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

TITLE (PROVISIONAL)	Proposals for enhanced health risk assessment and stratification in an integrated care scenario
AUTHORS	Dueñas-Espín, Iván; Vela, Emili; Pauws, Steffen; Bescos, Cristina; Cano, Isaac; Cleries, Montserrat; Contel, Joan Carles; de Manuel Keenoy, Esteban; Garcia-Aymerich, Judith; Gomez-Cabrero, David; Kaye, Rachelle; Lahr, Maarten; Lluch-Ariet, Magí; Moharra, Montserrat; Monterde, David; Mora, Joana; Nalin, Marco; Pavlickova, Andrea; Piera, Jordi; Ponce, Sara; Santaeugenia, Sebastià; Schonenberg, Helen; Störk, Stefan; Tegner, Jesper; Velickovski, Filip; Westerteicher, Christoph; Roca, Josep

VERSION 1 - REVIEW

REVIEWER	Peter Bower and Jonathan Stokes University of Manchester, United Kingdom
REVIEW RETURNED	18-Dec-2015

GENERAL COMMENTS	This paper explores an area of great interest in current health policy – the use of risk assessment and stratification tools in integrated care. The authors have presented some limited empirical data on current risk assessment strategies and used that as basis for a broader discussion of the issues related to their use. It seems more like a conceptual than an empirical contribution. The introduction could benefit from a clearer definition of integrated care. At present, it is not clear what they are classing in this category, and so how risk assessment directly relates (or not) as a vital aspect. The ACT programme should be briefly described so readers do not have to read other sources. The limitations of considering these issues in the context of a single programme such as ACT should be considered. There are some complex statements in the introduction. For example - 'We hypothesized that clinical prediction for any given specific medical purpose could be significantly improved by including the allocation of the individual in the population risk pyramid into the modelling approach [12].' These could be rephrased and the exact meaning clarified. The methods are not clearly described at present, and the start of the methods section would benefit from a clear statement of the design. Could details of the survey be added as an appendix? It is very unclear exactly what questions were asked, and how many

	people were actually involved in the survey. Was it essentially 5 respondents, or 5 groups of respondents? No response rates are given. Or is it 5 people asked/5 people answered for the whole study? How was the survey developed, in terms of quality and validity? The 'strengths and limitations' section could perhaps be clearer about the limited empirical content of the paper The section 'Elaboration of a proposal for enhanced clinical risk assessment' was very unclear. Exactly what is being done here? It is not clear what methods they use to "assess the flexibility" Table 3 should have another column for the evidence (and the strength) that these recommendations are based on. In the text at the top of p.8, one of the recommendations is justified by survey showing agreement between all (5) regions. Were the other recommendations based on consensus among the five partners? Or based on external evidence? Is Figure 1 utilising actual data from one of the partners, specifically developed for this project? Or is this generic data about GMA? This needs to be more clearly described and justified. The fact that these tools add value to prediction is not probably that contentious. The more complex issue is whether the tools serve as the basis for a cost-effective strategy of delivery of interventions. The paper seems to assume that is the case, but I think the evidence for this is far weaker. p.8: "We propose to incorporate…". This first sentence of this sentence seems out of place, and distorts the understanding of the whole of the section. Would recommend moving this first sentence to the end of the section, where they talk about "statistical refinement of the computational modelling". The authors should also expand on the "conceptual grounds and statistical feasibility" that apparently support the proposal. 'Enhanced clinical risk predictive modelling' section. Here they outline a single model as good practice. The paper would be improved with an assessment of the implementation of this 'gold standard' model, and how it relates to the health indicators they originally wanted to compare. Is it used for specific interventions currently? Targeting whom, and effectiveness in relation to these indicators? p.11 'Towards personalized medicine'. This section does not appear to be based on any of the methods outlined. Are the methods missing under this heading? In the discussion, the authors state that "For example, it might be most beneficial to target those individuals in the tip of the pyramid (~5%) accounting for high use of healthcare resources (~ 36% total healthcare costs, as assessed for Catalonia in 2014, Figure 2S)." This is not necessarily a good example of a preventive strategy that they are talking about in the sentence before. There is an assumption in this statement that we can reduce high use of

	healthcare resources by this group, but the evidence for this is not strong. An ability to predict does not imply an ability to intervene successfully. There is an assumption that this group of 'super-utiliser' patients stays consistent over time – but this is contradicted by one of the references this paper uses in the intro section (6. Johnson TL, Rinehart DJ, Durfee J, et al. For Many Patients Who Use Large Amounts of Health Care Services, The Need Is Intense Yet Temporary. Health Aff 2015;34:1312–9). Roland's paper in the BMJ also has a relevant critique of this model which the authors would do well to consider (http://www.bmj.com/content/345/bmj.e6017) The authors state that "the core lesson learnt from the current study is that a wider deployment of population-based risk assessment is needed and it constitutes a priority for any region planning adoption of integrated care." – but it is not clear that this is justified at all by the 'analysis' in this study. What is the evidence of the benefits, except that 5 regions in a study leading on risk stratification agreed on it being a good idea? Some of the statements made may need toning down to better represent the actual data presented in the paper, and to distinguish personal opinions from strong evidence. In summary, the paper deals with a high priority issue, but presents limited empirical data. The bulk of the paper involves discussions and recommendations about the use of these tools. Many of these are sensible, although they do assume the effectiveness these models, and that case may not be as clear as the authors imply. If the core contribution of the paper is conceptual (with the data serving as a platform to consider these issues), then the importance of clarity and an appropriate critical perspective on these issues becomes even more important. Given the limited data presented, the core of the contribution is around some of the ideas presented, and identification of priorities for policy and research. A revised paper should present a more detailed and balanced 'proposal for enhanced clinical risk assessment'.

REVIEWER	Tracy Johnson, Ph.D., M.A. Denver Health, United States
REVIEW RETURNED	23-Dec-2015

GENERAL COMMENTS	It was a privilege to review this very interesting and insightful paper on a topic that is of great international interest. I have recommended that this paper be accepted, with revision. Most of my comments relate to clarity of communication. The overall research questions and methods were not clearly communicated at the onset. Also, I observed some disconnects between the data presented in the Tables/Figures with Results/Discussion. Tighter organization and additional detail in suggested places would improve the paper. I have organized my specific comments according to page number. Please note, I was not able to access the referenced on-line

	supplemental material. P 4. "It is acknowledged that case finding fosters cost-effective preventive interventions despite compromised strength of prediction." [unclear what this means] P 5. "We hypothesized that clinical prediction … could be improved by including the allocation of the individual in the population risk pyramid into the modeling approach" I believe you sought to test this hypothesis in the analysis presented in Figure 1. But this is not clearly communicated in the two specific aims that are outlined in the next paragraph, nor in the methods and results sections. P 5. Nice description of the risk predictive modeling surveys and their purpose. I think you also assessed an ideal future state. If so, this should be noted here. P 6. Health indicators questionnaire description, purpose, timing were not clear. It is implied on p 8 that your purpose was to compare performance across regions, but that is not made clear in the methods or in the study aims. I was unable to access supplemental materials, so did not have access to Table 4S. I am very interested in this table and would review it if it were made available to me. P 6. "Elaboration of a proposal …" I believe this may be referencing Figure 1. If this is correct, some of the detail in the Figure 1 subhead should be communicated here as well. There is language on p. 8 that may be helpful as well. P 7. "The analysis of the risk-strata distributions resulting from the different regions showed poor comparability" (Table 2). While I believe this is a true statement based on the differences in populations selected and the variety of modeling tools used, Table 2 does not otherwise communicate this finding well. All of the regions except one have 4 levels, one has three levels. It would be helpful to present, for example, the percentage of patients assigned to each level or provide some other evidence that illustrates how you arrived at this conclusion. (You could also cite the literature. Authors at Mayo, for example, published a study illustrating how different risk tools identify different individuals as high-risk. P 7. "rigidities due to inclusion of expert-based criteria" This statement is made several times during the paper and deserves elaboration. Why does this inhibit transferability? Why is expert-based criteria included in some models? What is lost (if anything) by relying on purely statistical models? P 7. "we observed different degrees of maturity in the interactions with clinicians" Table 1 is cited here but identical language is used for the "clinical application" row in each table. Suggest including some of the nuance shared in the narrative in Table 1 as well.

	P 7. “a high acceptability of the population health approach for elaboration of health risk predictive model was confirmed”. It was not clear in your description of the survey in p.5 that you surveyed ACT regions about both their current practice and the ideal state, so initially this statement was confusing because Table 1 reveals large differences in current practice. P 8. Results of analysis summarized in Figure 1 are not summarized clearly in the text. Addition of GMA to the model appeared to improve R-squared in all cases. Is this correct? p 9. “Toward personalized medicine” This seems misplaced in a results section. What results are being presented here? Seems better placed in the discussion. P 10. “For example, it might be most beneficial to target for those individuals in the tip of the pyramid (~5%)…” There is a lively discussion about this as well. Some argue that the moderately-high risk (rising risk) group ought to be prioritized. More generally, I noted that the discussion did not address the subject of clinical impactability. Suggest a strong nod to this literature. Statistical measures of predictive value do not capture this dimension and yet when information is translated into clinical applications (e.g., patient lists for primary care physician use), this becomes a highly relevant consideration. This would fit well under the subhead, “clinical health-risk assessment”. P10-11. The implication of Figure 1 is not included in the discussion. Table 1- Source population in words as well as numbers would be helpful. Additional granularity on clinical applications would be helpful. Why is 3M CRG described as “explanatory” and ACG-PM as “predictive”, what is the difference? (Believe both are predictive.) Table 2 – Additional granularity (or a cite) would be helpful to allow the reader to understand why you conclude that the risk tools result in non-comparable population distributions. Table 3 – Omission of clinical impactability is surprising given that an intended use is “planning and commissioning health services” Figure 1 – Interesting data presented here that receives very little attention in the text.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

Reviewer Name
Peter Bower and Jonathan Stokes
Institution and Country
University of Manchester, United Kingdom

Please state any competing interests or state 'None declared':
None declared

Please leave your comments for the authors below
This paper explores an area of great interest in current health policy – the use of risk assessment and stratification tools in integrated care.

Q1. The authors have presented some limited empirical data on current risk assessment strategies and used that as basis for a broader discussion of the issues related to their use. It seems more like a conceptual than an empirical contribution.

R1 – Agreed. The manuscript builds on data obtained from the analysis of the five ACT regions and generates conceptual proposals to be assessed beyond the frame of the revised manuscript. The revised manuscript clearly addresses this issue through changes in: i) title; ii) abstract; iii) Strengths & limitations; and, iv) the entire main text of the revised manuscript. A revised version with marked changes has been submitted.

Q2. The introduction could benefit from a clearer definition of integrated care. At present, it is not clear what they are classing in this category, and so how risk assessment directly relates (or not) as a vital aspect.

R2 – The definition of integrated care adopted by the ACT project has been included in the Introduction (Para. 2). We also indicate that health risk assessment and stratification was one of the four key drivers for deployment of integrated analyzed in ACT (Para. 3). The role of health risk assessment in the deployment of integrated care is acknowledged by the European Innovation Partnership initiative for Active and Healthy Ageing (EIP-AHA). We understand the results of the current study may help to generate evidence in this area, as stated in the Discussion section.

Q3. The ACT programme should be briefly described so readers do not have to read other sources. The limitations of considering these issues in the context of a single programme such as ACT should be considered.

R3 – The main characteristics and aims of the ACT program are described in the Introduction (Para. 2 and 3). Also its intrinsic limitations are implicitly acknowledged in the last sentence of Para. 3.

Q4. There are some complex statements in the introduction. For example - 'We hypothesized that clinical prediction for any given specific medical purpose could be significantly improved by including the allocation of the individual in the population risk pyramid into the modelling approach [12].' These could be rephrased and the exact meaning clarified.

R4 – The specific statement mentioned by the reviewer has been deleted because it was considered outplaced in the Introduction. We have identified both complex statements and portions of text not properly allocated into the manuscript. Several changes have been introduced in the different sections of revised manuscript using the following criteria: i) clarify complex text that was considered part of a

core statement; ii) re-allocate text in the proper areas of the manuscript; and, iii) delete complex text that was marginal in the manuscript. Overall, the length of the revised manuscript has slight increased (from 3378 to 3578 words), but we believe that the revised version gained in clarity and facilitates a fluent reading.

Q5. The methods are not clearly described at present, and the start of the methods section would benefit from a clear statement of the design. Could details of the survey be added as an appendix? It is very unclear exactly what questions were asked, and how many people were actually involved in the survey. Was it essentially 5 respondents, or 5 groups of respondents? No response rates are given. Or is it 5 people asked/5 people answered for the whole study?

R5 – All the reviewer's suggestions have been followed. That is: i) a general statement of the three areas tackled in the Methods Section was included in Para. 1; ii) the characteristics of the survey have been detailed and an Annex has been included in the on-line supplementary material; iii) the purposes and design of the collection of health indicators has been clarified; and, iv) the last section linking empirical findings with conceptual proposals has been fully re-written.

Q6. How was the survey developed, in terms of quality and validity?

R6 – The empirical information of the manuscript was obtained from: i) the survey and, ii) questionnaire on health indicators

i) SURVEY - The five people responsible for risk assessment strategies in each region (named in the acknowledgments). Several iterations were done and all of them were fully accessible throughout the project lifetime. In order to simplify the description, we indicate that a two-phase survey was done which is essentially true. The fact is that the specific questions indicated in Opimec were enriched by several iterations to fully understand what was available in each region, as well as what was planned as future developments to reach and ideal setting.

ii) HEALTH INDICATORS – The revised version of the manuscript clearly indicates the original ACT aims regarding collection of health indicators, as well as the limitations encountered in the baseline assessment that precluded further progress. We understand that the information obtained is highly relevant for future studies. In this area, the information providers were the ACT coordinators in each region and both information and data analysis were done through intensive iterations throughout the first 18 months of the ACT project. The lack of comparability among regions, as explained in the text, did not allow progresses in the temporal analysis of the health indicators.

Q7. The 'strengths and limitations' section could perhaps be clearer about the limited empirical content of the paper

R7 – As mentioned above, changes in this section and all over the revised manuscript have been introduced to clarify this issue. The changes should help to distinguish between the empirical and conceptual parts of the study.

Q8. The section 'Elaboration of a proposal for enhanced clinical risk assessment' was very unclear. Exactly what is being done here? It is not clear what methods they use to "assess the flexibility"

R8 – Full agreement. Apologies for the confusion generated in the original manuscript. This section and the corresponding section in the Results have been completely re-written. We hope that the problem is solved in the revised manuscript.

Q9. Table 3 should have another column for the evidence (and the strength) that these

recommendations are based on.

R9 – Agreed and done. We hope that the proposal fulfill the reviewer's expectations.

Q10. In the text at the top of p.8, one of the recommendations is justified by survey showing agreement between all (5) regions. Were the other recommendations based on consensus among the five partners? Or based on external evidence?

R10 – Apologies again for the confusion generated on this issue (recommendation of reporting metrics on sensitivity/specificity of predictions). There was consensus on the need for reporting the metrics in an ideal scenario. It is also recommended by external sources (ref. 36). But, the reality was that two regions report the metrics (Scotland and Basque Country), one region does not use predictive modeling (Groningen) and two regions (Barcelona and Lombardy) have a pragmatic approach and use allocation of the individual in the risk-strata of the pyramid to support decisions/recommendations. We have deleted the confusing statement and have addressed the topic in Table 3 (see statement at the bottom).

Q11. Is Figure 1 utilising actual data from one of the partners, specifically developed for this project? Or is this generic data about GMA? This needs to be more clearly described and justified. The fact that these tools add value to prediction is not probably that contentious. The more complex issue is whether the tools serve as the basis for a cost-effective strategy of delivery of interventions. The paper seems to assume that is the case, but I think the evidence for this is far weaker.

R11 – We understand that the reviewer addresses two different issues deserving separate answers: i) several questions on GMA as a use case; and, ii) whether the tools serve as the basis for a cost-effective strategy of delivery of interventions.

The first issue (GMA) has been extensively explained, and hopefully clarified, in the revised manuscript in the Methods and Results section. Moreover, part of the conceptual analysis has been moved to the Discussion section as suggested by the reviewer.

The second issue (evidence of the tools serving as the basis for a cost-effective strategy) is more complex. Both ACT program and EIP-AHA assumes that risk assessment is a key driver for deployment of integrated care. But, we acknowledge the lack of high level evidence likely due to severe comparability problems (also seen in other aspects of integrated care). It is of note that the recommendations generated by the manuscript may help to partly solve this issue (lack of comparability). In the revised version manuscript, we have tune down the statements assuming evidence and we explicitly indicate that the study outcomes may be useful to generate (or disprove) such evidence. See also current version of the Discussion section.

Q12. p.8: "We propose to incorporate…". This first sentence of this sentence seems out of place, and distorts the understanding of the whole of the section. Would recommend moving this first sentence to the end of the section, where they talk about "statistical refinement of the computational modelling".

R12 – Agreed and done. The entire section has been re-written.

Q13. The authors should also expand on the "conceptual grounds and statistical feasibility" that apparently support the proposal.

R13 – The statement has been deleted and the entire text re-written.

Q14. 'Enhanced clinical risk predictive modelling' section. Here they outline a single model as good

practice. The paper would be improved with an assessment of the implementation of this 'gold standard' model, and how it relates to the health indicators they originally wanted to compare. Is it used for specific interventions currently? Targeting whom, and effectiveness in relation to these indicators?

R14 – As indicated above. The entire section has been re-written and hopefully we have been able to clarify the whole issue of choosing the GMA as a use case showing the potential of novel risk predictive tools to meet the evolving needs in the field because of its flexibility and transferability. Key aspects of the GMA and its clinical validation have been moved to both Methods and Results section. We believe that the ground used to build the conceptual proposal is now clear.

Since the GMA was developed independently from the ACT program, we believe that we cannot further expand the description, potential and plans within the frame of the current manuscript. Likewise, the selection and purposes of the health indicators in ACT had no relationship at all with the GMA.

In ACT, the work package on risk assessment and stratification (WP5) had two main focuses: population-based risk assessment and clinical risk assessment. The former addressed both risk prediction tools and health indicators. But, limitations described in the manuscript did not allow completion of the initial program in the area of risk assessment. The ACT design was done to explore the role of risk predictive tools for healthcare services optimization and integrated care deployment. Instead, the analyses done in ACT generated useful outcomes that may be helpful to overcome current limitations described in the text.

In Catalonia, the GMA is heavily used for service commission, adjust per capita financing scheme, cost-effectiveness assessment and monitoring of healthcare services in the real scenario. The links between population-based risk predictive modeling and clinical risk assessment constitute a research priority endorsed by the Catalan Department of Health. Some of the plans in this area are described in the conceptual part of the manuscript.

Q15. p.11 'Towards personalized medicine'. This section does not appear to be based on any of the methods outlined. Are the methods missing under this heading?

R15 – Agreed. The entire section has been moved to the Discussion section.

Q16. In the discussion, the authors state that "For example, it might be most beneficial to target those individuals in the tip of the pyramid (~5%) accounting for high use of healthcare resources (~ 36% total healthcare costs, as assessed for Catalonia in 2014, Figure 2S)." This is not necessarily a good example of a preventive strategy that they are talking about in the sentence before. There is an assumption in this statement that we can reduce high use of healthcare resources by this group, but the evidence for this is not strong. An ability to predict does not imply an ability to intervene successfully.

There is an assumption that this group of 'super-utiliser' patients stays consistent over time – but this is contradicted by one of the references this paper uses in the intro section (6. Johnson TL, Rinehart DJ, Durfee J, et al. For Many Patients Who Use Large Amounts of Health Care Services, The Need Is Intense Yet Temporary. Health Aff 2015;34:1312–9). Roland's paper in the BMJ also has a relevant critique of this model which the authors would do well to consider (http://www.bmj.com/content/345/bmj.e6017)

R16 – Agreed. The statement is not central in the manuscript and the debate on the issue can mislead the reader. We are inclined to believe that cost-effective preventive interventions should be

the choice. Consequently, the sentence has been deleted in the revised manuscript.

Q17. The authors state that “the core lesson learnt from the current study is that a wider deployment of population-based risk assessment is needed and it constitutes a priority for any region planning adoption of integrated care.” – but it is not clear that this is justified at all by the ‘analysis’ in this study. What is the evidence of the benefits, except that 5 regions in a study leading on risk stratification agreed on it being a good idea? Some of the statements made may need toning down to better represent the actual data presented in the paper, and to distinguish personal opinions from strong evidence.

R17 – Agreed. The statement has been tuned-down in the revised manuscript. We are stressing that the main aim is to overcome current limitations of the risk prediction tools in order to facilitate comparability such that evidence on usefulness can be generated in the future.

Q18. In summary, the paper deals with a high priority issue, but presents limited empirical data. The bulk of the paper involves discussions and recommendations about the use of these tools. Many of these are sensible, although they do assume the effectiveness these models, and that case may not be as clear as the authors imply. If the core contribution of the paper is conceptual (with the data serving as a platform to consider these issues), then the importance of clarity and an appropriate critical perspective on these issues becomes even more important.

Given the limited data presented, the core of the contribution is around some of the ideas presented, and identification of priorities for policy and research. A revised paper should present a more detailed and balanced ‘proposal for enhanced clinical risk assessment’.

R18 – Agreed. The revised version of the manuscript clearly distinguishes between the empirical component of the study and the conceptual part. We believe that the grounds to build the proposals are now well established. Moreover, effectiveness of the models is not assumed, as stated in the revised manuscript (Discussion section).

Reviewer: 2

Reviewer Name
Tracy Johnson, Ph.D., M.A.

Institution and Country
Denver Health, United States

Please state any competing interests or state ‘None declared’:
None declared

Please leave your comments for the authors below

Authors:

It was a privilege to review this very interesting and insightful paper on a topic that is of great international interest. I have recommended that this paper be accepted, with revision. Most of my comments relate to clarity of communication. The overall research questions and methods were not clearly communicated at the onset. Also, I observed some disconnects between the data presented in the Tables/Figures with Results/Discussion. Tighter organization and additional detail in suggested places would improve the paper.

R1 - We appreciate the reviewer's comments on the current study. The revised manuscript has been extensively remodeled trying to fulfill the reviewer's requirements.

I have organized my specific comments according to page number. Please note, I was not able to access the referenced on-line supplemental material.

R2 -Sorry for the problem to access the on-line supplementary material. Part I of the on-line supplementary material was relevant to capture the role of GMA as a use case to link the empirical part of the study (data collected in the ACT project) with the recommendations for enhancing clinical risk assessment. Please, let us know if you have again problems accessing the on-line supplementary material.

P 4. "It is acknowledged that case finding fosters cost-effective preventive interventions despite compromised strength of prediction." [unclear what this means]

R3 – Agreed. The aim of the statement was to indicate that case finding can generate healthcare efficiencies despite the robustness of the predictions are often rather poor. In the revised manuscript, the sentence has been deleted for clarity since it did not add useful information. The logics of the changes introduced in the revised version have been explained in the point-by-point responses to reviewer 1 (R4), reproduced below (see current R4)

P 5. "We hypothesized that clinical prediction … could be improved by including the allocation of the individual in the population risk pyramid into the modeling approach" I believe you sought to test this hypothesis in the analysis presented in Figure 1. But this is not clearly communicated in the two specific aims that are outlined in the next paragraph, nor in the methods and results sections.

R4 – Agreed. As indicated to reviewer 1 "the specific statement mentioned by the reviewer has been deleted because it was considered outplaced in the Introduction. We have identified both complex statements and portions of text not properly allocated into the manuscript. Several changes have been introduced in the different sections of the revised manuscript using the following criteria: i) clarify complex text that was considered part of a core statement; ii) re-allocate text in proper area of the manuscript; and, iii) delete complex text that was marginal in the manuscript. Overall, the length of the revised manuscript has slight increased (from 3378 to 3578 words), but we believe that the revised version gained in clarity and facilitates a fluent reading".

P 5. Nice description of the risk predictive modeling surveys and their purpose. I think you also assessed an ideal future state. If so, this should be noted here.

R5 - Thanks. We have introduced your suggestion in P5, final Para., lines 4-6.

P 6. Health indicators questionnaire description, purpose, timing were not clear. It is implied on p 8 that your purpose was to compare performance across regions, but that is not made clear in the methods or in the study aims. I was unable to access supplemental materials, so did not have access to Table 4S. I am very interested in this table and would review it if it were made available to me.

R6 – The Methods section has been substantially improved in the revised manuscript. The first paragraph has been expanded to describe the general structure of the entire section. Moreover, the different sections have been re-written to clarify purposes and methodological approaches. The original aim of the ACT program is now described in the subheading of health indicators. Unfortunately, the original plan could not be completed because the baseline assessment already showed lack of comparability among regions. The causes and recommendations are explained in the Results section. Moreover, specific recommendations on health indicators and on standardization of

calculations are provided in Part III of the on-line supplementary material.

P 6. “Elaboration of a proposal …” I believe this may be referencing Figure 1. If this is correct, some of the detail in the Figure 1 subhead should be communicated here as well. There is language on p. 8 that may be helpful as well.

R7 – Agreed. We acknowledge that the structure and wording in the original manuscript was confusing. In the revised version, the entire subheading has been reshaped in the Methods section, but also in the corresponding part of the Results section.

P 7. “The analysis of the risk-strata distributions resulting from the different regions showed poor comparability” (Table 2). While I believe this is a true statement based on the differences in populations selected and the variety of modeling tools used, Table 2 does not otherwise communicate this finding well. All of the regions except one have 4 levels, one has three levels. It would be helpful to present, for example, the percentage of patients assigned to each level or provide some other evidence that illustrates how you arrived at this conclusion. (You could also cite the literature. Authors at Mayo, for example, published a study illustrating how different risk tools identify different individuals as highrisk.

R8 – Agreed. Tables 1 and 2 have been modified following the reviewer’s suggestions. Moreover, the Mayo’s reference has been added to the text (ref 32)

P 7. “rigidities due to inclusion of expert-based criteria” This statement is made several times during the paper and deserves elaboration. Why does this inhibit transferability? Why is expert-based criteria included in some models? What is lost (if anything) by relying on purely statistical models?

R9 – The co-morbidity grouper of some commercial products includes expert-generated groups of specific co-morbidities. This fact may constitute a limiting factor for a generalized use of the predictive model. It may require adjustment of the coefficients for each geographical area due to differences between healthcare systems and/or in health information data collection.
The co morbidity groupers of the modern risk assessment tools (i.e. GMA) are based on statistical grounds such that it facilitates adaptability to different healthcare systems. This has been the experience of transferring the GMA from Catalonia (7.5 M people) to other thirteen regional health systems in Spain (approx. additional 38 M people). The answer to your question is that nothing is lost (a lot gained) by relying on statistical models. We believe that the issue is now clear in the Discussion section.

P 7. “we observed different degrees of maturity in the interactions with clinicians” Table 1 is cited here but identical language is used for the “clinical application” row in each table. Suggest including some of the nuance shared in the narrative in Table 1 as well.

R10 – We agree with the reviewer, but it was quite difficult to capture a detailed description allowing a clear distinction among: i) real implementation; ii) on going developments; and, iii) future plans. It is of note, however, that all the regions agreed on potential clinical usefulness of the population-based risk prediction tools.

P 7. “a high acceptability of the population health approach for elaboration of health risk predictive model was confirmed”. It was not clear in your description of the survey in p.5 that you surveyed ACT regions about both their current practice and the ideal state, so initially this statement was confusing because Table 1 reveals large differences in current practice.

R11 – Agreed. We believe that the issue is clarified in the revised manuscript. The second phase of

the survey included specific questions on what should be the ideal scenario for each region and how to reach it.

P 8. Results of analysis summarized in Figure 1 are not summarized clearly in the text. Addition of GMA to the model appeared to improve R-squared in all cases. Is this correct?

R12 – Agreed. This has been included in the revised manuscript. Also the performance of the GMA has been expanded by transferring a summary of the validation process from the on-line supplementary material to the Results section.

p 9. “Toward personalized medicine” This seems misplaced in a results section. What results are being presented here? Seems better placed in the discussion.

R13 – Agreed. It has been moved to the end of the Discussion section.

P 10. “For example, it might be most beneficial to target for those individuals in the tip of the pyramid (~5%)…” There is a lively discussion about this as well. Some argue that the moderately-high risk (rising risk) group ought to be prioritized. More generally, I noted that the discussion did not address the subject of clinical impactability. Suggest a strong nod to this literature. Statistical measures of predictive value do not capture this dimension and yet when information is translated into clinical applications (e.g., patient lists for primary care physician use), this becomes a highly relevant consideration. This would fit well under the subhead, “clinical health-risk assessment”.

R14 – We believe that this is a very relevant issue, but it may detract the reader from the main message of the manuscript. Consequently, we propose to delete it in the revised manuscript. We believe that the Discussion section is currently more balanced.

P10-11. The implication of Figure 1 is not included in the discussion.

R15 – Agreed. This has been now corrected in the Methods, Results and Discussion sections.

Table 1- Source population in words as well as numbers would be helpful. Additional granularity on clinical applications would be helpful. Why is 3M CRG described as “explanatory” and ACG-PM as “predictive”, what is the difference? (Believe both are predictive.)

R16 – Agreed and changed. The expression Explanatory model was confusing. The issue is related with the use of metrics to indicate the robustness of the predictions. In the revised manuscript, it has been explained at the bottom of Table 3.

Table 2 – Additional granularity (or a cite) would be helpful to allow the reader to understand why you conclude that the risk tools result in non-comparable population distributions.

R17 – Agreed and done.

Table 3 – Omission of clinical impactability is surprising given that an intended use is “planning and commissioning health services”

R18 – We believe that the item Clinical and social usefulness in Table 3 corresponds to the concept Clinical impactability

Figure 1 – Interesting data presented here that receives very little attention in the text.

R19 –Agreed. The entire section has been re-written in the revised manuscript.