Author's response to reviews

Title: Prevalence of depression in adults with type 2 diabetes in the Basque Country: relationship with glycaemic control and health care costs.

Authors:

   Edurne Alonso Morán (ealonso@bioef.org)
   Altynai Satylganova (altynai.satylganova@gmail.com)
   Juan F Orueta (jon.orueta@osakidetza.net)
   Roberto Nuño-Solinis (nuno@bioef.org)

Version: 2 Date: 17 April 2014

Author's response to reviews: see over
Dear Editor:

We submit reviewers’ answers of the paper entitled “Prevalence of depression in adults with type 2 diabetes in the Basque Country: relationship with glycaemic control and health care costs” (MS: 829338042104119). The review article is submitted through the electronic system of the Journal.

The authors have answered point by point all questions suggested by reviewers. You can find information on the second page of this letter.

Correspondence regarding this submission should be directed to me at the following address:

Edurne Alonso Morán
The Basque Foundation for Health Innovation and Research (BIOEF), The Basque Institute for Healthcare Innovation (O+berri)
Torre del BEC (Bilbao Exhibition Centre), Ronda de Azkue 1
48902 Barakaldo Bizkaia (Spain)
Tel: +34944538500
Fax: +34944530465
Email: edurne.almo@gmail.com

Thank you very much for all your comments about our submission.

Yours sincerely,

Edurne Alonso Morán.
Reviewer's report

**Title:** Prevalence of depression in adults with type 2 diabetes in the Basque Country: relationship with glycaemic control and health care costs.

**Version:** 1

**Date:** 28 February 2014

**Reviewer:** waleed sweileh

**Reviewer's report:**

**Major Compulsory**

1. Do authors have access to the type of anti-diabetic medications used by the patients? At least monotherapy versus combination therapy!!
   We don’t have access to the type of prescriptions used by patients

2. The age classification is not convincing! Why not just do a simple analysis of age category as > 65 and < 65 years old!
   Yes, I have changed the analysis and results using only 2 groups.

3. The authors used Berkowitz et al for classification of HbA1c level. The authors need to ADA recommendation for HbA1c classification of good versus poor glycemic control and show the statistical outcome again
   In the Basque Country we have our own clinical practice guideline for type 2 diabetes mellitus but in this case, our cut-off value is the same that ADAs’ ([http://care.diabetesjournals.org/content/37/Supplement_1/S14.full.pdf+html](http://care.diabetesjournals.org/content/37/Supplement_1/S14.full.pdf+html)), values equal or less than 7% are classified as a good control and values over 7% are classified as bad control.
   But the reason why we have used graduation of HbA1C that is higher than 7% (like we had groups <7%, 7%-8%, 8%-9% and >9%) is because literature shows that higher numbers of HbA1C is correlated with higher numbers of complications (for example page 12 of this [http://www.who.int/diabetes/publications/report-hba1c_2011.pdf](http://www.who.int/diabetes/publications/report-hba1c_2011.pdf)). Thus, our hypothesis (based on literature review) was that higher is the HbA1C, the higher is the possibility of having comorbid diabetes and depression, and incurred costs. Plus, according to the literature, the higher HbA1C values are associated the treatment due to complications. Thus, we think that the point is not in ADA recommendations, because in that case we will have only two groups (controlled and not-controlled), but to see how number of comorbidity grows together with growth of HbA1C. Moreover, we consider those below 7% well controlled, so we analyze what happens when HbA1C grows and if there is a direct relation with prevalence of depression and higher costs.

**Level of interest:**

An article whose findings are important to those with closely related research interests

**Quality of written English:** Acceptable
**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:** I declare that I have no competing interests

---

**Reviewer's report**

**Title:** Prevalence of depression in adults with type 2 diabetes in the Basque Country: relationship with glycaemic control and health care costs.

**Version:** 1

**Date:** 1 March 2014

**Reviewer:** Tiago N N Munhoz

**Reviewer's report:**

Research article

Prevalence of depression in adults with type 2 diabetes in the Basque Country: relationship with glycaemic control and health care costs.

Edurne Alonso Morán, Altnay Satylganova, Juan F Orueta and Roberto Nuño-Solinis

BMC Public Health (Submitted:2014-02-03)

**Discretionary Revision**

- It seems to me that it is not necessary to use Cramer’s V test. I recommend excluding this sentence from methodology: “Cramer’s V test proves the strength of such association with values between 0 and 1 (being 1 the perfect association and 0 no association).”
  
  We have been deleted from the analysis description.

- I suggest to use some of this references (and others) in your discussion:
  
  
  
  

We have added the second and third to discussion section.

Minor Essential Revisions

- I suggest to change “Measures of association” for “statistical significance”.
  We have changed it in analysis section by statistical significance
- one decimal places is enough for prevalences and means.
  We have changed by one decimal
- I suggest “associated factors” instead risk factors.
  We have changed by associated factors.

Major Compulsory Revisions

METHODS

- 1 Paragraph:
  - Was not possible to understand about PREST (Osakidetza): Does PREST cover all Basque population? Please make it clear at the beginning.
    We have clarified this information in the beginning of methods section: “the database (PREST) contains information collected since 2007 and includes every citizen covered by the public health insurance in the Basque Country, regardless of whether they had made any contact with or use of the Basque Health Service. That is, practically all the inhabitants of the Basque Country are included.”
  - The information about PREST is insufficient. And you should report that mental health services were not included in your analysis (you only talk about it in limitations section).
    We have added more information about PREST and we have also clarified that mental health services were not included (beginning of methods section). “PREST was assembled combining information on individuals from several sources (primary and specialized health care records, computerised files from day hospitals, emergency departments’ registries, hospital discharge reports, and census data) and involves the following data:
      - Demographic data: age, sex and census area of residence
      - A register for any contact of the patients with Osakidetza, containing the type of service provided, codes for diagnoses and information on significant procedures (such as dialysis, radiation therapy, chemotherapy or rehabilitation therapies).
      - Prescriptions in primary care
Services for which no information was available comprise: Mental Health (both admissions and outpatient visits), hospital-at-home services, medical transport, prostheses and other equipment delivered to patients at home”

- If I’m not mistaken, you are analysing data between 2010/2011. Why are you using ICD9? ICD 10 is available, at least, for the last 10 years (http://www.who.int/classifications/icd/en/)
  In Spain, according to the Health Ministry’s regulation (http://www.msssi.gob.es/en/estadEstudios/estadisticas/sisInfSanSNS/tablasEstatisticas/home.htm), diagnoses and procedures are codified by means of the ICD-9-CM system, while the Anatomical Therapeutic Chemical coding system (ATC) is used for drugs prescription

- 2, 3 and 4 paragraphs
  - You have to explain in details your definition of depression and diabetes (see 5 paragraph comments).
  
  We did it:
  “So, firstly, we identified diabetes mellitus patients who were ≥ 35 years old because diabetes type 2 is uncommon at younger ages (24). It was considered to have diabetes mellitus, those who had either being prescribed with oral anti-diabetics drugs or had any kind of diabetes-related diagnose or its complications, before 31st of August 2011; of them were excluded all people who had one or more diagnosis related to type 1 diabetes or with non-specified diabetes treated exclusively with insulin.
  Secondly, it was searched patients with depression at the discretion of GPs or other doctors outside the mental health services. Often these health professionals do not perform in their diagnoses a description precise enough to allow differentiation between different types of depressive disorders and, therefore, we accepted the presence of any of them. Since the database contains information of a period of 4 years, in order to exclude those episodes that are not currently actives, we looked only for patients who had a diagnosis of depression in last 12 months or had repeated prescriptions of antidepressant during at last four months in the last year. ”

- It is not clear what is meant by “whether or not they had had any contact with Osakidetza during such period. We have changed it by: “independently if patients had visited doctors or health centers of Osakidetza during such period”.

- Without clear description about your source of information, population, etc. it is difficult to determine internal and external validity.
  We agree that this information is really important, so we have added more information about it in methods section.

- You excluded individuals under 35 years old. You should, at least, discuss the implications and limitations about it. Depression onset, usually, occur before this age.
  Yes, but the study centers on people who have diabetes type 2. Second, we check how many of them have also depression.
Active episode? Any depression? It’s Major Depressive Episode (MDE), Major Depressive Disorder (MDD), dysthymia? Both? Do you have sufficient information to distinguish unipolar to bipolar depression?

We have explained before that “it was searched patients with depression at the discretion of GPs or other doctors outside the mental health services. Often these health professionals do not perform a sufficient description in their diagnoses to allow differentiation between different types of depressive disorders and, therefore, we accept the presence of any of them.”

Based on your data, is it possible to determine if an active depression, in your definition? “we looked for patients who had a diagnosis of depression in the previous year and had also repeated prescriptions of antidepressant over the last four months in the past year.”(comment 3 to 7)

As we answered before, in order to exclude those episodes that are not currently actives, we looked only for patients who had a diagnosis of depression in the last 12 months or had repeated prescriptions of antidepressant during at last four months in the last year.

Antidepressants are not prescribed exclusively for depression. You should, at least, discuss the implications and limitations about it.

It was added as a limitation of the study

Although diabetes is not my area of expertise but it is not clear to me your main reason of using the HbA definition from Berkowitz et al. paper.

In the Basque Country we have our own clinical practice guideline for type 2 diabetes mellitus but in this case, our cut-off value is the same that ADAs’ (http://care.diabetesjournals.org/content/37/Supplement_1/S14.full.pdf+html), values equal or less than 7% are classified as a good control and values over 7% are classified as bad control. But the reason why we have used graduation of HbA1c that is higher than 7% (like we had groups <7%, 7%-8%, 8%-9% and >9%) is because literature shows that higher numbers of HbA1c is correlated with higher numbers of complications (for example page 12 of this http://www.who.int/diabetes/publications/report-hba1c_2011.pdf). Thus, our hypothesis (based on literature review) was that higher is the HbA1c, the higher is the possibility of having comorbid diabetes and depression, and incurred costs. Plus, according to the literature, the higher HbA1C values are associated the cost treatment due to complications. Thus, we think that the point is not in ADA recommendations, because in that case we will have only two groups (controlled and not-controlled), but to see how number of comorbidity grows together with growth of HbA1C. Moreover, we consider those below 7% well controlled, so we analyze what happens when HbA1C grows and if there is a direct relation with prevalence of depression and higher costs.

What is the reason of using “the last measuring registered in the period analyzed was used” instead of the mean, for example?

It is easier for us to draw from health sources the latest data recorded in each period (one period for Osakidetza is from 1st September to 31st August).
Is the last measure of HbA1c inside the study period?
Yes, it is the last measure taken to the patients in the period analyzed.

I cannot understand how you created the ADG variable and why you described it in methodology if, seen to me, that it was not used/it usage was not explained.
In Methods section we have explained that “Basque citizens are annually classified using the Johns Hopkins Adjusted Clinical Groups (ACGs) case-mix system (23). Such system enables health problems to be identified from diagnoses and prescriptions and patients to be categorized according to their health care needs”. Moreover, in variables and analysis section it is described the ADG as “In order to take into account the presence of comorbidities, we used Aggregated Diagnosis Groups (ADGs), which is the component of the Johns Hopkins ACG case-mix system. ADG consists of 32 categories, specifically designed to aggregate diagnoses into groups with similar severity, duration of condition, and treatment needs. A more complete description of this methodology can be found in the bibliography (23).”

Please explain how "low educational attainment among young people (16–29 years)" was used to create the deprivation index when your studied sample included only those who were 35 years and older.
We have changed the definition to be more understandable “The deprivation index of the census tract (median population size is equal to 1,200 inhabitants) of residence proposed by the MEDEA project (26) was used as a proxy for individual socioeconomic position. Its design allows for the estimation of socioeconomic and environmental inequalities among inhabitants by census code in Spain. Its calculus takes into account the percentages of residents who are manual workers, unemployed, temporary employees, or have not finished primary school, overall and also specifically among young people, given the most recent Census (2001) available. The first quintile is richest people and the fifth quintile is poorest people. The deprivation index was elaborated and published in 2008”

Is this index generated with information measured in 2001? If yes, some population characteristics should be different from the study period. You should, at least, discuss the implications and limitations about it.
We have added as limitation that “Another limitation refers to the social variable used (deprivation index) which, given its ecological character, may underestimate the contribution of individual socioeconomic characteristics. Such index was used to classify geographic areas in quintiles and, although is probable that actual data relative to unemployment and others have changed since last census, we consider that the relative situation of each area with respect has to the other ones has not suffered a substantial alteration.”

Health cost was assessed for the mentioned health services. Which references do you use to estimate this cost? It is not clear.
We have rewritten both paragraphs explaining better how we have estimated them “Health care utilization costs were also estimated for the period of study. For variables such as visits to Accident & Emergency (A&E), rehabilitation sessions, outpatient care, primary care visits, laboratory tests and radiological examinations ordered by primary care, and various outpatient procedures (dialysis, radiotherapy and chemotherapy); the number of services used by each patient was multiplied by standardized costs (the average cost of each service provided to a patient treated in Osakidetza, according to calculations made by the aforementioned organisation). We estimated healthcare costs of primary care prescriptions recorded in electronic health records based on the market value of the drugs. The costs of hospitalisation and outpatient surgery were calculated in relation to their cost-weights in the corresponding Diagnosis-Related groups (DRGs). ”

- If mental health services were excluded, you should, at least, discuss the implications and limitations about it on your “health costs” variable, analysis and results.
  We have included this in limitation of the study “The non-acute mental health services are not included so this could affect the total cost estimated.”

- **10 paragraph**
  - Same comments for 9 paragraph.
    The same answer as two before
  - Standardized cost? Based in what?
    The same answer as two before

- **11 paragraph**
  - It is not clear why and when you used Wilcoxon Mann Whitney test.
    To check if there are differences, in the comorbidity distribution of depression and type 2 diabetes, between genders. Wilcoxon Mann Whitney test is used when the categorical variable (in this case gender) is two groups (males and females)
  - If socioeconomic status (SES) are in quintiles, why to use Kruskal Wallis test?
    To check if there are differences, in the comorbidity distribution of depression and type 2 diabetes, between SES groups. If categorical variable (in this case SES) was only two groups it uses Wilcoxon Mann Whitney test but if there are more than 2 groups, we have to use Kruskal Wallis test.
  - Comorbidity between which variable? Depression and diabetes? Depression, diabetes and SES?
    Between depression as comorbidity of type 2 diabetes and SES groups.
  - This sentence is wrong. Recommend to exclude. “Pearson test and Fisher’s exact test prove if it exist an association between the two variables and if p value obtained from test is smaller than 0.05 there is not an association”. We have deleted from analysis section.
  - Was adjusted analysis used (like you said in abstract?)? If you do so, please explain in methodology and indicate at tables/results.
    We explained at analysis section that both regression analyses were adjusted by gender, SES, agegroup, hba1c levels and ADG (variable for level of illness).
In the logistic regression, the dependant variable is having or not depression as comorbidity and the independent variables (variables for adjustment) are gender, SES, agegroup, hba1c levels and ADG. In the linear regression analysis, the dependant variable is the total cost and as independent variables: having or not depression as comorbidity, gender, SES, agegroup, hba1c levels and ADG.

**RESULTS**

- 1 paragraph
  - Is 1473943 the whole population of Basque Country who were 35 years old?
    It’s not clear to me if you used the whole study population.
    1,473,943 people living in the Basque Country are 35 years or more, from them 126,894 have diabetes type 2 and finally from the 126,894 diabetics, 12,392 have depression as comorbidity. So, my study population are the 126,894 (with or without depression).
  - You should report the prevalence of type 2 diabetes AND depression [0.8% CI95% ? (12,393 / 1,473,943).
    I did it. “12,392 patients had both diagnoses at the same time in the period analyzed (prevalence 0.84%, CI 95% [0.826, 0.856])”
  - I could not understand what the denominator for this prevalence: 9.76%
    As we said in the manuscript “the 9.76% of patients with type 2 diabetes had also depression as comorbidity”. We speak about percentage, so from all diabetics, 9.76% have depression. (12,393/126,894).
  - Tables 1 and 2. I suggest showing all characteristic under study on Table 1 (sex, gender, deprivation index, etc.)
    We joined table 1 and 2 for all characteristics.
  - Table 3 is unnecessary.
    We have deleted it from manuscript but I spoke about it in results (as before but without table and Cramers’ v test).
  - Table 4 and 5 have insufficient titles.
    We have explained more the tables 4 and 5, now are named tables 2 and 3.

- 2 paragraph
  - Table 1 - Suggest to describe all characteristics under study and P value for each /whole variable (not for subgroups).
    We have described all characteristics from table 1 and we have added CI for prevalence and age

- 3 paragraph
  - It is incorrect to say incidence. You are measuring prevalence.
    We have changed it by prevalence.
  - P value under 0,001 should be reported p<0,001
    Yes, it was a mistake, we have changed by p<0.001

- 4 paragraph
  - I suggest to remove Table 2. If the authors consider that the information is really relevant (I think that is not), please, show significance tests. At least, CI95%.
We have joined to table 1, characteristics of the population.

- “Course of type 2 diabetes” sounds like incidence rate. I suggest using other expression.
  We have changed the expression by: There was a tendency of men to be able to adhere to recommended HbA1C values better than women in all defined groups and in both populations.

- 3 decimal places is enough for P values
  We have changed by 3 decimal.

- Table 3 is unnecessary. Why is this information important? I cannot understand.
  We have deleted it, we only mention at results.

- Suggest removing “p value of Pearson chi2 test was 0.365” for p=0.365.
  We have changed it.

- Suggest removing “We also saw with Cramer’s test that there was no association between both variables because the value is too near to 0.”
  We have removed it.

- 5 paragraph

  - This sentence (or similar one) should be described in the method section: Logistic regression analysis was performed to explain depression in type 2 diabetes population by the factors: HbA1c, sex, age band, ADG and deprivation index.

  It was explained at analysis section before, we paste it:
  “Logistic regression analysis was performed to explain depression in patients with diabetes type 2 through HbA1c categories, sex, age band, ADGs and deprivation index. Finally, linear regression was used to explain the total cost by depression, HbA1c categories, sex, age band, ADGs and deprivation index”.

  - Probability is different from odds. You are modelling odds.

  - You should not report p values for each subgroup. Or at least should also include your overall model p-value. (Table 4 and 5)
    We have removed p values from tables; only appear values and confidence intervals.

  - “[...] was a risk factor of having depression (OR=1.1; p=0.029)”. Please report CI95% instead P value.
    We have changed it.

- 6 paragraph

  - “A regression analysis was performed in order to see if there were statistically significant differences in mean cost between groups with and without comorbidity of depression.”

  Where are the groups with and without comorbidity? You are using depression as an dependent variable.

  We use depression as a binary variable 0 and 1, 0 are people with type 2 diabetes but without depression and 1 are people with type 2 diabetes and with depression.
Completely unnecessary: “Total cost as dependent variable and gender, deprivation index, age band, ADGs and HbA1c as independent variables (see table 5).”
We have deleted it.

“The R2 of the model was 0.4125 (41.25% of the variance explained).” Please explain to your readers what it means to your results.
We have explained what this means “The $R^2$, which indicates how well data points fit a statistical model, was 0.413. Then, independent variables explained the 41.25% of the total variance.”

Are all variables in the model? Are you adjusting for all variables? Please explain this in the method section. You should not report p values for each subgroup. Please, report P value for whole variable. (Table 4 and 5)
We have explained at analysis section which variables were used “Logistic regression analysis was performed to see the associated factors to depression in patients with diabetes type 2 using as independent variables HbA1c categories, sex, age band, ADGs and deprivation index. Finally, linear regression was used to see how many, on average, of the total cost is imputable to depression, HbA1c categories, sex, age band, ADGs and deprivation index.”

You should correct the title of Table 5 to "Linear regression..." and change the header of the Table 5 from “odds ratio” to cost or a similar term.
We agree; we copied from the other table. We have changed both, the title and the header.

7 paragraph

“The factors age band and deprivation index were not statistically significant for explaining the total cost, except from ages between 70 and 84 years with 35-39 age band as reference”.
You are talking about linear regression. The word “explaining” seems incorrect to me.
We have changed it by “The factors age band and deprivation index were not statistically significant in the model, except from...”

Moreover, women spent 164€ less than men, being this difference statistically significant (p<0.001).”
Please, explain in the discussion section why it is interesting.
We have added to discussion that “It is interesting to say that: although women showed more percentage of having depression than men, women, on average, spent less.”

**DISCUSSION**

Discussion is insufficient. You have to explain your findings. You have to discuss about bidirectional association between diabetes and depression (even in the introduction section)
We have extended discussion section and we have spoken about bidirectional association of both.

1 paragraph:
• Course: it’s not a prospective study
  We have paraphrased the sentence, now it reads: “Moreover, men showed better HbA1C values and thus better controlled Type 2 Diabetes than women in all age groups.”
• 2 paragraph
  o Probability is different from odds. You are modelling odds.
    We agree, we have changed by “Also, it was observed that for every man who suffered comorbidity, there were 2.6 women who suffered it, which was consistent with the prevalence obtained.”
• 3 paragraph
  o Please, explain these results. You are just repeating the results section.
    We have explained our findings trying not to repeat the results
• 4 paragraph
  o There is other explanation in the literature (see references at beginning)
    We have added some references
• 5 paragraph
  o Please, explain these results. You are just repeating the results section
    We have rewritten the paragraph “The total cost was associated with a bad control of HbA1c, sex, age and depression as comorbidity in patients who have diabetes type 2. It is interesting to note that: although women showed more percentage of having depression than men; women, on average, accounted for less healthcare costs.”
  o Completely unnecessary: (pearsons x 2)
    We have removed
• 6 paragraph
  o In my opinion, your result show that, on average, people with depression spend 510€ more than people without depression. This is different from “With regard to cost analysis, the linear regression has shown that depression as comorbidity explained 510€ the total cost incurred by patient.”
    We agree, we have changed it by “on average, people with depression spend 510€ more than people without depression”
  o Also, I could not understand your explanation: “Since, regression included the burden of diseases (ADG); this difference could be attributed to depression and it could be explained by number of reasons such as cost implications of depression treatment, higher occurrence and consequent cost of type 2 diabetes complications in comorbid group (28).”
    We have explained more this “Since, regression included ADGs which control the burden of diseases; this difference could be really attributed to depression (other illness which can have patients are control by these variables). So, this difference could be explained by number of reasons such as cost implications of depression treatment, higher occurrence and consequent cost of type 2 diabetes complications in comorbid group [30]. “
    What we meant is that ADG controls in regression analysis other illness that can have patients, so if depression as alone variable, on average, was 516€, we
can really attributed this difference to depression and not to other illness which patients can have

- You say: “Since, regression included the burden of diseases (ADG);” but your Tables says: “ADGs are not represented in this table.” I am confused. Please explain in details, including in methods section.
  - Yes, they were included in the analysis but they were not represented in the tables because they are 32 variables. I have explained it above tables of regressions.

- 7 paragraph
  - “This gave us the possibility to include the entire known diabetic population in the geographical area of our study, thus avoiding selection bias.”
  - For me, it’s the opposite of “The limitations of this study were that administrative databases only contained information about conditions for which people looked for medical attention” (9 paragraph) Please, explain.
    - This is a limitation for all studies because always there are people who have some illness but they do not know but we have explained this in limitations: “The limitations of this study were that administrative databases only contained information about conditions for which people looked for medical attention. Therefore, the prevalence of diabetes and depression may have reflected only known cases, and excluded the presence of the cases, that were not known by patients or doctors.”

- 8 paragraph
  - “regressions were controlled through inclusion ADGs in our analysis, variable that controls the burden of diseases.” Were they adjusted for? I am confused. Please explain in details, including in methods section.
    - We have explained in variables their definition and in analysis that we use them in regressions. But they are 32 variables so we did not include their values (of the regression analysis) in the tables.

- 9 paragraph
  - You have many limitations. Some was pointed out, some are impossible to determine without more clarifications.
    - We tried to add all limitations that the study has.
  - Like you said in Introduction “Several studies have shown that the following socio-demographic risk factors can be predictive of having depression among type 2 diabetes patients: female sex, younger age, lower socioeconomic status, being unmarried, poor social support, certain ethnicities.” However, being unmarried or certain ethnicities and having poor social support are not analysed. It could bias your result? How? You also should discuss this.
    - We did not choose the population based on these characteristics. We do not have access to this type of variables, so we could not include them in the analysis. We have added this to the limitations of the study: “Finally, we had not access to information about being married, social support and ethnicity.”

CONCLUSION
• Your data and analysis are insufficient to support the inferences and conclusions that you make.
  
  We have changed conclusion section, we have rewritten it.

**Level of interest:** An article of limited interest

**Quality of written English:** Acceptable

**Statistical review:** Yes, and I have assessed the statistics in my report.

**Declaration of competing interests:** I declare that I have no conflict of interest