

## PEER REVIEW HISTORY

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

### ARTICLE DETAILS

<b>TITLE (PROVISIONAL)</b>	Calibrated prevalence of disabling chronic pain according to different approaches: a face-to-face cross-sectional population-based study in Southern Spain
<b>AUTHORS</b>	Cabrera-León, Andrés; Rueda, María; Cantero-Braojos, Miguel

### VERSION 1 - REVIEW

<b>REVIEWER</b>	Pier Francesco Perri Department of Economics, Statistics and Finance University of Calabria - Italy
<b>REVIEW RETURNED</b>	24-Sep-2016

<b>GENERAL COMMENTS</b>	<p>In the submitted paper, the Authors discuss the use of calibration to obtain Disabling Chronic Pain (DCP) prevalence estimates in Andalusia for the entire population and certain domains of interest. The calibration estimates - obtained under three different approaches - are compared with direct estimates obtained by the traditional design-based Horvitz-Thompson estimator.</p> <p>All in all, the paper is well written and motivated, the methodology well stated as well as the benefits, requirements and limitation of the calibration techniques. My opinion is favorable to the publication. I have just three minor questions for the Authors:</p> <p>1) Page 6, line 32. It is stated that "One of the most recent techniques to make the estimation...is calibration, [15]...". Indeed, calibration is not a recent technique since it has been introduced by Deville and Sarndal in 1992. Also model-calibration is a bit dated (Wu &amp; Sitter, JASA, 2001)</p> <p>2) When using linear calibration and the Chi-square metric, negative weights might occur. Did you find negative weights in your study and, eventually, how did you treat them?</p> <p>3) Looking at the results displayed in Figure 3 (page 30) and Figure 1 (supplementary material, page 33), I wonder whether the prevalence estimates are statistically different across the four different approaches for the domains and the population, respectively. Some estimates appear very close to each other. If not different, final conclusions should be modified. I suggest the Authors to consider/comment this point.</p>
-------------------------	--

<b>REVIEWER</b>	Failde, Immaculada University of Cadiz. Spain
<b>REVIEW RETURNED</b>	26-Oct-2016

<b>GENERAL COMMENTS</b>	<p>This is a well written paper, which addresses an important issue such as disability associated with the presence of chronic pain (DCP), with specific objectives, and which provides interesting information on different calibration methods that can be helpful to get a better estimation of the prevalence of DCP. Despite its interest, the article presents some aspects that need to be clarified.</p> <p>1.- The title might be confusing, as it could be thought that the data is obtained directly from the authors, but in fact what the authors actually do is to apply different calibration procedures to previous data from the Andalusian Health Survey.</p> <p>2.- In "Material and Methods", the authors describe the dependent variable according to the disability self-reported by the subjects interviewed in response to several questions from the survey. However, I believe that it would be helpful to include in Table 1 how this information was collected in the survey. This would clarify what is being asked. Also, when this variable is defined on page 8, the 3 categories that have been considered in the sample (Disabling Chronic Pain, Non Disabling Chronic Pain and Non Chronic Pain) should be defined, as it appears later on page 12.</p> <p>3.- On page 10, the authors provide a detailed description of the procedure for the selection of the calibration variables. However, it is unclear why after selecting the variables age and sex from a first model, they perform another multinomial logistic model from which educational level and employment status are also selected. Moreover, in this second model, a third variable for the adjustment is mentioned, but not specified. In the text (top of page 11), the authors talk about "other variables". This should be clarified.</p> <p>4.- The authors should include a table with the characteristics of the population according to the 3 categories or domains of the dependent variable. This could reinforce the decision of grouping the dimensions "nDCP" and "non CP" which, a priori, might not be similar.</p>
-------------------------	--

### VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

Reviewer Name: Pier Francesco Perri

Institution and Country: Department of Economics, Statistics and Finance, University of Calabria - Italy

Please state any competing interests: None declared

Please leave your comments for the authors below

In the submitted paper, the Authors discuss the use of calibration to obtain Disabling Chronic Pain (DCP) prevalence estimates in Andalusia for the entire population and certain domains of interest. The calibration estimates - obtained under three different approaches - are compared with direct estimates obtained by the traditional design-based Horvitz-Thompson estimator.

All in all, the paper is well written and motivated, the methodology well stated as well as the benefits, requirements and limitation of the calibration techniques. My opinion is favorable to the publication. I have just three minor questions for the Authors:

COMMENT 1: Page 6, line 32. It is stated that “One of the most recent techniques to make the estimation...is calibration, [15]...”. Indeed, calibration is not a recent technique since it has been introduced by Deville and Sarndal in 1992. Also model-calibration is a bit dated (Wu & Sitter, JASA, 2001)

RESPONSE 1: Thank you for your comments. We have changed that sentence as follows: “One of the most used techniques to make the estimation of prevalence more valid and accurate is calibration,[15]...”

COMMENT 2: When using linear calibration and the Chi-square metric, negative weights might occur. Did you find negative weights in your study and, eventually, how did you treat them?

RESPONSE 2: As we exposed in the first sentence of the sixth paragraph of the Statistical analysis section and of the third paragraph of the Discussion section, we applied Chi-square (lineal) and raking distances to calculate the calibrated weights for the three calibration approaches studied (marginal, crossing and model-assisted). We obtained practically the same estimations for marginal calibration and exactly the same results for crossing and model-assisted calibration. Effectively, using Chi-square, negative calibrated weights might occur. But, in our study, all the calibrated weights obtained using Chi-square distance were positive (the minimum values were 0.12878 for marginal calibration, 0.30422 for crossing calibration and 0.87844 for model-assisted calibration). Thus, we have included (marked) text in the third paragraph of the Discussion.

COMMENT 3: Looking at the results displayed in Figure 3 (page 30) and Figure 1 (supplementary material, page 33), I wonder whether the prevalence estimates are statistically different across the four different approaches for the domains and the population, respectively. Some estimates appear very close to each other. If not different, final conclusions should be modified. I suggest the Authors to consider/comment this point.

RESPONSE 3: Effectively, there are no statistically significant differences in the prevalence estimates (neither total nor domains). That may be due to the fact that the sample design of the Andalusian Health Survey, which is a very complete, intensive and large population survey, has been checked over the years (there are five editions of the survey) so that it represents population properly. However, as showed in our study, any calibration approach used produces better estimations in the sense of both adjusting the sample to the study population and reducing the sampling error. Therefore, we have included (marked) text in the first paragraph of the Discussion. We described the pro and cons of each calibration approach in the third, fourth and fifth paragraphs.

Reviewer: 2

Reviewer Name: Failde I

Institution and Country: University of Cadiz. Spain

Please state any competing interests: None declared

Please leave your comments for the authors below

This is a well written paper, which addresses an important issue such as disability associated with the presence of chronic pain (DCP), with specific objectives, and which provides interesting information on different calibration methods that can be helpful to get a better estimation of the prevalence of DCP. Despite its interest, the article presents some aspects that need to be clarified.

COMMENT 1: The title might be confusing, as it could be thought that the data is obtained directly from the authors, but in fact what the authors actually do is to apply different calibration procedures to previous data from the Andalusian Health Survey.

RESPONSE 1: Thank you for your comments. Effectively, the Andalusian Health Survey is a

multipurpose survey designed to evaluate the health of non-institutionalized adults in Andalusia and their utilization of health services. Since its first edition in 1999, Andrés Cabrera-León has participated in the design and development of this survey and, in its last edition, he promoted the addition of the items necessary to identify people with chronic pain and disability. Moreover, to avoid any confusion we propose erasing the second part of the title, rewriting it as follows: "Calibrated prevalence of disabling chronic pain according to different approaches". We hope this addresses your comment.

COMMENT 2: In "Material and Methods", the authors describe the dependent variable according to the disability self-reported by the subjects interviewed in response to several questions from the survey. However, I believe that it would be helpful to include in Table 1 how this information was collected in the survey. This would clarify what is being asked. Also, when this variable is defined on page 8, the 3 categories that have been considered in the sample (Disabling Chronic Pain, Non Disabling Chronic Pain and Non Chronic Pain) should be defined, as it appears later on page 12.

RESPONSE 2: Following your suggestion we have included a footnote in Table 1 about the dependent variable and more information about how chronic diseases were collected. We have also added a (marked) sentence regarding the categorization of the dependent variable.

COMMENT 3: On page 10, the authors provide a detailed description of the procedure for the selection of the calibration variables. However, it is unclear why after selecting the variables age and sex from a first model, they perform another multinomial logistic model from which educational level and employment status are also selected. Moreover, in this second model, a third variable for the adjustment is mentioned, but not specified. In the text (top of page 11), the authors talk about "other variables". This should be clarified.

RESPONSE 3: After selecting the variables age and sex from the first general model as auxiliary variables, we executed other alternative models to identify more auxiliary variables. Those models had as independent variables sex, age and other (third) variables not included previously in the general model. In this way, we were able to identify two other auxiliary variables (educational level and employment status). Unfortunately, only educational level (as well as sex and age) satisfied the rest of the calibration conditions. To make this clearer, we have added (marked) texts in the seventh and ninth paragraphs of the Statistical analysis section. We have also specified the third variables we used as independent ones in the alternative models (ninth paragraph of the Statistical analysis section)

COMMENT 4: The authors should include a table with the characteristics of the population according to the 3 categories or domains of the dependent variable. This could reinforce the decision of grouping the dimensions "nDCP" and "non CP" which, a priori, might not be similar.

RESPONSE 4: The decision of grouping the dimensions nDCP and non CP is based on the results of the multinomial logistic models. Those results, among others, are showed in the article "Cabrera-León A, Cantero-Braojos M, Garcia-Fernandez LI, Guerra de Hoyos JA. Living with disabling chronic pain: results from a face-to-face cross-sectional population-based study". That work was partially presented at the Congress of the Spanish Society of Rehabilitation and Physical Medicine (Cádiz, Spain; Jun 2015). One of its conclusions was that no statistically significant differences were found between the non-CP and nDCP populations. That article is under review by another scientific journal. We have referenced it as a working paper in the penultimate paragraph of the Statistical analysis section (reference 36).

## VERSION 2 – REVIEW

<b>REVIEWER</b>	Pier Francesco Perri Department of Economics, Statistics and Finance, University of Calabria - Italy
<b>REVIEW RETURNED</b>	26-Nov-2016

<b>GENERAL COMMENTS</b>	The authors have considered the point I raised and the paper is now suitable for the publication.
-------------------------	---

<b>REVIEWER</b>	Failde Inmaculada. MD University of Cadiz
<b>REVIEW RETURNED</b>	28-Nov-2016

<b>GENERAL COMMENTS</b>	The answer included by the authors in the revision process are according to the comments required by the reviewer.
-------------------------	--