PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	EFFECT OF CARDIOVASCULAR PREVENTION STRATEGIES ON INCIDENT CORONARY DISEASE HOSPITALISATION RATES IN SPAIN. AN ECOLOGICAL TIME SERIES ANALYSIS
AUTHORS	Medrano, María José; Alcalde-Cabero, Enrique; Ortiz, Cristina; Galán, Iñaki

VERSION 1 - REVIEW

REVIEWER	Maria Guzman Castillo Department of Public Health and Policy University of Liverpool United Kingdom
REVIEW RETURNED	11-Nov-2013

GENERAL COMMENTS	A fully detailed technical appendix should be included describing the models (Poisson regression and non-parametric GAM). In the main text, it is not clearly stated why two different models were used (Poisson regression and non-parametric GAM). I would like to see a justification for the use of GAM instead of the traditional time series models such as ARIMA. For the sake of clarity, I suggest a table describing the independent variables and their units of measurement. In page 7, the authors stated: "the dependent variable was the number of incident hospitalisations in each sex and age stratum". Firstly, the age strata are not defined in the document. Secondly, this definition of the outcome variable differs from the one in page 5 and 6 which is defined as age-adjusted incidence rates. Another limitation of this manuscript is that it does not look at the variability explained by sex and age. I would suggest to include age and sex also as random-effect variables. Generally speaking the paper is well-written and structured, it flows well.

REVIEWER	Jianhua Wu

	Wolfson Institute of Preventive Medicine, Queen Mary University of London, UK
REVIEW RETURNED	13-Nov-2013

GENERAL COMMENTS	This is an interesting paper to assess the impact of primary intervention of IHD in Spain at a population level. Its results have indicated the association between reduction of incident hospitalization due to IHD and the decline in smoking and the increase in vascular risk drug therapy. It could shed light on further experimental study to confirm the effectiveness of the two primary interventions at a population level. The figures in the paper need to be revised. Figure 1 should be 4x2 panels rather than 2x4. The fonts are not clearly visible. In Figure 2, it is better to indicate the meaning of the solid curve and dashed curves (95% CI?). Then it is not clear in left panel of Figure 2, the solid curve and dashed curve almost overlap.

VERSION 1 – AUTHOR RESPONSE

Reviewer Name : Maria Guzman Castillo
Comments:
1.- A fully detailed technical appendix should be included describing the models (Poisson regression and non-parametric GAM).
Response: We agree that readers of a medical journal may be not familiar with the methods used in this highly specific epidemiological study. A technical appendix has been included as supplementary material, describing these models and their applications in epidemiology, as well as the details and syntax of all fitted models.
We believe that in depth mathematical characteristics of these models lay out of the scope of the journal and its readers, however we have included bibliographical references in case someone is interested in their study.

2.- In the main text, it is not clearly stated why two different models were used (Poisson regression and non-parametric GAM).
R.: We are sorry we did not clearly mention in the manuscript that both, GAM and mixed-effects models, were Poisson models: in the GAM formula we included a family object specifying the distribution to use in fitting was Poisson; for mixed-effects models we used directly the command 'xtmepoisson' implemented in Stata, that fits mixed-effects models for count responses assuming a Poisson distribution of the data. The rationale for using two different Poisson models is that GAM models further allows to graphically depict the relationship including both smoothing and also a non-parametric fit, with no a priori assumptions on the actual relationship between response and predictor. As time is used as the predictor, the result is a smoothed time series of the response.
We have added two new sentences in the methods section of the manuscript, further explaining these points.

3.- I would like to see a justification for the use of GAM instead of the traditional time series models

such as ARIMA.
R.: The annual incidence time series is too short (10 points) to allow fitting ARIMA models. The alternative would have been using shorter time periods (day, week or month) in the aggregation of data, but explanatory variables on these basis are not available, and moreover this shorter aggregation would have little sense given the study objective. Finally, ARIMA models are more suitable for predictive than for explanatory purposes.
4.- For the sake of clarity, I suggest a table describing the independent variables and their units of measurement.
R.: These data were represented graphically in figure 1, but we agree that this figure is not very clear, and so believes the second reviewer. Following both reviewers suggestions we have included a new table, and we have also changed the orientation of the panels in figure as well as the font size.
5.- In page 7, the authors stated: "the dependent variable was the number of incident hospitalisations in each sex and age stratum". Firstly, the age strata are not defined in the document. Secondly, this definition of the outcome variable differs from the one in page 5 and 6 which is defined as age-adjusted incidence rates.
R.: The age strata were 5-year age groups starting from 30-34 to 85 and older. This sentence has been added to the text.
Respecting the second point, age adjusted rates were used for the univariate description of the time trends. However, for analytical purposes, this is, in modelling of the effects of explanatory variables, the outcome variable was the count of incident cases in each age, sex and year stratum; these models included the census population of each stratum as the exposed population and age as an adjusting variable. In fact this is equivalent to having age adjusted rates as the dependent variable; due to the log link in Poisson regression
$\log(\text{n cases}) = \log(\text{p population}) + \sum\beta(x)$;
$\log(\text{n cases}) - \log(\text{p population} = \sum\beta(x)$;
$\log(\text{n cases/p population}) = \sum\beta(x)$;
We have modified the wording in the corresponding paragraphs in pages 5, 6 and 7 to improve clarity and make it more understandable for the reader

6.- Another limitation of this manuscript is that it does not look at the variability explained by sex and age. I would suggest to include age and sex also as random-effect variables.
Age and sex were significantly associated with incident hospitalization in all analyses. IRR (95% CI) for age, ranked from 1.36 (1.35-1.36) for each 5-year age increase in analyses with full multivariate adjustments, to 1.40 (1.40-1.41) in less adjusted models. In the same manner, sex IRR was 0.53 (0.52-0.54) for female with respect to male. Age and sex structure of the Spanish population did vary slightly in the analyzed 10-year period, although this variation was towards older ages; on the contrary, incident coronary disease hospitalizations decreased strongly, which is the opposite direction to what one would expect in an ageing population. Nevertheless, the % of variability in annual rates explained by prevention variables raised from 92% with respect to the empty model, to 97% when calculated with respect to the model adjusted by age and sex, thus meaning a 5% variability due to these.
We have modified the results section of the main text and the abstract to include these.
Respecting the last point, including age and sex as random effect variables would build a multi-level hierarchical model, while the structure of our data is not strictly nested and hierarchical: for example, exposed individuals in one year are also considered in other years (unless dead) and in

some cases in different age group, and IHD cases in each of the sexes are classified also by the same age groups. One alternative would be to fit cross-classified models which would be far more complex. Moreover, again explanatory variables are not available at the individual level for the entire Spanish population, so the individual level cannot be considered; this is one of the study limitations commented on in the discussion section. However, age and sex maybe also considered not as groups in a multilevel approach, but as categories of factor variables considered for confounding control, as fixed effects variables (models used in the original analysis) or, as an alternative to cross-classified or to hierarchical random coefficients and intercepts, taking a random effect of year among the values of age and sex as factor variables. So, following your suggestion we have repeated the models in table 1 including age and sex in the random effects equation as factors and year as level variable, but the results were almost identical to those obtained in the original models.

Reviewer Name: Jianhua Wu
Comments:
1.- Figure 1 should be 4x2 panels rather than 2x4. The fonts are not clearly visible.
Response: The figure was ment to lie in landscape orientation and to be full page size, but it was impossible to keep these characteristics in the pdf file creation of the manuscript. Indeed, the solution is as simple as you suggest, thank you.
Figure has been modified accordingly, changing the orientation of the panels in the figure as well as the titles font size. Following the other reviewer suggestion, we have also included a new table describing the independent variables and their units of measurement.

2.- Figure 2, it is better to indicate the meaning of the solid curve and dashed curves (95% CI?). Then it is not clear in left panel of Figure 2, the solid curve and dashed curve almost overlap.
R.: The figure legend has been modified to include that solid line represents the incidence rate ratios (IRRs) and dashed lines are the upper and lower limits of its 95% confidence interval.
Solid and dashed curves in left panel are almost overlapped because the confidence interval is very narrow, as also seen in the Poisson models in tables 1 and 2. This is due to:
A) Observed data are very large numbers, with nearly one hundred thousand cases and millions of population in each point of the curve
B) The relation of incidence rates with time is very strong and uniform
C) The model fits very accurately the observed data, capturing almost all variability.
This explanation has been included in the correspondin

VERSION 2 – REVIEW

REVIEWER	Maria Guzman Castillo Department of Public Health and Policy University of Liverpool United Kingdom
REVIEW RETURNED	03-Jan-2014

- The reviewer completed the checklist but made no further comments.