PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	The validity of four clinical prediction scores for pulmonary embolism in a sub-Saharan African setting: a protocol for a multicentre Cameroonian cross-sectional study.
AUTHORS	Esiéné, Agnès; Owono Etoundi, Paul; Tochie, Joel Noutakdie; Mbengono Metogo, Junette Arlette; Ze Minkande, Jacqueline

VERSION 1 – REVIEW

REVIEWER	Özlem Köksal Bursa Uludag University Faculty of Medicine Emergency Department
REVIEW RETURNED	22-May-2019

GENERAL COMMENTS	Dear author(s), Firstly thank you for share this paper with us. I have some suggestions; 1) You mentioned the study carried out between the period of July 2019 and December 2020 in the study design section. But it's impossible I think. You should be explained this situation. 2) The discussion section is very short. You should be extended this section support with using more literature. Best regards,

REVIEWER	Sander van Doorn Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University
REVIEW RETURNED	06-Jun-2019

GENERAL COMMENTS	Thank you very much for allowing me to review the manuscript "Validaty [sic] of four clinical prediction scores for pulmonary embolism in a sub-Saharan African setting: a protocol for a multicentre Cameroonian cross-sectional study" in which Esiéné and colleagues describe the protocol for a cross-sectional diagnostic study in the emergency care setting in Cameroon. They plan to validate four commonly used clinical probability scores (CPS) including the Wells rule (original and simplified) and Revised Geneva rule (original and simplified). This seems a highly relevant study, in a healthcare setting where this has clearly important implications. Overall, the rationale and methodology are well described. Please find my specific comments below. In general, the language is clear though the manuscript does contain many typo's. I believe the overall quality will be increased if the English language is critically revised. More specific, a 'gold standard' (or, as the manuscript erroneously states, 'goal standard') often does not exist, also in pulmonary embolism diagnosis. Generally, the term 'reference test' is preferred.

	Major comments Foremost, the manuscript does not describe well enough if and how D-dimer measures are incorporated in the diagnostic pathway. Nowadays most CPS are coupled with D-dimer measures and, as shown by the multiple studies cited in this manuscript, patients with both a low CPS score and negative D-dimer can safely be excluded from further testing. Using D-dimer therefore is of critical importance in a study such as the one described in the protocol, but how will this be dealt with? The authors plan to validate four CPS including the Wells rule (original and simplified) and Revised Geneva rule (original and simplified). Recently, the YEARS algorithm has been shown as a promising and very 'efficient' CPS in the emergency setting. See: van der Hulle et al., https://doi.org/10.1016/S0140-6736(17)30885-1. Have the authors considered including the novel score as well? To "ascertain the most accurate CPS amongst the four assessed" the AUC will be used, though it has been shown that the AUC often lacks power to do so. Other measures, including net benefit or decision curves should be considered. For a clearer picture of the study procedures, a flow chart may be considered. In thrombosis research, a follow-up of 3 months in often used as an additional 'reference test' in patients who cannot undergo CPA. Has this been considered for this protocol? Minor comments Contrary to most common research in adults, patients aged 15-18 years are eligible for inclusion. What was the rationale for this, and what are possible ethical consequences of including minors? This may be elaborated in the manuscript. The authors do not specifically state whether pregnancy is considered an exclusion criteria while I suspect this is the case. This may be added. As a sampling method, the authors state 'consecutive convenience' sampling will be used. Why is 'convenience' added here? Aiming for strict consecutive sampling will of course reduce possible selection bias and increase validity. Have any considerations or expectations on sample size been made? Please discuss these.

REVIEWER	James Bentham University of Kent, UK
REVIEW RETURNED	31-Jul-2019

GENERAL COMMENTS	The authors describe a very valuable study. I have carried out a statistical review, and have a number of suggestions. The authors should describe checks of heterogeneity in results across the seven hospitals, across age groups and between men and women. Also, is there any published evidence from previous studies that any score performs better or worse in older patients, or in men or women, for example? If so, this should be described.

	The authors should list the summary statistics that they will report. There should at least be breakdowns by age and sex of the numbers of patients in each of the seven hospitals. If there are other variables that have been shown previously to affect the performance of the scores they should also be reported. In "Methods and analysis" in the abstract, it is stated that the ROC curve will be used to select the best test. I think this should be reported, but that it should not be the only consideration of what is the best test, unless this is specifically justified. Are the dichotomisations of the four scores standard practice? If not, the authors should plan to report statistics such as sensitivity and specificity for other threshold values. The word "accuracy" is used in two senses - the statistical sense "the accuracy of each clinical score will be calculated..." and a more general sense "insights on the test with the best pretest accuracy". I think the latter should be changed to "performance" throughout, and accuracy should only be used in the statistical sense. The paper should be thoroughly proof-read and revised accordingly.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

1) You mentioned the study carried out between the period of July 2019 and December 2020 in the study design section. But it's impossible I think. You should be explained this situation.

Authors' response: PE is relatively scare in Africa with prevalence rate reported between 0.14% to –61.5%. Following our sample size calculation of a minimum of 364 participants, we think the study period mentioned in the study period will be enough to attend this sample size, because since July 2019, we have enrolled already 31 patients. If we enrolled averagely 20 patients per month till December 31, 2020 we will reach or exceed the calculated minimum sample size mentioned in Page 6, line 16-17. We wished to finalize the study before publishing it. Also, we wish to publish the protocol first for transperency purposes.

2) The discussion section is very short. You should be extended this section support with using more literature.

Authors' response: Page 9, line 5-19, the discussion has been more elaboratated

Reviewer: 2

1: In general, the language is clear though the manuscript does contain many typo's. I believe the overall quality will be increased if the English language is critically revised. More specific, a 'gold standard' (or, as the manuscript erroneously states, 'goal standard') often does not exist, also in pulmonary embolism diagnosis. Generally, the term 'reference test' is preferred.

Authors' response: we have improved the quality of english throughout the manuscript by proof-reading by an english native speaker

2: Foremost, the manuscript does not describe well enough if and how D-dimer measures are incorporated in the diagnostic pathway. Nowadays most CPS are coupled with D-dimer

measures and, as shown by the multiple studies cited in this manuscript, patients with both a low CPS score and negative D-dimer can safely be excluded from further testing. Using Ddimer therefore is of critical importance in a study such as the one described in the protocol, but how will this be dealt with?

Authors' response: Page 4, line 14: we statetd theat '' Current guidelines recommend their use coupled with D-dimer to preclude patients with a low PE probability from further diagnostic tests, without compromising the patient's safety''. However, due to limited laboratory resources in sub-Saharan Africa, especially in remote areas of Cameroon the measurment of D-dimers is unvailable. Hence, I this study we decided to focus on the performance of clinical prediction scores for pulmonary embolism which can be filled at the bedside of patients with suspicion of pulmonary embolism.

3: The authors plan to validate four CPS including the Wells rule (original and simplified) and Revised Geneva rule (original and simplified). Recently, the YEARS algorithm has been shown as a promising and very 'efficient' CPS in the emergency setting. See: van der Hulle et al., https://doi.org/10.1016/S0140-6736(17)30885-1. Have the authors considered including the novel score as well?

Authors' response: Page 6, line 23-24: the YEARS algorithm will not be studied because its entails the mearsurment of D-dimers which is not available in all sub-Saharan Africa laboratory settings. However in Page 4, line 12, the YEARS algorithm was stated.

4: To "ascertain the most accurate CPS amongst the four assessed" the AUC will be used, though it has been shown that the AUC often lacks power to do so. Other measures, including net benefit or decision curves should be considered.

Authors' response: Page 2, line 18 and Page 8, line 9: we stated that we will also measure net benefit or decision curves as well as measurements of calibrations such as calibration plots, Hosmer and Lemeshow statistics, observed/expected event rates

5: For a clearer picture of the study procedures, a flow chart may be considered.

Authors' response: Page 7, line 1 and 2: we stated a flow chart illustrating the study procedure has been added.

6 : In thrombosis research, a follow-up of 3 months in often used as an additional 'reference test' in patients who cannot undergo CPA. Has this been considered for this protocol?

Authors' response: We agree with you that a follow-up of 3 months in thrombosis research, how over our study is just a cross-sectional design to determine the diagnostic performance of four clinical prediction scores for pulmonary embolism. Hence, follow-up is not needed. The cross-sectional design limits us to interview or see patients only once. However, they will be follow-up clinically till 3 months but not for the purpose of this study.

7 : Contrary to most common research in adults, patients aged 15-18 years are eligible for inclusion. What was the rationale for this, and what are possible ethical consequences of including minors? This may be elaborated in the manuscript.

Authors' response: We did include patients aged above 15 years. Hence minors between 16 and 18 years will be included and we have no rational for not including minors

8 : The authors do not specifically state whether pregnancy is considered an exclusion criteria

while I suspect this is the case. This may be added.

Authors' response: Page 6, line 9 : Pregnant women were included.

9: As a sampling method, the authors state ‘consecutive convenience’ sampling will be used. Why is ‘convenience’ added here? Aiming for strict consecutive sampling will of course reduce possible selection bias and increase validity

Authors' response: Page 6, line 17: we deleted the word convenience

10: Have any considerations or expectations on sample size been made? Please discuss these

Authors' response: Page 6, line 16-17 : we have discussed the sample size calculation (n=364)

Reviewer: 3
11: The authors should describe checks of heterogeneity in results across the seven hospitals, across age groups and between men and women. Also, is there any published evidence from previous studies that any score performs better or worse in older patients, or in men or women, for example? If so, this should be described.

Authors' response: Page 6, line 1-6: The seven partaking hospitals in this study have little or no heterogeneity across age groups and between men and women. This has been discussed in this page and line

12: The authors should list the summary statistics that they will report. There should at least be breakdowns by age and sex of the numbers of patients in each of the seven hospitals. If there are other variables that have been shown previously to affect the performance of the scores they should also be reported.

Authors' response: Page 6, line 1-64 : This has been discussed in this page and line

13: In "Methods and analysis" in the abstract, it is stated that the ROC curve will be used to select the best test. I think this should be reported, but that it should not be the only consideration of what is the best test, unless this is specifically justified.

Authors' response: Page 2, line 17-18 and Page 7 to 9: This has been rectified accordingly

14: Are the dichotomisations of the four scores standard practice? If not, the authors should plan to report statistics such as sensitivity and specificity for other threshold values.

Authors' response: dichotomisations of the four scores is standard practice. However, we equally plan to report sensitivity and specificity for other threshold values as mention in Page 2, line 12 and Page 7, line 19 to 21

15: The word "accuracy" is used in two senses - the statistical sense "the accuracy of each clinical score will be calculated..." and a more general sense "insights on the test with the best pretest accuracy". I think the latter should be changed to "performance" throughout, and accuracy should only be used in the statistical sense.

Authors' response: This has been rectified accordingly

16: The paper should be thoroughly proof-read and revised accordingly.

Authors' response: This has been rectified accordingly

VERSION 2 – REVIEW

REVIEWER	S. van Doorn Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University
REVIEW RETURNED	12-Aug-2019

GENERAL COMMENTS	The authors addressed most of my comments. They may, however, consider explicitly mentioning the omission of d-dimer measurements in the Strengths/Limitations section or in the Methods section to avoid any confusion (as d-dimer measurements are so omnipresent in risk stratification in thrombosis research).

REVIEWER	James Bentham University of Kent, UK
REVIEW RETURNED	15-Aug-2019

GENERAL COMMENTS	I think the authors have improved the paper, and it is now ready to be published. The flowchart makes the study design clear, and the description of the statistical analyses is improved. I think the data collection form is very thorough, and I hope these statistics will be reported in detail in the paper describing the results. I don't think I need to see the paper again. My only comment is that there are still a number of spelling mistakes. Ones that I caught are p2, line 5 "algorithm... pulmonary", missing word here p2, line 19 should be "each of the" p3, line 19 should be "of confirmed PE cases" p3, line 22 should be singular "sequela" p5, line 1 should be "where they were designed" p5, line 20 should be "hospitals" p6, line 3 should be "hospitals" p6, line 21 should be "assessed for PE clinical probability" p8, line 14 should be "and SRG scores" p9, line 7: the - should be removed before "61.5%" p9, line 18 should be "appropriate" p10, line 7 should be "sub-Saharan" p10, line 11 should be "hospitals"

VERSION 2 – AUTHOR RESPONSE

Reviewer 2:

They may, however, consider explicitly mentioning the omission of d-dimer measurements in strengths/limitations sections or in the Methods section to avoid any confusion (as d-dimer measurements are so omnipresent in risk stratifications in thrombosis research)

Authors' response: Page 3, line 18-19: this has been revised accordingly.

Reviewer 3:

My only comment is that there are still a number of spelling mistakes.

Authors' response: spelling mistakes have been corrected throughout the manuscript as requested