PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Covidogram as a simple tool for predicting severe course of COVID-19: population based study
AUTHORS	Jarkovský, Jiří; Benešová, Klára; Cerny, Vladimir; Razova, Jarmila; Kala, petr; Dolina, Jiri; Majek, Ondrej; Sebestova, Silvie; Bezdekova, Monika; Melicharova, Hana; Snajdrova, Lenka; Dusek, Ladislav; Parenica, Jiri

VERSION 1 – REVIEW

REVIEWER	Matthew Sperrin University of Manchester, UK
REVIEW RETURNED	26-Oct-2020

GENERAL COMMENTS	In this paper, a risk score is developed for prognosis of covid-19, to be applied (I assume) to patients who have received a positive test. The setting is the Czech Republic. The sample size is moderate, but likely to be sufficient to avoid any serious model overfitting. There are a number of analytical decisions that are either poorly reported, and/or place the resulting model at high risk of bias. Details are given below. 1. The aim of the study does not match either the analyses or conclusions. The aim is stated as 'to determine factors for the development of...' which suggests a prognostic factor study. However, the paper then goes on to develop a prognostic model, which is not the same thing. 2. There is incorrect conflation with a positive covid test, and patients with covid-19. We know that a substantial proportion of patients with covid-19 do not get tested. E.g. P11 L 34 'it is a population-based analysis of all COVID-19 cases in the Czech Republic' is not correct. For the same reason, the comparison in Table S2 is far more nuanced than the authors acknowledge. It seems likely that the population of patients diagnosed with covid-19 is older and with more comorbidities, not because these patients are more likely to get covid-19, but they are less likely to be asymptomatic or have minor symptoms and hence not report for a test. 3. Exclusion of 464 patients with 'still unknown history' and 565 patients with 'follow-up period shorter than 14 days', and 'analysis without censoring' - all of these issues suggest that selection was done post-baseline, which may substantially bias the results. Disregarding censoring to 'simplify visualisation and interpretation' does not seem a valid justification since a model incorporating censoring can still output probabilities (e.g. of severe symptoms

	within N days), and indeed hazard ratios (for example), which actually have a clearer interpretation than odds ratios. 4. P12 L46 and elsewhere: 'multivariate' should be 'multivariable'. 5. It appears from Table 2 that age was categorised (although this is not described in the methods): why was this decision taken, and why these cutpoints? Generally it is better not to categorise continuous variables. 6. What was the criteria used for the backward stepwise algorithm when selecting the optimum model? 7. How many predictors were considered for inclusion in the model? Based on Table 1, plus the categorisation of age, I am assuming about 25 effective number of predictors? 8. It isn't clear what the authors did when they say 'A 10-fold cross-validation was performed to reduce the likelihood of model overfitting'. Cross validation doesn't present overfitting - but it can, used correctly, give estimates of e.g. AUC that are adjusted for in-sample optimism. 9. What cut-offs to the reported sensitivity and specificity refer to? 10. No results are presented on calibration of the model. 11. Tables 2 and S1 need to include the intercepts of the fitted models: as without this absolute risks cannot be calculated. 12. There is no discussion on how missing data are handled. Given the predictors used this may not be an issue - but it still needs to be stated. 13. An alternative 'simpler' model is presented in Table S1. Its performance is identical to the original moel (except for a difference in the third decimal place on one of the confidence limits) which seems surprising, and should be double checked. Which risk score are the authors actually recommending? The one in Table 2 or in Table S1? 14. What are the confidence intervals for in Figure 1? This Figure reports risk probabilities for individuals, which as such already reflect uncertainty. 15. There is no mention of TRIPOD, which is the standard checklist for reporting on the development/validation of a risk prediction model.

REVIEWER	Benjamin Gallo Marin Warren Alpert Medical School of Brown University, United States
REVIEW RETURNED	24-Nov-2020

GENERAL COMMENTS	Dear authors, Thank you for a fantastic manuscript. Below are some recommendations to further strengthen the paper. "More frequently, the disease affects the respiratory system.....". This sentence may benefit from more citations to ensure readers are able to learn more about extra-respiratory manifestations of

	COVID-19 on their own. Consider citing reviews here, perhaps. If you are having trouble finding papers to cite here, feel free to cite one of these that I have had the pleasure to contribute to - of course, any comparable publication will serve its purpose here. Aghagoli, G., Gallo Marin, B., Katchur, N.J., Chaves-Sell, F., Asaad, W.F., Murphy, S.A. Neurological Involvement in COVID-19 and Potential Mechanisms: A Review. Neurocrit Care (2020). https://doi.org/10.1007/s12028-020-01049-4 Aghagoli G, Gallo Marin B, Soliman LB, Sellke FW. Cardiac involvement in COVID-19 patients: Risk factors, predictors, and complications: A review. Journal of Cardiac Surgery. April 19, 2020. doi: 10.1111/jocs.14538 "In a number of patients, there is a risk" - can you report in the manuscript what is the estimate proportion of these patients? The number may certainly be evolving or be different across different populations, so reporting this as a range would be appropriate if needed. I think it is important to do so to provide context to the reader. "However, the clinical course might progress over time" - I think this statement requires some modification. It is expected that the clinical course of most pathologies will progress in some way - Perhaps what you would like to convey is that the clinical picture of COVID-19 can quickly turn an unfavorable clinical course in a patient. "It was repeatedly demonstrated that older age..." - you may want to say in parenthesis that age > 55 is what we think of in this context. I would definitely include obesity (BMI > 30) , moderate-to-severe asthma, sickle cell disease, and immunosuppression. You may find these characteristics in the CDC website, which should be cited You mention PCR testing and RT-PCR testing in the methods. I think it would be best to state "RT-PCR" throughout the manuscript, for consistency. "Of the total 7,455 evaluated patients, 1,182..." - Could you articulate what exactly this primary endpoint means? Patients treated with Proton Pump Inhibitors - I appreciate that you included previous data suggesting that PPI use may be linked to an increased risk in pneumonias. This is interesting. I worry however that this conclusion might be a bit too stretched - for instance, does the observation still hold true for patients with acid-related disorders who are not taking proton pumps inhibitors? The observation is complicated by the fact that some patients may not be adherent to their PPI regimen, and that there is likely great variability in the amount of time that they've been on PPIs - Therefore I think that while these conclusions are interesting, they should be frame with caution. I would qualify these statements in your manuscripts by saying that theoretically an apparent link to severity may be explained by pharmacology OR by underlying pathology of acid-related disorders - whether it be controlled or not through medication. Nevertheless, the finding is interesting - just be a bit more conservative about how you report it. Do you have

	any hypotheses for patients who take medications that induce acid-related disorders (i.e. biphosphonates, some SSRIs, etc?) ? Excellent Limitations section and interesting Covidogram. I look forward to seeing this report in print.

VERSION 1 – AUTHOR RESPONSE

1.

The aim of the study does not match either the analyses or conclusions. The aim is stated as 'to determine factors for the development of...' which suggests a prognostic factor study. However, the paper then goes on to develop a prognostic model, which is not the same thing.
Thank you, we corrected this inaccuracy. The aim of the paper was to develop a prognostic model.

2.

There is incorrect conflation with a positive covid test, and patients with covid-19. We know that a substantial proportion of patients with covid-19 do not get tested. E.g. P11 L 34 'it is a population-based analysis of all COVID-19 cases in the Czech Republic' is not correct. For the same reason, the comparison in Table S2 is far more nuanced than the authors acknowledge. It seems likely that the population of patients diagnosed with covid-19 is older and with more comorbidities, not because these patients are more likely to get covid-19, but they are less likely to be asymptomatic or have minor symptoms and hence not report for a test.
Thank you for this important comment. We added this limitation into the Methods section – P 11 –
… it is a population-based analysis of all diagnosed COVID-19 cases in the Czech Republic. The cohort covers also asymptomatic and oligosymptomatic patients identified thanks to epidemiological monitoring.
We also discuss this issue in the limitations:
..The cohort does not include strictly all COVID-19 cases in the Czech Republic because some patients are asymptomatic and have not been tested. Older peoples with comorbidities are probably more likely to have a symptomatic course of COVID-19. It could be also a reason why the population of patients diagnosed with Covid-19 is older and with more comorbidities in comparison with the Czech Republic population (Table S2)

3.

Exclusion of 464 patients with 'still unknown history' and 565 patients with 'follow-up period shorter than 14 days', and 'analysis without censoring' - all of these issues suggest that selection was done post-baseline, which may substantially bias the results. Disregarding censoring to 'simplify visualisation and interpretation' does not seem a valid justification since a model incorporating censoring can still output probabilities (e.g. of severe symptoms within N days), and indeed hazard ratios (for example), which actually have a clearer interpretation than odds ratios.

There are two reasons for the selection of patients. First, there are “464 patients with still unknown history” caused by the process of dataset integration. The final dataset was integrated from two datasets connected through pseudonymous personal code, the primary dataset contained COVID19 cases and secondary dataset of health insurance companies provided medical history of patients; patients with no match consisted of two groups, i) foreigners who are not included in public health insurance in the Czech Republic and ii) patients with not match due to mistyping in personal identification code. We are unable to correct these problems as we do not have access to primary personal identification codes which were replaced by pseudonymous codes by standardized algorithm in all our databases. Second, there are “565 patients with follow-up period shorter than 14 days”;

these patients were included in the dataset as the last with very few data available because the data were obtained from the real-time data collection on COVID19 epidemiology in population of the Czech Republic; we decided to remove these patients due to following reasons: i) some records of these patients were still incomplete as they were newly added to the database in the time of export of our data, ii) because the characteristics of newly positive patients in time were almost constant in this period of epidemy there was no risk of selection bias and iii) the survival analysis methodology make sense only for this small subset and due to large enough sample size, lower quality of just recently added data and no risk of sampling bias we decided to remove these patients and continue with logistic regression methodology with endpoint occurrence in constant time period (long enough to show for endpoint).

4.

P12 L46 and elsewhere: 'multivariate' should be 'multivariable'.

Corrected, thank you.

5.

It appears from Table 2 that age was categorised (although this is not described in the methods): why was this decision taken, and why these cutpoints? Generally it is better not to categorise continuous variables.

The categorisation was adopted for a purpose of better practical interpretation of results by clinicians; the cut-offs are based on standard 10 years demographical stratification with aggregation of all patients younger than 40 years due their sample size.

6.

What was the criteria used for the backward stepwise algorithm when selecting the optimum model?

Probability for removal in backward stepwise was set to 0.10.

7.

How many predictors were considered for inclusion in the model? Based on Table 1, plus the categorisation of age, I am assuming about 25 effective number of predictors?

Yes, 25.

8.

It isn't clear what the authors did when they say 'A 10-fold cross-validation was performed to reduce the likelihood of model overfitting'. Cross validation doesn't present overfitting - but it can, used correctly, give estimates of e.g. AUC that are adjusted for in-sample optimism.

Thank you for your methodological comment, you are right. We changed the description to “A 10-fold cross-validation was performed to obtain estimates of model performance that are adjusted for in-sample optimism.”

9.

What cut-offs to the reported sensitivity and specificity refer to?

This is the cut-off with the highest sum of sensitivity and specificity.

10.

No results are presented on calibration of the model.

We added results of Hosmer-Lemeshow test.

11.

Tables 2 and S1 need to include the intercepts of the fitted models: as without this absolute risks cannot be calculated.

The intercepts are added to tables.

12.

There is no discussion on how missing data are handled. Given the predictors used this may not be an issue - but it still needs to be stated.

The basic characteristics of patients (age, gender) were provided for all patients, data on comorbidities were available for all patients with match between COVID19 and health insurance companies datasets; patients without match between datasets were excluded from the analysis, no other missing data handling was necessary.
13.
An alternative 'simpler' model is presented in Table S1. Its performance is identical to the original model (except for a difference in the third decimal place on one of the confidence limits) which seems surprising, and should be double checked. Which risk score are the authors actually recommending? The one in Table 2 or in Table S1?

We are recommending simplified model in Table S1 due to its almost identical performance with the primary model in Table 2 and easier interpretation in clinical practice; the covidogram in Figure 1 is based on simplified model too.
14.
What are the confidence intervals for in Figure 1? This Figure reports risk probabilities for individuals, which as such already reflect uncertainty.
The GENLIN (Generalized Linear Models) procedure in IBM SPSS Statistics was used to compute the 95% confidence interval for the mean of the patients' response.

15.
There is no mention of TRIPOD, which is the standard checklist for reporting on the development/validation of a risk prediction model.
TRIPOD is newly mentioned, please, see the section of statistical analysis.

Reviewer: 2

Comments to the Author
Dear authors,

Thank you for a fantastic manuscript. Below are some recommendations to further strengthen the paper.

"More frequently, the disease affects the respiratory system.....". This sentence may benefit from more citations to ensure readers are able to learn more about extra-respiratory manifestations of COVID-19 on their own. Consider citing reviews here, perhaps. If you are having trouble finding papers to cite here, feel free to cite one of these that I have had the pleasure to contribute to - of course, any comparable publication will serve its purpose here.

Aghagoli, G., Gallo Marin, B., Katchur, N.J., Chaves-Sell, F., Asaad, W.F., Murphy, S.A. Neurological Involvement in COVID-19 and Potential Mechanisms: A Review. Neurocrit Care (2020). https://doi.org/10.1007/s12028-020-01049-4

Aghagoli G, Gallo Marin B, Soliman LB, Sellke FW. Cardiac involvement in COVID-19 patients: Risk factors, predictors, and complications: A review. Journal of Cardiac Surgery. April 19, 2020. doi: 10.1111/jocs.14538
Thank you, we agree. The proposed references have been added.

"In a number of patients, there is a risk" - can you report in the manuscript what is the estimate

proportion of these patients? The number may certainly be evolving or be different across different populations, so reporting this as a range would be appropriate if needed. I think it is important to do so to provide context to the reader.
Thank you, it is a very important information. We think that the best information is from WHO – total mortality is about 2.28%. The information was added in the text.

"However, the clinical course might progress over time" - I think this statement requires some modification. It is expected that the clinical course of most pathologies will progress in some way - Perhaps what you would like to convey is that the clinical picture of COVID-19 can quickly turn an unfavorable clinical course in a patient.
Thanks, we changed this, please see the text.

"It was repeatedly demonstrated that older age..." - you may want to say in parenthesis that age > 55 is what we think of in this context. I would definitely include obesity (BMI > 30) , moderate-to-severe asthma, sickle cell disease, and immunosuppression. You may find these characteristics in the CDC website, which should be cited
https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-with-medical-conditions.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fcoronavirus%2F2019-ncov%2Fneed-extra-precautions%2Fgroups-at-higher-risk.html
Thank you, we have added this information into the text.

You mention PCR testing and RT-PCR testing in the methods. I think it would be best to state "RT-PCR" throughout the manuscript, for consistency.
We use RT-PCR in the text, thank you.

"Of the total 7,455 evaluated patients, 1,182..." - Could you articulate what exactly this primary endpoint means?
Primary endpoint was defined as death or necessity of hospitalisation in an ICU, including the use of mechanical ventilation or ECMO, please see the text P 14.

Patients treated with Proton Pump Inhibitors - I appreciate that you included previous data suggesting that PPI use may be linked to an increased risk in pneumonias. This is interesting. I worry however that this conclusion might be a bit too stretched - for instance, does the observation still hold true for patients with acid-related disorders who are not taking proton pumps inhibitors? The observation is complicated by the fact that some patients may not be adherent to their PPI regimen, and that there is likely great variability in the amount of time that they've been on PPIs - Therefore I think that while these conclusions are interesting, they should be frame with caution. I would qualify these statements in your manuscripts by saying that theoretically an apparent link to severity may be explained by pharmacology OR by underlying pathology of acid-related disorders - whether it be controlled or not through medication. Nevertheless, the finding is interesting - just be a bit more conservative about how you report it. Do you have any hypotheses for patients who take medications that induce acid-related disorders (i.e. biphosphonates, some SSRIs, etc?) ?
Thank you for very important comment. We realize that, as you exactly expressed, our data suggest that acid related disorders treated with PPIs might be theoretically linked to severity of COVID-19.
We added next comments, as you suggested.

..for example P 18

…The observation is complicated also by the fact that some patients may not be adherent to their PPI regimen and there was great variability in the amount of time that they have been on PPIs (from one to twelfth month within 2019). Based on our analysis we are not able to decide whether severity of disease COVID-19 might be theoretically explained by pharmacology or by underlying pathology of acid related disorders.

P 17:

Surprisingly, our analysis revealed that the presence of acid-related disorders might be theoretically linked to a severe course of COVID-19

And in conclusion:

Finally, the analysis has shown, for the first time, that acid-related disorders treated with proton-pump inhibitors might also be theoretically associated with a severe course of the disease

Excellent Limitations section and interesting Covidogram.

I look forward to seeing this report in print.

VERSION 2 – REVIEW

REVIEWER	Matthew Sperrin University of Manchester, UK
REVIEW RETURNED	05-Jan-2021

GENERAL COMMENTS	The authors have done a good job of addressing my comments. I do, however, have some remaining concerns. The numbering refers to my previous comments. As a general point, many of the responses to my comments do not state specifically what change was made in the paper, which is very helpful information for a reviewer. 3. I'm happy with the explanation on the unknown history patients, however further clarification is still required on the patients with follow-up period shorter than 14 days. What if one the patients with 'follow-up period shorter than 14 days' had had the outcome while still being followed up? Would they have been included? If so, inclusion/exclusion depends on outcome, which will clearly introduce bias. One way around this would be to use a date cutoff, i.e. exclude all of the recently added data after a certain date. 10. Hosmer-Lemeshow test is not a good way to assess calibration. Please include calibration plots and compute calibration in the large and calibration slopes, both in- and out-of-sample. See https://pubmed.ncbi.nlm.nih.gov/26772608/ 13. If Table S1 is the coefficients for the final recommended model, this should be a main table, not a supplement. 14. These 'confidence intervals' should be removed. Just because a procedure in SPSS outputs something does not mean that it is meaningful for a specific analysis. For a brief discussion see https://www.europeanurology.com/article/S0302-2838(10)01053-5/fulltext

REVIEWER	Benjamin Gallo Marin Warren Alpert Medical School of Brown University, United States
REVIEW RETURNED	22-Dec-2020

GENERAL COMMENTS	Dear Authors, It is my opinion that you have appropriately responded to my comments and those of the other reviewer. I have read your text in its entirely and believe that the manuscript is now well suited for publication.

VERSION 2 – AUTHOR RESPONSE

1.

I'm happy with the explanation on the unknown history patients, however further clarification is still required on the patients with follow-up period shorter than 14 days. What if one the patients with 'follow-up period shorter than 14 days' had had the outcome while still being followed up? Would they have been included? If so, inclusion/exclusion depends on outcome, which will clearly introduce bias. One way around this would be to use a date cutoff, i.e. exclude all of the recently added data after a certain date.

Yes, we used a cut-off date of 3 May 2020. All patients diagnosed after this date were excluded because of short follow-up whether they experienced an event or not.

Please, see the text, p. 10

2.

Hosmer-Lemeshow test is not a good way to assess calibration. Please include calibration plots and compute calibration in the large and calibration slopes, both in- and out-of-sample. See https://pubmed.ncbi.nlm.nih.gov/26772608/

Thank you for your comment. We replaced Hosmer-Lemeshow test by suggested methodology and attached calibration plots for the original model and the simplified model (figure S2A and S2B). Flexible calibration curves based on local regression confirm that both predictive models are well-calibrated.

We definitely agree with reviewer the necessity of both in and out of sample calibration, nevertheless the out of sample calibration is currently not available as data of large sample of patients from the second wave COVID-19 in the Czech are still under preparation.

3.

Table S1 is the coefficients for the final recommended model, this should be a main table, not a supplement.

Thank you, we added the table previously named Table S1 into the main manuscript – please see the Table 3

4.

These 'confidence intervals' should be removed. Just because a procedure in SPSS outputs something does not mean that it is meaningful for a specific analysis. For a brief discussion see https://www.europeanurology.com/article/S0302-2838(10)01053-5/fulltext

Thank you for the comment, confidence intervals were deleted.

VERSION 3 – REVIEW

REVIEWER	Matthew Sperrin University of Manchester, UK
REVIEW RETURNED	21-Jan-2021

GENERAL COMMENTS	I am satisfied with the responses to my queries.