Author's response to reviews

Title: Validation of Diabetes Mellitus and Hypertension diagnosis in computerized medical records in Primary Health Care.

Authors:

Carmen de Burgos-Lunar (carmenblunar@hotmail.com)
Miguel A Salinero-Fort (Miguel.salinero@salud.madrid.org)
Juan Cárdenas-Valladolid (jcardenas.gapm04@salud.madrid.org)
Sonia Soto-Díaz (sosoto.gapm04@salud.madrid.org)
Carmen Y Fuentes-Rodríguez (carmeny.fuentes@salud.madrid.org)
Juan C Abánades-Herranz (jabanades.gapm04@salud.madrid.org)
Isabel del Cura-González (icura.gapm09@salud.madrid.org)

Version: 2 Date: 3 June 2011

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

Please find attached the corrections for the manuscript entitled “Validation of Diabetes Mellitus and Hypertension diagnosis in computerized medical records in Primary Health Care”.

According to the comments from the reviewers, we have made a substantial revision and reworked the manuscript. Patients who did not have enough information to validate the diagnoses were excluded from statistical analysis in order to avoid misclassification bias. The calculations were redone and the results are slightly different from the first version of the manuscript. The wording was clarified and we have improved the style of written English. The authors accept the uniform requirements for biomedical journals. All authors have seen and approved the content, and contributed significantly to the work.

Sincerely yours,

Carmen de Burgos Lunar, MD. MPH.

Hospital Carlos III
Calle Sinesio Delgado 10
28029 Madrid, Spain
Reviewer's report
Title: Validation of Diabetes Mellitus and Hypertension diagnosis in computerized medical records in Primary Health Care.
Version: 1 Date: 25 January 2011
Reviewer: Claudio Cricelli

Reviewer's report:

Major
According to the applied validation method it is not clear n how many cases (%) the available information were sufficient to perform an acceptable validation (e.g. with regard to DM2 in how many cases there were data regarding blood glycemia levels)

We now provide this information in the sections of methods and results, in new paragraphs:

“Furthermore, some patients also had to be excluded from our analysis because even though they met the inclusion criteria, they had not at least one plasma glucose measurement (7.3%) or two BP measurements (22.9%) in their CCR for the validation of DM and HT, respectively.”

“The 7.3% of patients from sample 2 (without DM code) had to be excluded because there was not at least one fasting plasma glucose. They were 64.5% males with a mean age of 73.6 (SD 14.4) years. There were significant differences in mean age between patients excluded and not excluded.
The 22.9% of patients from sample 4 (without HT code) had to be excluded because there were not at least two BP measurements. They were 36.1% males with a mean age of 69.1 (SD 11.6) years. There were no significant differences in mean age and female proportion between patients excluded and not excluded from the sample.”

The validation algorithm is not specified. The clinical definition identifies the criterias, the algorythm clarifies the imputation pathways of a specified diagnosis

We have included the validation algorithm in Figure 1.

Minor
The authors apply a formula to estimate the Real Prevalence (Real Prevalence = (registered prevalence +E - 1 )/ (S +E - 1 ) of a pathology starting from estimated prevalence and from sensitivity and specificity data . This is very interesting but these formula needs validation through at least on bibliographical reference.
We provide a justification of adjustment and a bibliographical reference.

“When diagnostic tests are applied in population, the proportion of those testing positive (apparent prevalence) can not be used as an estimation of the prevalence of a disease in that population, because the Sn and Sp of these tests are usually less than 100%. Thus, the proportion of individuals with a positive result includes false positive cases and excludes cases that are false negatives, so in order to estimate the true prevalence of a disease from diagnostic tests it is required to adjust for the misclassification resulting from the Sn and Sp of the used test. In this study we have used the formula proposed by Rogan and Gladen for this adjustment. True prevalence = (apparent prevalence + Sp - 1) / (Sn + Sp -1) “


With reference to the results the general population Hypertension prevalence is substantially lower than most current studies’ estimates. Explanation provided in the article are not fully convincing.

We have provided additional arguments to explain our results in the prevalence of hypertension in the general population

OLD PARAGRAPH “The real prevalence of HT obtained in our study, 17.69%, is also lower than the findings for all of Spain, situated around 35% in the adult population and in the Madrid Community around 29.3%. The differences observed in the magnitude of the prevalences could reflect, not only differences in other factors such as the prevalence of obesity, but also important methodological differences. For this reason, the PREDIMERC study has been done with volunteers, with a global response rate of 56.4%, which, as showed before, was able to produce a selection bias. Furthermore, it should be taken into account that the standard of reference of our study is the fulfillment of the diagnostic criteria that can be tested with information collected in the CCR. The false positives could be assumed to be diagnostic errors, that will have not been verified, leading to a subestimation of the prevalence in our study.”

“The real prevalence of HT in our study (17.14%) is similar than the obtained with the BIFAP database (16.1%) and in the Spanish National Health Survey (18.89%) but also
lower than the findings in other studies in Spain, around 35% in the adult population 30 and 29.3% in the PREDIMERC study 27.

This differences observed in the magnitude of the prevalences could be due to patients included in PEDIMERC were aged between 30 and 74, whereas our study includes all those aged 18 and over. If we had selected people aged 30 or over in our database, the true prevalence of DM would have been 6.87% and 23.72% of HT.

Moreover, PREDIMERC has been done with volunteers, with a global response rate of 56.4%, which may produce a selection bias as it was mentioned before. Furthermore, we can not assure that false positives were misdiagnosis, simply that the verification of diagnostic criteria could not be met which may lead to an underestimation of the prevalence in our study.“

The article’s introduction and the methods applied need to be expressed in a linear way. (less confusion )

The wording was clarified and we have improved the style of written English.

No information is given regarding the minimum criteria for the data registration required for selecting the clinical records (...The study presents some limitations which are, on one hand, that the information included may not be completely exhaustive) For example : in order to include clinical patients records in a study at least one or two years active data recordings are basically required.

Considering this suggestion, we have included the paragraph: “According to this situation, at least one or two years of active data are basically required in order to include clinical patient records in a study.”
Reviewer: Jennifer H Garvin  
Reviewer's report:  

Major Compulsory Revisions  
The manuscript requires substantial revision and the quality of the research cannot be assessed as written. Please find high level comments that should be used to rework the publication. The abstract should be revised with the aim of making a good initial impression which will engage the reader. The abstract should be substantially revised. The same kinds of issues (listed below) are found throughout the rest of the manuscript as well and should be revised as applicable.

1. Assertions made such as "To obtain epidemiological information from representative samples of the population, it is necessary to turn to secondary sources..." should be tempered to say something such as "To obtain epidemiological information from representative samples of the population, it may be necessary to turn to secondary sources...". Please review similar strong assertions throughout the paper and make revisions as suggested.

2. Proper names should be capitalized and it is not clear why "Primary Health Care" is capitalized. If it is a proper name, please add clarification to that effect in the publication.

3. The abstract contains a sentence which says "contain clinical-administrative information of an elevated percentage of the population". I am not sure what is meant by this statement.

4. In sum, the publication begins with confusing and grammatically incorrect sentences and it is not possible to comment on the research with the abstract and the subsequent sections of the manuscript in its current form. Generally, wording should be clarified and punctuation should be edited. For example, "diagnosis registered" should likely be something to the effect of "diagnosis documented". Similarly, the phrase "random sampling" should likely be "random sample". In terms of punctuation, the use of the comma in statistical values should likely be a period (99,53 vs. 99.53). Abbreviations are used at times (OMI-AP, BIFAP) without being initially spelled out. Attention should be given to appropriateness of wording. Some words such as "exhausivity" are not correct.

The description of the methods is not clear. For example, in the section related to the study population the authors state, "We selected the episodes by means of diagnostic codes, without paying attention to the labels. Therefore, selection bias may occur in those cases in which the professional registering the episodes modified the label." It is unclear what is meant by "label" and how the label would impact the research methods.
The wording was clarified and we have improved the style of written English.

- We have changed the comma to a period for decimal values.
- We have removed the confusing and incorrect terms like "diagnosis registered", "diagnosis label" and "exhausivity".
- We have spelled out the abbreviations used in the manuscript: BIFAP (Base de Datos para la Investigación Farmacoepidemiológica en Atención Primaria), GPRD (General Practice Research Database)
- Capital letters are removed from where they ought not to be used
- Strong assertions were tempered