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.

This paper was submitted to the JECH but declined for publication following peer review. The authors addressed the reviewers' comments and submitted the revised paper to BMJ Open. The paper was subsequently accepted for publication at BMJ Open.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Predicting Risk of Hospitalization or Death: A Retrospective Population Based Analysis
AUTHORS	Louis, Daniel; Robeson, Mary; McAna, John; Maio, Vittorio; Keith, Scott; Liu, Mengdan; Gonnella, Joseph; Grilli, Roberto

VERSION 1 - REVIEW

REVIEWER	Carlos Manuel Morais da Costa Escola Nacional de Saúde Pública - Universidade Nova de Lisboa Portugal
REVIEW RETURNED	20-Apr-2014

GENERAL COMMENTS	Checklist 1 and 6: The title of the study is: "Predicting Risk of Hospitalization Using a Population Based Longitudinal Database", but, the main objective is: "Develop predictive models to identify patients at high risk of a progression of their medical problems or who are at risk of developing new medical problems", which are different things and should be analyzed differently, while the first one is more related with the risk of hospitalization, the second one is more focused with the identification of conditions / diseases / patients with a faster progression of the medical problems than expected (depending of each natural disease evolution), with the need of using different dependent variables for each situation. I think that the first issue is the one that should be analyzed on the paper. However, it seems that the study is focused on the second issue: "The dependent variable was defined as the occurrence of a hospitalization for problems that are potentially avoidable, or whose progression may have been avoided or delayed through appropriate patient care, or the death of the individual for any reason in 2012". With fine tuning we see that this "problem" related with the objectives definition of the study are still present in the criteria for the population to be analyzed. For instance, when the authors mentioned: "We felt that inclusion of hospitalization for cancer in the dependent variable should depend on the ability to either prevent or avoid progression of the disease. We therefore included colon cancer and cervical cancer in the definition because they are potentially preventable but excluded all other cancers where prevention/prediction is not currently possible.", it seems that they are more concerned with hospitalization and its timeliness, and less with the risk of death, because they are

	excluding a large number of deaths. For instance, in Portugal in 2012 there were 10,722 in-hospital deaths for all the cancers, and for the colon cancer and cervical cancer we found 1,685 deaths (circa 15.7%). Taking these issues into account, the paper should be more precise with the definition of the objectives, and in my opinion mainly focused on risk of hospitalization, not considering avoidable hospitalizations and the risk of death. Checklist 6 and 7: For the objectives defined: risk of hospitalization or risk of hospitalization and in-hospital death, the dependent variable is different, and much more important is that the independent variables or covariates included in the models could be different and even if they are the same its influence in these two risks should be also different. Suggestions: run two separates models, one for risk of hospitalization and the second only for risk of in-hospital death, or just run the first one. This is very important since in-hospital mortality or mortality rate must be risk adjusted, what can be easily done using Disease Staging. Also there is sound evidence that there is a variation between hospitals practices, also including the in-hospital death and then for the number of potential avoidable deaths, which could introduce some bias on the study. Accordingly, assuming equal performance among hospitals should be avoided in this study and the risk of death should not be included in the dependent variable. Although the statistics are fully described, still I am not sure what risk is being assessed. There are good news with the model discrimination, since the "c" statistic had a very good value (even when using the model of the previous year), and is seems that the model it is calibrated, taking in account the ratio observed/predicted hospitalization or deaths for each decile group, so I don't understand the reason for not including the Hosmer-Lemeshow test. On the other hand, in my opinion, it will be interesting: (1) to explain better the reason for the cut-off used for very high users and high users and to discuss the consequences of these choices; (2) to mention the rationale of the p-values used for inclusion and retention of the covariates, and also if the stepwise models are forward or backward. Since these issues can modify the covariates included in the model and also each odds-ratio, I think that it is an issue that should be better explained and should be more discussed, and (3) to explain the reasons for not using some interactions, for instance age and number of chronic condition, and to see if this procedure increases or not the sensitivity, which is poor for the very high risk users. Checklist 10: The results are interesting, although no one knows if they are biased or not, depending of the behavior and importance of in-hospital deaths on each covariate. This is relevant, since if there is some evidence that the analysis is biased, the results of the manuscript could be limited. This issue should be very deeply analyzed and discussed in order to improve the health policy and management. On the other hand the results are descriptive, and in my opinion the covariates should also include the odds-ratio for hospitalization, which can be even more relevant for strategic planning and management. Moreover this odds-ratio can be also used as an indicator for the assessment of good practices.

	Finally there are no results for the combination of covariates. For example, the importance of the cardiovascular disease it is not broken down by age and gender, which could be important for policy and management strategies. In my opinion this type of analysis should be more used for predicting and assessing the risk of hospitalization. I had mentioned before the most relevant issues, but I point out the following: - using only risk of hospitalization as dependent variable; - using the Hosmer-Lemeshow Test for see the model calibration or then explain the reasons for not using it; - using more evidence based cut-off points for identifying very high users and high users; - explain the rationale for the p-values for the inclusion and retention of covariates and the stepwise meyhodology used; - using interactions for model improvement; - using the odds-ratio as a tool for policy definition and follow-up management.

REVIEWER	Ian Blunt Nuffield Trust, UK
REVIEW RETURNED	22-May-2014

GENERAL COMMENTS	This is a well put together paper and I have confidence in the results. However, I find it slightly unusual that the authors do not provide the final model in the paper or appendices - that is, the final selected variables, their coefficients and p values. This would allow it to be used in other settings that had similar data, and be compared with the many other publicly available models. It would also be useful to be more explicit about the basis on which admissions are deemed potentially avoidable. The principle and examples are well documented, but who is making these decisions - an expert panel, personal judgement by the authors or something else? It might be described in refs 7+8, but a quick reminder sentence in the paper would be useful. Again, this is a well put together paper and nicely described. The model produced has an impressive c statistic and high sensitivity . However, I note that its PPV is relatively poor compared with other similar models. It would be interesting to see the authors expand on the implications of this in the discussion - it is particularly important as when a model is applied in practice it is the performance at set thresholds, and not the c statistic, that is key. Finally, it's always worth pointing out that any predictive model needs to be paired with an effective intervention if it is to be of any use. What interventions might the authors recommend this model be used to trigger?

VERSION 1 – AUTHOR RESPONSE

Reviewer(s)' Comments to Author:
Reviewer: 1
Reviewer Name Carlos Manuel Morais da Costa
Institution and Country Escola Nacional de Saúde Pública - Universidade Nova de Lisboa

Portugal
Please state any competing interests or state 'None declared': I know Prof. Daniel Louis and Dr. Joseph Gonnella
I am using Disease Staging in Portugal to analyze hospital utilization and performance. I have several projects funded by public hospitals, private companies (pharmaceutical, non governmental organizations and media, for example) using Disease Staging, but none related with the issue analyzed on the manuscript
I am preparing throughout Escola Nacional de Saúde Pública some research projects with the Center for Research in Medical Education and Health Care, Jefferson Medical College of Thomas Jefferson University - Dr. Joseph Gonnella: Director and Prof. Daniel Louis: Managing Director

Checklist 1 and 6:

The title of the study is: "Predicting Risk of Hospitalization Using a Population Based Longitudinal Database", but, the main objective is: "Develop predictive models to identify patients at high risk of a progression of their medical problems or who are at risk of developing new medical problems", which are different things and should be analyzed differently, while the first one is more related with the risk of hospitalization, the second one is more focused with the identification of conditions / diseases / patients with a faster progression of the medical problems than expected (depending of each natural disease evolution), with the need of using different dependent variables for each situation. I think that the first issue is the one that should be analyzed on the paper. However, it seems that the study is focused on the second issue: "The dependent variable was defined as the occurrence of a hospitalization for problems that are potentially avoidable, or whose progression may have been avoided or delayed through appropriate patient care, or the death of the individual for any reason in 2012".With fine tuning we see that this "problem" related with the objectives definition of the study are still present in the criteria for the population to be analyzed.

We recognize that hospitalization and progression of a medical problem are not synonymous. While we would have preferred to measure disease progression, that is not possible using the administrative data available to us. We therefore made this choice, as have others developing similar risk models. We have modified the Introduction section of our manuscript to clarify the project goals.

For instance, when the authors mentioned: "We felt that inclusion of hospitalization for cancer in the dependent variable should depend on the ability to either prevent or avoid progression of the disease. We therefore included colon cancer and cervical cancer in the definition because they are potentially preventable but excluded all other cancers where prevention/prediction is not currently possible.", it seems that they are more concerned with hospitalization and its timeliness, and less with the risk of death, because they are excluding a large number of deaths. For instance, in Portugal in 2012 there were 10,722 in-hospital deaths for all the cancers, and for the colon cancer and cervical cancer we found 1,685 deaths (circa 15.7%).

Death, either in or out of the hospital, for any cause, was included in our dependent variable definition. We have clarified this in the methods section

Taking these issues into account, the paper should be more precise with the definition of the objectives, and in my opinion mainly focused on risk of hospitalization, not considering avoidable

hospitalizations and the risk of death.

While imperfect, we are not trying to predict hospitalization for issues such as childbirth or for medical or surgical problems that are unavoidable given current medical knowledge, such as appendicitis. Furthermore, we believe the definition of the dependent variable in our model makes it more likely that we are identifying high risk patients with “impactible” medical problems; those that may possibly be avoided with high quality care.

Checklist 6 and 7:

For the objectives defined: risk of hospitalization or risk of hospitalization and in-hospital death, the dependent variable is different, and much more important is that the independent variables or covariates included in the models could be different and even if they are the same its influence in these two risks should be also different. Suggestions: run two separates models, one for risk of hospitalization and the second only for risk of in-hospital death, or just run the first one.

Perhaps, we were not clear in the Methods section. We did not limit deaths to those patients who died in the hospital. The database used in this study allowed the identification of mortality whether the patient died in the hospital or in another location We have clarified this in the Methods section Ideally, we would have liked to include only deaths resulting from the same medical problems include in our hospitalization variable definition. However, cause of death data were not available to us. We have added a note about this in the Limitations section of the manuscript.

This is very important since in-hospital mortality or mortality rate must be risk adjusted, what can be easily done using Disease Staging. Also there is sound evidence that there is a variation between hospitals practices, also including the in-hospital death and then for the number of potential avoidable deaths, which could introduce some bias on the study. Accordingly, assuming equal performance among hospitals should be avoided in this study and the risk of death should not be included in the dependent variable.
Although the statistics are fully described, still I am not sure what risk is being assessed. There are good news with the model discrimination, since the "c" statistic had a very good value (even when using the model of the previous year), and is seems that the model it is calibrated, taking in account the ratio observed/predicted hospitalization or deaths for each decile group, so I don’t understand the reason for not including the Hosmer-Lemeshow test.

We have not included the Hosmer-Lemeshow test because our study is extremely overpowered for that test. In a situation like this, trivially small differences in the average prevalence estimates will result in large numbers in the numerator of the test statistic and cause it to inflate out of proportion with the importance of any lack of fit detected. This is a well-known property of the test that has been evaluated well in a simulation study by Kramer and Zimmerman (2007 Crit Care Med; 35:2052-6). Their study showed that when there was only a slight deviation from perfect fit, the Hosmer-Lemeshow test was statistically significant in 100% of their 1,000 simulated datasets. This finding suggests that it would likely be impossible that any study with a sample size as large as ours would produce models capable of passing the Hosmer-Lemeshow test – unless the fit were essentially perfect. Of course, we cannot be sure of their motivation, this perhaps explains why other researchers using large data sets also chose to not use the Hosmer-Lemeshow test. (for example, Billings et al – ref 14 in our manuscript)
The calibration of our modeling is on display in the Figure. We feel that this accurately portrays that our modeling has fit the data very well.

On the other hand, in my opinion, it will be interesting: (1) to explain better the reason for the cut-off

used for very high users and high users and to discuss the consequences of these choices;

Of course, there is a trade-off in using our model, or any predictive model, between the threshold and the proportion of patients exceeding that threshold. These risk thresholds were selected after consultation with physicians practicing in the medical homes to yield a total of about 10% of the 1,500 patients enrolled with a typical primary care physician. This is stated in the Methods section.

(2) to mention the rationale of the p-values used for inclusion and retention of the covariates, and also if the stepwise models are forward or backward. Since these issues can modify the covariates included in the model and also each odds-ratio, I think that it is an issue that should be better explained and should be more discussed, and (3) to explain the reasons for not using some interactions, for instance age and number of chronic condition, and to see if this procedure increases or not the sensitivity, which is poor for the very high risk users.

The stepwise model selection procedure we applied is neither forward or backward, just stepwise. Variables could be entered or withdrawn during the process. In stepwise selection, an attempt is made to remove any insignificant variables from the model before adding a significant variable to the model. Each addition or deletion of a variable to or from a model is a separate step in the selection process, and at each step a new model is fitted.

We considered including interaction terms but their main effects would need to be forced into models. As such, model selection would have to be conducted in stages: first, to select interactions and second to select main effect terms while forcing in the interactions and their main effects. Given the large number of predictor variables, we felt that the possible improvement in the models did not warrant the effort necessary to conduct this two-stage selection procedure.

With regards to the last comment, the sensitivity of our modeling is not poor for the very high risk users. When we place the threshold for defining very high risk at 25% (or more) estimated risk of hospitalization, we are not actually implying that nobody outside that group is predicted to experience death or hospitalization – which is precisely what would result in a perfect sensitivity (assuming we make at least one true positive). On the contrary, by the definition of the threshold and our prevalence estimated-based risk scores, we expect that many such events will occur among those with an estimated risk below 25%. The PPV estimate is essentially reflecting the estimated prevalence of hospitalization or death among those with an estimated risk of 25% or more. After applying Bayes' theorem and a few steps of algebra, it also becomes clear that sensitivity is directly related to prevalence as well:
sensitivity=(PPV×(1-prevalence))/((1-PPV)×prevalence)×(1-specificity). Feedback from the physicians suggests that our models excel at identifying their high risk patients.

Checklist 10:

The results are interesting, although no one knows if they are biased or not, depending of the behavior and importance of in-hospital deaths on each covariate. This is relevant, since if there is some evidence that the analysis is biased, the results of the manuscript could be limited. This issue should be very deeply analyzed and discussed in order to improve the health policy and management.
On the other hand the results are descriptive, and in my opinion the covariates should also include the odds-ratio for hospitalization, which can be even more relevant for strategic planning and management. Moreover this odds-ratio can be also used as an indicator for the assessment of good practices.
Finally there are no results for the combination of covariates. For example, the importance of the

cardiovascular disease it is not broken down by age and gender, which could be important for policy and management strategies. In my opinion this type of analysis should be more used for predicting and assessing the risk of hospitalization.

We agree that age is an important factor in risk prediction and that other predictor variables may behave differently for different age groups. As described in the Modeling section regression models were fit in each of 14 gender and age strata. We have now included modelling coefficients in the supplemental materials to our manuscript. We believe that our risk predictions are accurate and useful. However, we are reluctant to attempt to interpret specific odds ratios since that is beyond the scope of this study.

I had mentioned before the most relevant issues, but I point out the following:
- using only risk of hospitalization as dependent variable;
- using the Hosmer-Lemeshow Test for see the model calibration or then explain the reasons for not using it;
- using more evidence based cut-off points for identifying very high users and high users;
- explain the rationale for the p-values for the inclusion and retention of covariates and the stepwise meyhodology used;
- using interactions for model improvement;
- using the odds-ratio as a tool for policy definition and follow-up management.

Reviewer: 2
Reviewer Name Ian Blunt
Institution and Country Nuffield Trust, UK
Please state any competing interests or state ‘None declared’: None declared

This is a well put together paper and I have confidence in the results.

Thank you for your positive response.

However, I find it slightly unusual that the authors do not provide the final model in the paper or appendices - that is, the final selected variables, their coefficients and p values. This would allow it to be used in other settings that had similar data, and be compared with the many other publicly available models.

As described in the Methods section of our manuscript. We fit models for 14 age/sex strata and therefore we have not one, but 14 models. Nevertheless, we have taken the advice of this reviewer and now included as appendix materials all 14 models.

It would also be useful to be more explicit about the basis on which admissions are deemed potentially avoidable. The principle and examples are well documented, but who is making these decisions - an expert panel, personal judgement by the authors or something else? It might be described in refs 7+8, but a quick reminder sentence in the paper would be useful.

We modified the Methods to make it clear that this was s judgment of the authors.

Again, this is a well put together paper and nicely described. The model produced has an impressive c statistic and high sensitivity . However, I note that its PPV is relatively poor compared with other similar models. It would be interesting to see the authors expand on the implications of this in the

discussion - it is particularly important as when a model is applied in practice it is the performance at set thresholds, and not the c statistic, that is key.

At a predicted risk of >25% our model had a PPV of .411. In other words, the prevalence of hospitalization or death in this risk group was 41.1%. We do not believe that this is poor compared to other models. Billings et al (BMJ Open 2013) reported a PPV of .417 at a risk threshold of 30. As noted above in our response to reviewer 1, there is a trade-off in using our model, or any predictive model, between the threshold for follow-up and predictive accuracy. We have added some comments to the Discussion section to reflect this.

Finally, it's always worth pointing out that any predictive model needs to be paired with an effective intervention if it is to be of any use. What interventions might the authors recommend this model be used to trigger?

We strongly agree with this comment. In fact, we are working with the physicians, nurses, and administration of the newly formed Medical Homes in Parma to assist them in understanding how to use the results of these models and in developing potentially effective interventions. We have added some comments to the Discussion section to expand on this.

VERSION 2 – REVIEW

REVIEWER	Ian Blunt Nuffield Trust
REVIEW RETURNED	14-Jul-2014

GENERAL COMMENTS	None

REVIEWER	Charles Maynard University of Washington Seattle, Washington USA
REVIEW RETURNED	28-Jul-2014

GENERAL COMMENTS	This paper presents a statistical model for predicting death or hospitalization in residents of Emilia-Romagna, Italy. page 4. The introduction is quite lengthy and could be shortened. page 7. There probably needs to be better justification of the method for defining avoidable hospitalizations. Was consideration given to using ambulatory care sensitive conditions? On the other hand, if the interest is in identifying hospitalizations, whether or not they were avoidable may be irrelevant. page 9, line 54. Which stepwise method was used: Backwards or forwards? Page 16, line 44. The correct name is United States Department of Veterans Affairs. Page 19. It will be interesting to see how these models are used in

	real life situations.

REVIEWER	David Pilcher ANZICS Centre for Outcome and Resource Evaluation
REVIEW RETURNED	08-Aug-2014

GENERAL COMMENTS	The authors describe the use of an extensive administrative database containing demographic and medical information to develop a prediction model which estimates the risk of "unnecessary" (my choice of words not theirs) hospitalisation or death in the general population of a large region of Italy. It is a generally well written and interesting study. Its novelty appears to be the attempt by the authors to identify individuals whose hospitalisations could potentially be prevented in future. This is clinically relevant. Main comments: The authors have used subjective criteria to define their main outcome by picking only the hospitalisations which they considered relevant and by excluding others. While there may be clinical and public health validity in this approach, the authors should recognise that this may bias their findings and potentially artificially elevate the discriminatory performance (C statistic) of their modelling. Was any attempt made to determine whether there was agreement between the choice of variables to include? Since a greater amount of information about co-morbidities, diagnoses and drug usage was available from those patients who had been admitted to hospital, and they are trying to predict hospitalisation, could this have influenced/biased their results? I presume individuals where there was no data about the presence of a specific condition were consider as not to have the condition, rather than as unknowns. Modelling such as this could potentially identify areas within their region (e.g. metropolitan v rural) where there are a greater risk adjusted number of hospitalisations/deaths than others. Has there been any attempt to do this? Was data about socio-economic status available? Many other countries have this available and linked to postal areas which has been shown to be related to Although diagnoses and pharmacy information from previous hospitalisations, the number of previous hospitalisations does not appear to have been entered into the modelling? Is this correct & if so why not? Did the authors consider calculation or estimation of standard co-morbidity indices such as Charlson or Elixhauer? Although overall discrimination and calibration information is provided, did the authors consider looking at specific subgroups of patients, to determine model performance within these groups (e.g. patients with previous hospitalisations, those with cancer or with a history of psychiatric disorders). Their dataset should be big enough to assess performance within these groups.

REVIEWER	Ruth Masterson Creber, PhD, RN University of Pennsylvania, Philadelphia PA USA
REVIEW RETURNED	09-Aug-2014

GENERAL COMMENTS	There was a clean copy then a marked copy. I wasn't sure why the marked copy was included unless it was a response to reviewers, but if that was the case I assumed I would also see a response to reviewers document. Was this a primary or secondary review?

VERSION 2 – AUTHOR RESPONSE

Reviewer: 2
Reviewer Name Ian Blunt
Institution and Country Nuffield Trust
Please state any competing interests or state 'None declared': None declared

The authors have responded appropriately to the points raised, and the paper has been strengthened as a result.

We appreciate reviewer 2's suggestions and positive response to our modified manuscript.

Reviewer: 3
Reviewer Name Charles Maynard
Institution and Country University of Washington
Seattle, Washington
USA
Please state any competing interests or state 'None declared': None declared

This paper presents a statistical model for predicting death or hospitalization in residents of Emilia-Romagna, Italy.

page 4. The introduction is quite lengthy and could be shortened.

We have edited and shortened the introduction.

page 7. There probably needs to be better justification of the method for defining avoidable hospitalizations. Was consideration given to using ambulatory care sensitive conditions? On the other hand, if the interest is in identifying hospitalizations, whether or not they were avoidable may be irrelevant.

We consulted the literature concerning ambulatory care sensitive conditions. However, many of those are focused on a pediatric population which was excluded from our project. (AHRQ Quality Indicators, Hospital Admission for Ambulatory Care Sensitive Conditions, Department of Health and Human Services, Agency for Healthcare Research and Quality, AHRQ Pub. No. 02-R0203 Revision 1 (April 17, 2002)
Some users (such as the Canadian Institute for Health Information (CIHI) Health indicators 2013: http://www.qualitymeasures.ahrq.gov/content.aspx?id=47604) exclude patients 75 or older. Much of our population of interest is age 75 or older.
After consultation with our clients in the Emilia-Romagna Region and Parma Local Health Authority we decided to use a broader definition which we have described in the definition of our dependent

variable part of the Methods section.

page 9, line 54. Which stepwise method was used: Backwards or forwards?

The stepwise model selection procedure we applied is neither forward nor backward, just stepwise. Variables could be entered or withdrawn during the process. In stepwise selection, an attempt is made to remove any insignificant variables from the model before adding a significant variable to the model. Each addition or deletion of a variable to or from a model is a separate step in the selection process, and at each step a new model is fitted.

Page 16, line 44. The correct name is United States Department of Veterans Affairs.

We have made the correction in the manuscript.

Page 19. It will be interesting to see how these models are used in real life situations.

We agree completely. So far, the models and resulting profiles have been very well received in the Medical Homes in the Parma Local Health Authority. The Emilia-Romagna Region has decided to expand the project to three additional health authorities over the next 2 years. Of course, evaluation of impact in terms of reducing the rate of potentially avoidable hospitalization will take time.

Thank you for the opportunity to review this paper.

Thank you for taking the time to read and review our manuscript.

Reviewer: 4
Reviewer Name David Pilcher
Institution and Country ANZICS Centre for Outcome and Resource Evaluation
Please state any competing interests or state ‘None declared’: None

The authors describe the use of an extensive administrative database containing demographic and medical information to develop a prediction model which estimates the risk of “unnecessary” (my choice of words not theirs) hospitalisation or death in the general population of a large region of Italy. It is a generally well written and interesting study. Its novelty appears to be the attempt by the authors to identify individuals whose hospitalisations could potentially be prevented in future. This is clinically relevant.
Thank you.

Main comments:
The authors have used subjective criteria to define their main outcome by picking only the hospitalisations which they considered relevant and by excluding others. While there may be clinical and public health validity in this approach, the authors should recognise that this may bias their findings and potentially artificially elevate the discriminatory performance (C statistic) of their modelling.

In our manuscript, we describe the development of a predictive model using the RER's regional longitudinal administrative health care database to help identify patients who are most at risk of hospitalization for conditions that may be impacted through improved patient care. This choice would only bias our findings if we were to suggest that the findings relate to hospitalizations other than those included in our definition of the dependent variable. We have been careful not imply that, so we believe that concern over bias from this source should be minimal.

Was any attempt made to determine whether there was agreement between the choice of variables to include?

We are not sure from this comment if it refers to the dependent variable or the independent variables. We have attempted to clearly define these variables and have included as appendix material the specific results from all regression models which were fit in each of 14 gender and age strata. We have not made specific comparisons among the variables selected included in each of the 14 regressions.

Since a greater amount of information about co-morbidities, diagnoses and drug usage was available from those patients who had been admitted to hospital, and they are trying to predict hospitalisation, could this have influenced/biased their results? I presume individuals where there was no data about the presence of a specific condition were consider as not to have the condition, rather than as unknowns.

Clearly, our models rely on prior utilization to inform the variables related to conditions. There were no attempts made to sample the population to verify condition status among either those with a history of utilization or those without. That would, of course, be very expensive and was beyond the scope of what we were able to do for this project.

The variables do not indicate the presence of the condition, but rather health care utilization attributed to the presence of the condition. So, we may miss patients who have not received treatment for a specific condition. We have modified the limitations paragraph in the Discussion section to reflect this.

We cannot know whether or not this potential source of misclassification has biased the results, but it has almost certainly increased the uncertainty of the predictions. That is reflected in the high, but imperfect performances of the models in predicting hospitalizations for the selected conditions or death.

While the reviewer is correct that we have more specific diagnostic information available for patients with prior hospitalization, the drug usage data available to us is from outpatient prescription data which is available whether or not a patient has been hospitalized. We believe that this is a key strength of the Emilia-Romagna database.

Modelling such as this could potentially identify areas within their region (e.g. metropolitan v rural) where there are a greater risk adjusted number of hospitalisations/deaths than others. Has there been any attempt to do this?

Yes. The Emilia-Romagna region has 11 geographically defined Local Health Authorities (recently some have been combined.) At a meeting organized by the region and including the directors general and medical directors of the Local Health Authorities we presented information about the observed and "expected" rates of hospitalization using the results of the models described in this manuscript.

Was data about socio-economic status available? Many other countries have this available and linked to postal areas which has been shown to be related to

(This comment seems to be have been cut off, but we think it is clear.) Unfortunately, we did not have socio-economic status data available to us. We agree that it could be an important predictor. We also think that living situation (alone, with spouse, with or near other family members) would be very helpful. We have had some discussions about the possibility of collecting and including some of these data in future versions of these models.

Although diagnoses and pharmacy information from previous hospitalisations, the number of previous hospitalisations does not appear to have been entered into the modelling? Is this correct & if so why not?

The number of previous hospitalizations was a variable used in building our models. Regression models were fit in each of 14 gender and age strata using a stepwise process. This variable was selected for inclusion in 12 of the gender and age-specific strata modeled. It was excluded from only the youngest female (ages 18-34) and the oldest male (ages 85 and older) categories.

Did the authors consider calculation or estimation of standard co-morbidity indices such as Charlson or Elixhauer?

Yes we did consider these, but we believe that the variables that we chose are more relevant to the specific databases and variable definitions we used in this project. The Charlson index, while used for other types of problems, was initially developed for cancer patients with weights estimated for specific populations. The Elixhauser index does not use weights so is more general. However, we felt that since we have data such as pharmacy data and home health data in addition to the hospitalization information that is typically used with the Charlson and Elixhauser indices our approach to variable definition was more suitable for this project.

Although overall discrimination and calibration information is provided, did the authors consider looking at specific subgroups of patients, to determine model performance within these groups (e.g. patients with previous hospitalisations, those with cancer or with a history of psychiatric disorders). Their dataset should be big enough to assess performance within these groups.

Our initial charge from the Emilia-Romagna was to develop the models applicable to the entire adult population which we have reported in our manuscript. We concur with their assessment that this provides the best basis for the initial profiles provided to the physicians in the medical homes. However, we would be very interested in the future to develop models that are more focused on specific population groups such as those suggested by the reviewer.

Reviewer: 5
Reviewer Name Ruth Masterson Creber, PhD, RN
Institution and Country University of Pennsylvania,
Philadelphia PA
USA
Please state any competing interests or state ‘None declared’: None declared

Abstract and early in methods: In a simple sentence or two, please clearly define that the entire dataset included data from 2004-2012, the model was built in 2004-2011 data and tested in 2012 data (if that is correct). If the outcome was any hospitalization over the course of a year that should be explicitly stated in the abstract and earlier in the methods.

We have clarified this in both the abstract and in the second paragraph of the Methods section.

Pg 30. Line 19: Clarify “deliveries”

We have changed the wording to make this clear.

In the methods section please provide a clear conceptual rationale for why hospitalization and deaths were included as a single outcome. From the perspective of the PCMH, interventions would be very different for a patient who died of terminal cancer versus an acute exacerbation of heart failure that could have been avoided or delayed with better preventative care. There were enough hospitalizations in 2012, why not just make that the outcome? Also, was the length of time for hospitalization one year for everyone? Did you explore risk of 30-day or 60-day hospitalizations as well?

Of course, the reviewer is correct that the management of a patient with terminal cancer is different than the management of a patient with heart failure. However, we believe that a patient with coronary artery disease, for example, who dies of an acute myocardial infarction, should be included in the dependent variable even if the death is out of the hospital. We have added a sentence to the Methods section to clarify our reasoning. We recognize that inclusion of patients who died, either in or out of the hospital, has led to including some patients in the dependent variable whose deaths could not have been prevented or delayed and may have increased the uncertainty of the predictions. This potential problem has been noted in the limitations paragraph of the Discussion section. Nevertheless, we believe that our models perform well in predicting the dependent variable of hospitalization for the selected conditions or death.

Yes, the length of time for hospitalization was one year. We did not explore 30-day or 60-day hospitalizations.

Discussion:Adding a paragraph about the Italian medical system and structure would be helpful to a reader who is not familiar with it. That will also help to contextualize how incentives are aligned differently in Italy for the prevention of hospitalizations. With a single-payer system does that change incentives to prevent/reduce hospitalizations?

We appreciate this suggestion and have added a paragraph to the Discussion section.

VERSION 3 – REVIEW

REVIEWER	Charles Maynard University of Washington Seattle, WA USA
REVIEW RETURNED	02-Sep-2014
GENERAL COMMENTS	The authors have been very responsive to the initial review.

REVIEWER	Dr. Ruth Masterson Creber University of Pennsylvania
REVIEW RETURNED	03-Sep-2014
GENERAL COMMENTS	The authors have sufficiently addressed each of my comments on the previous draft.