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

TITLE (PROVISIONAL)	A prospective validation study of prognostic biomarkers to predict adverse outcomes in COVID-19 patients: a study protocol
AUTHORS	Tang, Benjamin; Shojaei, Maryam; Wang, Ya; Nalos, Marek; Mclean, Anthony; Afrasiabi, Ali; Kwan, Tim; Kuan, Win Sen; Zerbib, Yoann; Herwanto, Velma; Gunawan, Gunawan; Bedognetti, Davide; Zoppoli, Gabriele; Ballestrero, Alberto; Rinchai, Darawan; Cremonesi, Paolo; Bedognetti, Michele; Matejovic, Martin; Karvunidis, Thomas; Macdonald, Stephen; Cox, Amanda; West, Nicholas; Cripps, Allan; Schughart, Klaus; Maria, Andrea; Chaussabel, Damien; Iredell, Jonathan; Weng, Stephen

VERSION 1 – REVIEW

REVIEWER	TOH LEONG, TAN UNIVERSITI KEBANGSAAN MALAYSIA, MALAYSIA
REVIEW RETURNED	23-Sep-2020

GENERAL COMMENTS	Thank you for the opportunity to review the protocol on “A prospective validation study of prognostic biomarkers to predict adverse outcomes in COVID-19 patients: a study protocol”. The protocol describes a validation study to evaluate several biomarkers’ performance in predicting clinical outcomes of COVID-19 patients. This protocol will address the urgent need for biomarkers for early prediction of COVID-19 complications. Nevertheless, there are a few issues that need to be addressed in its current form. Below are my comments to the authors: Method The prospective study design ensures solid validation of the tested biomarkers’ performance. I highly recommend the authors to include the below information: 1. The study protocol shall include the method of patient sampling. eg: simple randomization, convenience, systematic, cluster, etc. 2. The authors are very thoughtful to consider recruiting patients from the heterogeneous clinical setting. However, selective bias will occur if the clinical settings are not properly pre-selected. Furthermore, it will be good to mention the exact number of recruitment sites in the protocol. 3. The exact sample size calculation/formula needs to be described fully. 4. It is crucial to detail out the processes and methods for blood sample handling. The protocol should include the processes of handling, the methods of storage if they are not processed immediately and the method of processing targetted biomarkers.

	5. The authors shall include a conceptual figure to describe the flow of recruitment, up to the follow-up period for better understand. 6. The authors use RT-PCR for virology testing but also, accept, where appropriate, other tests (e.g. rapid antigen assay, viral culture, or serology). A well-planned prospective study shall use the gold standard virology testing (RT-PCR) which instead of a rapid test assay; serological tests are acceptable. Rapid Test Assay reported false positives in many COVID-19 negative cases. Authors may refer to the below: a) Nagura-Ikeda M, Imai K, Tabata S, Miyoshi K, Murahara N, Mizuno T, Horiuchi M, Kato K, Imoto Y, Iwata M, Mimura S. Clinical evaluation of self-collected saliva by RT-qPCR, direct RT-qPCR, RT-LAMP, and a rapid antigen test to diagnose COVID-19. Journal of clinical microbiology. 2020 Jul 7. b) Rashid ZZ, Othman SN, Samat MN, Ali UK, Wong KK. Diagnostic performance of COVID-19 serology assays. The Malaysian Journal of Pathology. 2020 Apr 1;42(1):13-21. c) Bastos ML, Tavaziva G, Abidi SK, Campbell JR, Haraoui LP, Johnston JC, Lan Z, Law S, MacLean E, Trajman A, Menzies D. Diagnostic accuracy of serological tests for COVID-19: systematic review and meta-analysis. BMJ. 2020 Jul 1;370. Minor corrections Introduction 1. Page 5 lines 35-36, “Second, these biomarkers cannot reliably distinguish between bacterial and viral pneumonia” may be incorrect. Based on many studies, CRP and IL6 are good for differentiating bacterial and viral infection, eg: a) Gao L, Liu X, Zhang D, Xu F, Chen Q, Hong Y, Feng G, Shi Q, Yang B, Xu L. Early diagnosis of bacterial infection in patients with septicemia by laboratory analysis of PCT, CRP, and IL-6. Experimental and Therapeutic Medicine. 2017 Jun 1;13(6):3479-83; b) Simon L, Gauvin F, Amre DK, Saint-Louis P, Lacroix J. Serum procalcitonin, and C-reactive protein levels as markers of bacterial infection: a systematic review and meta-analysis. Clinical infectious diseases. 2004 Jul 15;39(2):206-17. 2. Page 9 line 50: Patient and public “involvement” is miss-spelled.

REVIEWER	Nuala Meyer University of Pennsylvania, USA
REVIEW RETURNED	16-Oct-2020

GENERAL COMMENTS	The authors propose a prospective validation study of 8 candidate mRNA transcripts from whole blood, several of which can be combined into a summary genomic score, as a diagnostic test for COVID-19 outcomes (pneumonia, ARDS, bacterial pneumonia superinfection). The protocol is clearly written and the authors have delineated their study population, blood collection, and primary transcripts to analyze. They will test the diagnostic performance characteristics of each transcript and the genomic score, and will test the incremental value using net reclassification improvement and decision curve analysis. Minor comments: 1. The abstract “Ethics and dissemination” section mentions both influenza and COVID-19, but this protocol seems exclusively

	focused on COVID-19 subjects. Would remove the reference to influenza here as it is confusing. 2. The analytic plan could be further clarified, specifically: what degree of improvement in net reclassification will the authors consider significant? Similarly, the description of the decision curve analysis could be improved. Can the authors provide more explanation of the 'optimal decision threshold' and which variables they imagine contributing to this?

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

This protocol will address the urgent need for biomarkers for early prediction of COVID-19 complications. Nevertheless, there are a few issues that need to be addressed in its current form.

Below are my comments to the authors:

Method

The prospective study design ensures solid validation of the tested biomarkers' performance. I highly recommend the authors to include the below information:

1. The study protocol shall include the method of patient sampling. e.g.: simple randomization, convenience, systematic, cluster, etc.

Author response: We have added the following text to indicate the method of patient sampling (page 5).

"To capture the full spectrum of disease severity, we will recruit a heterogenous patient population using convenience sampling…"

2. The authors are very thoughtful to consider recruiting patients from the heterogeneous clinical setting. However, selective bias will occur if the clinical settings are not properly pre-selected. Furthermore, it will be good to mention the exact number of recruitment sites in the protocol.

Author response: The sites are pre-selected and they do represent a full range of clinical setting. In the revised text, we have specified the number of recruitment sites, as below;

"The target populations will be drawn from pre-selected sites representing six clinical settings, including;

- *The community*
- *Outpatient clinics (e.g. "Fever clinic")*
- *Hospital wards*
- *Emergency departments*
- *Field clinics (e.g. "Cruise ship clinic")*
- *Intensive care units*"

3. The exact sample size calculation/formula needs to be described fully.

Author response: We have re-written the section on sample size calculation (page 8). The new text now reads as below;

"Assuming an event rate of 0.2 (e.g. 20% of recruited subjects develops COVID-19 complications such as respiratory failure) and a sample sensitivity of 0.80, the sample size needed for a two-sided 95% sensitivity confidence interval with a width of at most 0.2 is 350. Assuming an event rate of 0.2 and a sample specificity of 0.8, the sample size needed for a two-sided 95% specificity confidence interval with a width of at most 0.2 is 88. The whole table sample size required so that both confidence intervals have widths less than 0.2 is 350, the larger of the two sample sizes. Therefore, the appropriate sample size for this study is 350 patients. The sample size calculation was estimated using PASS 15 (NCSS, LLC. Kaysville, Utah, U.S.A.)."

4. It is crucial to detail out the processes and methods for blood sample handling. The protocol should include the processes of handling, the methods of storage if they are not processed immediately and the method of processing targeted biomarkers.

Author response: We have added a new paragraph (page 6) to include details relating to sample handling, storage and biomarker measurement, as follows:

"Blood samples will be collected from individuals using the PAXgene/Tempus blood RNA system. Collected tubes will be incubated at room temperature for 4h following blood collection and then stored at -80 °C. Prior to RNA isolation, tubes will be removed from -80°C and allowed to thaw at room temperature overnight. Total RNA will be isolated following the manufacturer's recommended protocol (PAXgene Blood RNA Kit; QIAGEN/ Tempus Spin RNA Isolation;Thermo Fisher). Quality of the resulting RNA samples will be verified on an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA); RNA concentrations will be determined using a NanoDrop® ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE). Reverse transcription will be performed using the qScript cDNA SuperMix per the manufacturer's protocol (Gene Target Solutions, Australia). QPCR will be carried out using TaqMan gene expression Master Mix (Thermo Fisher Scientific, Australia) on a CFX384 Real-Time PCR detection system (Bio-Rad Laboratories, Hercules, CA). CFX Maestro Software will be used for gene expression analysis."

5. The authors shall include a conceptual figure to describe the flow of recruitment, up to the follow-up period for better understand.

Author response: As requested, a new figure (Figure 1) has been added on page 14.

6. The authors use RT-PCR for virology testing but also, accept, where appropriate, other tests (e.g. rapid antigen assay, viral culture, or serology). A well-planned prospective study shall use the gold standard virology testing (RT-PCR) which instead of a rapid test assay; serological tests are acceptable.

Rapid Test Assay reported false positives in many COVID-19 negative cases. Authors may refer to the below:

a). Nagura-Ikeda M, Imai K, Tabata S, Miyoshi K, Murahara N, Mizuno T, Horiuchi M, Kato K, Imoto Y, Iwata M, Mimura S. Clinical evaluation of self-collected saliva by RT-qPCR, direct RT-qPCR, RT-LAMP, and a rapid antigen test to diagnose COVID-19. Journal of clinical microbiology. 2020 Jul 7.

b). Rashid ZZ, Othman SN, Samat MN, Ali UK, Wong KK. Diagnostic performance of COVID-19 serology assays. The Malaysian Journal of Pathology. 2020 Apr 1;42(1):13-21.

c). Bastos ML, Tavaziva G, Abidi SK, Campbell JR, Haraoui LP, Johnston JC, Lan Z, Law S, MacLean E, Trajman A, Menzies D. Diagnostic accuracy of serological tests for COVID-19: systematic review and meta-analysis. BMJ. 2020 Jul 1;370.

> **Author response:** Thank you for the useful references. We agree with the Reviewer that RT-PCR should be adopted as the gold standard virology test. This is consistent with the CDC definitions, which states a positive RT-PCR is considered as "confirmed" COVID-19 cases, positive rapid antigen test as "probable" and positive serological test as "suspect". We have adopted this definition in the current study.

Minor corrections

Introduction

1. Page 5 lines 35-36, "Second, these biomarkers cannot reliably distinguish between bacterial and viral pneumonia" may be incorrect. Based on many studies, CRP and IL6 are good for differentiating bacterial and viral infection, eg:

a) Gao L, Liu X, Zhang D, Xu F, Chen Q, Hong Y, Feng G, Shi Q, Yang B, Xu L. Early diagnosis of bacterial infection in patients with septicemia by laboratory analysis of PCT, CRP, and IL-6. Experimental and Therapeutic Medicine. 2017 Jun 1;13(6):3479-83;

b) Simon L, Gauvin F, Amre DK, Saint-Louis P, Lacroix J. Serum procalcitonin, and C-reactive protein levels as markers of bacterial infection: a systematic review and meta-analysis. Clinical infectious diseases. 2004 Jul 15;39(2):206-17.

2. Page 9 line 50: Patient and public "involvement" is miss-spelled.

Author response: We have removed the sentence "…these biomarkers cannot reliably distinguish between bacterial and viral pneumonia". We have also corrected all misspelling.

Reviewer: 2

The authors propose a prospective validation study of 8 candidate mRNA transcripts from whole blood, several of which can be combined into a summary genomic score, as a diagnostic test for COVID-19

outcomes (pneumonia, ARDS, bacterial pneumonia superinfection). The protocol is clearly written and the authors have delineated their study population, blood collection, and primary transcripts to analyse. They will test the diagnostic performance characteristics of each transcript and the genomic score, and will test the incremental value using net reclassification improvement and decision curve analysis.

Minor comments:

1. The abstract “Ethics and dissemination” section mentions both influenza and COVID-19, but this protocol seems exclusively focused on COVID-19 subjects. Would remove the reference to influenza here as it is confusing.

Author response: As requested, we have removed the wording “influenza” from the text.

2. The analytic plan could be further clarified, specifically: what degree of improvement in net reclassification will the authors consider significant? Similarly, the description of the decision curve analysis could be improved. Can the authors provide more explanation of the ‘optimal decision threshold’ and which variables they imagine contributing to this?

Author response:

Regarding the net reclassification improvement (NRI), we would like to clarify that this index will NOT be used for variable selection. Therefore, we will not test statistical significance between any increments in NRI. Generally speaking, NRI is not an appropriate parameter for hypothesis testing because;

(1). NRI does not incorporate an individual position on a risk distribution (i.e. relative movement on the risk position axis has little impact on treatment decisions),

(2). NRI can make uninformative new markers appear more predictive (e.g. this occurs when risk models are not well-calibrated).

(3). NRI weights reclassifications indiscriminately.

Based on the above limitations related to NRI, we therefore recommend not adding additional text since it may confuse rather than clarify the description of NRI.

Regarding decision curve analysis, we have revised the paragraph to incorporate the Reviewer’s suggestion. The revised paragraph now reads as follow:

"…This is done by calculating the net benefit of one or more models (for instance, with and without predictive biomarker), in comparison to a default strategy of treating all or no patients. Net benefit can be defined as: the sensitivity x prevalence - (1-specificity) x (1-prevalence) x w, where w is the odds at threshold probability. By defining a low threshold (treat many), an assumption can be made that harms arising false-positive is relatively limited but may increase costs and capacity, whereas defining a high threshold (treat few) could result in fewer false-positives but introduce more false-negatives, individuals who are still at high risk"

VERSION 2 – REVIEW

REVIEWER	Nuala Meyer University of Pennsylvania, USA
REVIEW RETURNED	09-Dec-2020

GENERAL COMMENTS	The authors have addressed this reviewer's concerns.