

## 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

<b>TITLE (PROVISIONAL)</b>	The Association Between Primary Care Physician Diagnostic Knowledge and Death, Hospitalization and Emergency Department Visits Following an Outpatient Visit at Risk for Diagnostic Error: A Retrospective Cohort Study Using Medicare Claims
<b>AUTHORS</b>	Gray, Bradley; Vandergrift, Jonathan; McCoy, Rozalina; Lipner, Rebecca; Landon, Bruce

### VERSION 1 – REVIEW

<b>REVIEWER</b>	Rita Fernholm Karolinska Institutet, Sweden
<b>REVIEW RETURNED</b>	26-Aug-2020

<b>GENERAL COMMENTS</b>	Thank you for a very interesting manuscript. Important subject. Have you thought about that if the certification is not mandatory, you will probably omit the doctors that might have the lowest competence? Did you test to include that group to see what results they would have?
-------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

<b>REVIEWER</b>	Dr Clare Goyder Nuffield Department of Primary Care Health Sciences University of Oxford
<b>REVIEW RETURNED</b>	12-Oct-2020

<b>GENERAL COMMENTS</b>	<p>I found this paper very interesting and agree diagnostic error is a neglected area for research and found it fascinating that the authors have attempted to link diagnostic errors specifically with deficiencies in diagnostic knowledge. The authors have access to a large data-set which is a strength of the study. I would like to see more evidence to support their novel diagnostic knowledge measure as this is central to the whole study.</p> <p>I am concern that the abstract is not as clearly written as it could be. Care with "whose cause was at risk for" as this phrase was not clear to me. Please consider re-writing abstract from line 25 to 30 and 50 to 51 so that this reads more clearly, so many words have been removed that the clarity of meaning is lost. Line 45 - "compliant" I think needs to be edited to "complaint" I think - please review this -also the word "compliant" repeated later in paper and I am concerned that this is an error The strengths and limitations section - lines 54-61 is too brief Need to think about link with clinical reasoning - is it actually that physicians who score more highly on knowledge are actually safer in their approach to clinical reasoning and referring/investigating when can't rule out more serious pathology?</p>
-------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

	<p>Patient's presentations can change and conditions can evolve - can we really be certain that diagnostic 'errors' identified are indeed 'errors' and not just conditions that worsened over time? Hospitalizations and ED visits after a consultation could be linked to appropriate safety netting at the time of the consultation and I wonder how appropriate it is to use these as outcomes? The mortality data is however more convincing although I would like a full statistical review prior to acceptance.</p> <p>Can the authors improve the implications section - vitally important that these findings are acted on and systems changed to prevent these errors continuing. Could these specific conditions be fed into future exams perhaps through case vignettes?</p> <p>I could not see any ethical approval referred to - was this not required given the nature of this study?</p> <p>The style of writing could be improved to removing reference to authors names themselves eg. line 72 - "Singh et al" could be removed and state "it is estimated that at least..." and then reference Singh at the end of the sentence.</p> <p>Overall an important study so I hope with some revision and full statistical review could be published.</p>
--	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

<b>REVIEWER</b>	Saul Weingart Tufts Medical Center and Tufts University School of Medicine Boston, USA
<b>REVIEW RETURNED</b>	18-Oct-2020

<b>GENERAL COMMENTS</b>	<p>The authors describe a rigorous and meticulous analysis of the relationship between primary care internists' performance on diagnosis-related exam questions on the ABIM internal medicine recertification exam and three outcomes (death, ED visit, hospital admission) within 90 days of an ambulatory visit for a symptom associated with a "diagnosis-sensitive" condition such as stroke or spinal cord compression. Diagnostic error in ambulatory settings is a topic of considerable research and policy interest.</p> <p>The strengths of the paper include its large sample size, detailed methodology, and performance of multiple sensitivity analyses -- together yielding a plausible statistically significant association between poor performance on ABIM certifying exam questions and higher mortality and healthcare utilization for diagnosis-sensitive conditions.</p> <p>The authors flag key weaknesses of the paper, including lack of medical record validation of the presence of diagnostic error. It would be particularly useful in future studies to perform this analysis, even on a small sample set. The authors also note the possibility of omitted variable bias and unmeasured confounders. I am concerned about the possibility that lower-performing MDs on the ABIM exam (who we know are more likely to work in small or solo practice based on Table 2) may have limited cross-coverage or less access to specialists and other professional colleagues for informal consultation or referral. Together, this might lead to poorer access to specialty care and a lower threshold to refer patients to the ED for care. I wondered if the authors have information about whether there is also an urban/suburban/rural location difference between high- and low-performing PCPs, further accentuating these barriers?</p> <p>It would also be helpful to know if the authors disaggregated the "diagnostic knowledge" (a term that I was not familiar with)</p>
-------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

	<p>questions into components such as fund of knowledge, diagnostic reasoning, and test interpretation, since one or another component may be most relevant to the analysis. I have real concern about the assertion that poor performance on diagnosis-related ABIM questions translates into poor clinical judgment, given the artificiality of pen-and-paper tests, lack of access to medical reference materials, and some individuals' ability to excel in that format.</p> <p>Overall, the authors present a rigorous and thought-provoking analysis that is hypothesis-generating about the the relationship between standardized test performance and clinical outcomes. There are many missing relationships that need to be assessed and validated. My primary recommendation to the authors to is temper their assertion in the discussion that "diagnostic knowledge... is a risk factor for outcomes" (p. 18) and conclusion (p. 19) to reflect the tentative nature of the these provocative but somewhat speculative relationships.</p>
--	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

<b>REVIEWER</b>	Deborah Constant University of Cape Town
<b>REVIEW RETURNED</b>	24-Dec-2020

<b>GENERAL COMMENTS</b>	I have notified the Editors that I do not have sufficient clinical background to assess this manuscript fairly. I am satisfied that the statistical analysis is appropriate and the conclusions aligned with the results.
-------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

### VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

Dr. Rita Fernholm, Karolinska Institute

Comments to the Author:

Reviewer 1 Comment 1: Thank you for a very interesting manuscript. Important subject. Have you thought about that if the certification is not mandatory, you will probably omit the doctors that might have the lowest competence?

Response: We agree that this is an important limitation of our study. Another is that it may not apply to older or younger populations of physicians. Considering this, we added the following to the limitations section:

“In some cases, physicians elected not to pursue certification and our results do not generalize to those who did not pursue MOC.” We now mention this limitation in the second limitations paragraph in the Discussion section.

“Our findings might also not be applicable to older physicians who certified before 2000 or younger physicians who certified after 2000 as well as physicians who choose not to attempt an exam. While a physician’s clinical knowledge might be related to their decision to not take the MOC exam therefore not maintaining their certification, other factors certainly play a role in this decision.”

Reviewer 1 Comment 2: Did you test to include that group to see what results they would have?

Response: The reviewer makes an interesting suggestion. We did not run such a test because we don’t have diagnostic exam performance for these physicians and even if we included them as a separate category we would not know which visit time period to apply. In addition we do not have a practice characteristic measure for this group since the practice data is entered as part of the

maintenance of certification process. We felt that omitting practice characteristic or treatment score as regression control variables could result in a biased estimation of the association between knowledge and our outcome measures.

Reviewer: 2

Dr. Clare Goyder, University of Oxford

Comments to the Author:

I found this paper very interesting and agree diagnostic error is a neglected area for research and found it fascinating that the authors have attempted to link diagnostic errors specifically with deficiencies in diagnostic knowledge. The authors have access to a large data-set which is a strength of the study.

Reviewer 2 Comment 1: I would like to see more evidence to support their novel diagnostic knowledge measure as this is central to the whole study.

Response: We agree that the exam-based assessment of diagnostic knowledge that we used is a key aspect of this study. The ABIM's approach to classifying and creating questions for the MOC exam is well established and documented. To create the composite measure used in this study, we relied upon a pre-existing classification system used by the ABIM to identify "diagnosis-related" questions. These comprise about half of the questions on the MOC exam. To create our composite score, we simply totaled the percentage correct for each physician.

To address the reviewers concerns, we include psychometric evidence of the internal validity of this measure in the Appendix Section 3 that shows that the measure holds together well as a measure of a single underlying construct. In addition to a more complete explanation of the measure, we also included more statistics in the Measure of Diagnostic Knowledge subsection of the Methods section including Alpha Cronbach scores and measures of correlation between diagnostic scores and both treatment exam scores and all non-diagnostic knowledge exam scores.

"Psychometric analysis indicates that scores on exam questions related to diagnosis were meaningfully correlated (i.e., Cronbach's alpha score of 0.84), and thereby represent an independent underlying construct that could be interpreted as diagnostic knowledge (see Appendix Section 3 for more details). Similarly, this analysis indicated that questions coded as treatment-related also represent an independent underlying construct (i.e., Cronbach's alpha score of 0.75). Although performance on diagnosis and treatment related questions were correlated (Pearson Correlation=0.62), 59.5% of the variation in diagnosis exam performance for the physician study sample was not explained by performance on other parts of the exam."

It should also be pointed out that the study itself provides validity for this measure as we found that scoring well on the diagnoses questions was predictive of better outcomes for diagnostic error sensitive conditions, but that same relationship was not present for the similarly constructed treatment composite. We provide a full discussion of the measure in the section describing the "measure of diagnostic knowledge" with further information in Appendix 3.

Also, we now include this in the limitations section of the Discussion section.

"Another limitation of our study is that the IM-MOC exam was specifically designed to measure clinical knowledge in general, it was not designed to measure diagnostic knowledge specifically. That said, diagnostic knowledge is a major component of the exam and was found to meet the criteria for measuring this underlying construct".

Reviewer 2 Comment 2: I am concern that the abstract is not as clearly written as it could be. Care with "whose cause was at risk for" as this phrase was not clear to me. Please consider re-writing abstract from line 25 to 30 and 50 to 51 so that this reads more clearly, so many words have been removed that the clarity of meaning is lost.

Response: As requested, we attempted to clarify the writing in the abstract and made substantial edits throughout. As you can appreciate, this is a complicated concept to get across in just a few words and we are willing to work with the editors if you have additional suggestions.

Reviewer 2 Comment 3: Line 45 - "compliant" I think needs to be edited to "complaint" I think - please review this -also the word "compliant" repeated later in paper and I am concerned that this is an error  
Response: We thank the reviewer for catching this error. We made these suggested changes.

The strengths and limitations section - lines 54-61 is too brief

Need to think about link with clinical reasoning - is it actually that physicians who score more highly on knowledge are actually safer in their approach to clinical reasoning and referring/investigating when can't rule out more serious pathology? Patient's presentations can change and conditions can evolve - can we really be certain that diagnostic 'errors' identified are indeed 'errors' and not just conditions that worsened over time?

Response: Because our study is cross sectional, it is subject to selection bias. For instance, if the patients presenting to lower knowledge physicians were sicker on average that could explain our results. The same might be true if the mix of visits was different. We therefore used regression analyses that controlled for an extensive set of both patient and physician level variables that might be associated with the outcomes of interest and most notably performance on parts of the exam unrelated to diagnostic knowledge. With regards to the specific point raised above—that conditions might evolve over time—our logic model suggests that in cases where the physician made the wrong initial diagnosis or initiated the wrong work up or referral, those decisions would contribute to clinical deterioration and, therefore, the outcomes of interest. Thus, we actually believe that worsening of conditions is along the pathway linking our independent predictor and the outcomes of interest. As noted above, the cross-sectional design of our study precludes us from drawing causal conclusion, which is why we note our findings are associations. We have also added additional limitations up front and have greatly expanded the limitations section of the discussion.

Additional strengths and limitations:

Strengths and limitations of this study

- o Unique diagnostic knowledge measure linking diagnostic knowledge with adverse outcomes
- o Scalable adverse outcome measures and extensive sensitivity analyses
- o Our assessment of diagnostic error is indirect (as indicated by adverse outcomes)
- o Results are subject to selection bias if the mix of index visits or the severity of the patients or practice support differed for physicians with different levels of diagnostic knowledge.
- o Results are only generalizable to physicians who elected to attempt ABIM's certification exam and were about 10 years past initial certification and patients older than 65.

Additional response: The reviewer makes an interesting point regarding the validity of our outcome measures. Given the lack of detailed data, we certainly cannot say that each outcome examined (death, hospitalization, or ED visit) was preceded by a diagnostic error on the part of the physician. What we are arguing is that an association with diagnostic knowledge and outcomes that are at greater risk for diagnostic error simply implies that better diagnostic knowledge may result in better outcomes on average. That is to say, some subset of the patients may have had a better outcome if they had been treated by a physician with better diagnostic knowledge but certainly not all of them. Of course, as we now point out more clearly, this is an exploratory study so the underlying causal links we have just implied need further study.

See second to last paragraph in the manuscripts:

“In this exploratory analysis, we found evidence that diagnostic knowledge of primary care physicians seeing a patient for an index visit for a complaint that is at heightened risk of diagnostic error is associated with adverse outcomes.”

Reviewer 2 Comment 4: Hospitalizations and ED visits after a consultation could be linked to appropriate safety netting at the time of the consultation and I wonder how appropriate it is to use these as outcomes?

The mortality data is however more convincing although I would like a full statistical review prior to acceptance.

Response: The reviewer makes a good point that hospitalizations and ED visits could indicate a correct action after a diagnosis of conditions such as stroke. In the base case analysis we considered this by not counting as a bad outcome hospitalization and ED visits on the same day as the index visit.

To continue to explore this possibility we extended this to next day outcomes. Reclassifying outcomes had no material impact on associations with diagnostic errors. This sensitivity analysis is now described in the Sensitivity subsection of the Methodology section and in the Sensitivity subsection of the Results section.

We also agree that the mortality measure has the advantage the reviewer points out. Mortality however, is a relatively rare outcome. Moreover, we cannot say with any certainty whether the death was related to the "index" condition for which the patient initially presented as cause of death is not identified in Medicare data. An advantage unique to the hospitalizations and ED visit outcome measures is that we can limit them to hospitalizations that could plausibly have been related to the initial index visit at risk for diagnostic error (e.g., we would include a hospitalization for pneumonia for someone who initially presented with shortness of breath, but not one for hip fracture). This also allows us to match antecedent index visit diagnoses to related conditions diagnosed at a later hospital stay or ED visit.

Reviewer 2 Comment 5: Can the authors improve the implications section - vitally important that these findings are acted on and systems changed to prevent these errors continuing. Could these specific conditions be fed into future exams perhaps through case vignettes?

Response: We discuss implications of our findings in the second to last paragraph in the discussion. "Our results are important for two additional reasons. First, these results provide evidence that board certification and maintenance of certification, which involves lifelong learning directed at maintaining medical knowledge, might, in fact, be a valid approach to assuring the delivery of high quality care. Many in the US complain about the time and expense of MOC and often point to the lack of rigorous assessment between aspects of MOC and outcomes of interest to patients. These findings suggest that processes such as MOC may translate into meaningful improvements in outcomes because they can provide incentives for meaningful learning. This learning also could be enhanced through exam feedback targeted at diagnostic knowledge. Second, the findings also suggest that interventions aimed at improving diagnostic skills, whether knowledge-based or through, for instance, delivery of relevant information at the point of care [this is in response to system changes] might be approaches that might be worthwhile if the findings of this study are validated with additional research."

In terms of system changes we included:

"During visits identified as being at risk for diagnostic errors, physicians could be given related information at the point of care including suggestions for specialty consultation."

Reviewer 2 Comment 6: I could not see any ethical approval referred to - was this not required given the nature of this study?

Response: This was included in the last paragraph of the Methodology section.

"The Advarra Institutional Review Board approved our study protocol"

Reviewer 2 Comment 7: The style of writing could be improved to removing reference to authors names themselves eg. line 72 - "Singh et al" could be removed and state "it is estimated that at least..." and then reference Singh at the end of the sentence.

Response: We made the suggested changes

Reviewer: 3

Dr. Saul N Weingart, Tufts Medical Center

Comments to the Author:

The authors describe a rigorous and meticulous analysis of the relationship between primary care internists' performance on diagnosis-related exam questions on the ABIM internal medicine recertification exam and three outcomes (death, ED visit, hospital admission) within 90 days of an ambulatory visit for a symptom associated with a "diagnosis-sensitive" condition such as stroke or spinal cord compression. Diagnostic error in ambulatory settings is a topic of considerable research and policy interest.

The strengths of the paper include its large sample size, detailed methodology, and performance of multiple sensitivity analyses -- together yielding a plausible statistically significant association between poor performance on ABIM certifying exam questions and higher mortality and healthcare utilization for diagnosis-sensitive conditions.

Reviewer 3 Comment 1: The authors flag key weaknesses of the paper, including lack of medical record validation of the presence of diagnostic error. It would be particularly useful in future studies to perform this analysis, even on a small sample set. The authors also note the possibility of omitted variable bias and unmeasured confounders. I am concerned about the possibility that lower-performing MDs on the ABIM exam (who we know are more likely to work in small or solo practice based on Table 2) may have limited cross-coverage or less access to specialists and other professional colleagues for informal consultation or referral. Together, this might lead to poorer access to specialty care and a lower threshold to refer patients to the ED for care. I wondered if the authors have information about whether there is also an urban/suburban/rural location difference between high- and low-performing PCPs, further accentuating these barriers?

Response: The reviewer makes some excellent points. In our original submission, we had addressed the latter concern to some extent by including a sensitivity where we excluded physicians who worked in academic medical centers. However, considering the reviewer's comment, we replaced this one sensitivity analysis with two additional sensitivity analyses. In the first, we limited the analysis sample to physicians in small practices (10 physicians or less, 54.5% of physicians and 62% of the visit sample) based on information on practice location maintained by the ABIM. In the second, we limited the analysis sample to physicians in large practices ( $\geq 50$  physicians)/academic medical center practices, 23.7% of physicians and 16% of the visit sample) where physicians likely had better access to specialists. Associations with diagnostic knowledge estimated from these regressions were similar to the base case, suggesting that differences in practice setting was not driving our results.

We excluded the death outcome measure from these sensitivity analyses because the large practice subsample only included 39 death events. Because of this, our model did not converge with the full set of controls. This was less of a problem for the small practice subsample. However, although the results for small practices were similar to the base case and statistically significant, for balance we also do not include this outcome for that sensitivity analysis.

We describe this approach in the sensitivity subsection of the Methodology section

"Fourth, we considered the possibility that our results were biased due to omitted variables correlated with practice size. To examine this possibility, we estimated associations with knowledge and our two utilization measures across a sample of physicians in either small ( $\leq 10$  physicians, 54.5% (768/1,410) of physicians) or large practices ( $> 50$  or in academic medical centers, 23.7% (334/1,410) of physicians). We did not conduct these sensitivities for death because there were too few deaths in the subgroups to allow us to reliably estimate the associations (e.g., 39 deaths for physicians in large practices)."

And as a limitation in the Discussion section:

"That said, the fact that practice size was found to be correlated with diagnostic exam performance is concerning. For example, practice size could be correlated with access to specialists or informal consultations with colleagues that intern might be related to our outcome measures. However, sensitive analyses indicate that associations with knowledge and our utilization adverse outcome measures were fairly similar across physicians practice size/type (small, and large or academic)."

Reviewer 3 Comment 2: It would also be helpful to know if the authors disaggregated the "diagnostic knowledge" (a term that I was not familiar with) questions into components such as fund of knowledge, diagnostic reasoning, and test interpretation, since one or another component may be most relevant to the analysis. I have real concern about the assertion that poor performance on diagnosis-related ABIM questions translates into poor clinical judgment, given the artificiality of pen-and-paper tests, lack of access to medical reference materials, and some individuals' ability to excel in that format.

Response: This is a good suggestion. Unfortunately exam questions were not coded by the ABIM exam committee MDs and experts in this area with the level of specificity suggested by the reviewer so we were unable to disaggregate the diagnostic knowledge measure further. We also agree with the concern that poor performance on the exam might not translate into poor decision making. As noted above, we were not able to perform a randomized controlled experiment, so we took great pains to account for selection bias and other analytic issues that could have produced our findings.

Nonetheless, we acknowledge this as an important limitation of our work that is clearly stated in the limitation paragraph of the discussion section.

To address the reviewer's latter point we added the following:

"Finally, some might assert that a standardized exam without access to medical reference material might be more a reflection of a physician's rote memory and ability to recall medical facts than a test of their clinical knowledge and judgement. Although this is a fundamental limitation of our study, it should be noted that the exam is designed to mimic decision making in real life situations including having such things as lab values and reference material embedded in questions and past research indicates that an "open" book format that allows physicians access to reference material did not materially impact exam performance. It should also be noted that the necessary rapidity of decision making by primary care physicians who have limited time per encounter might fairly be represented by an exam with time constraints."

Reviewer 3 Comment 3: Overall, the authors present a rigorous and thought-provoking analysis that is hypothesis-generating about the the relationship between standardized test performance and clinical outcomes. There are many missing relationships that need to be assessed and validated. My primary recommendation to the authors to is temper their assertion in the discussion that "diagnostic knowledge... is a risk factor for outcomes" (p. 18) and conclusion (p. 19) to reflect the tentative nature of the these provocative but somewhat speculative relationships.

Response: The reviewer makes some excellent points and in response we made several changes to the Discussion section to emphasize that our findings cannot be interpreted as causal given the cross sectional nature of our study:

At the beginning of the 3rd to last paragraph in the paper:

" In this exploratory analysis, we found evidence that diagnostic knowledge of primary care physicians seeing a patient for an index visit for a complaint that is at heightened risk of diagnostic error is associated with adverse outcomes."

And in the last sentence of the second to last paragraph of the paper:

"Yet more research is needed to better and more rigorously explore the link between diagnostic knowledge and diagnostic errors that are identified through chart review or other methods of direct ascertainment and the extent to which such errors result in adverse clinical outcomes."

Reviewer: 4

Ms. Deborah Constant, University of Cape Town Faculty of Health Sciences

Comments to the Author:



I have notified the Editors that I do not have sufficient clinical background to assess this manuscript fairly. I am satisfied that the statistical analysis is appropriate and the conclusions aligned with the results.

Response: Thanks very much.

#### VERSION 2 – REVIEW

<b>REVIEWER</b>	Clare Goyder University of Oxford UK
<b>REVIEW RETURNED</b>	25-Feb-2021

<b>GENERAL COMMENTS</b>	Thank you very much for your extensive revisions and responses. I congratulate the authors on a thorough response. I had a couple of minor issues in the discussion line 356: I didn't understand why 'contacts' in this sentence - would it be better to say "by requiring that the patient had not had an ED visit over the previous 3 months." line 378: is 'intern' meant to be 'in turn'? line 413: the reason for 'have' here is not clear to me and 'lab values' is understood less internationally- suggest better to say "and include patient's laboratory results and reference material embedded" Otherwise well done, very interesting work
-------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

#### VERSION 2 – AUTHOR RESPONSE

**Reviewer: 2**

Dr. Clare Goyder, University of Oxford

**Comments to the Author:**

Thank you very much for your extensive revisions and responses. I congratulate the authors on a thorough response.

**Response:** We thank you for your excellent suggestions

**Comment:** I had a couple of minor issues in the discussion line 356: I didn't understand why 'contacts' in this sentence - would it be better to say "by requiring that the patient had not had an ED visit over the previous 3 months."

**Response:** We agree with your suggestion and re-wrote the sentence as follows:

...we assured that we were studying new problems by requiring that the patient had not had an ED, hospital or outpatient visit over the previous 3 months.

**Comment:** Line 378: is 'intern' meant to be 'in turn'?

**Response:** We made the suggested change

**Comment:** Line 413: the reason for 'have' here is not clear to me and 'lab values' is understood less internationally- suggest better to say "and include patient's laboratory results and reference material embedded"

Otherwise well done, very interesting work

**Response:** We agree with your suggestion and rewrote the sentence as follows:

“including such things as patient's laboratory results and reference material impeded in the exam”