Validation of existing COVID-19 prognostic models using ACTT-1 trial data (#48985)

1) Have any data been collected for this study already?
It's complicated. We have already collected some data but explain in Question 8 why readers may consider this a valid pre-registration nevertheless.

2) What's the main question being asked or hypothesis being tested in this study?
Since the beginning of the COVID-19 pandemic, there have been at least 40 published prognostic models for predicting COVID-19 progression. A living systematic review (Wynants et al. 2020) has concluded that the majority of the published models are not suitable for clinical practice. A systematic validation of 22 prognostic models (Gupta et al. 2020) concluded that the new models did not appear to outperform the established NEWS2 score. Furthermore, the new models did not offer added prognostic performance over single variables such as age or oxygen saturation.

There is a need to both cull and validate the existing collection of COVID-19 prognostic risk models before the models can be implemented in practice. Potential applications include supporting clinical decision making and enriching treatment clinical trials to target higher risk individuals.

Additionally, as the standard of care evolves in response to the rapidly growing body of knowledge around COVID-19, there is a need to both monitor the durability of the accuracy of existing prognostic models and assess whether modifications to the models can improve predictive accuracy.

Question: How well do the established and best-performing of the existing prognostic risk scores, some of which were meant for more general settings, apply to a population of COVID-19 patients on placebo?

We will use data from the adaptive COVID-19 treatment trial (ACTT-1) of remdesivir vs. placebo to answer our question. Though the trial is completed, none of the co-authors have seen the data yet.

3) Describe the key dependent variable(s) specifying how they will be measured.
1. Time to in-hospital mortality starting from treatment initiation
2. Time to progression to ordinal 6/7 on pre-specified severity scale (e.g. will they need ICU/vent/machine)

4) How many and which conditions will participants be assigned to?
List of models we would like to evaluate: NEWS, NEWS2, NEWS + age, MEWS, and a Carr et al. extension of NEWS2. We have confirmation that the data required for NEWS, NEWS + age, and MEWS are available. It is likely that we will not have necessary data for NEWS2 and the Carr model; if so, we will not analyze them. Given the pace at which new models are being published, we will attempt to also evaluate any new models for which our dataset contains the variables necessary to validate them.

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.
The primary analysis will be to compare the performance of the above mentioned models in different ways. Overall discriminatory performance will be summarized by the C-index. Discriminatory performance at set times after baseline scores are measured will be measured using the cumulative/dynamic area under the ROC curve, specifically at 14 days and 28 days of follow-up. Time-varying discrimination will be measured using the incident/dynamic area under the ROC curve.

We will analyze the placebo arm and the remdesivir arms separately, and differences will then be qualitatively assessed and interpreted. Analyzing the remdesivir arm by itself may facilitate the ability to validate any novel findings using data from the ACTT-2 remdesivir arm.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.
Outliers will be kept in the analysis unless there is strong evidence for data corruption, in which case they will be treated as missing data. Missing data will be handled using multiple imputation.

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.
1063 patients were randomized 1:1 to remdesivir and placebo.

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)
At the point of submitting this registration, ACTT-1 was completed and preliminary results were published in the New England Journal of Medicine (Beigel et al. 2020). None of the authors of this project participated in the analysis of the results of the paper. Additionally, prior to submission of this registration, none of the authors accessed the data.

As more of an exploratory exercise, we will consider several variations on the above mentioned risk scores including:

1. Performance of time-averaged risk scores over different averaging windows
2. The slope of risk scores over the first two visits
3. “Smoother” versions of NEWS and NEWS2. For example, the elements of NEWS will be included as continuous predictors in a Cox model with splines to assess how much predictive performance is lost due to categorization of predictors.
4. Since the exploratory risk scores will be newly generated on this dataset, their performance will be evaluated using cross validation.

For racial subgroups that have a sufficient number of individuals, the prognostic accuracy of risk scores will be separately evaluated for each subgroup to provide a description of how the accuracy varies by race.