Qeios

Peer Review

Review of: "Correlation Between the SARS-CoV-2 Seroprevalence and the Theoretical Occupational Risk (TOR) and COVID-19 Morbidity Score (MBS) Among Local Public Workers in the Centre-Val de Loire Region – CovidOr Study"

Rodolfo Rossi[1]

1. Center for Operational Research, International Committee of the Red Cross, Geneva, Switzerland

Dear Authors,

Thank you for your manuscript and the opportunity for me to review it. In general terms, your manuscript is well designed, methodologically sound, and quite clear. You are submitting this manuscript as a “short communication,” while in my opinion, this is a full “research article.”

There are, however, some points for which I propose some amendments. These aim at making your manuscript clearer and potentially stronger.

Since there are no line numbers, I try to explain here below my proposals, hoping that these are clear for you:

Introduction

The objectives of your study are clear. However, it is not clear why you are conducting this study now, about 5 years after the data were collected. Please explain this timing: what relevance does this study have after 5 years since the beginning of the pandemic? You may have valid reasons for this; otherwise, your manuscript may look a bit outdated. Still in the introduction, you make reference to a clinical study (CovidOr), which, if I understand correctly, was the source of your data. Could you expand a bit on the summary of this study? What type of “clinical” study was it? A Cohort? A Case-control? How did it trigger your analysis?

doi.org/10.32388/BT13HW

Methods

The subchapter on Statistical Analysis contains more than the statistical analysis. I would add a subchapter called something like “exposure and outcome variables” or similar. This subchapter should be placed before the one on statistical analysis, and here you can explain the TOR, the MBS, and other variables you consider relevant. In the part on statistical analysis, you mention “A multivariate binary logistic regression model was created...”. Probably it is a typing mistake: a multivariate logistic regression cannot be binary, and vice versa; the binary logistic regression includes only 2 variables: the outcome and one exposure, and gives crude odds ratios (OR); a multivariable, then, is performed to calculate adjusted OR in which different exposure variables are assessed simultaneously. Generally speaking, the analysis starts with binary models, in which the association with the outcome variable is explored for each exposure, one by one; then, the exposures that are statistically associated with the outcome are included in a multivariable model (an additional note here: your title speaks about “correlation,” which is usually calculated in “linear regression”; in logistic regression, one assesses the “association” rather than correlation. In statistical terms, “association” is stronger than “correlation” – You may want to correct the title). With a sample of over 3600 participants, it is not clear why you needed to change the alpha, even if you include many variables in the same model. I would maintain the alpha at 0.05 (or 5%), which means that you will consider a variable as significant if there is a maximum of 5% probability that the association is given by chance. State also which statistical test you used to derive the p-values; usually, it can either be the Wald Test (personally, my preferred) or the Likelihood Ratio Test; mention also the statistical package you used (e.g., STATA, SPSS, R, etc.). These statements may not look important but add transparency to your manuscript: if other researchers wished to perform the same analysis on the same dataset, these pieces of information are essential.

In the sample size subchapter, you can include that the 3600 participants were selected from a workforce of about 17000 and which criteria were used to select them.

Results

At the beginning of this section, you start by describing the dataset, which is very good. I would move the sentence “There was no significant correlation between age and seropositivity (OR=1.008, CI= [0.988:1.029], p=0.415)” later in the text, after you describe the rest of the dataset (gender, seropositivity distribution, etc.). Again, a note on the word “correlation”: I would replace it with “association”. It is excellent that you included a figure with the age distribution of participants; however, Figure 2 is, in my opinion, unnecessary: sex distribution can easily be mentioned in the text and avoid a figure. Figure 3

gives the numbers of seropositive per centre, but without denominators; I would replace it with the prevalence of seropositivity per centre.

Just after Fig. 3, you write: “To adjust for possible confounding factors related to the comorbidities of the workers and to reduce selection bias, an MBS variable was introduced in the model.” While the inclusion of MBS in a model definitely helps to control for confounders, unfortunately, it does very little to reduce selection bias: selection bias is reduced at the design stage, by ensuring that the sample is representative of the population from which the sample was selected.

There is one sentence that is not clear: “The mean MBS was 0.73. Compared to seronegative participants, seropositive participants had around a 25% decrease in their MBS (OR=0.752, CI= [0.582:0.971], p=0.029). However, after adjusting the risk alpha for multiple tests (adjusted alpha=0.625%), this decrease in MBS turned out to be statistically insignificant.”

“The MBS mean”: if the MBS is a count (i.e., the number of risk factors a person has), then it may be clearer to present this with range, median, and quartiles, rather than the mean.

“Compared to seronegative participants, seropositive participants had around a 25% decrease in their MBS (OR=0.752, CI= [0.582:0.971], p=0.029)”. As explained here, it seems that being seropositive helps decrease the MBS, which is not possible. Since seropositivity is your outcome variable, the OR should be interpreted as that MBS was negatively associated with seropositivity (also to clarify if the logistic regression assessed for a linear trend).

“However, after adjusting the risk alpha for multiple tests (adjusted alpha=0.625%), this decrease in MBS turned out to be statistically insignificant”. Here comes my previous point on changing the alpha: is it really by changing the test that makes this association insignificant, or by adjusting by other variables in a multivariable model that makes this association insignificant?

The results section ends with this sentence: “There was no significant association between SARS-CoV-2 seroprevalence and the overall TOR score for an adjusted risk alpha of 0.625% (OR=1.7, CI= [1.146:2.549], p=0.009) (Table 4). Now, in my opinion, a p=0.009 is highly significant; if you use a linear trend approach, this means that for each increase in TOR score, the odds of being seropositive increase by 70%. Again, not clear the point of adjusted risk alpha.

The tables

Table 1 is definitely useful. Probably a bit too long, but useful. Is there a way that you cluster the professions? You may have the full list in an annex, rather than the main text.

Table 2 is also useful, but this summary of evidence on the risk factors should be part of the literature review and included in the introduction. Then, the table can also be put in an annex.

Table 3: it is not immediately clear that you only put in this table the n and % of seropositive participants. You make no reference in the text (results part) to symptoms. You can add a simple sentence to say that the most reported symptoms were fatigue and fever.

Table 4 is the one that needs more revision. In the legend below, a part titled “model results”, please explain the table, i.e., “association between seropositivity and various exposure factors”; add whether the OR is crude or adjusted (and if adjusted, by which variable); state the statistical test (Wald Test? LRT?). When a statistical package gives an output with p=0.000, it means p<0.001 (there is no zero probability). This should be corrected. As a matter of “style” (more than content), the table should start with the OR, then the 95% CI, and the p-values in the last column. Also, state if some variables were treated with a linear trend approach (mostly for the MBS and TOR). In the table, there is a p-value for Center; I guess this is an error, since Center is the overall variable for location, while Orléans, Région Centre, and CCTVL are the 3 variable levels, in which you treat CCTVL as the baseline.

Discussion

You start the discussion by stating that TOR was not associated with seropositivity, while the increase in TOR increased the odds of seropositivity by 70% (see my comment above).

You compare your results with only one other study on occupational risk for COVID-19. There are some available in the literature. Comparison with various studies gives more strength to your paper. Here is an example: https://bmcinfectdis.biomedcentral.com/articles/10.1186/s12879-023-08674-2

On the basis of the results, which recommendations do you propose?

Looking forward to reading your revision

Cordialement

Declarations

Potential competing interests: No potential competing interests to declare.