PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (see an example) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below. Some articles will have been accepted based in part or entirely on reviews undertaken for other BMJ Group journals. These will be reproduced where possible.

ARTICLE DETAILS

TITLE (PROVISIONAL)	An Ecological Study on the Geographic patterns of Ischemic Heart Disease in Portugal and its association with demography, economic factors and health resources distribution
AUTHORS	Luís Ferreira-Pinto, Francisco Rocha-Gonçalves and Armando Teixeira-Pinto

VERSION 1 - REVIEW

REVIEWER	I have no conflict of interest. Dr. Leonie Sundmacher Senior Research Fellow Managing Editor Health Policy Department of Health Care Management WHO Collaborating Centre for Health Systems Research and Management Berlin University of Technology H 80, Straße des 17. Juni 135 10623 Berlin
REVIEW RETURNED	01/02/2012

THE STUDY	It is not a strict "no". I would like to suggest the following: incentives may have existed for physicians to move to areas with high or low rates of cardiovascular diseases (e.g. higher demand, nicer areas), in which case the causality between cardiovascular diseases and health care resources would be reversed in the regression. Furthermore, omitted variables that correlate to both cardiovascular diseases and health care resources could have confounded the relationship and biased the results (you speak about urbanity). To preclude the possibility of reversed causality and confounding, you could use an instrumental variable (IV) approach to investigate potential endogeneity of health care resources to cardiovascular diseases. Moreover, simple linear regressions may not be adequate if the dependent variables do not follow a normal distribution. I would guess that they are a right-skewed.
RESULTS & CONCLUSIONS	The results are credible but a stronger statistical analysis would provide more reliable results.
GENERAL COMMENTS	I think it is a nice paper! I did not recommend minor revision because a change of the statistical analysis is a major task. If you decide to keep the statistical analysis as it is, I would suggest that you discuss the potential for endogeneity bias in the method section. Following this discussion, you can only discuss correlations and no effects in the discussion section. You should still check if a linear regression is adequate.

REVIEWER	Julian Perelman Assistant Professor Escola Nacional de Saúde Pública, Universidade Nova de Lisboa Portugal No competing interests.
REVIEW RETURNED	05/02/2012

THE STUDY	
RESULTS & CONCLUSIONS	
REPORTING & ETHICS	
GENERAL COMMENTS	This paper analyzes geographical variations in incidence and mortality by IHD in Portugal and aims at explaining these variations by socioeconomic factors and availability of health resources. The subject is certainly relevant, at least for Portugal, and the attempt to explain variations is interesting for an international audience. However, results are quite limited and subject to serious methodological problems; authors do not provide convincing explanations for their results, nor do they relate them to the current literature. Also I don't agree with authors when they claim the methodology is innovative (they perform a simple linear regression using socioeconomic factors and health resources as explanatory variables – many others have done this before). Finally, authors refer they use specific methods ("ecological study", "demand/supply approach") while this is not the case. I feel the contribute of the paper is relatively limited for an international audience; by contrast, the representation of geographical variations of mortality and admissions could be very helpful for decision-makers in Portugal. I detail these issues here-below. **Introduction** Main comments The paper lacks of a theoretical background and clear statement of assumptions. Authors should tell from the beginning what they expect to find and why, and how this contributes to the literature, e.g., why they expect regional differences in Portugal and how they expect them to be related to socioeconomic factors or to healthcare resources. They should specify the causal pathway between these factors and IHD, which are certainly not limited to healthstyle. These aspects should be much more documented. There is no serious review of the literature. The reader should be told what has been studied so far, how this has been studied, and in particular what have been the results, going into more details. The authors refer other papers but in very vague terms, stating the relationship with risk factors. Authors should explain what previous papers have been missing, and how they expect improve previous knowledge. The reference to a supply/demand approach is misleading. A real supply/demand approach requires a deeper and more sophisticate analysis, I suggest authors refer e.g. to Gravelle et al (Health Economics 2003, vol 12, issue 12). One of the main methodological problems is endogeneity, which may occur in this paper. That is, health resources are allocated based on needs, so that authors may be using an explanatory factor that is itself explained by the dependent variable (SAR or SMR). Previous papers studied risk factors which may also be considered as components of demand,

	it's just they did not call it "demand". That is, calling the analysis "supply/demand approach" does not bring anything specific as compared to previous analyses. Other comments The review of the literature should clearly distinguish what has been observed for incidence and for mortality, and how the paper contributes on both issues. The introduction is somewhat confuse in that many objectives are mixed along the text. Main objectives and contributes should appear at first very clearly. **Methodology** Main comments The number of hospital admissions is a very problematic proxy for incidence. First, some events can occur without hospitalization, in particular when the person dies. Second, incidence refers to new cases. A patient admitted at hospital may have been suffering from IHD for many years, so it is not necessarily a new case. Third, from my knowledge of the Portuguese NHS hospital database, it is not always possible to link patients across admissions, so that a patient admitted several times during a given year might be erroneously considered as several new cases. Fourth, values should be divided not by the total population of the county but by those older than 18. The PPI variable is a composite indicator based on other socioeconomic indicators, and has been created by the National Institute for Statistics. Authors should provide some details about how PPI was built, for readers to know what is used exactly. Explanation should be given about how to interpret specific values of this index (what does it mean e.g. that PPI is 236 for Lisbon?). Many other indicators could be relevant to reflect socioeconomic factors, which are available (on income, education, occupation, employment status…etc.). Authors should explain why PPI was preferred, although I think a deeper analyses should be performed to test other factors. I don't understand why PPI, number of physicians…etc. were log-transformed. This is what we do with dependent variables to obtain normality, not with explanatory variables. By contrast, normality of SAR and SMR should be discussed; my intuition is that these variables do not fulfill the normality condition (e.g. they are non-negative) and that simple linear regression is not appropriate. The sample is quite small (278 observations) and certainly requires the use of more sophisticate techniques. The potential endogeneity problem should be addressed (see above). Other comments I don't think the patient's residency address is available in the database, but well the name of their county of residence. Authors should clarify this point. Authors should explain what is a semi-parametric regression for latitude and longitude (why semi-parametric, in particular). **Results and discussion** Main comments The major result of multivariate analysis is that mortality and admissions are higher in richer counties, which contradicts previous papers and expectations. Authors do not provide any explanation for this result. We feel some additional analyses are missing, e.g. using other socioeconomic indicators. The lack of references and

	theoretical background is particularly problematic when it comes to discuss this result. The descriptive analysis should be oriented towards the focus of the study, that is, a county-level multivariate analysis. Hence there should be an exhaustive description of counties, for dependent and independent variables included in further multivariate analysis. E.g. the average/variation of PPI and other explanatory variables per county (and for the complete country like in Table 1), the average/variation in SAR and SMR per county….etc. It is absolutely misleading to start the Results section with a description of the hospital individual data while only aggregate data will be used in the analysis. The abstract is also very misleading because readers may think the work is performed on a sample of 250,000 persons while it is done on a sample of 278 observations. On Tables 2 and Table 3 all results should be displayed, not only those with significant values. Information should be provided about the goodness-of-fit of the models. Readers should have the possibility to assess the validity of the techniques and models. The discussion should carefully mention the limitations of the paper and explain how they may influence results. Other comments Naming specific Portuguese counties is not particularly relevant for an international audience. What matters is to tell there are variations and explain why. The univariate analysis does not provide a significant contribute to the paper. This may be helpful for the authors when doing the analysis but does not bring interesting information to the reader. Only the multivariate analysis is relevant in terms of results for discussion.

VERSION 1 – AUTHOR RESPONSE

Reviewer1: Dr. Leonie Sundmacher
Senior Research Fellow
Managing Editor Health Policy
Department of Health Care Management
WHO Collaborating Centre for Health Systems
Research and Management
Berlin University of Technology

Reviewer1: I would like to suggest the following: incentives may have existed for physicians to move to areas with high or low rates of cardiovascular diseases (e.g. higher demand, nicer areas), in which case the causality between cardiovascular diseases and health care resources would be reversed in the regression.
Furthermore, omitted variables that correlate to both cardiovascular diseases and health care resources could have confounded the relationship and biased the results (you speak about urbanity).
To preclude the possibility of reversed causality and confounding, you could use an instrumental variable (IV) approach to investigate potential endogeneity of health care resources to cardiovascular diseases.
Authors: We acknowledge the comments from the reviewer and agree that an IV approach could be helpful. However, we have very limited data that we could use, in particular to choose an appropriate instrumental variable.
We should underpin, however, that it is not our purpose to draw causal conclusions from our study. We fully understand the limitations of using secondary and aggregated data and it was not our

intention to claim causality from our findings. We added a paragraph in the discussion to address this issue and included a reference to the limitations of the Ecological design.

Reviewer1: Moreover, simple linear regressions may not be adequate if the dependent variables do not follow a normal distribution. I would guess that they are a right-skewed. The results are credible but a stronger statistical analysis would provide more reliable results.
Authors: We thank the reviewer for pointing this out. In fact we do need normality in the dependent variable but on the error term of the regression. By mistake, we did not mention that SAR and SMR were log-transformed as well. We added this point to the methods section.

Reviewer1: I think it is a nice paper! I did not recommend minor revision because a change of the statistical analysis is a major task.
If you decide to keep the statistical analysis as it is, I would suggest that you discuss the potential for endogeneity bias in the method section. Following this discussion, you can only discuss correlations and no effects in the discussion section. You should still check if a linear regression is adequate.
Authors: Thank you for the comment. As mentioned before we added a paragraph in the discussion about the limitations of the study design.

Reviewer2: Julian Perelman
Assistant Professor
Escola Nacional de Saúde Pública, Universidade Nova de Lisboa
Portugal
No competing interests.

Reviewer2: This paper analyzes geographical variations in incidence and mortality by IHD in Portugal and aims at explaining these variations by socioeconomic factors and availability of health resources. The subject is certainly relevant, at least for Portugal, and the attempt to explain variations is interesting for an international audience. However, results are quite limited and subject to serious methodological problems; authors do not provide convincing explanations for their results, nor do they relate them to the current literature. Also I don't agree with authors when they claim the methodology is innovative (they perform a simple linear regression using socioeconomic factors and health resources as explanatory variables – many others have done this before). Finally, authors refer they use specific methods ("ecological study", "demand/supply approach") while this is not the case.
I feel the contribute of the paper is relatively limited for an international audience; by contrast, the representation of geographical variations of mortality and admissions could be very helpful for decision-makers in Portugal. I detail these issues here-below.

Authors: We believe that some of the points raised by the reviewer may be explained by different understanding of study design methodology. Ecologic (or Ecological) study is a particular design of Epidemiological observational studies in which "the unit of observation is a group of people, rather that an individual" (Rothman, Modern Epidemiology.) The associations found in such studies cannot be interpreted as causal associations. Even if all the confounders were available, the results could still be biased due to the Ecological fallacy (also known as Ecological bias or Aggregation bias) (Greenland, 1989). However, these studies have been useful in "describing different populations, even if confounded by unknown or uncontrollable factors, such difference at least signal the presence of effects worthy of further investigation" (Rothman, Modern Epidemiology).

According to the previous definition, the study reported in this manuscript in fact an Ecological study.

Additionally, we do not pursue causal interpretation from the associations found in the analysis as we agree with the idea above that one should not infer causality from this type of studies. We added a paragraph to the discussion to clarify this point.

Regarding the “demand/supply approach” we acknowledge the reviewer’s comment. Our intention was to study the association between the health resources available in the different regions and the SAR and SMR. We changed the introduction accordingly.

The innovative part of the methodology does not refer to the linear regression but to the semiparametric regression that we applied to model the geographical variation of SMR and SAR.

Finally, we do not agree with the comment “I feel the contribute of the paper is relatively limited for an international audience” but we think this is an Editorial decision.

Reviewer2: Introduction
Main comments
The paper lacks of a theoretical background and clear statement of assumptions. Authors should tell from the beginning what they expect to find and why, and how this contributes to the literature, e.g., why they expect regional differences in Portugal and how they expect them to be related to socioeconomic factors or to healthcare resources. They should specify the causal pathway between these factors and IHD, which are certainly not limited to healthstyle. These aspects should be much more documented.

Authors: As discussed before, this is an ecological study that does not intend to draw causal interpretation from the results. The main objective of the paper is to describe the variation of SAR and SMR associated with IHD in Portugal.

Reviewer2: There is no serious review of the literature. The reader should be told what has been studied so far, how this has been studied, and in particular what have been the results, going into more details. The authors refer other papers but in very vague terms, stating the relationship with risk factors. Authors should explain what previous papers have been missing, and how they expect improve previous knowledge.

Authors: The first objective of this study is to describe the regional variation of IHD in Portugal. We do mention that regional heterogeneity was observed in other national and international studies. As far as we know, there are not many studies addressing this issue in Portugal.

Reviewer2: The reference to a supply/demand approach is misleading. A real supply/demand approach requires a deeper and more sophisticate analysis, I suggest authors refer e.g. to Gravelle et al (Health Economics 2003, vol 12, issue 12). One of the main methodological problems is endogeneity, which may occur in this paper. That is, health resources are allocated based on needs, so that authors may be using an explanatory factor that is itself explained by the dependent variable (SAR or SMR). Previous papers studied risk factors which may also be considered as components of demand, it’s just they did not call it “demand”. That is, calling the analysis “supply/demand approach” does not bring anything specific as compared to previous analyses.

Authors: We acknowledge the reviewer’s comment. We removed the reference to supply/demand approach.

Reviewer2: Other comments
The review of the literature should clearly distinguish what has been observed for incidence and for mortality, and how the paper contributes on both issues.
The introduction is somewhat confuse in that many objectives are mixed along the text. Main objectives and contributes should appear at first very clearly.

Authors: We thank the reviewer's comment. We stated the objective at the beginning of the introduction.

Reviewer2: Methodology
Main comments
The number of hospital admissions is a very problematic proxy for incidence. First, some events can occur without hospitalization, in particular when the person dies. Second, incidence refers to new cases. A patient admitted at hospital may have been suffering from IHD for many years, so it is not necessarily a new case.

Authors: Thank you for point this out. We agree with the reviewer that the number of hospital admissions is not equivalent to incidence and there are several issues with its use as a proxy of incidence. We added this as a limitation in the discussion section

Reviewer2: Third, from my knowledge of the Portuguese NHS hospital database, it is not always possible to link patients across admissions, so that a patient admitted several times during a given year might be erroneously considered as several new cases.

Authors: The reviewer is correct. One patient may be admitted multiple times and we are not able to identify re-admission. This was added as a limitation in the discussion section

Reviewer2: Fourth, values should be divided not by the total population of the county but by those older than 18.

Authors: We agree with the reviewer. By mistake this was not explicitly in the methods section. We did not use the county's population but the population above 18 yrs old. This was corrected in the methods section

Reviewer2: The PPI variable is a composite indicator based on other socioeconomic indicators, and has been created by the National Institute for Statistics. Authors should provide some details about how PPI was built, for readers to know what is used exactly. Explanation should be given about how to interpret specific values of this index (what does it mean e.g. that PPI is 236 for Lisbon?).
Many other indicators could be relevant to reflect socioeconomic factors, which are available (on income, education, occupation, employment status…etc.). Authors should explain why PPI was preferred, although I think a deeper analyses should be performed to test other factors.

Authors: Thank you for the suggestion. We added the sentence "It is a summary index of 17 economical variables that include, for example, income per capita, electric consumption, taxes and number of vehicles per capita, among others. It is represented as a base 100 index, meaning that if a region as a PPI of 110 it is 10% above the national average" to the Methods section.

Reviewer2: I don't understand why PPI, number of physicians…etc. were log-transformed.

This is what we do with dependent variables to obtain normality, not with explanatory variables. By contrast, normality of SAR and SMR should be discussed; my intuition is that these variables do not fulfill the normality condition (e.g. they are non-negative) and that simple linear regression is not appropriate. The sample is quite small (278 observations) and certainly requires the use of more sophisticate techniques.

Authors: The linear regression assumes normality of the error term and not on the dependent variable. By mistake we omitted that the SAR and SMR were log-transformed but this was not enough to obtain normality in the residuals. The log-transformation of some covariates worked well in “normalizing” the residuals.

Reviewer2: The potential endogeneity problem should be addressed (see above).
Authors: We do not try to establish causality. This was added as a limitation of the study.

Reviewer2: Other comments
I don’t think the patient’s residency address is available in the database, but well the name of their county of residence. Authors should clarify this point.
Authors should explain what is a semi-parametric regression for latitude and longitude (why semi-parametric, in particular).
Authors: The database indicates the postal code of the individual. This is enough to identify the county. We added a reference to the semiparametric regression

Reviewer2: Results and discussion
Main comments
The major result of multivariate analysis is that mortality and admissions are higher in richer counties, which contradicts previous papers and expectations. Authors do not provide any explanation for this result. We feel some additional analyses are missing, e.g. using other socioeconomic indicators. The lack of references and theoretical background is particularly problematic when it comes to discuss this result.
Authors: The reverse association between mortality and regional economical development has been observed in similar studies. A simple example at Europe level is the fact that Portugal has one of the lowest IHD mortality rates and countries like Finland and Austria have a much higher rate. Once again, these associations should not be interpreted as causal relations. We added another reference with similar results to ours.

The descriptive analysis should be oriented towards the focus of the study, that is, a county-level multivariate analysis. Hence there should be an exhaustive description of counties, for dependent and independent variables included in further multivariate analysis. E.g. the average/variation of PPI and other explanatory variables per county (and for the complete country like in Table 1), the average/variation in SAR and SMR per county….etc.
Authors: There are 278 counties. We think such table would not be publishable. However, we are willing to provide it as supplementary material.

It is absolutely misleading to start the Results section with a description of the hospital individual data while only aggregate data will be used in the analysis. The abstract is also very misleading because readers may think the work is performed on a sample of 250,000 persons while it is done on a sample of 278 observations.
Authors: Thank you for the suggestion. We corrected the abstract accordingly.

On Tables 2 and Table 3 all results should be displayed, not only those with significant values. Information should be provided about the goodness-of-fit of the models. Readers should have the possibility to assess the validity of the techniques and models.
Authors: The variables with no p-values were not included in the final model. We added the r2 statistics for the multivariable models.

The discussion should carefully mention the limitations of the paper and explain how they may influence results.
Authors: We improved the discussion on the limitations of this study.

Other comments
Naming specific Portuguese counties is not particularly relevant for an international audience. What matters is to tell there are variations and explain why.
The univariate analysis does not provide a significant contribute to the paper. This may be helpful for the authors when doing the analysis but does not bring interesting information to the reader. Only the multivariate analysis is relevant in terms of results for discussion.
Authors: We do not agree with the reviewer in this point. In our opinion, the univariate analysis should be presented together with the multivariable (not multivariate) analysis. This is a common practice in Health related articles and we agree with it as it is relevant to see how crude estimates are affected by adjustment to other covariates.

VERSION 2 – REVIEW

REVIEWER	Leonie Sundmacher Professor for Primary Healthcare Economics Berlin University of Technology Germany I have no conflict of interests
REVIEW RETURNED	25/05/2012

The reviewer completed the checklist but made no further comments.