feature for the usual care arm? Is the cost reflected?

All arms had the same intensive follow-up. What we wanted to highlight here is that follow-up was as intensive as to have had an effect, in all the arms, but especially in the only arm that we did not want to affect. Following the reviewer's recommendation we elaborate on this in the Discussion section.

R5.2 What's the health care and patient profile setting for Canary Islands? Is it similar to Spain? Shall the readers interpret the results just for Canary Islands or for a broader region?

The National Health System in Spain is comprised of 17 Regional Health Systems. Canary Islands is one of those regions They are all quite similar in their organization. Although not all systems offer the same support to patients with diabetes, primary care is very homogenous and the interventions could be implemented with few modifications in regions other than the Canary Islands. We have added some of these ideas in the Discussion.

Cost
R5.3 The cost of INDICA interventions include a lot of fixed costs, e.g. material development time for patients; material development time for GP; development of computer system, etc. Such costs will dimmish very quickly when we include more patients. The current cost calculation is appropriate for the experiment population, but it's not appropriate if readers want to extend it to the whole population of the Canary Islands. I suspect this will change the cost calculation and is more appropriate for the goal of this paper.

We conducted a budget impact analysis of the implementation of any of these interventions and included their results in a report for the public health authorities. We could have certainly reported this to complement the within-trial cost-effective analysis. However, we believe this would require much more material than that presented here and it was not our intention in the first place. For example, the resources and costs needed to implement the intervention for patients in the whole region will not be the multiplication of resources just based on the number of people with diabetes in the region, but will require having in mind the number of people in each island, the adherence in each area, and some scales and access equity consideration, etc. That is, the number of nurses and laptos will not be proportional to the population as there are many scarcely inhabited areas that should be serviced anyway. So laptops will last more than two years but their cost should also be assigned to more patients. In summary, we decided to report the within-trial costs here and explore these issues by means of the sensitivity analysis, which we now report in the online supplemental Appendix 1. We discuss all this in the Discussion section.

Statistical methods:
R5.4 Where is the sensitivity analysis? I can't find it the in the paper.

We are very sorry we did not include the results of the one-way sensitivity analysis in the Supplementary material. We varied the costs by +-20%. We have added this detail in Methods and a table with the results in the online supplemental Appendix 1.

R5.5 Attrition rate isn't clear. How many participants did the study lose over the 24 period? What are the baseline stats for the population who remained at the end? Is the analysis result sensitive to imputation method?

Thank you for your comment, other reviewers have also shown interest in this topic.
In regard to attrition rate, as explained in comment R1.17 we have added to the results section a phrase referring to the article for evaluation of clinical results where the flowchart in which the follow-

up rates by arms is explained in detail, can be found. The same flowchart shows that at 24 months the percentage losses in the intervention arm and control group are approximately 30% and 23%, respectively.
In regard to the imputation method, the multiple imputation method has been used for the imputation of lost data as explained in comments R1.2 and R3.3. In the papers already published on clinical efficacy [1] and self-reported by patients [2] from this study a multiple imputation procedure was also performed.
Costs and QALYs estimate from imputed and non-imputed data are similar and lead us to the same conclusions. The imputation improves the accuracy of the estimate and reduces the possible bias associated with data loss. This was also verified for both clinical and self-reported perceived efficacy measures by patients. A phrase on this has been included in the results section.
The imputation model for variables included in the cost-effectiveness evaluation has been included as Supplemental Appendix 2 in this publication. Moreover, reference has been made in the Statistical Methods section to the analysis on the loss pattern already published in the paper on evaluation of the clinical efficacy of ECA [1].

[1] Y. Ramallo-Fariña, M.A. García-Bello, L. García-Pérez, et al. Effectiveness of internet- based multicomponent interventions, for patients and health care professionals, to improve clinical outcomes in type 2 diabetes. The INDICA study: a multiarm cluster randomized controlled trial, JMIR MHealth UHealth (2020). https://doi.org/10.2196/18922.

Reviewer: 4
Dr. Samuel Sarmiento, Johns Hopkins University School of Medicine
General
I like that you designed your trial with a cost-effectiveness (CE) analysis in mind, as well as a potential qualitative study in the future. I think this makes your results easier to interpret, even for those without a background in health economics.

R4.1. Please proofread the document again and correct minor grammatical errors throughout the text. I'll point out some of them in the appropriate sections. Attention to such details will give more credibility to your study.

The study has been edited by a native biomedical translator. His certificate of native editing is attached to the submission.

R4.2. Minor grammatical adjustments:
All references to lines and pages I make here are based on the proof you submitted to BMJ, not on the original manuscript.
Abstract
• Line 7 & elsewhere: In primary care.
• Line 11: Within-trial period.
• Line 14 & elsewhere: In the Canary Islands.
• Line 37: Aimed at professionals or usual care.
• Line 46: Keywords: primary care.
• Don't forget to leave a space before symbols such as "greater than" or "less than" (>, <) throughout the document. For instance, "p value <0.05; HbA1c >7%." Please be consistent with this and punctuation in general.
Strengths & Limitations
• Line 10: From a healthcare perspective.

• Line 23: The main limitation is the relatively short duration of the trial, two years.
Introduction
• Line 9: Comma after "…patients who have already developed T2DM".
Methods
• Line 7: Controlled clinical trial.
• Line 13: Was the main outcome.
• Line 13: If quality-adjusted life years is in the plural, you can abbreviate as "QALYs."
Patient involvement:
• Line 38: Canary Islands.
• Line 43: In the preparation of the protocol.
• Line 48: "…interviews that will be the subject of another study." Modified

Health care utilization and costs:
• Line 14: "…and medications."
Results
CE analysis: base case:
• Line 15: No space after €.
Discussion
• Line 27: "The differences are explained by the differences…" Rephrase to improve clarity.
• Line 56: "Non-significant."
• Page 14, line 3: Change "moreover" to "however." You are about to introduce a statement that contrasts with the previous one, that is, despite a lower effectiveness, PTI was dominant in patients with poorly controlled diabetes.
• Page 14, line 9: Change to "patients with poorly controlled blood glucose levels."
• Page 14, line 27: "Potential efficacy," not efficiency.
• Page 14, line 32: "Some degree of missing data."
• Page 15, line 9: "With the highest needs…" "…its impact on the…"

Thank you for reviewing the English. The suggested changes have been made.

R4.3 Line 29: Did you consider adjusting for inflation and reporting costs for the year you analyzed the data?

All unit costs are expressed in Euros year 2017. The unit costs were adjusted for inflation when needed. We have added a sentence in the Methods section.

Outcomes:
R4.4 Line 53: It's been a while since I reviewed the EQ instrument, but can a health state really be worse than death? The answer is obviously no. Negative scores simply represent decreases in life quality, and not health states worse than death.

Thank you for the comment. Now we say that "negative scores represent health states perceived as worse than death".

R4.5 Line 17: This is the most important line in the document and you must elaborate on this in the discussion. In most health systems, an ICER of €30,000 can be considered cost-effective.
R4.6 However, in this case, this number is affected by the intervention being cheaper than the standard of care, even though it was also slightly less effective.
However, in the subgroup analysis among those with an HbA1c >7%, this was clearly the dominant strategy. You have to drive this point home in the discussion.

Taking into account all the reviewer's recommendations, we believe we have improved the Discussion. The base case and the subgroup analysis results are further commented now compared to before.

R4.7 Please create subheadings for "strengths" and "limitations." This will improve readability.

Following the reviewer's recommendations, we have added a new subheading.

R4.8 In the section "Strengths & Limitations," you present the duration of the trial as a limitation. Therefore, you cannot also include it as a strength in the discussion. This is either a strength or a limitation, but not both. In my opinion, given the long-term effect of certain interventions on quality of life, two years is a relatively short period of time to draw any definitive conclusions.

Thank you for pointing this out. A follow-up of two years is an achievement from the point of view of a trial (most trials do not have such a long follow-up), but not enough from the point of view of a cost-utility analysis. We have rewritten this section in the Discussion to clarify this point.

Graphs & Tables

R4.9 Table 2: Please add an asterisk or another symbol next to the statistically significant results so that readers can more easily find them or put them in bold.

Thank you very much, we have taken your suggestion into consideration and p-values associated to statistically significant results are now in bold in Table 2.

R4.10 Table 3: I don't have access to your raw numbers, but regarding the ICER between PTI and UC, going by the information in this table:
…
Am I missing something?

We have added all the decimals in the table as they are determinant in the estimation of the ICER (especially the decimals of the QALYs: the difference between arms is not 0.1 but 0.00465). We have also added the incremental costs and QALYs as the reviewer suggested; the previous values were those obtained by means of the adjusted model. Nevertheless, the ICER remains above 25,000 Euros per QALY when we use all the decimals.

VERSION 2 – REVIEW

REVIEWER	Oyagüez, Itziar Pharmacoeconomics & Outcomes Research Iberia
REVIEW RETURNED	04-Feb-2022

GENERAL COMMENTS	None to be added

REVIEWER	Pollard, Daniel University of Sheffield, School of Health and Related Research
REVIEW RETURNED	16-Feb-2022

GENERAL COMMENTS	The authors have done a great deal of work to improve the mansucript and respond to the inital round of reviwer comments. It reads very well and I have no additional comments to make.

REVIEWER	Sarmiento, Samuel Johns Hopkins University School of Medicine, Plastic and Reconstructive Surgery
REVIEW RETURNED	15-Feb-2022

GENERAL COMMENTS	Thank you for addressing the comments of the editor and the reviewers. I think the study is acceptable for publication now.

REVIEWER	Chen, Yiwei Stanford University, Department of Economics
REVIEW RETURNED	13-Feb-2022

GENERAL COMMENTS	The authors have covered many of my previous concerns, but there remain some minor questions to be addressed. Statistical methods: 1. Can the authors perform another robustness check with complete data only? If the results are consistent, I have more confidence in the multiple imputation analysis the authors performed. If not, the authors shall investigate and give a reason why a bias exists in the complete-data-only version and why the particular multiple imputation analysis the authors performed can fix it. I suspect the results won't differ too much as the attrition rate is moderate. Minor comments on writings 1. The authors have addressed my question regarding whether other arms have an intensive follow up just as the control group. The answer is yes but the writing in the discussion didn't make it explicit. Another confusing sentence is that the authors wrote that the intensive trial follow up can be seen itself as a treatment. I think what the authors meant are that PTI, PFI, and CBI may not do much more than UC which already has an intensive follow up component. 2. The discussion on generalizability and transferability isn't satisfactory as the author stopped at saying it's not straightforward. Can the authors give a few hypotheses on why the findings may not carry over to broader regions? E.g. is the digitalization level higher or lower in the rest of Spain, making PTI potentially less or more expansive?

VERSION 2 – AUTHOR RESPONSE

We are very grateful to all the reviewers.

Reviewer: 5

Dr. Yiwei Chen, Stanford University

Comments to the Author:

The authors have covered many of my previous concerns, but there remain some minor questions to be addressed.

Statistical methods:

1. Can the authors perform another robustness check with complete data only? If the results are consistent, I have more confidence in the multiple imputation analysis the authors performed. If not, the authors shall investigate and give a reason why a bias exists in the complete-data-only version and why the particular multiple imputation analysis the authors performed can fix it. I suspect the results won't differ too much as the attrition rate is moderate.

Answer: Thank you for this suggestion. We have conducted the analysis with complete data. The direction of the results does not change (see Table below). Consequently, we have amended the last paragraph in the Results section:

Page 11:

The estimate of costs and QALYs was similar for all imputed, non-imputed and completed data. The same arms stayed as dominant and the same conclusion with regard to ICER was upheld.

		PTI	PFI	CBI	UC
Imputed data					
	QALYs	1.71	1.69	1.63	1.72
	COST	2571.53	2793.91	3025.12	2750.44
Non-imputed data					
	QALYs	1.77	1.76	1.69	1.77
	COST	2550.8	2778.26	3317.36	2759.03
Completed data					
	QALYs	1.70	1.69	1.59	1.73

	COST	2674.03	2771.55	3317.26	2758.96

Minor comments on writings

1. The authors have addressed my question regarding whether other arms have an intensive follow up just as the control group. The answer is yes but the writing in the discussion didn't make it explicit. Another confusing sentence is that the authors wrote that the intensive trial follow up can be seen itself as a treatment. I think what the authors meant are that PTI, PFI, and CBI may not do much more than UC which already has an intensive follow up component.

Answer: The reviewer is right; we were not sufficiently explicit in the text. We have rewritten the paragraph in the Discussion section to say that:

Page 15:

Similarly, the unexpected results with regard to the outcomes measured in the usual care arm might be accounted for by the intensive trial follow-up that all the arms experienced (i.e. answering questions about diet, physical activity and self-care six times in two years, plus blood tests and other examinations) that could be seen as a kind of intervention [references]. Therefore, the intensity of the follow-up in the study might have also impacted patient behaviour in the usual care arm, to a point of reducing the differences in effects at the end of the two-year period.

2. The discussion on generalizability and transferability isn’t satisfactory as the author stopped at saying it’s not straightforward. Can the authors give a few hypotheses on why the findings may not carry over to broader regions? E.g. is the digitalization level higher or lower in the rest of Spain, making PTI potentially less or more expansive?

Answer: We are grateful to the reviewer because, certainly, we did not elaborate about the digital skills in the region. Now, the text says:

Page 12:

Transferability to real clinical practice of cost-effective interventions could be even more efficient as their application can be extended to thousands of T2DM patients, with minimal cost increases.

Page 13:

Finally, the generalizability of the INDICA findings and the transferability of its results to other settings are not straightforward. Interventions were designed and implemented considering the level of health and digital literacy of the population in the Canary Islands, that is quite similar to the average in Spain (and above the EU mean) [2 new references: 41 & 42], and the organization of the primary healthcare provision by the public system in the region. Although not all regions in Spain offer the same support to patients with diabetes, primary healthcare is quite homogenous throughout the country so the interventions could be implemented with few modifications in regions other than the Canary Islands. Therefore, we could conclude that the intervention and the cost-effectiveness results could be transferable to other regions in Spain, but the transferability to other countries would need a thorough analysis of the care for T2DM in other foreign settings.

New references:

-Eurostat. ICT usage in households and by individuals. 2021.https://ec.europa.eu/eurostat/databrowser/view/isoc_ci_ifp_iu/default/table?lang=en (accessed 22 Feb 2022).

-Instituto Nacional de Estadística. Survey on Equipment and Use of Information and Communication Technologies (ICT) in Households. 2021.https://www.ine.es/en/prensa/tich_2021_en.pdf (accessed 22 Feb 2022).

VERSION 3 – REVIEW

REVIEWER	Chen, Yiwei Stanford University, Department of Economics
REVIEW RETURNED	06-Mar-2022

GENERAL COMMENTS	The authors have successfully answered all my questions and this looks a very good paper.