

## 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>	Development of a risk prediction score for screening for HBV, HCV, and HIV among migrants in France: results from a multicentre observational study (STRADA study)
<b>AUTHORS</b>	Duracinsky, Martin; Yaya, Issifou; Yombo-Kokule, Lisa; Bessonneau, Pascal; Thonon, Frédérique; Rousset-Torrente, Olivia; Roudot-Thoraval, Françoise; Lert, France; Zucman, David; CHASSANY, Olivier

### VERSION 1 – REVIEW

<b>REVIEWER</b>	Feleke, Sefineh Woldia University, Public health
<b>REVIEW RETURNED</b>	31-Jul-2023

<b>GENERAL COMMENTS</b>	<p>Overall, it is well written To improve the manuscript I put the following general comments. In the method section</p> <ul style="list-style-type: none"><li><input type="checkbox"/> Do you think confounding is a deal during the prediction model? In the univariate analysis, independent variables with p-value less than 0.25 were included in multivariable logistic regression analysis in order to control potential confounders. Or how you handle confounding by performing univariate analysis?</li><li><input type="checkbox"/> Have you done any validation? Or was it internally valid? If so how you assessed the validation of your model?</li><li><input type="checkbox"/> How you develop the risk score? And why you developed it? You may develop like this “We transformed each coefficient of the model into a rounded number by dividing it by the lowest coefficient. The number of points was subsequently rounded to the nearest integer. We determined the total score for each individual by assigning points for each variable present and adding them up”</li><li><input type="checkbox"/> Have you assessed the ROC for the coefficients and the risk scores?</li><li><input type="checkbox"/> As we know HBV, HCV, and HIV are highly interrelated, especially in their mode transmission, hence have developed a model that predicts in combination of them.</li></ul>
-------------------------	--

<b>REVIEWER</b>	Botelho, Eliã Federal University of Para, Programa de Pós-Graduação em Enfermagem
<b>REVIEW RETURNED</b>	16-Aug-2023

<b>GENERAL COMMENTS</b>	Dear authors, I read you paper with high interest and really enjoy it. Congrats for your beautiful work!!! Just have some few considerations that are listed below:
-------------------------	--

	<p>Please, replace all "Test Rapide d'Orientation Diagnostique" for English;</p> <p>You should describe in more details the purpose of TROD because we understand that it is a simple sociodemographic and risk behavior questionnaire.</p> <p>How many OFII there are in France? If there are more than 21, why did you chose these 21? Mention it in text.</p> <p>Also, in design yet, you should mention the quantity of immigrant that France received per year and informing what continent are most of them from;</p> <p>"The Akaike information criterion (AIC) was to select the final model." Please, mention the considered limit (e.g.: &lt;5, &lt;10,...).</p> <p>I got confused in this point "using the predictive value of the questionnaire, its sensitivity, its specificity as well as the 95% confidence intervals". If the questionnaire was already validated, you should refer it.</p> <p>Just on page 10 I could understand the above question. Please, bring these descriptions to methods and let results to results only.</p> <p>"The <math>\beta</math>-coefficients of each variable were multiplied by a constant (we have chosen 5)... " What were the criterium employed? Did you took some previous reference? You should describe it in more details;</p> <p>Please, correct all numbers (e.g.: 21133 to 21,133);</p> <p>Discussion is OK as well as conclusion. My suggestion is that you should put more emphasis on the possibility of other countries applying the TROD.</p> <p>Congrats once again!!!</p>
--	---

### VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

Mr. Sefineh Feleke, Woldia University

Comments to the Author:

Overall, it is well written

To improve the manuscript I put the following general comments.

In the method section

Do you think confounding is a deal during the prediction model? In the univariate analysis, independent variables with p-value less than 0.25 were included in multivariable logistic regression analysis in order to control potential confounders. Or how you handle confounding by performing univariate analysis?

Authors' response : Thank you very much for this valuable comment. As mentioned in the manuscript, univariate analysis was used to select variables that could be associated with the outcome, with an alpha error of around 25%. In the next step, given the categorical nature of the outcome and the number of covariates, we managed the confounding factors in a multivariate logistic regression model, which is one of the best statistical analysis to

control confounding effects.

Have you done any validation? Or was it internally valid? If so how you assessed the validation of your model?

Authors' response : Thank you so much for your comments. In the current manuscript, no validation analysis has been carried out. For validation purposes, we plan to submit this developed risk score to an independent sample with the same characteristics.

How you develop the risk score? And why you developed it? You may develop like this "We transformed each coefficient of the model into a rounded number by dividing it by the lowest coefficient. The number of points was subsequently rounded to the nearest integer. We determined the total score for each individual by assigning points for each variable present and adding them up"

Authors' response : Thank you so much for your comments. We describe the development method in the manuscript as following : "Model performance was evaluated in terms of discrimination. The discriminating capacity of the TROD Screen questionnaire for HBV, HCV and HIV was evaluated by using the predictive value of the questionnaire, its sensitivity, its specificity as well as the 95% confidence intervals. A ROC curve was used quantify discrimination and determine the cutoff score of the questionnaire for each infection. That also assesses whether those with higher predicted risks are more likely to have a HBV, HCV or HIV infection. To make the models easier to use in clinical practice, we created a risk score for evaluating the likelihood of HBV, HCV or HIV infection, based on multivariable regression coefficients, which were rescaled and rounded to the whole number. To determine the score for each level of the variables, weighted points were assigned to each of the final associated factors. The  $\beta$ -coefficients of each variable were multiplied by a constant (we have chosen 5), and rounded to the nearest integer [18,19]. Based on the sensitivity–specificity curve analysis (Figure 1), the cut-offs point was chosen for the risk score for HBV (both in men, and women), HCV and for HIV infections, with maximum sensitivity and specificity."

We developed this risk score to identify participants most at risk of having HIV, HBV or HCV infection and offer them a rapid screening test.

Have you assessed the ROC for the coefficients and the risk scores?

Authors' response : Thank you so much for your comment. The ROC curve shows the true-positive rate versus the false-positive rate for rules that classify individuals using risk thresholds that vary over all possible values.

As we know HBV, HCV, and HIV are highly interrelated, especially in their mode transmission, hence have developed a model that predicts in combination of them.

Authors' response : Thank you so much for your comments. Although HIV, HBV and HCV infections are interrelated in their transmission modes, in our analysis, we found that the predictive factors were not identical for the three infections, so we developed a score for each infection.

Reviewer: 2

Dr. Eliã Botelho, Federal University of Para

Comments to the Author:

Dear authors,

I read you paper with high interest and really enjoy it. Congrats for your beautiful work!!! Just have some few considerations that are listed below:

Please, replace all "Test Rapide d'Orientation Diagnostique" for English;

Authors' response : Thank you so much for your suggestion. We have replaced all "Test Rapide d'Orientation Diagnostique" in the manuscript by "Rapid Screening Test".

However, we did not change the term "TROD Screen questionnaire" as it is the name given to that questionnaire.

You should describe in more details the purpose of TROD because we understand that it is a simple sociodemographic and risk behavior questionnaire.

Authors' response : The TROD screen questionnaire is an adapted questionnaire built for migrants seen un the OFII, which collected sociodemographic data and migratory trajectories. It is included additional items from the risk questionnaire for HIV, HBV and HCV infection previously published in a German study (cohabitation with an infected person, blood transfusion before 1992, tattoo or piercing, surgery, hemodialysis or transplantation in the past).

How many OFII there are in France? If there are more than 21, why did you chose these 21? Mention it in text.

Authors' response : Thank you so much for your suggestion. There are a total of 32 OFII centers, including 28 in mainland France and 4 overseas. During our study all OFII centers were invited to participate in the study, only 21 OFII centers have accepted, they had actually participated in the study.

Also, in design yet, you should mention the quantity of immigrant that France received per year and informing what continent are most of them from;

Authors' response :

"The Akaike information criterion (AIC) was to select the final model." Please, mention the considered limit (e.g.: <5, <10,...).

Authors' response : Thank you so much for your suggestion. For the selection of the final model, we chose the most parsimonious model with the lowest AIC.

I got confused in this point "using the predictive value of the questionnaire, its sensitivity, its specificity as well as the 95% confidence intervals". If the questionnaire was already validated, you should refer it.

Authors' response : Thank you so much for this comment. This questionnaire have not been validated yet. The validation of this tool is in the perspectives of our future research. Just on page 10 I could understand the above question. Please, bring these descriptions to methods and let results to results only.

Authors' response : We would like to thank you for this suggestion. The paragraph have been moved in the methods section.

"The  $\beta$ -coefficients of each variable were multiplied by a constant (we have chosen 5)... " What were the criterium employed? Did you took some previous reference? You should describe it in more details;

Authors' response : Thank you so much for this comment. This method used in our study to calculate the score was described in previous studies by Austin et al, Madan et al and Solomon et al, which I referenced in the manuscript.

Please, correct all numbers (e.g.: 21133 to 21,133);

Authors' response : Thank you for this suggestion, we have corrected all numbers in the manuscript.

Discussion is OK as well as conclusion. My suggestion is that you should put more emphasis on the possibility of other countries applying the TROD.

Authors' response : We appreciate your suggestion. We have included it in our upcoming research, but first we need to validate this risk score.

Congrats once again!!!

Authors' response : Thank you so much.

## VERSION 2 – REVIEW

<b>REVIEWER</b>	Botelho, Eliã Federal University of Para, Programa de Pós-Graduação em Enfermagem
<b>REVIEW RETURNED</b>	02-Nov-2023

<b>GENERAL COMMENTS</b>	The authors have attended all of my suggestions. Congrats again for this beautiful work!
-------------------------	--

**VERSION 2 – AUTHOR RESPONSE**