Author’s response to reviews

Title: A Personalized Intervention to Prevent Depression in Primary Care: Cost-effectiveness Study Nested into a Clustered Randomized Trial

Authors:

Ana Fernandez (anafezsez@gmail.com)
Juan Mendive (juanmmendive@gmail.com)
Sonia Conejo-Ceron (soniafundacionimabis@hotmail.com)
Patricia Moreno-Peral (predictmalaga@hotmail.com)
Michael King (michael.king@ucl.ac.uk)
Irwin Nazareth (i.nazareth@ucl.ac.uk)
Carlos Martín-Pérez (med000261@hotmail.com)
Carmen Fernández-Alonso (carmenferal@gmail.com)
Antonina Rodríguez-Bayón (antoininabaeza@gmail.com)
Jose Aiartzaguena (jmaiarzaguena@euskalnet.net)
Carmen Montón-Franco (carmenmontonf@ono.com)
Antoni Serrano-Blanco (aserrano@pssjd.org)
Inmaculada Ibañez-Casas (iibanez@ugr.es)
Emiliano Rodríguez-Sánchez (emilianorodriguesanchez@yahoo.es)
Luis Salvador-Carulla (luis.salvador-carulla@sydney.edu.au)
Paola Bully-Garay (PAOLA.BULLYGARAY@osakidetza.net)
Maria Ballestar-Rodríguez (isabelballesta@gmail.com)
Pilar Lafuente (lafuentepilar13@gmail.com)
Maria del Mar Muñoz-García (marmg@ugr.es)
Pilar Minguez-Gonzalo (pilargon@ono.com)
Dear Professor Samuel,

Please, find enclosed a revised version of our manuscript entitled “A Personalized Intervention to Prevent Depression in Primary Care: Cost-effectiveness Study Nested into a Clustered Randomized Trial”. The manuscript has been revised, and we hope that in the present form it fulfils the comments made by the reviewers.

We have formatted all revisions with track changes to allow for easy recognition of modifications.

We also would like to express our thanks to the reviewers for their careful review of the manuscript and kind comments about the quality of our study.

Yours sincerely,

Ana Fernandez, on behalf of all the authors
Reviewer #1:

The manuscript "A Personalized Intervention to Prevent Depression in Primary Care: Cost-effectiveness Study Nested into a Clustered Randomized Trial" addresses an important and interesting topic, the prevention of depressive episodes by a personalized approach. In general, the paper is well written and all relevant aspects are presented. The design of the economic evaluation conforms to the international standards of a state-of-the-art analysis. The authors report the ICER, the CE-plane and the CEAC. The statistical analysis considers thoroughly the specific distributional characteristics of cost data. The authors perform different GLM and evaluate the quality of the models. Finally, the authors decide to use GLM with gamma-log for costs and gaussian-identity for QALY. Based on experience from previous analyses, this appears to me as a suitable and appropriate choice. The results section is well structured and comprehensive. The discussion section presents the most important result, discusses strength and limitations and shows implications for practice. The conclusions are supported by the results. This is the second economic evaluation of an intervention based on the PredictD risk algorithm. The personalized intervention is an innovative aspect. Hence, the manuscript represents a progress.

RESPONSE: We want to thank the reviewer for his/her positive comments.

Apart from the positive aspects mentioned above, there are some concern I have related to the manuscript, especially in context with the analytical approach.

In the Introduction, the authors stress the point that primary prevention is the most important approach to reduce the incidence of depression and it seems to me from the content of the analysis that the authors see their intervention as primary prevention. However, in the first paragraph of the result section the authors declare that the number of patients with a previous episode of depression was higher in the intervention group. Hence, the study considers participants unaffected by depression and those affected by depression. It is not clear to me if and I am in doubt that the mechanisms of the intervention are the same for people without previous episodes and those with previous episodes. Based on the logic of prevention the first group is a target group of primary prevention and the second group is a target group of tertiary prevention or maintenance of remission. Both prevention types use different instruments and have other objectives, so it is probable that the intervention has different implication in both groups. To receive a clearer and more robust picture of the true cost-effectiveness of the intervention, I want to ask the authors to perform subgroup analyses for participants without prior episodes and for those patients with previous episodes. Additionally the authors should make clearer which approach to prevention they foster with their intervention.
RESPONSE: Thank you for your comments, which are pertinent and interesting.

Primary prevention is one that avoids occurrence of disease either through eliminating their risk factors or increasing resistance to disease; so primary prevention requires that its target population do not suffer from the disease to be prevented. However, primary prevention of depression differs from other diseases such as diabetes (that generally never heals) in which fortunately people suffering from an episode of depression can be cured quite frequently. From this point of view there are three theoretical approach of primary prevention of depression: 1) primary prevention of the onset of depression (first episode); 2) primary prevention of the other episodes (excluding the first episode); and 3) primary prevention of depression of any episode (including the first episode). We stress that these three types of primary prevention of depression must comply with the condition that their target population does not suffer from depression when it is recruited to prevent. From this theoretical point of view, in our study we would have used the third approach (primary prevention of depression of any episode of depression) as one of our inclusion criteria was that all participants were free of depression for at least the 6 months before recruitment. Unlike primary prevention, the target population of tertiary prevention must suffer from depression when it is recruited to prevent. Tertiary prevention aims to soften the impact of an ongoing illness, by helping people manage long-term problems or permanent impairments in order to improve as much as possible their ability to function, their quality of life and their life expectancy.

We agree that it would be very interesting to separate our results according to the prevention of the first episode of depression or the follow episodes. However, why we cannot classify our patients in those who suffered the first episode of depression or any other episode?

We used a lifetime screen for depression based on positive response to the first two questions of the Composite International Diagnostic Interview (CIDI): 1) “Have you ever, except in the last 6 months, had nearly two weeks or longer when nearly every day you felt sad, empty or depressed for most of the day?”; 2) “Have you ever, except in the last 6 months, had 2 weeks or longer when you lost interest in most things like work, hobbies and other things that you usually enjoyed?”

In the study of Arroll et al. (Arroll B, Khin N, Kerse N. Screening for depression in primary care with two verbally asked questions: cross sectional study. BMJ. 2003;327:1144–1146) the positive predictive value when patients responded yes for both questions, compared with the CIDI for major depression, was 18%. This means that 82% of patients who were considered depressed with these two questions really were not depressed (82% of false positives). The true
usefulness of the two questions is when they are used as a screening instrument since its negative predictive value is almost 100% (less than 1% of false negatives). In other words, its sensitivity is very high but its specificity very low. We have used the two questions as a lifetime screen for depression, not as a diagnostic tool. Our interest in using the lifetime screen for depression was not to get a validate diagnosis of lifetime depression; we wanted to find a variable that, along with the others included in our predictive model, would improve our prediction of future episodes of major depression (King et al. Development and validation of an international risk prediction algorithm for episodes of major depression in general practice attendees: the PredictD study. Arch Gen Psychiatry. 2008;65(12):1368-76).

However, why we think our study is largely based on primary prevention of the onset of depression.

In our study, only 32.4% and 36.6% of patients, in the control and intervention group respectively (see Bellón et al. Intervention to Prevent Major Depression in Primary Care: A Cluster Randomized Trial. Ann Intern Med. 2016;164(10):656-65), responded yes to the two questions of lifetime screen for depression. If we assumed, as we said before, that the predictive positive value of the two questions was 18% (Arroll et al.), in our study the proportion of patients who truly suffered a first episode of depression before recruitment (except in the 6 months prior to the baseline interview) was 5.8% and 6.6% in the control and intervention group respectively. Therefore, from this point of view our study of primary prevention of depression is largely based (approximately 94% of participants) on primary prevention of the onset of depression. We also want to emphasize that these calculations can be made for the entire population of our study, but cannot be applied individually to each patient.

To be clearer on our approach to prevent depression, we have added to the introduction the following paragraph:

“Primary prevention is one that avoids occurrence of disease either through eliminating disease risk factors or increasing resistance to disease, so to be applied requires that its target population does not have the disease (depression in our case). Classically, primary prevention of depression is classified as ‘universal’, when it is applied to general population; ‘selective’ to participants with some risk factor (-s) for depression; and ‘indicated’ to patients with subthreshold depression (they have some symptoms of depression but do not meet the criteria for diagnosis)” And for the same reason, in the following sentence we have added the word “primary”: “The best primary
prevention program is likely to be one which targets modifiable risk factors; empowers individuals to address their risks; and is inexpensive and capable of large-scale dissemination. [10].

Finally, we also have added the following paragraph in the discussion (limitations section):

“Fifth, in our study, only 32.4% and 36.6% of patients, in the control and intervention group respectively [13] responded yes to the two questions of lifetime screen for depression (except in the 6 months prior to the baseline interview, in which no patient suffered major depression according to the CIDI). As the predictive positive value of responding yes to these two questions is 18% (Arroll et al.), in our study the proportion of patients who truly suffered a first episode of depression before recruitment was 5.8% and 6.6% in the control and intervention group respectively. Therefore, from this point of view our study of primary prevention of depression is largely based (approximately 94% of participants) on primary prevention of the onset of depression (first episode). ”

Related to the statistical analyses, I have two specific concerns that need to be addressed by the authors. Firstly, the authors write that the participants in the intervention group had a higher risk of depression than their counterparts in the control group. As discussed by the authors, imbalances can appear because of the cluster-randomization. However, this specific imbalance is quite problematic as it is directly connected to the aim of the study. Patients with a higher risk of depression receive an intervention to prevent depressive episodes. This group has better chances to show improvements. It is essential to consider this aspect in the statistical analysis, either by controlling for this risk or by matching/balancing the groups.

RESPONSE: We agree, adjusted analysis is essential in our study. Maybe we have not been able to explain it properly in the text. All the analyses were adjusted by baseline risk of depression and the unbalanced variables at baseline. We have now made this clearer in the statistical analysis section:

“All models were adjusted by their respective baseline values (QALY and costs), the individual risk of depression (i.e. risk score at the PredictD algorithm), and the following variables that were unbalanced at baseline (and were not included in the PredictD algorithm): employment
status, owner/occupier accommodation, perception of safety inside/outside the home, anxiety disorder, experiences of discrimination and city”.

Secondly, the authors recruited patients, treated by different physicians, working at different primary care centers, which were located in different cities. This is a high number of clusters. The authors control for some of these effects but not for all, e.g. city. At least the authors have to confirm that this is not necessary.

RESPONSE: Yes, we agree. We have used multilevel GLM models using as random component primary care centers or physicians when it was indicated. We explain it in the text (at the end of page 10 and at the top of page 11): “…We calculated the intraclass correlation coefficients (ICCs) of the health centre, GP and both taking the costs and QALYs as dependent variables. The ICCs for health centre were significant for the effect, while the ICC for GP was significant for the costs. Thus, we used multilevel GLM models to account for such clustering effects.”

In addition, all the analyses were also adjusted by city including this categorical variable (7 categories) in the fix component of the models. We have highlighted this: “All models were adjusted by their respective baseline values (QALYs and cost), the individual risk of depression (i.e. risk score at the PredictD algorithm), and the following variables that were unbalanced at baseline (and were not included in the PredictD algorithm): employment status, owner/occupier accommodation, perception of safety inside/outside the home, anxiety disorder, experiences of discrimination and city.”

Concerning the level of the economic evaluation, I have one concern that the authors should consider. The authors calculated ICER based on adjusted costs and QALY. This is not proper without further adjustments. If the authors want to present such an adjusted ICER, they have to perform a Cholesky decomposition (or a comparable approach). Otherwise, the authors should present and discuss the unadjusted ICER and draw their conclusions from the CEAC.

RESPONSE: To the best of our knowledge Cholesky decomposition is usually used in modelling where you want to predict a parameter value and that parameter value is derived from a regression model. However, when economic evaluations are based on a trial it is better to bootstrap the trial data, as we have done. The advantages of bootstrapping are: a) it captures the correlation in costs and outcomes; and b) it makes no parametric assumptions about the distribution of the data.

There are further minor aspects the authors should correct.

- How did the research assistants randomly select the patients?

RESPONSE: The research assistants randomly select the patients using random starting points for each day, generated using a random number generator. We have added this information at the method section.

- The authors should give at least a short description of the components of the PredictD-algorithm.

RESPONSE: We have added the following text at the method section: “The PredictD algorithm is composed of 12 risk factors: six are patient characteristics or past events (sex, age, sex*age interaction, education, childhood physical abuse, and lifetime depression) and six refer to current status (SF-12 physical score, SF-12 mental score, dissatisfaction with unpaid work, number of serious problems in very close persons, dissatisfaction with living together at home, and taking medication for stress, anxiety or depression). The "predictD algorithm" provides, in addition to the quantification of the overall risk of depression, knowledge of those risk factors influencing a given patient that could guide a possible preventive intervention” (Page 7)

- The authors should cite a source for the unit costs for the calculation of the indirect costs.

RESPONSE: Units costs were retrieved from the Oblikue company, which has a database (“esalud”) with all the Spanish healthcare costs retrieved from the official health services tariffs of the Autonomous Communities. We have added information about this database in the method section: “This unit cost was retrieved from ‘Oblikue dataset (esalud)’ (www.oblikue.com) which includes the official health services tariffs of the different Spanish Autonomous Communities”.

- Presenteeism costs were calculated. The authors should give some information on the assessment and valuation.

RESPONSE: We have added the following text at the method section: “Self-reported presenteeism was assessed using some questions from the WHO Health and Work Performance Questionnaire (HPQ) [20]. For this assessment, respondents first estimated how many days
during the past 4 weeks they had been at work not being able to perform their job as usual (A),
and then they rated their overall work performance during these days using a 0–100 scale where
0 corresponds to doing no work at all (while at work) and 100 signifies top work performance
(B). Then a score was calculated as follows: ((100 – B)*A)*6. Where 6 represented 6 months”

- The authors should state that the analyses were performed based on the ITT principle.
 RESPONSE: We have added this new text at the beginning of the statistical analysis section:
“Analyses were performed based on the Intention to Treat (ITT) principle, analyzing all
participants according to their randomized treatment and using multiple imputations when
outcomes were missing”

- There is no description of the calculation and construction of the CEAC (NBR approach). This
should be changed.
 RESPONSE: We have added the following text, to the existing one: “Cost Effectiveness
Acceptability Curves (CEACs) were then constructed. Each CEAC was derived from the net
benefit approach: Net monetary benefit=λ×(Δ Effect)−( Δ Cost), where λ represents the amount
of money society is willing to pay to gain one extra unit of effect. Each bootstrapped pairs of Δ
Effect and Δ Cost (i.e. 50000) were used to calculate the CEACs”.

- In table 2, the upper limits of the CI for the QALY differences are different between societal
and National Health perspective.
 RESPONSE: Thank you so much for pointing out this mistake. Numbers have been changed.

- The authors should discuss that they did not consider costs for informal care and general
medication. Both are relevant categories. The authors said that the patients in the intervention
group had a different perception of safety inside-outside the home. These patients are prone to
use supportive services delivered by relatives. On the other hand, has depression an influence on
somatic health too. On this way, it influences the utilization of medication in general.
 RESPONSE: Thank you so much, we have added this as a limitation: “In addition, we have not
taking into account informal care related costs and costs from general medication. As depression
has an impact on physical health, it is possible that this has been affected, making the costs
associated with depression possibly higher than we have calculated in our study”
Reviewer #2:

This is a clearly written article on the cost-effectiveness of a personalized intervention to prevent depression in primary care.

RESPONSE: Thank you so much for this comment.

I only have a few questions:

- Costs of the intervention included costs of a booklet for the patients and the cost associated with the training of the physicians. How have these latter costs been translated to patient level? Are they divided by the number of participants in the trial or by the total number of patients in a GP practice eligible for the intervention, as in real practice this would be the target population.

RESPONSE: The costs associated with the training of the physicians were divided by the number of participants in the trial. We have added a comment in the discussion section highlighting that the cost associated with the intervention would be even lower: “On the other hand, the cost associated with the training of the physician was translated to patient level by dividing by the number of participants in the trial and not by the total number of patients, which would be more appropriate in real practice. Consequently, the costs associated with the intervention in real practice would be even lower”

- How is this intervention going to be implemented in practice? Are eligible GP patients screened once during lifetime with the PredictD algorithm, or are they screened yearly? In the latter case, the current first screening might be more effective than subsequent screenings as (most) patients at risk will already be detected at a previous screening. What would be the consequences for the cost-effectiveness? Please elaborate on this.

RESPONSE: Thanks for this comment, that we think it is an interesting reasoning. To carry out our trial, primary care attendees were randomly selected and all were evaluated for their risk of depression every 6 months, prior to the respective intervention visit with their GP. It is expected that because of the preventive intervention predictD, both the risk probability and the specific modifiable risk factors will vary over time. We still do not have data to give a recommended frequency to estimate the risk of depression (3, 6, 12 months or more?) in the practice; this will depend on the sensitivity to change of the instrument (predictD risk algorithm), its acceptance by patients and GPs and their associated costs. We have added a commentary to the discussion (Implications for practice section) about what you are commenting on:
“Our study showed that universal prevention of depression in adults, through the predictD intervention and implemented by GPs, has a high probability of being cost-effective compared to usual care. However, with our study we cannot know if this type of universal prevention would be more cost-effective than other types of primary prevention (selective or indicated). Further studies comparing these types of prevention and different frequencies of risk evaluation are need. GPs generally perceive that have very little time to do preventive activities, and they may agree to do so only for high-risk patients. Another possibility would be to provide universal prevention to patients by mean of a scalable and cheap strategy (e.g. through APPs and smartphones), reserving the intervention of GPs only in cases of high risk of depression. Our research team is currently conducting a new trial with the latter strategy (the e-predictD study).”

- Do all patients have individual risk factors after screening with the PredictD algorithm? In a general population I would not expect this. If there are no risk factors, I assume that no plan have to be made. Which percentage of the patients do have risk factors and need a plan to manage those individual risk factors? How much time does it take to make such a plan following a biopsychosocial framework? Is this time included in the cost of the intervention shown in table 1? Please give a more detailed description of these intervention costs.

RESPONSE: We have added the following text at the method section: “The PredictD algorithm is composed of 12 risk factors: six are patient characteristics or past events (sex, age, sex*age interaction, education, childhood physical abuse, and lifetime depression) and six refer to current status (SF-12 physical score, SF-12 mental score, dissatisfaction with unpaid work, number of serious problems in very close persons, dissatisfaction with living together at home, and taking medication for stress, anxiety or depression). The "predictD algorithm" provides, in addition to the quantification of the overall risk of depression, knowledge of those risk factors influencing a given patient that could guide a possible preventive intervention” (Page 7). All the participants presented different levels of risk (from almost 0 to 1). For instance, just by the fact of being a women the risk will be higher than being a man, in spite of having the same score in the other factors. Similarly, older age was a protective factor. We based our study on the level of risk together with an explanation of which factors explained this level. Each GP received (at baseline, 6 and 12 months) reports with information on risk from each of approximately 24 (non-depressed) patients enrolled at baseline. If a patient had a moderate or high risk (2nd or 3rd tertile), the GP arranged an appointment with the patient aiming to communicate his or her risk of depression. If a patient had a low risk (1st tertile), the GP could choose to have a telephone or face-to-face consultation in accordance with the patient’s preference. Only 4(1.4%) patients at baseline, 99(3.2%) at 6-months and 115(3.9%) at 12-months received it by telephone. Face-to-face visits lasted approximately from 5 to 15 minutes and this time was proportional to the level of risk. Therefore, only 3 visits were mandatory. If the GP considered that the complexity of the case would require more visits, it was proposed to the patient. The patient at his own request could also propose to his GP new visits. All visits that were made, compulsory and optional,
were taken into account for costs. In addition to communicating the risk of depression to the patient both, patient and GP, tried to agree on a personalized plan for prevention of depression and the GP aimed to draw on and integrate various aspects of primary care that contribute to prevention of depression. These include: previous knowledge and relationship with the patient, establishment of a basic psychotherapeutic relationship, family-oriented practice, social interventions, community referral by GPs, and management of physical problems. We encouraged GPs to invite patients to make suggestions during the interview about strategies, attitudes and behaviors they were already using to prevent depression. GPs then positively reinforced those they considered the most appropriate to prevent depression for each patient. Even when the risk was very low, the GP and the participant discussed why the risk was low and the GP reinforced those behaviors and/or attitudes that were protective factors. A more detailed explanation of the predictD intervention can be found in the protocol (reference 14) and an example of interview GP-patient can be viewed on a 5-minute videotape. The participants in the video were researchers simulating this interview: https://www.youtube.com/watch?v=kbFhVVDNFGE

We have added a more detailed description of these intervention costs at the end of the cost section: “The intervention was embedded in current practice. Participants in the intervention group were required to meet at least 3 times during the intervention (at baseline, 6, and 12 months): in each of the three GP-patient interviews the GP communicated to the patient specific and updated information on his/her risk of depression at any time and both, patient and GP, tried to agree on a personalized plan for prevention of depression. These visits lasted approximately from 5 to 15 minutes and this time generally was proportional to the level of risk. If the GP considered that the complexity of the case would require more visits, it was proposed to the patient. The patient at his own request could also propose to his GP new visits. All visits that were made during follow-up, compulsory and optional, were taken into account for costs.

- Table 1: why do the individual cost items not sum up to the total costs?

RESPONSE: This is because we have calculated the mean costs with GLM, using a gamma and log approach, in order to take into account the skewed distribution of the costs.

Especially in case of the mental health care costs in the control group, as both the outpatient mental health visit costs and antidepressant costs are lower than in the intervention group, but total costs are much higher?
RESPONSE: This is because the total mental health costs included the costs related to inpatient hospitalization. There are only 8 patients who were admitted, 7 for the control group and 1 for the interventions group.

What is meant by total public health costs, are these total hospital costs or total healthcare costs?
RESPONSE: We meant total healthcare costs. We have changed this.

The number of patients admitted in an inpatient psychiatric care unit are low, but for how long are they admitted, as long admission may still result in considerable costs, please add to the footnote in table 1.
RESPONSE: We have added this to the footnote. The mean length of stay was 10 days.

- Table 2 and page 12: One or two decimal places in reporting ICERs suggests too much precision, round to whole numbers. In table 2 'Dominant' should be added for the ICER (cost per QALY gained) from the societal perspective.
RESPONSE: We have changed this.

- The combination of table 2, figure 2 and 3 seems a bit overdone as they all have overlapping information. In figure 2 titles on the axes are missing. What kind of scale is used on the x-axis of figure 3? It seems to start as a logarithmic scale, and to continue as a normal scale? This is somewhat confusing.
RESPONSE: The x axis does not use any scale. We have just placed different values of Willingness to Pay.

- In the conclusion it is stated the PredictedD intervention is likely to be perceived as cost-effective. However, an intervention is not cost-effective in itself, but in comparison with another intervention. Please add 'compared to usual care'.
RESPONSE: We have added this. Thank you so much.
Editorial Concerns

In addition to addressing the reviewers comments, please address the following editorial concerns:

a. In accordance with our polices (https://www.biomedcentral.com/getpublished/editorial-policies), all manuscripts need to include the required declarations. Please add declarations pertaining to 'Consent to publish'

b. The declaration pertaining to 'Availability of data and materials' refers to availability to the public, not the authors. Please amend accordingly

c. Please add a list of abbreviations

RESPONSE: We have added this information.