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)	Functional Health Literacy in a population-based sample in Florence: a cross-sectional study using the Newest Vital Sign
AUTHORS	Bonaccorsi, Guglielmo; Lastrucci, Vieri; Vettori, Virginia; Florence Health Literacy, Research Group; Lorini, Chiara

VERSION 1 - REVIEW

REVIEWER	Gillian Rowlands Newcastle University UK At one point in the paper I suggest that the authors cite a publication of mine, but have made this clear in the text
REVIEW RETURNED	23-Oct-2018

GENERAL COMMENTS	Introduction • Page 2 line 43: ‘provide’ should be ’provides’ • Page 3 lines 44 to 47 needs rewording English not clear. ‘Whereas a study conducted in Australian population reported also male gender, foreign nationality and socioeconomic status as predictors of inadequate HL, and, for what concern HL consequences, the study found that low HL levels were associated with a higher risk of chronic diseases and a lower access to primary care services’. Methods: Page 4: • The sample was recruited from General practices. Can this be described as a population sample? The% of the population registered with GPs should be described, together with a description of the characteristics of non-registered people. • GPs recruited using convenience criteria – what criteria were employed? (locality etc). What was the characteristics of the populations of the practices compared with the general population? How did GPs randomly choose the 80 patients? Random number table? Other? • Given my concerns about recruitment, it would be helpful to show the proportions of the sampled population with key characteristics (such as age, gender, ethnicity, education) compared to the general Italian population • Has the NVS IT been published anywhere? Was in based on the US NVS (as cited – Weiss et al) or the European version used in the HLS EU survey (Rowlands et al) (1) – which uses a European-type label. In fact Palumbo is cited in the discussion as having published the NVS-IT – this paper should also be cited in the

methods (conflict of interest statement – I led on the Rowlands et al paper).
Page 4: Procedures:
• Impact of postal invite method?
• Page 5: lines 4 – 11: service use data – is length of recall too long, especially for GP visits. This was the length of time used in the HLS EU study, but should be discussed

Results
• Page 5 lines 19 – 25: Pre-testing of NVS to check the impact of data collection method (online) good but was sample size large enough? How was the decision about sample size reached? Did the sample include people with low skills? If so what % compared to the % of low skills in the general population?
• Page 5 line 27: Statystical should read 'statistical'
• Page 5 line 52 'follow' should read 'follows'
• Page 6 line 5 'is' should read 'was'
• Page 6 line 40 reword
• Page 6 – statistical comparison of responders and non-responders?
• General comment: use a consistent (past) tense.
• Page 6 lines 47 to 50 – 'graduated participants' should read 'participants who had graduated from…'
• Table 1 could BMI groups be merged as per text? Is it possible to reduce table 1 to one page?
• P9 line 7 please provide these results as a supplementary table to be made available online

Discussion
• P9 line 34 should read 'and DESCRIBE OR EXPLORE the association of ….'
• I am concerned about the extrapolation of the findings to the whole of the Italian population given my concerns about the recruitment method. If my suggestion about comparing the sampled population with the general Italian population were carried out, I would be less concerned about this.
• Throughout the discussion there is an assumption that functional HL (as measured by the NVS) is the same as the wider definition (Sorensen et al) used in the HLS-EU-47. This should be made clearer throughout the discussion. Functional 'test' measures like the NVS are much more closely linked with education level, and less linked with aspects of health literacy such as systems navigation etc. This doesn't negate the discussion, but it should be clear that the authors are talking about only functional HL.
• Page 10 lines 20 – 24 – the section about cognitive decline should reference the findings of Kobayashi et al (2) and incorporate this into this section of the discussion
• Page 10 lines 29 – 38 – the discussion about socioeconomic status and HL – I do not this these hypotheses about causality can be drawn from these data. This is a really complex area- I think the association should just be described, together with, perhaps, a reflection on how HL can be seen as an additional social determinant of health.
• Page 10 line 53 'resulted to be' should read 'were found to be'
• Page 11 line 33 – social desirability & recall bias should be referenced.

Additional suggested references

	1. Rowlands G, Protheroe J, Winkley J, et al. A mismatch between population health literacy and the complexity of health information: an observational study. Br J Gen Pract. 2015;65(635):e379-86. 2. Kobayashi LC, Smith SG, O'Conor R, et al. The role of cognitive function in the relationship between age and health literacy: a cross-sectional analysis of older adults in Chicago, USA. BMJ open. 2015;5(4):e007222.

REVIEWER	Dr. Elizabeth Mansfield Trillium Health Partners -- Institute for Better Health
REVIEW RETURNED	17-Dec-2018

GENERAL COMMENTS	4. Are the methods described sufficiently to allow the study to be repeated? P 4, Line 13 -- Please provide a bit more information about the methods rather than simply referring the reader to a published protocol. P 4,Line 23-25 -- Stating that a sampling method has been used by others is not a strong rationale for selecting a specific approach. Please elaborate. P 4,Line 26 - Unclear as to why it is necessary to mention that the president of the Provincial Medical Council "informed their colleagues to join the study." P 4,Line 29 -- I think it is important when describing a population-based sample to provide some contextual information about the districts of Florence included in the 11 GPs' practices. P 4, Line 33 - 38 -- Unclear as to who applied exclusionary criteria -- GPs or was this determined by the researchers -- Unclear as to what "selected subject" means. Prefer the language of "participants" to subjects. 5. Are research ethics (e.g. participant consent, ethics approval) addressed appropriately? P 6, Line 38 -- I am uncomfortable with reporting reasons for nonparticipation as the described participants did not consent to participate in the study. 15. Is the standard of written English acceptable for publication? P 6, Line 43 -- There are different places in this manuscript where "gender" is used incorrectly. Here you are referring to "sex." Sex refers to biological, anatomical characteristics whereas gender refers to social roles or personal identity based on an individual's sex. There are grammar errors throughout this manuscript and other areas where the sentence structure needs to improved and sentence meaning improved. The paper is weakened by a number of instances where grammatical errors and/or writing style issues are present. Below are a few examples selected from the text: P 2, Line 30 -- Capitalize Findings P 2, Line 36 -- Capitalize Hospital P 3, Line 5 -- Stylistically, prefer that a manuscript does not begin with a definitional quote -- put in your own words P 3, Line 26 -- What does "Until today" mean? Incorrect usage here. Page 3, Line 37 -- "resulted to significantly predict" -- needs to be restated

	P 3, Line 44 -- Was it male gender or sex? How were the researchers getting at gender? P 3, Line 50 -- "conducted in small, convenience samples" change to "conducted with" P 4, Line 23 -- "study at hand" -- very casual language, would change P 5, Line 8 -- "the use of other..." subject verb agreement incorrect P 5, Line 18 -- "Due to that" Due to what? P 5, Line 23 -- What is a "washout period" Multiple instances where articles are missing -- e.g. P 9, Line 8 "as dependent variable" change to "as the dependent variable" Page 9, Line 49 -- not sure what the means "relatively homogeneity" P 10, Line 7 -- meaning of this sentence is unclear -- "no consistency in discussing..." P 10, Line 12 -- "contribute with evidences" unclear P 10, Line 15 -- first sentence -- "confirmed it" state more formally what was confirmed P 10, Line 24 -- Please rewrite this sentence "As older and less-educated people are those who experiment the highest burden...." P 10, Line 50 -- "none of these models has been..." P 10, Line 53 -- "no other health outcomes resulted to be...." Entire paragraph needs clarity and to be meaningfully connected to examples from the literature. Other comments: P 3, Line 7 -- If HL is a major public health problem, please explain why. Think a more persuasive case for HL needs to be made in the introduction. Noting associations of HL with health behaviors and outcomes does not provide theoretical support as to why this is a major public health problem -- a little more context here would be helpful and will strengthen the research rationale for the reader. P 3 Line 18 -- Do you mean several minutes for the time to administer the tool? This seems a little strange given that the telephone interviews for administering the shortened version of the tool take 20 to 25 minutes.

REVIEWER	Richard Osborne Deakin University
REVIEW RETURNED	02-Jan-2019

GENERAL COMMENTS	This is an innovative paper from Italy which explores health literacy in patients attending GP clinics. It uses an early generation functional health literacy test. It is a cross sectional study that explores correlations between health literacy and demographic and health variables. Overall it is not a big advancement of the field. The paper provides limited insight on exactly what can be done to improve primary care, including the causes and solutions regarding the challenges posed to clinical practice and public health that are related to health literacy. The authors claim the study is population-based but it is not the case. From the methods section and previous paper, GPs were selected using convenience sampling – first come basis – such clinics are not at all likely to be representative. While the patients may have been selected at random from the GP clinics, this is not a population-based sampling method. It may be representative of

the particular patients attending particular clinics, but that is all. This misunderstanding was not identified in the previous BMJ Open publication.

Introduction

P3 L24 – The statement ‘The NVS shows high sensitivity in detecting limited literacy’ is not correct. The reference cited [8] provides weak evidence to this effect, and probably says the opposite, i.e., that the NVS misclassifies some people. As there is no clear empirical work on what is high/low HL cut off, sensitivity cannot be estimated for any HL test. The authors need to reconsider the evidence regarding data collected from the NVS – they claim the test is ‘validated’ however it is important to understand that it is not a questionnaire that is validated, rather it is the data from a questionnaire for use for a particular purpose in a particular setting that is validated.

P3 L29-32. These sentences do not make sense. On review of ref 12, this paper doesn’t seem to say anything about over representation of disadvantaged groups. Typically, surveys, such as the one conducted, have under representation of the people the researchers seek to study. This paper does say “A weakness in the health literacy field is that the most commonly used tools mainly test reading, comprehension and numeracy skills, and some cognitive tasks, rather than the broad range of issues included in modern definitions of health literacy.” Which is a critical weakness of the current study. I would expect clinical practice needs much more information than a person’s reading and writing ability – i.e., the full range of skills related to accessing, understanding and using health information and services.

P3 L53 – Why do the authors expect to see a consensus on antecedents and consequences of HL? HL is well known to be dependent on contextual factors, e.g., the ease of navigation of local healthcare systems, the attitudes and skills of local healthcare professionals, social connectedness, education and poverty, etc. The field is profoundly variable for this reason.

Methods

P4 Did the method of recruitment specifically exclude illiterate people? Did GPs select in people who could read? Were people with other reading difficulties (e.g., sight problems) supported to participate? The exclusion of people through written consent procedures is another reason for this not being representative. It certainly is not representative if less than 50% of the sample invited actually took part.

P5 L4. Explain “referred weight”.

P5 L17. The testing of telephone vs face to face administration of the NVS is interesting and novel. The random sequence is a strong design. A statistical test to show no difference (without a sample size estimate) is not sufficient. What is the power to show equivalence? Do not undertake hypothesis testing of demographic differences – present the group differences and whether these differences are clinically or socially meaningful. It is likely that ROC (Receiver Operating Curves) is better statistical procedure to explore equivalence across scale scores. While the mean differences are not statistically significant (using conservative non-parametric tests on a small sample) the absolute score differences seem large. The mean difference is 0.65 (4.76 – 4.11) which is more that 10% of the scale range (the range is 0 to 6). Another issue is that the scale score is 'lumpy' due to only 7 questions – so a small change in an average score can lead to augmented misclassification (i.e., the categories of high/low HL), i.e., do the differences between administration methods lead to clinically/socially important differences in the number of people misclassified? This is important as there appears to be few people in the low category – and this is where stronger evidence is needed about the equivalence. A scatterplot of the two methods with cut offs marked would help the reader understand equivalence (and possibly a ROC curve).

P6 L8 Patient and public involvement. It seem that only professionals were involved in this study. Most of the content related to his section is therefore not relevant to the section. Consider removing.

Table 1.
Tables should be standalone. Include the full term for HL.
Include all categories for the Self-reported health status
Include units for Long-term illness – this probably should be N (%)
BMI in full
For family members in household – does this exclude other non-family members living in the household?
I think it table would more insightful if the % were calculated for columns, rather than rows. What is the research question here? Is it, among those people in the HL categories, what is the frequency of people with high/low education etc.?

P8 L28 It is inadequate to simply say an association was present. The sample size is large, and the direction of the association could be positive or negative by a clinically irrelevant or important amount, but still be statistically associated. The uncertainty and direction and magnitude should be provided.

Table 2. Provide unadjusted estimates. It is not clear what is in the model and what has been adjusted for.

Table 3. This is the most important and interesting results. Include unadjusted estimates and exactly what was included in the model.

P9 L33 Note that this is not a population-based sample, and the data cannot be used to generalise about the general Italian population, nor the general population of people attending GP practices. The data are relevant to <50% of people (who are likely to have higher HL than the general population due to the recruitment process – i.e., having to read the consent form / survey) attending selected GP practices.

P9 L42 It is important that like is being compared with like – did the EU survey use the NVS or the HLS? These are not comparable. If different sampling strategies were used then the differences are likely to be sampling variations, not any population level differences. This is eluded to in P9L50 but is not clear. If the data are not comparable (and a strong argument that the data are comparable needs to be made so as to not mislead the reader) then the findings should not be compared. The authors need to

	consider internal and external validity. Given the sampling, and potential for misleading findings, the paper should mainly focus on results that arguably have internal validity – i.e., the antecedent analysis. The authors should refer to the findings of a recent BMJ Open paper that suggests the HL tests (such as NVS) may be more related to cognitive ability than to HL per se. https://bmjopen.bmj.com/content/8/9/e022502 This study focuses on functional HL as measured by the NVS, so the Discussion needs to carefully reflect this unidimensional aspect of HL, primarily stating 'functional HL' when referring to their results. Readers need to be kept aware of all the other important elements of HL not measured. P10L45 The Sorensen model, with 4 competencies generated in three domains has only been posited, and, to the knowledge of this reviewer, not yet tested let alone partially validated. P11 L11 – this paragraph will need substantial revision given the analysis noted above. Also, it is the experience of interviewers using the NVS that the application of NVS can induce stigma, shame and stress in people with low literacy/numeracy. People who received the invitation to take part who are at risk of this may not take part – this is a major methodological concern, and potentially greatly limits the clinical insights from this study. P11 L31 – the convenience sampling and the way people were recruited needs to be listed as a major limitation for external comparisons – as noted above. The age, education, literacy levels etc could be compared with national norms to explore more accurately whether the data are at least comparable. This paragraph also brings new data in about the sampling strategy which should not happen in the Discussion. The authors should discuss how robust the cut offs of the categories of HL are using the NVS – have they yet been tested against any socially or clinically relevant/meaningful indicators?

REVIEWER	Delphine Courvoisier HUG Switzerland
REVIEW RETURNED	10-Jan-2019

GENERAL COMMENTS	This article presents the associations between functional HL as measured by the NVS with antecedents and consequences, using a cross-sectional design. Abstract : design : the design is not randomized, which is usually used when an intervention is randomized, except for the small comparison of face-to-face vs. Telephone interview. P4, line 30 : asking GP to recruit randomly without providing method to do a random selection does not yield a random sample. The authors do acknowledge that it is a convenience sample but the risk of bias, especially on the prevalence of limited HL is high.

	Statistical analysis : for inter-method reliability (NVS by phone or by face-to-face), the correct method of analysis is ICC(2,1). A non-significant paired t-test only says that the sample is small. Table 1 : please justify why long-term illness does not have percentages. Table 1 : why indicate tests in the note below the table, since you do not report any p-value. Tests are usually indicated in the methods, or could be indicated in p8,line29 when you report the univariable associations. Model selection : the selection of variables to include in the multivariable model by taking only the significant univariable associations is not recommended. It may lead to excluding variables that would have been relevant and were non significant due to confouding. A better selection method is the LASSO, or since your sample size is relatively large, you could include all predictors in a first multivariable model. Table 2 : p-values lower than 0.0000 are usually indicated as <0.001 Table 3 : the pseudo R2 is greater than 1. Please check your metric. P10,line14. Education has changed a lot over time, and age could be an independent predictor of HL because it is a proxy for receiving an education that never talked about health (for instance no sex education), but not because of cognitive decline

VERSION 1 – AUTHOR RESPONSE

Reviewers' Comments to Author:

Reviewer: 1

Reviewer Name: Gillian Rowlands

Introduction

- Page 2 line 43: ‘provide’ should be ’provides’

Reply: we have corrected the error in the revised manuscript as suggested.

- Page 3 lines 44 to 47 needs rewording English not clear. ‘Whereas a study conducted in Australian population reported also male gender, foreign nationality and socioeconomic status as predictors of inadequate HL, and, for what concern HL consequences, the study found that low HL levels were associated with a higher risk of chronic diseases and a lower access to primary care services’.

Reply: we have reworded the sentence in the revised manuscript (please see p. 3 ll. 37 - 39).

Methods:

Page 4:

- The sample was recruited from General practices. Can this be described as a population sample? The% of the population registered with GPs should be described, together with a description of the characteristics of non-registered people.

Reply: The sample should be considered population-based as it was recruited from a list of residents available from the registers of general practices of the municipality of Florence. We have better specified the characteristics of these registers as well as the % of population registered in the revised manuscript (please see p. 4 ll. 20-28). The percentage of non-registered people is really very small and is mainly represented by people who are in the process to be registered in a general practice (i.e. new residents or people who have recently changed the place of residence).

Please note that the sample was not designed to be representative of the overall Italian or Florentine population. Indeed, the population-based sample was obtained with a combination of convenience and probability sampling procedures: GPs were recruited with convenience criteria, and each recruited GP subsequently selected 80 subjects from its register through a random number generator (this has been better detailed in the methods section of the revised manuscript, please see p. 4 ll. 29-38).Thus, the sample cannot be considered representative. We have acknowledged this issue in the revised manuscript, and we have revised the discussion section in order to avoid any possible misunderstandings about the extrapolation of the findings to the whole Italian population (please see p. 12 ll. 10-12, and p. 14 ll. 5-8).

- GPs recruited using convenience criteria – what criteria were employed? (locality etc). What was the characteristics of the populations of the practices compared with the general population? How did GPs randomly choose the 80 patients? Random number table? Other?

Reply: All the GPs of the municipality were invited to join the study; and GPs were recruited on a first-come basis. The GPs selected the subjects from their registers through a random number generator. We have better detailed the recruitment procedures in the revised manuscript (please see p. 4 ll. 29-38).

As for the study population, please note that it is not a patients' population, but it is composed by Italian and Foreign people that reside in the area served by the practices. We have better clarified this issue in the revised manuscript (please see p. 4 ll. 20-28). Furthermore, we have reported the characteristics of the areas of Florence included in the study in the revised manuscript (please see p. 4 ll. 33-34). As the sample was not designed to be representative of the overall Italian or Florentine population, we have supposed that it is not necessary to provide further comparison with the Italian general population.

- Given my concerns about recruitment, it would be helpful to show the proportions of the sampled population with key characteristics (such as age, gender, ethnicity, education) compared to the general Italian population

Reply: As noted above, although the sample was population-based, it was not designed to be representative of the overall Italian or Florentine adult population. We have better specified the sampling frame and the recruitment procedures in the methods section (please see p. 4 ll. 20-38) and acknowledged this issue in the discussion section of the revised manuscript (please see p. 12 ll. 10-12, and p. 14 ll. 5-8).

- Has the NVS IT been published anywhere? Was in based on the US NVS (as cited – Weiss et al) or the European version used in the HLS EU survey (Rowlands et al) (1) – which uses a European-type label. In fact Palumbo is cited in the discussion as having published the NVS-IT – this paper should also be cited in the methods (conflict of interest statement – I led on the Rowlands et al paper).

Reply: Yes, the validation study of the NVS-IT was published on a peer-reviewed journal (please see reference n. 18 of the revised manuscript). The NVS-IT was developed from the European version used in the HLS EU survey (Rowlands et al, reference n. 17 of the revised manuscript). We have better specified these points in the introduction section of the revised manuscript (please see p. 3 l.40 and p.4 l.1). Please note that the NVS-IT validation study was conducted by our research group, the research group of Palumbo did not published any validation study of the NVS-IT.

Page 4: Procedures:

- Impact of postal invite method?

Reply: we have better detailed the study recruitment procedure(please see p. 5 ll. 7-10), and we have reported the impact of postal invitation in the result section of the revised manuscript (please see p. 8 ll.5-9).

- Page 5: lines 4 – 11: service use data – is length of recall too long, especially for GP visits. This was the length of time used in the HLS EU study, but should be discussed

Reply: We have acknowledged this issue as a limitation of the study in the revised manuscript (please see p.14 ll. 12-15).

Results

- Page 5 lines 19 – 25: Pre-testing of NVS to check the impact of data collection method (online) good but was sample size large enough? How was the decision about sample size reached?

Did the sample include people with low skills? If so what % compared to the % of low skills in the general population?

Reply: No agreement exists on how to determine the sample size for test-retest studies. In our study, sample size was established considering the suggestion of Parker (2018), considering a number of participants of about 5 times the number of the items of the tool (i.e. NVS-IT). Furthermore, this sample size was chosen in line with the sample size requirements for estimating the value of intraclass correlation coefficient and the Cohen's kappa agreement test as proposed by Bujang (2017, two articles). These considerations have been added to the revised manuscript (please see p. 5 ll. 35-38).

As for the people with low skills, the sample includes people with low educational level, and the distribution of educational level in the pre-testing sample was similar to the educational level distribution in the population-based sample of the study. We have reported the distribution of educational level of the pre-testing sample in the result section of the revised manuscript (please see p. 7 ll. 10-11).

- Park MS, Kang KJ, Jang SJ, Lee JY, Chang SJ. Evaluating test-retest reliability in patient-reported outcome measures for older people: A systematic review. Int J Nurs Stud 2018;79:58-69. doi: 10.1016/j.ijnurstu.2017.11.003.
- Bujanga MA, Baharum N. Guidelines of the minimum sample size requirements for Cohen's Kappa. Epidemiology Biostatistics and Public Health 2017;e12267:1-10. doi: 10.2427/12267
- Bujanga MA, Baharum N. A simplified guide to determination of sample size requirements for estimating the value of intraclass correlation coefficient: a review. Arch Orofac Sci 2017;12(1):1-11.

- Page 5 line 27: Statystical should read 'statistical'

Reply: we have corrected the error in the revised manuscript as suggested.

- Page 5 line 52 'follow' should read 'follows'

Reply: we have corrected the error in the revised manuscript as suggested.

- Page 6 line 5 'is' should read 'was'

Reply: we have corrected the error in the revised manuscript as suggested.

- Page 6 line 40 reword

Reply: we have reworded the sentence in the revised manuscript (please see p.8 ll. 5-9).

- Page 6 – statistical comparison of responders and non-responders?

Reply: statistical comparisons on age and gender were already reported in the manuscript (please see p. 8 ll. 9-11). Unfortunately, we do not have any other information on the non-responders, therefore we are unable to provide any further statistical comparisons.

- General comment: use a consistent (past) tense.

Reply: we have revised the result section using a consistent past tense in the revised manuscript.

- Page 6 lines 47 to 50 – 'graduated participants' should read 'participants who had graduated from…'

Reply: we have reworded the phrase in the revised manuscript (please see p. 8 l. 15).

- Table 1 could BMI groups be merged as per text? Is it possible to reduce table 1 to one page?

Reply: we have merged BMI groups as per text in the revised manuscript. Unfortunately, it was not possible to reduce the table to one page without removing several variables. It would be possible to split the table into two tables (one for the antecedents of HL and one for outcomes of HL), please do not hesitate to let us know if this is a more suitable option.

- P9 line 7 please provide these results as a supplementary table to be made available online

Reply: the supplementary table was provided as suggested.

Discussion

- P9 line 34 should read 'and DESCRIBE OR EXPLORE the association of ….'

Reply: the suggested correction has been made in the revised manuscript.

- I am concerned about the extrapolation of the findings to the whole of the Italian population given my concerns about the recruitment method. If my suggestion about comparing the sampled population with the general Italian population were carried out, I would be less concerned about this.

Reply: as noted in previous comments/responses, the sample was not designed to be representative of the Italian or Florentine adult population. We have better specified the sampling frame and recruitment procedures in the methods section of the revised manuscript (please see p. 4 ll. 20-38). Furthermore, we have revised the discussion section in order to avoid any possible misunderstandings about the extrapolation of the findings to the whole Italian or Florentine population (please see p. 12 ll. 10-12 and p. 14 ll. 5-8 of the revised manuscript).

- Throughout the discussion there is an assumption that functional HL (as measured by the NVS) is the same as the wider definition (Sorensen et al) used in the HLS-EU-47. This should be made clearer throughout the discussion. Functional 'test' measures like the NVS are much more closely linked with education level, and less linked with aspects of health literacy such as systems navigation etc. This doesn't negate the discussion, but it should be clear that the authors are talking about only functional HL.

Reply: we agree with the reviewer. In the discussion section of the revised manuscript we have stated "functional HL" when referring to our results in order to make clear throughout the discussion that the study focused on functional health literacy.

- Page 10 lines 20 – 24 – the section about cognitive decline should reference the findings of Kobayashi et al (2) and incorporate this into this section of the discussion

Reply: the findings of the suggested study were cited in the discussion section of the revised manuscript (please see p. 12 ll. 33-35 and reference n. 31).

- Page 10 lines 29 – 38 – the discussion about socioeconomic status and HL – I do not this these hypotheses about causality can be drawn from these data. This is a really complex area- I think the association should just be described, together with, perhaps, a reflection on how HL can be seen as an additional social determinant of health.

Reply: we have revised the paragraph removing causality hypotheses as suggested by the reviewer (please see p. 13 ll. 1-5 of the revised manuscript).

- Page 10 line 53 'resulted to be' should read 'were found to be'

Reply: we have corrected the error in the revised manuscript as suggested.

- Page 11 line 33 – social desirability & recall bias should be referenced.

Reply: we have referenced recall and social desirability bias in the revised manuscript (please see reference n. 45 of the revised manuscript).

- Additional suggested references

1. Rowlands G, Protheroe J, Winkley J, et al. A mismatch between population health literacy and the complexity of health information: an observational study. Br J Gen Pract. 2015;65(635):e379-86.

2. Kobayashi LC, Smith SG, O'Conor R, et al. The role of cognitive function in the relationship between age and health literacy: a cross-sectional analysis of older adults in Chicago, USA. BMJ open. 2015;5(4):e007222.

Reviewer: 2

Reviewer Name: Dr. Elizabeth Mansfield

- P 4, Line 13 -- Please provide a bit more information about the methods rather than simply referring the reader to a published protocol.

Reply: we have provided more information about the methods in the revised manuscript (please see p. 4 ll. 20 -38, p. 5 ll. 7-10, and p. 5 ll.-35-38).

- P 4, Line 23-25 -- Stating that a sampling method has been used by others is not a strong rationale for selecting a specific approach. Please elaborate.

Reply: we have described the rationale about the sampling method adopted in the revised manuscript (please see p. 4 ll. 26-28).

- P 4, Line 26 - Unclear as to why it is necessary to mention that the president of the Provincial Medical Council "informed their colleagues to join the study."

Reply: We deemed necessary to mention that the president of the Provincial Medical Council ”informed their colleagues to join the study” because it highlights the fact that all the GPs of the municipality of Florence were invited to join the study, and this imply that the same possibility to join the study was given to all the GPs of the city of Florence. Furthermore, it provides information about the way the GPs were invited. We think that these are a relevant aspects of the study methodology. We have reformulated the sentence in the method section to better explain these purposes in the revised manuscript (please see p. 4 ll. 29-30). Furthermore, we have avoided the direct mention of the president in the revised manuscript.

- P 4, Line 29 -- I think it is important when describing a population-based sample to provide some contextual information about the districts of Florence included in the 11 GPs' practices.

Reply: we have provided more contextual information about the districts of Florence included in the GPs practices in the revised manuscript (please see p.4 ll. 33-34).

- P 4, Line 33 - 38 -- Unclear as to who applied exclusionary criteria -- GPs or was this determined by the researchers -- Unclear as to what "selected subject" means. Prefer the language of "participants" to subjects.

Reply: Exclusion criteria were applied by the GPs. We have better clarified this aspect in the revised manuscript (please see p. 4 l. 38).

The “selected subjects” are the people that were randomly chosen by the GPs and invited to participate to the study. Please note that these people cannot defined “participants” as -at this stage- they were not invited to participate. In this cases we have changed the word “subjects” to the word “people”, while in the remaining cases we have used the word “participants” instead of the word “subjects” in the revised manuscript, as suggested.

- P 6, Line 38 -- I am uncomfortable with reporting reasons for nonparticipation as the described participants did not consent to participate in the study.

Reply: non-participation reasons have been reported in a general and aggregate way in the revised manuscript (please see p.8 ll. 6-9). We think that it is important to distinguish between people who refused to participate in the study and people who were unreachable. To the best of our knowledge, this is compliant with privacy and ethical regulations.

- P 6, Line 43 -- There are different places in this manuscript where "gender" is used incorrectly. Here you are referring to "sex." Sex refers to biological, anatomical characteristics whereas gender refers to social roles or personal identity based on an individual's sex.

Reply: we have changed the word “gender” to “sex” in the revised manuscript as suggested

- There are grammar errors throughout this manuscript and other areas where the sentence structure needs to improved and sentence meaning improved. The paper is weakened by a number of instances where grammatical errors and/or writing style issues are present. Below are a few examples selected from the text:

P 2, Line 30 -- Capitalize Findings

P 2, Line 36 -- Capitalize Hospital

P 3, Line 5 -- Stylistically, prefer that a manuscript does not begin with a definitional quote -- put in your own words

P 3, Line 26 -- What does "Until today" mean? Incorrect usage here.

Page 3, Line 37 -- "resulted to significantly predict" -- needs to be restated

P 3, Line 44 -- Was it male gender or sex? How were the researchers getting at gender?

P 3, Line 50 -- "conducted in small, convenience samples" change to "conducted with"

P 4, Line 23 -- "study at hand" -- very casual language, would change

P 5, Line 8 -- "the use of other..." subject verb agreement incorrect

P 5, Line 18 -- "Due to that" Due to what?

P 5, Line 23 -- What is a "washout period"

Multiple instances where articles are missing -- e.g. P 9, Line 8 "as dependent variable" change to "as the dependent variable"

Page 9, Line 49 -- not sure what the means "relatively homogeneity"

P 10, Line 7 -- meaning of this sentence is unclear -- "no consistency in discussing..."

P 10, Line 12 -- "contribute with evidences" unclear

P 10, Line 15 -- first sentence -- "confirmed it" state more formally what was confirmed

P 10, Line 24 -- Please rewrite this sentence "As older and less-educated people are those who experiment the highest burden...."

P 10, Line 50 -- "none of these models has been..."

P 10, Line 53 -- "no other health outcomes resulted to be...." Entire paragraph needs clarity and to be meaningfully connected to examples from the literature.

Reply: We regret there were problems with the English. The manuscript has been carefully revised to improve the grammar and readability.

Other comments:

- P 3, Line 7 -- If HL is a major public health problem, please explain why. Think a more persuasive case for HL needs to be made in the introduction. Noting associations of HL with health behaviors and outcomes does not provide theoretical support as to why this is a major public health problem -- a little more context here would be helpful and will strengthen the research rationale for the reader.

Reply: We have provided more context on the topic in the revised manuscript as suggested (please see p. 3 ll. 5-7).

- P 3 Line 18 -- Do you mean several minutes for the time to administer the tool? This seems a little strange given that the telephone interviews for administering the shortened version of the tool take 20 to 25 minutes.

Reply: Yes, various functional HL tools take several minutes to administer (e.g. TOFHLA, S-TOFHLA and METER). Please note that the telephone interview of our study also included the presentation of the interviewer and the investigation of antecedents and outcomes. We have better specified this aspect in the revised manuscript (please see p. 5 ll.13-14).

Reviewer: 3

Reviewer Name: Richard Osborne

This is an innovative paper from Italy which explores health literacy in patients attending GP clinics. It uses an early generation functional health literacy test. It is a cross sectional study that explores correlations between health literacy and demographic and health variables. Overall it is not a big advancement of the field. The paper provides limited insight on exactly what can be done to improve primary care, including the causes and solutions regarding the challenges posed to clinical practice and public health that are related to health literacy.

- The authors claim the study is population-based but it is not the case. From the methods section and previous paper, GPs were selected using convenience sampling – first come basis – such clinics are not at all likely to be representative. While the patients may have been selected at random from the GP clinics, this is not a population-based sampling method. It may be representative of the particular patients attending particular clinics, but that is all. This misunderstanding was not identified in the previous BMJ Open publication.

Reply: we regret not having better detailed the sampling frame and the recruitment procedures in the original manuscript. Indeed, the study population should be considered population-based as it was selected from registers of Italian and Foreign residents of the area served by the practices. The characteristics of the registers from which the study population was chosen were not clearly explained in the original manuscript; we have better detailed these characteristics in the revised manuscript (please see p. 4 ll. 20-28). Furthermore, please note that the sample was not designed to be representative of the overall Italian or Florentine population. We have acknowledged this issue in the revised manuscript, and we have revised the discussion section in order to avoid any possible misunderstandings about the extrapolation of the findings to the whole Italian population (please see p. 12 ll. 10-12 and p. 14 ll. 5-8).

Introduction

- P3 L24 – The statement ‘The NVS shows high sensitivity in detecting limited literacy’ is not correct. The reference cited [8] provides weak evidence to this effect, and probably says the opposite, i.e., that the NVS misclassifies some people. As there is no clear empirical work on what is high/low HL cut off, sensitivity cannot be estimated for any HL test. The authors need to reconsider the evidence regarding data collected from the NVS – they claim the test is ‘validated’ however it is important to understand that it is not a questionnaire that is validated, rather it is the data from a questionnaire for use for a particular purpose in a particular setting that is validated.

Reply: we agree with the reviewer on the fact that “sensitivity cannot be estimated for any HL test”; we have revised the sentence in the revised manuscript (please see p. 3 ll. 20-21).

As far as the term “validated” is concerned, we have avoided the inappropriate use of this term in the revised manuscript.

- P3 L29-32. These sentences do not make sense. On review of ref 12, this paper doesn’t seem to say anything about over representation of disadvantaged groups. Typically, surveys, such as the one conducted, have under representation of the people the researchers seek to study. This paper does say “A weakness in the health literacy field is that the most commonly used tools mainly test reading, comprehension and numeracy skills, and some cognitive tasks, rather than the broad range of issues included in modern definitions of health literacy.” Which is a critical weakness of the current study. I would expect clinical practice needs much more information than a person’s reading and writing ability – i.e., the full range of skills related to accessing, understanding and using health information and services.

Reply: There was an error, we wanted to say “under-representation” instead of “over-representation”. We have corrected the error in the revised manuscript (please see p. 3 l. 26). The corrected sentence is now in line with what is reported at page 7 of the cited study: “One of the problems with many approaches to health literacy interventions in healthcare settings is that they focus only on those patients who are already accessing health services. However, the overall effectiveness of a health service organization is largely dependent on whether or not the people who need it most actually access the service. Low health literacy - as represented by such issues as low educational attainment and low socio-economic position - is a major barrier to access for many people“.

As noted above, our sample should be considered population-based, and thus it should not be considered per se as affected by the issue of under-representation of people with low SES.

- P3 L53 – Why do the authors expect to see a consensus on antecedents and consequences of HL? HL is well known to be dependent on contextual factors, e.g., the ease of navigation of local healthcare systems, the attitudes and skills of local healthcare professionals, social connectedness, education and poverty, etc. The field is profoundly variable for this reason.

Reply: we agree with the reviewer’s comment; the sentence has been modified in the revised manuscript (please see p.4 ll. 3-4).

Methods

- P4 Did the method of recruitment specifically exclude illiterate people? Did GPs select in people who could read? Were people with other reading difficulties (e.g., sight problems) supported to participate? The exclusion of people through written consent procedures is another reason for this not being representative. It certainly is not representative if less than 50% of the sample invited actually took part.

Reply: No, the method of recruitment did not exclude illiterate people. GPs randomly selected the participants through a random number generator from a list of residents. As for people with reading difficulties, the written consent form is required by the Italian law, however a follow-up phone call was made to all the invited people in order to clarify any questions and to provide assistance to any people with reading difficulties. Thus, the exclusion of people through written consent procedures is very limited in our study. Furthermore, the study was designed to facilitate the participation of people with reading difficulties as the survey was carried out with an oral interview and the NVS label was designed to be easy readable (i.e. large font size and line-spacing). We have better detailed the recruitment procedures in the revised manuscript (please see p.4 l. 35 and p.5 ll. 6-10).

As noted in the previous comments/responses, our study was not designed to be representative of the overall Italian or Florentine population.

- P5 L4. Explain “referred weight”.

Reply: we meant “self-reported weight”; we have better specified this in the revised manuscript.

- P5 L17. The testing of telephone vs face to face administration of the NVS is interesting and novel. The random sequence is a strong design. A statistical test to show no difference (without a sample size estimate) is not sufficient. What is the power to show equivalence? Do not undertake hypothesis testing of demographic differences – present the group differences and whether these differences are clinically or socially meaningful. It is likely that ROC (Receiver Operating Curves) is better statistical procedure to explore equivalence across scale scores. While the mean differences are not statistically significant (using conservative non-parametric tests on a small sample) the absolute score differences seem large. The mean difference is 0.65 (4.76 – 4.11) which is more that 10% of the scale range (the range is 0 to 6). Another issue is that the scale score is ‘lumpy’ due to only 7 questions – so a small change in an average score can lead to augmented misclassification (i.e., the categories of high/low HL), i.e., do the differences between administration methods lead to clinically/socially important differences in the number of people misclassified? This is important as there appears to be few people in the low category – and this is where stronger evidence is needed about the equivalence. A scatterplot of the two methods with cut offs marked would help the reader understand equivalence (and possibly a ROC curve).

Reply: Statistical analysis of data coming from the test-retest phase has been improved in the revised manuscript. Specifically: a) the descriptive statistics has been improved; b) the Intraclass correlation coefficient (ICC) has been calculated for the whole dataset and separately in the two subgroups; c) two scatterplots have been added; d) Chi2 test was used to evaluate the association between the classification into two groups of HL (inadequate and at risk HL vs adequate HL) and the mode of administration at the first and at the second interview, respectively; e) Cohen's kappa to assess the agreement in the classification into two groups of HL (inadequate and at risk HL vs adequate HL) at T0 and T1 has been added. Furthermore, sample size decision has been discussed in the methods of the revised manuscript (please see p. 5 ll. 35-38).

Regarding ROC curves, in our opinion this method is not adequate here since we do not have a gold standard diagnostic method that classify the subjects into two groups, and whose results have to be predicted by another diagnostic method that give continuous values. Indeed, we are comparing scores (from 1 to 6) obtained administering the same test with different modes of administration. For this reason, according also with the request of the referee n. 4, we have calculated the ICC and the Cohen's kappa.

- P6 L8 Patient and public involvement. It seem that only professionals were involved in this study. Most of the content related to his section is therefore not relevant to the section. Consider removing.

Reply: We have revised the section and removed the non-relevant content (please see p.7 ll. 5-6).

Table 1.

- Tables should be standalone. Include the full term for HL.

Reply: The table has been modified as suggested.

- Include all categories for the Self-reported health status

Reply: All categories have been included in the revised table

- Include units for Long-term illness – this probably should be N (%)

Reply: The table has been modified as suggested.

- BMI in full

Reply: The table has been modified as suggested.

- For family members in household – does this exclude other non-family members living in the household?

Reply: it also includes the non-family members living in the household; this characteristic of the variable has been better specified in the main text and in the table of the revised manuscript (please see p.5 ll.16-17).

- I think it table would more insightful if the % were calculated for columns, rather than rows. What is the research question here? Is it, among those people in the HL categories, what is the frequency of people with high/low education etc.?

Reply: The percentage was calculated for column only for the total. As for the HL levels columns, we think it is more insightful to report the percentages calculated for rows in order to show the frequencies of HL levels within each variable class.

- P8 L28 It is inadequate to simply say an association was present. The sample size is large, and the direction of the association could be positive or negative by a clinically irrelevant or important amount, but still be statistically associated. The uncertainty and direction and magnitude should be provided.

Reply: Please note that, according to what has been suggested by the reviewer n. 4, the approach to the regression analysis has been changed, and the univariate logistic regression models were not performed in this case. We have better specified the direction of the association between HL level and the variables in the main text of the revised manuscript (please see p. 10 ll. 12-15).

- Table 2. Provide unadjusted estimates. It is not clear what is in the model and what has been adjusted for.

Reply: The estimates are unadjusted. We have better specified what was included in the model in the methods section of the revised manuscript (please see p. 6 ll. 27-32). Furthermore, please note that, according to what has been request by the referee n 4, the approach we have used in the multiple regression in this revised manuscript is different from the one we had used in the previous version of our manuscript.

- Table 3. This is the most important and interesting results. Include unadjusted estimates and exactly what was included in the model.

Reply: the estimates are unadjusted. We have better specified what was included in the model in the methods section of the revised manuscript (please see p. 6 ll. 39-40). Furthermore, please note that, according to what has been request by the referee n 4, the approach we have used in the multiple regression in this revised manuscript is different from the one we had used in the previous version of our manuscript.

- P9 L33 Note that this is not a population-based sample, and the data cannot be used to generalise about the general Italian population, nor the general population of people attending GP practices. The data are relevant to <50% of people (who are likely to have higher HL than the general population due to the recruitment process – i.e., having to read the consent form / survey) attending selected GP practices.

Reply: As noted in previous comments/responses, the sample should be considered population-based as it was recruited from a list of residents available from the registers of general practices of the municipality of Florence. We have better specified the characteristics of these registers in the revised manuscript (please see p. 4 ll. 20-28).

Please note that although our sample should be considered population-based, it cannot be considered representative of the overall Italian or Florentine adult population. Indeed, the population-based sample was obtained with a combination of convenience and probability sampling procedures: GPs were recruited with convenience criteria, and each recruited GPs subsequently selected 80 subjects from their registers through a random number generator (we have better specified these aspects at p. 4 ll. 29-35 of the revised manuscript). We have acknowledged that our sample was not designed to be representative in the revised manuscript, and we have revised the discussion section in order to avoid any possible misunderstandings about the extrapolation of the findings to the whole Italian population (please see p. 12 ll. 10-12 and p. 14 ll. 5-8).

As for the recruitment process, as noted in the previous comment/response, in our study people with reading difficulties were supported to participate in the study. Furthermore, the survey was carried out with an oral interview and the NVS label was designed to be easy readable (i.e. large font size and line-spacing). We have specified these aspects in the revised manuscript (please see p 5 ll. 5-10).

- P9 L42 It is important that like is being compared with like – did the EU survey use the NVS or the HLS? These are not comparable. If different sampling strategies were used then the differences are likely to be sampling variations, not any population level differences. This is eluded to in P9L50 but is not clear. If the data are not comparable (and a strong argument that the data are comparable needs to be made so as to not mislead the reader) then the findings should not be compared. The authors need to consider internal and external validity. Given the sampling, and potential for misleading findings, the paper should mainly focus on results that arguably have internal validity – i.e., the antecedent analysis.

Reply: we agree with the reviewer's comment; data from the EU survey are not comparable with our data as our study was based on a convenience population-based sample. We have removed the comparison of our findings with those of the EU survey from the discussion section of revised manuscript. Furthermore, as noted in previous comments/responses, we have revised the discussion section in order to avoid any possible misunderstandings about the extrapolation of the findings to the overall Italian population, and we have acknowledged the limited external validity of the study in the revised manuscript (please see p. 12 ll. 10-12 and p. 14 ll. 5-8).

- The authors should refer to the findings of a recent BMJ Open paper that suggests the HL tests (such as NVS) may be more related to cognitive ability than to HL per se. https://bmjopen.bmj.com/content/8/9/e022502

Reply: the findings of the suggested reference were mentioned in the discussion section of the revised manuscript (please see p. 12 ll. 32-35 and reference n. 30).

- This study focuses on functional HL as measured by the NVS, so the Discussion needs to carefully reflect this unidimensional aspect of HL, primarily stating ‘functional HL’ when referring to their results. Readers need to be kept aware of all the other important elements of HL not measured.

Reply: we have stated “functional HL” for referring to our results in the discussion section of the revised manuscript as suggested.

- P10L45 The Sorensen model, with 4 competencies generated in three domains has only been posited, and, to the knowledge of this reviewer, not yet tested let alone partially validated.

Reply: the direct mention of the Sorensen model was removed from the paragraph in the revised manuscript.

- P11 L11 – this paragraph will need substantial revision given the analysis noted above. Also, it is the experience of interviewers using the NVS that the application of NVS can induce stigma, shame and stress in people with low literacy/numeracy. People who received the invitation to take part who are at risk of this may not take part – this is a major methodological concern, and potentially greatly limits the clinical insights from this study.

Reply: we have slightly revised the comment of the analyses in the paragraph as the results of the new analyses are in line with what was previously reported in the original manuscript (please see p. 13 l. 27).

As for the shame and stress issues, generally speaking, we suppose that the willingness to participate to an interview that may cause shame and stress is higher if the interview is phone-administered (or at least similar between the phone and face-to-face modes). In our opinion, this should also be valid for the specific case of the NVS interview. We have acknowledged this issue and the need to investigate this hypothesis in the revised manuscript (please see p. 13 ll. 30-37).

- P11 L31 – the convenience sampling and the way people were recruited needs to be listed as a major limitation for external comparisons – as noted above. The age, education, literacy levels etc could be compared with national norms to explore more accurately whether the data are at least comparable. This paragraph also brings new data in about the sampling strategy which should not happen in the Discussion.

Reply: we have acknowledged the convenience sampling and the limited external validity of the study in the revised manuscript as suggested (please see p.14 ll. 5-8). Data about the sampling strategy were moved in the methods section of the revised manuscript (please see p. 4 ll. 31-33).

- The authors should discuss how robust the cut offs of the categories of HL are using the NVS – have they yet been tested against any socially or clinically relevant/meaningful indicators?

Reply: in our study, the cut-offs of the categories of the NVS were not tested against any socially or clinically relevant/meaningful indicators; we have used the cut-offs that are generally used in the studies that applied the NVS.

Reviewer: 4

Reviewer Name: Delphine Courvoisier

Institution and Country: HUG - Switzerland

This article presents the associations between functional HL as measured by the NVS with antecedents and consequences, using a cross-sectional design.

- Abstract : design : the design is not randomized, which is usually used when an intervention is randomized, except for the small comparison of face-to-face vs. Telephone interview.

Reply: this error was removed in the revised manuscript.

- P4, line 30 : asking GP to recruit randomly without providing method to do a random selection does not yield a random sample. The authors do acknowledge that it is a convenience sample but the risk of bias, especially on the prevalence of limited HL is high.

Reply: we regret not having better detailed the recruitment procedures in the original manuscript. Please note that the sample was obtained with a combination of convenience and probability sampling procedures. Indeed, GPs were recruited with convenience criteria, and each recruited GP subsequently selected 80 subjects from its register through a random number generator. We have better detailed these procedures in the revised manuscript (please see p. 4 ll. 29-35). Furthermore, as for the risk of bias on the prevalence of limited HL, please note that the method of recruitment have foreseen the assistance to people with reading difficulties. We have better specified this aspect in the revised manuscript (please see p. 5 ll. 5-10).

- Statistical analysis : for inter-method reliability (NVS by phone or by face-to-face), the correct method of analysis is ICC(2,1). A non-significant paired t-test only says that the sample is small.

Reply: the ICC (2,1) analysis has been added in the revised manuscript.

- Table 1 : please justify why long-term illness does not have percentages.

Reply: percentages for long-term illness were added in the revised manuscript

- Table 1 : why indicate tests in the note below the table, since you do not report any p-value. Tests are usually indicated in the methods, or could be indicated in p8,line29 when you report the univariable associations.

Reply: tests were specified in the method section of the revised manuscript, and the note below the table was removed.

- Model selection : the selection of variables to include in the multivariable model by taking only the significant univariable associations is not recommended. It may lead to excluding variables that would have been relevant and were non significant due to confouding. A better selection method is the LASSO, or since your sample size is relatively large, you could include all predictors in a first multivariable model.

Reply: we agree with the reviewer. The model selection method was modified in the revised manuscript, and all the predictors were included in a first multivariable model. Nonetheless, the final results of the models did not change.

- Table 2 : p-values lower than 0.0000 are usually indicated as <0.001

Reply: p-values lower than 0.000 in table 2 and 3 were modified in the revised manuscript as suggested.

- Table 3 : the pseudo R2 is greater than 1. Please check your metric.

Reply: we regret the error; the correct value has been reported in the revised manuscript.

- P10,line14. Education has changed a lot over time, and age could be an independent predictor of HL because it is a proxy for receiving an education that never talked about health (for instance no sex education), but not because of cognitive decline

Reply: the topic of education and age-related differences in HL was included in the discussion section of the revised manuscript (please see p 12 ll. 35-36). As far as the role of cognitive ageing in age-related differences in HL is concerned, we think that this topic should be considered a relevant issue and discussed. We have better detailed the discussion of this topic in the revised manuscript as suggested also by reviewers n. 1 and n. 3 (please see p. 12 ll. 32-35).

VERSION 2 – REVIEW

REVIEWER	Delphine Courvoisier HUG - Switzerland
REVIEW RETURNED	12-Mar-2019

GENERAL COMMENTS	The authors have done an excellent job adressing the concerns of all the reviewers.