medRxiv preprint doi: https://doi.org/10.1101/2020.12.03.20239863; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Detailed disease progression of 213 patients hospitalized with Covid-19 in the Czech Republic: An exploratory analysis

Martin Modrák[1], Paul-Christian Bürkner[2], Tomáš Sieger[3], Tomáš Slisz[4], Martina Vašáková[4], Grigorij Mesežnikov[5], Luis Fernando Casas-Mendez[5], Jaromír Vajter[5], Jan Táborský[6], Viktor Kubricht[7], Daniel Suk[8], Jan Horejsek[8], Martin Jedlička[9], Adriana Mifková[9], Adam Jaroš[10], Miroslav Kubiska[11], Jana Váchalová[11], Robin Šín[11], Markéta Veverková[12], Zbyšek Pospíšil[13], Julie Vohryzková[14], Rebeka Pokrievková[15], Kristián Hrušák[14], Kristína Christozova[14], Vianey Leos-Barajas[16], Karel Fišer[17], Tomáš Hyánek[10]

[1] Institute of Microbiology of the Czech Academy of Sciences, Prague, Czech Republic
[2] University of Stuttgart, Germany
[3] Dept. of Cybernetics, Faculty of Electrical Engineering, Czech Technical University in Prague, Czech Republic
[4] Department of Respiratory Medicine, 1st Faculty of Medicine, Charles University; Thomayer Hospital, Prague, Czech Republic
[5] Motol University Hospital, Prague, Czech Republic
[6] AGEL Hospital Nový Jičín, Nový Jičín, Czech Republic
[7] Kralovské Vinohrady University Hospital, Prague, Czech Republic
[8] General University Hospital in Prague, Czech Republic
[9] Military Hospital Olomouc, Olomouc, Czech Republic
[10] Na Homolce Hospital, Prague, Czech Republic
[11] Department of Infectious Diseases and Travel Medicine, Faculty of Medicine in Pilsen, Charles University, University Hospital in Pilsen, Pilsen, Czech Republic
[12] Hořovice Hospital, Hořovice, Czech Republic
[13] Třebíč Hospital, Třebíč, Czech Republic
[14] 2nd Faculty of Medicine, Charles University in Prague
[15] 3rd Faculty of Medicine, Charles University in Prague
[16] Department of Statistical Sciences, University of Toronto
[17] Department of Bioinformatics, 2nd Faculty of Medicine, Charles University in Prague

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.12.03.20239863; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Abstract: We collected a multi-centric retrospective dataset of patients (N = 213) who were admitted to ten hospitals in Czech Republic and tested positive for SARS-CoV-2. The dataset contains baseline patient characteristics, breathing support required, pharmacological treatment received and multiple markers on daily resolution. Patients in the dataset were treated with hydroxychloroquine (N = 108), azithromycin (N = 72), favipiravir (N = 9), convalescent plasma (N = 7), dexamethasone (N = 4) and remdesivir (N = 3), often in combination. Most patients were admitted during the first wave of the epidemic. To explore association between treatments and patient outcomes we performed multiverse analysis, observing how the conclusions change between defensible choices of statistical model, predictors included in the model and other analytical degrees of freedom. Weak evidence to constrain the potential efficacy of azithromycin and favipiravir can be extracted from the data. Additionally, we performed external validation of several proposed prognostic models for Covid-19 severity showing that they mostly perform unsatisfactorily on our dataset.

Introduction

The Covid-19 pandemic caused by severe acute respiratory syndrome coronavirus (SARS-CoV-2) has, as of November 2020, led to over 46 million cases and over 1.1 million deaths. Proposed treatments include antivirals approved for other indications (chloroquine, hydroxychloroquine, lopinavir/ritonavir, remdesivir, favipiravir, umifenovir), azithromycin, corticosteroids, immunoglobulins, tocilizumab and convalescent plasma.[1,2] As of this writing, the only treatment that exhibited positive effects in randomized trials is the corticosteroid dexamethasone for severe cases.[3,4] Remdesivir, previously designed and approved for Ebola, SARS and MERS, is currently the only antiviral against SARS-CoV-2 approved by the FDA. However, during spring and summer it was available only in limited amounts via compassionate use programme. Randomized clinical trials on remdesivir have provided conflicting evidence.[5,6] Despite initial promise, hydroxychloroquine has failed to show benefit for hospitalized patients,[6,7] as an early treatment,[8] as well as pre- and post-exposure prophylaxis.[9–11] Further, no improvement could be discerned when adding azithromycin to hydroxychloroquine therapy.[12] Similarly, lopinavir/ritonavir did not show benefit.[13] No benefit was observed for convalescent plasma.[14,15]

Interestingly, in observational studies, hydroxychloroquine was often found to be associated with better outcomes.[16–18]

Favipiravir is one of the less explored experimental treatments, previously used for influenza including avian types, with only one larger RCT in Covid- 19 known to us,[19] the study however has many potentially severe problems.[20] Other studies are scarce - even when considering very small studies and case series.[21–23] None of the results allow strong conclusions about benefits or harms of favipiravir, although a metaanalysis of all available small studies showed an effect of favipiravir on clinical improvement in mild and moderate disease.[24] For additional summary of treatments under investigation see Siemieniuk et al. [25]

High IL-6, D-dimer values were observed to be associated with worse outcome and increased disease severity.[26] Large study of electronic health records[27] showed an

medRxiv preprint doi: https://doi.org/10.1101/2020.12.03.20239863; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

increase in C-reactive protein in early disease and increase of D-dimer and white blood cell count in later stages of the disease.

An ongoing challenge in evaluating Covid-19 treatments is that the analysis and interpretation of the data is often inappropriate or misleading, most notably interpreting lack of evidence due to small sample size as evidence of no effect.[28,29]

Additionally, many methods for predicting disease severity of Covid-19 were published, but the methods are at high risk of bias and lack external validation.[30]

The present study aims to describe the outcomes and disease course of hospitalized patients with mild to severe clinical presentation in a multicentric Czech cohort, explore the association between the outcomes and pharmacological interventions and to provide external validation to previously published prognostic models for Covid-19 severity.

Patients and Methods

Data Collection

A convenience sample of patients from 10 sites was collected. For each site, the dataset contains all patients hospitalized in the participating wards over the data collection period. The data collection started at the onset of the Covid-19 pandemic (except for one site where some older records were inaccessible), but the end date for collection differs between sites due to time constraints of the participating physicians. Three sites included total of 23 patients that could be considered part of "second wave", i.e. admitted after September 1st, last patient included in the dataset was admitted on October 12th. See Supplementary Figure 1 for per-site data collection periods. Patients over the age of 18 were included if they had PCR-confirmed infection of SARS-CoV-2 and were not participating in a clinical trial of any Covid-19 pharmacotherapy.

We collected data on comorbidities and information about disease progression on daily resolution including breathing support required, oxygen flow rate, experimental anti-Covid-19 and antimicrobial drugs taken and several laboratory markers (PCR positivity for SARS-CoV-2, C-reactive protein, D-dimer, Interleukin 6, Ferritin, lymphocyte count). Full protocol for data collection is attached in the supplementary material. Due to very low number of patients using extra-corporeal membrane oxygenation (N = 1) or non-invasive positive pressure ventilation (N = 6) in our sample, we merged those categories with mechanical ventilation.

Not all patients developed pneumonia or other symptoms of Covid-19. All patients received the standard of care which could include supplemental oxygen and ventilation and antibiotics for bacterial superinfections, as determined by the attending physician. Some patients were not indicated for all treatment modalities (especially mechanical ventilation) based on decision of the attending physician and

medRxiv preprint doi: https://doi.org/10.1101/2020.12.03.20239863; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

underlying patient condition. We note that the participating sites were not homogeneous in either patient population or treatment protocols. The choice of pharmacological treatment was based on the decision of the attending clinician and its availability.

The study was approved by the Ethical committees of General University Hospital, Hospital Nový Jíčín, Motol University Hospital, Thomayer Hospital, University Hospital Vinohrady, Military Hospital Olomouc, Na Homolce Hospital, University Hospital in Pilsen, Hořovice Hospital, Jihlava Hospital.

Statistical analysis

The character of the convenience sample does not allow for a proper assessment of the association between treatments and patient outcomes, because the treatments had not been assigned to patients at random, but were only observed retrospectively. This can be partially remedied by adjusting for patient characteristics in the analysis, but such adjustments will always be imperfect and the analysis needs to be treated as exploratory and interpreted cautiously.

Since many details of analysis may influence the conclusions made, we performed multiverse analysis[31] and report results for all the hypothesis tested across multiple different models using both frequentist and Bayesian paradigms. For each model class we worked with several possible sets of adjustments. All analyses were performed in the R language,[32] visualizations and data cleaning was run with the `tidyverse` package.[33]

First class of models are frequentist survival and multistate models under the proportional hazards assumption as implemented in the `coxph` function from the `survival` package.[34] We primarily use a model with competing risks for death and discharge from hospital (see Figure 1a).

Second class of models are Bayesian hidden Markov models (HMM) of disease progression implemented via a custom extension to the `brms` package.[35] The parametrization of the HMM is inspired by Williams et al.[36]: the actual process of disease is assumed to be continuous and allow only for transitions between neighboring states (as shown in Figure 1b, c). The total probability of transition between any two states over the period of a day is then computed as the total probability of transition across all possible paths. This class of models does not satisfy the proportional hazards assumption, instead, it is assumed the process has the Markov property - i.e. that the (potentially unobserved) state and the covariates at a given day contain all the information to determine probabilities of the states on the next day. We use two versions of such models, one working solely with the observed breathing support and one assuming a hidden improving/worsening distinction. All of the hidden Markov models take into account whether best supportive care was initiated and a patient was thus not indicated to progress to more intensive treatment modalities.

medRxiv preprint doi: https://doi.org/10.1101/2020.12.03.20239863; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

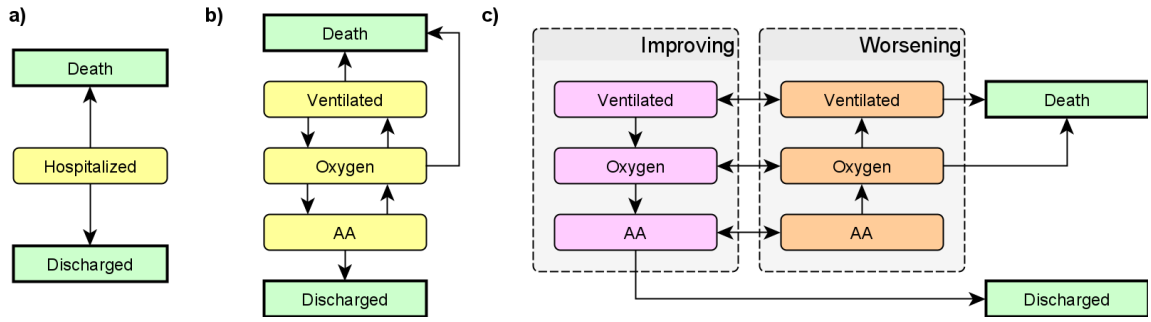

Figure 1: States used in the competing risk model (a) and in the two hidden Markov model variants (b,c). AA = Ambient air, Oxygen = Nasal oxygen, Ventilated = any form of ventilation (non-invasive positive-pressure ventilation, mechanical ventilation and extra-corporeal membrane oxygenation). In all models the 'Death' and 'Discharged' states are terminal. In the second hidden Markov model (c), the 'Improving' and 'Worsening' variants of each non-terminal state are not observable - only the breathing support is observed and improving/worsening is inferred from progression of the disease.

Finally we used a set of Bayesian regression models implemented with the `brms` package.[35] Those included overall survival, state at day 7 or 28 as either binary or categorical outcome and a Bayesian version of the Cox proportional-hazards model.

Except for age, sex and comorbidities, all covariates are treated as time-varying, e.g., the effect of taking a drug is only included for the days after the drug was taken. More details on the exact model formulations can be found in the supplementary statistical analysis.

Evaluating prognostic models

We searched the living systematic review of Covid-19 prognostic models[30] for those that could be applied to our dataset (i.e. where we have gathered all the input features). We primarily focused on the Area Under Receiver Operating Characteristic Curve (AUC), and its bootstrapped 95% confidence intervals which we computed using the `pROC` package.[37] When there were multiple reasonable ways to evaluate the outcome or a predictor in our dataset, we computed and reported all of those options. We used two simple scores with age or the decade of age as the sole predictor to have a baseline to compare the scores against.

Complete code for all analyses is available at https://github.com/cas-bioinf/covid19retrospective/.

Results

In total, we were able to gather data for 213 patients, see Table 1 for the overall characteristics of the patient sample and several subgroups we used in the analysis,

medRxiv preprint doi: https://doi.org/10.1101/2020.12.03.20239863; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

including treatments taken. Counts of all treatment combinations are shown in Supplementary Figure 2 and Supplementary Figure 3 shows outcomes by study site, demonstrating quite large hospital-specific differences. The dataset includes 19 patients already reported in a study of inflammatory signatures of Covid-19.[38]

Table 1: Patient characteristics for the overall sample and treatment subgroups. Note that the favipiravir subgroup is not exclusive with either the HCQ or No HCQ group. IQR = interquartile range, COPD = Chronic obstructive pulmonary disease, BMI = body-mass index, Best supportive care = patient was not indicated to undergo more intensive treatment modality.

	All	HCQ	No HCQ	Favipiravir
N	213	108	105	9
Distinct sites	10	10	10	1
Male	105 (49%)	53 (49%)	52 (50%)	4 (44%)
Age (mean, IQR)	69 (58 - 80)	67 (56 - 80)	70 (64 - 82)	59 (51 - 68)
Admitted for Covid	172 (81%)	96 (89%)	76 (72%)	8 (89%)
Took azithromycin	72 (34%)	63 (58%)	9 (9%)	8 (89%)
Took dexamethasone	4 (2%)	0 (0%)	4 (4%)	0 (0%)
Took favipiravir	9 (4%)	8 (7%)	1 (1%)	9 (100%)
Took remdesivir	3 (1%)	0 (0%)	3 (3%)	0 (0%)
Convalescent plasma	7 (3%)	6 (6%)	1 (1%)	3 (33%)
Ischemic Heart Disease	43 (20%)	15 (14%)	28 (27%)	1 (11%)
Takes antihypertensives	141 (66%)	66 (61%)	75 (71%)	4 (44%)
Heart Failure	34 (16%)	15 (14%)	19 (18%)	0 (0%)
COPD	21 (10%)	9 (8%)	12 (11%)	1 (11%)
Asthma	18 (8%)	9 (8%)	9 (9%)	0 (0%)
Other lung disease	14 (7%)	7 (6%)	7 (7%)	0 (0%)
Diabetes	51 (24%)	18 (17%)	33 (31%)	3 (33%)
Renal Disease	43 (20%)	19 (18%)	24 (23%)	2 (22%)
Liver Disease	14 (7%)	8 (7%)	6 (6%)	1 (11%)
Smoking	28 (13%)	15 (14%)	13 (12%)	1 (11%)
BMI (mean, IQR)	28 (24 - 31)	28 (24 - 30)	28 (24 - 31)	32 (27 - 35)
Best supportive care	58 (27%)	22 (20%)	36 (34%)	0 (0%)
Deceased	42 (20%)	15 (14%)	27 (26%)	0 (0%)
Discharged	122 (57%)	76 (70%)	46 (44%)	8 (89%)

In Figure 2 we show the overall disease progression for all patients and in Figure 3 we show the time-course of a subset of the markers we have measured. The data show some interesting patterns: patients with low Interleukin-6 or D-dimer values are overrepresented among patients with better outcomes, most patients had high CRP upon admission and for many the CRP levels stayed elevated over the whole hospitalization. However, the limited nature of the data does not allow for any statistically robust conclusions. We also see that the marker levels were not

medRxiv preprint doi: https://doi.org/10.1101/2020.12.03.20239863; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

substantially stratified by study site. Those patterns should however be interpreted with care due to systematic missingness of the data - in particular, patients that fared worse were probably more likely to have the markers measured. However we believe this kind of patient-level view is useful to appreciate the extent of both between-patient and within-patient variability.

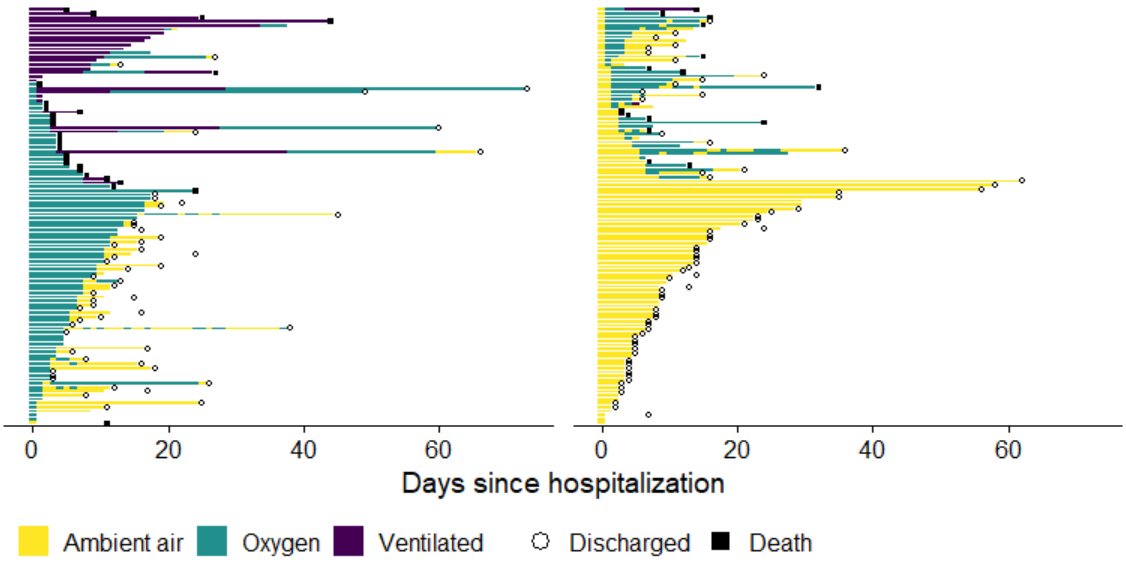

Figure 2: Disease progression for all patients included in the study as determined by breathing support required. Each horizontal strip is a single patient, the ordering on the vertical axis is by disease severity, the split into two panels is for visual purposes only. Ventilated = any form of ventilation (non-invasive positive-pressure ventilation, mechanical ventilation and extra-corporeal membrane oxygenation).

Association between patients' characteristics and treatments

As noted above, the nature of the convenience sample did not enforce random assignment of treatments to patients. In fact, patients with worse baseline characteristics, which lead to worse outcomes, were less likely to receive hydroxychloroquine (see Supplementary Figure 4). This clearly creates a bias towards a positive effect of hydroxychloroquine on the outcome (and potentially for other treatments as well - most were used in combination with hydroxychloroquine), which, however, could be false.

medRxiv preprint doi: https://doi.org/10.1101/2020.12.03.20239863; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

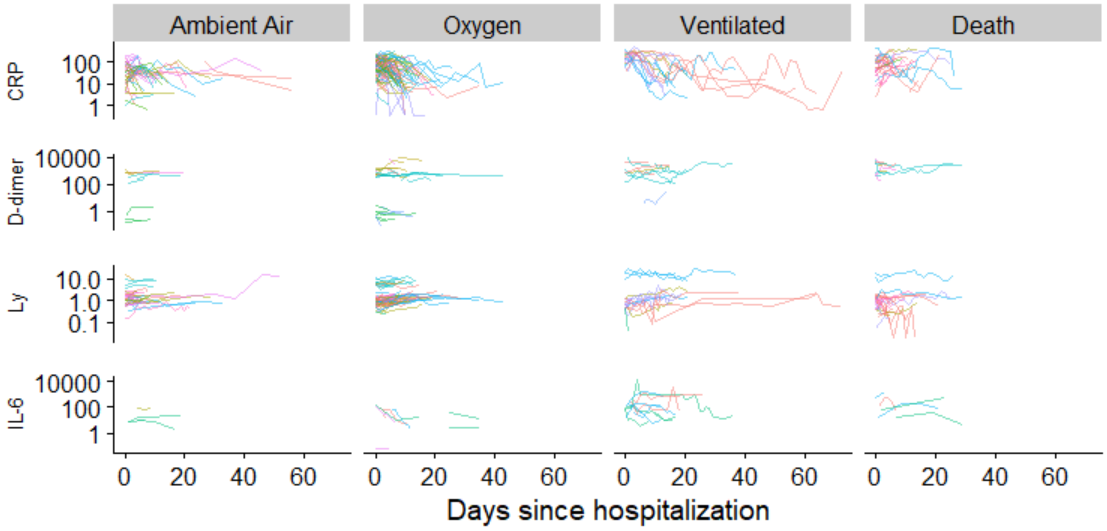

Figure 3: Values of selected markers over the course of the disease. Each line represents a patient, stratified by the worst breathing support required over the hospitalization. Color indicates study sites. The vertical scale is logarithmic. Ventilated = any form of ventilation (non-invasive positive-pressure ventilation, mechanical ventilation and extra-corporeal membrane oxygenation), CRP = C-reactive protein [mg/l], D-dimer [ng/ml DDU], Ly = lymphocyte count [10^9/l], IL-6 = Interleukin 6 [ng/l].

Taken quantitatively, the comorbidities known upon hospitalization were informative with respect to the future hydroxychloroquine treatment: the score representing the cumulative presence of ischemic heart disease, hypertension drugs, former heart failure, COPD, other lung diseases, renal disease, or high creatinine was associated with a lower chance of taking hydroxychloroquine over the course of the hospitalization (the chance was only 79.9%, 95% confidence interval (65.3, 97)%, Chi-square test in the logistic regression model, χ^2=5.18, df=1, P=0.023).

Association between treatments and outcomes

Here, we focus on hydroxychloroquine and azithromycin as those are the only treatments with larger sample size. We also investigate favipiravir as it is less well reported in the literature. Hydroxychloroquine was dosed almost exclusively in a 5-day regime starting with a loading dose of 800mg on the first day and followed by 400mg. Majority of patients complemented hydroxychloroquine with azithromycin while azithromycin was rarely used alone (see Table 1). Azithromycin was most frequently dosed 250 or 500mg/day, but doses ranging from 100mg/day to 1500 mg/day were observed. Favipiravir was used only at one site with a loading dose of 3600mg on the first day, followed by at most 9 days with a 1600mg dose. All but one of the patients receiving favipiravir also received hydroxychloroquine. Treatment was initiated mostly within two days of admission (see Supplementary Figure 5).

medRxiv preprint doi: https://doi.org/10.1101/2020.12.03.20239863; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

The results of the multiverse analysis for association between hydroxochloroquine, azithromycin and favipiravir usage and death is shown in Figure 4 - here, we only show models that were not found to have immediate problems representing the data well or computational issues (see Supplementary statistical analysis for details). Results for all models we tested are reported in supplementary Figures 6-8, with additional details in supplementary statistical analysis. The results do not change noticeably when only patients from the first wave are included (supplementary Figures 6-8).

Most models report that using hydroxychloroquine is associated with lower risk of death. We must however bear in mind the potential bias noted in the previous section. Also, we see that for the HMM models, as we add adjustments the credible intervals do not widen but instead shift towards zero. This is a weak indication that further adjustments could drive the effect towards zero. We did not attempt to model additional adjustments as the models became computationally unstable. The case of hydroxychloroquine serves as a "control group" for our other results - since randomized trials give us high confidence that hydroxychloroquine does not substantially reduce mortality, we can be quite certain the associations we observe for hydroxychloroquine are just a measure of bias in the data. Additionally, our models either cannot determine the sign of association between azithromycin and risk of death or even show an increase in risk of death. This serves as a weak evidence that a substantial improvement in mortality from azithromycin is unlikely.

Most models exclude very strong association between increased risk of death and using favipiravir, but our data are necessarily quite limited, which is reflected in the very wide uncertainties around estimates. We also cannot put strict bounds on the association between favipiravir and length of hospitalization.

We also examined the association between treatments and length of hospital stay for all the patients that survived. Almost all models cannot discern the sign of the association for all treatments examined (Supplementary Figures 6-8). Similarly, we studied the association between D-dimer and Interleukin 6 and outcomes, with unconclusive results as well (Supplementary Figure 9)

medRxiv preprint doi: https://doi.org/10.1101/2020.12.03.20239863; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

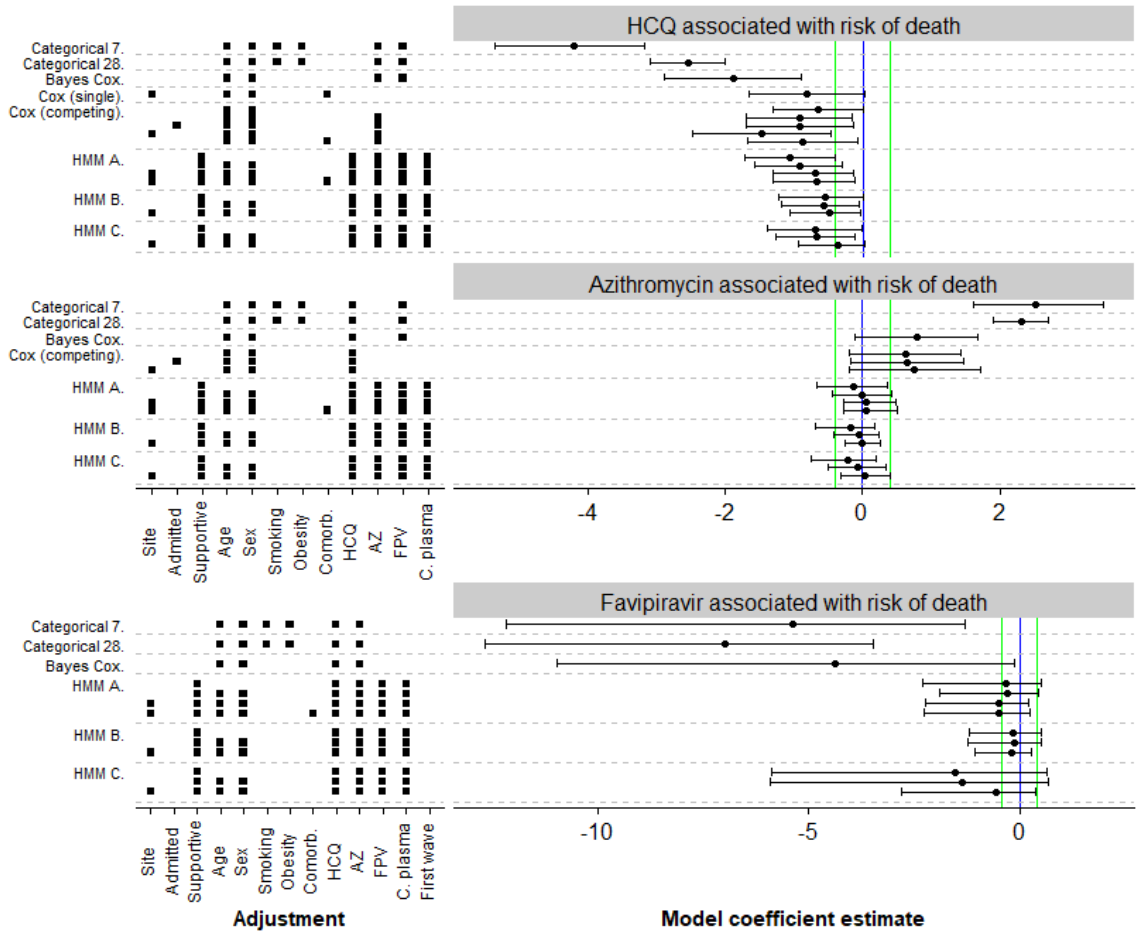

Figure 4: Estimates of model coefficients for association between treatments and main outcomes. Each row represents a model - Categorical 7/28 = Bayesian categorical regression for state at day 7/28, Bayes Cox = Bayesian version of the Cox proportional hazards model with a binary outcome, Cox (single) = frequentist Cox model with a binary outcome, Cox (competing) = frequentist Cox model using competing risks (as in Figure 1a), HMM A = Bayesian hidden-Markov model as in Figure 1b with predictors for rate groups, HMM B = Bayesian hidden-markov model as in Figure 1b with predictors for individual rates, HMM C = Bayesian hidden-Markov model as in Figure 1c. For frequentist models, we show maximum likelihood estimate and 95% confidence interval, for Bayesian models we show posterior mean and 95% credible interval. The estimands are either log odds-ratio (Categorical, HMM) or log hazard ratio (Cox variants). In all cases coefficient < 0 means better patient outcome in the treatment group. Vertical lines indicate zero (blue) and substantial increase or decrease with odds or hazard ratio of 3:2 or 2:3 (green). Additionally the factors the model adjusted for are listed - Site = the study site, admitted = Admitted for Covid-19, Supportive = best supportive care initiated, Comorb. = total number of comorbidities, AZ = took azithromycin, HCQ = took hydroxychloroquine, FPV = took favipiravir, C. plasma = received convalescent plasma.

medRxiv preprint doi: https://doi.org/10.1101/2020.12.03.20239863; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Published prognostic models are not better than using age as the sole predictor of outcome

Following Wynants et al.[30] we found five prediction models we were able to recompute: Li et al. report the ACP index[39] combining CRP and age to form 3 grades, Chen & Liu[40] report a continuous score using age, CRP, D-dimer and lymphocyte count, Shi et al. [41] use age, sex and hypertension to form 4 grades, Caramelo et al. use age, sex, hypertension, diabetes, cardiac disease, chronic respiratory disease and cancer to form a continuous score,[42] Bello-Chavolla et al. [43] use age, diabetes, obesity, pneumonia, chronic kidney disease, COPD and immunosuppression to build a score ranging from -6 to 22. For the latter two scores we had to impute some of the predictors as they had no immediate equivalent in our dataset. The outcomes present in the studies were: 12-day mortality, 30-day mortality and mortality without any further details, here we report results for both 12-day and 30-day mortality. Full details on the scores and how we used our dataset to compute them is given in the supplementary statistical analysis.

All prognostic models we tested performed similarly to or notably worse than using age as the only predictor and also worse than originally reported (Figure 5). Additionally, some publications did not provide enough detail to unambiguously reconstruct how the score and/or outcome was assessed. We thus concur with Wynants et al. [30] that reported prediction scores are at high risk of bias and need additional careful evaluation.

medRxiv preprint doi: https://doi.org/10.1101/2020.12.03.20239863; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

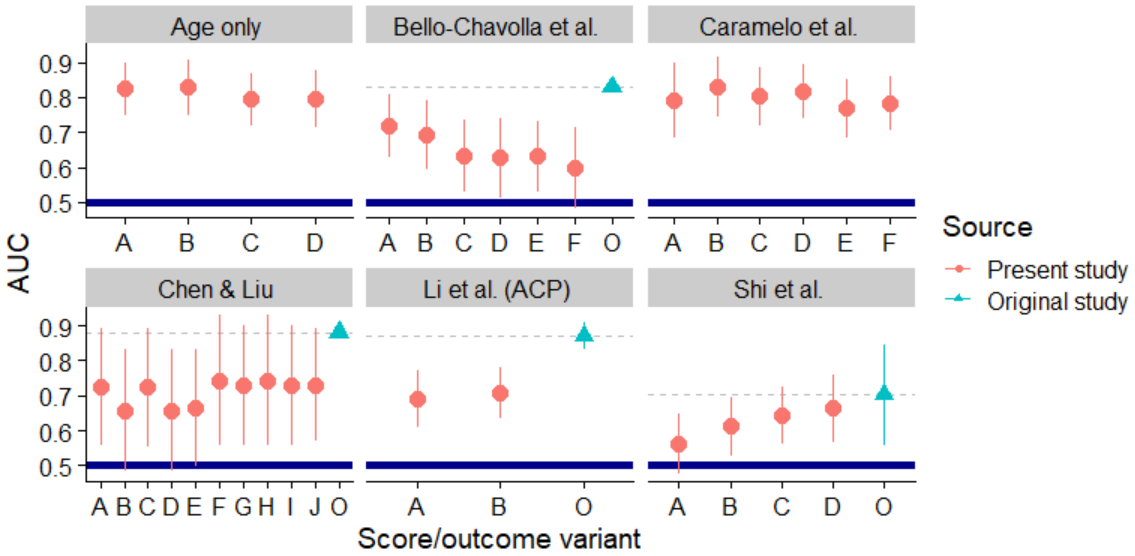

Figure 5: Performance of tested prediction scores as measured by AUC. AUC = 1 means perfect prediction while AUC ≤ 0.5 means that the score is worse than random guess and a better prediction would be obtained by reversing the score (marked by thick blue line). The line ranges represent the bootstrapped 95% confidence intervals. Red dots show results computed in present study - model variants (horizontal axis) vary in the outcome measured (12-day or 30-day mortality, severe disease) and potentially on how ambiguities in score computation were resolved, although this rarely makes a big difference - see supplementary statistical analysis for details. Cyan triangles show AUC as reported by the original authors or recomputed based on their published data. When the confidence interval or the AUC of the original study is not shown, it means that the value was not reported by the authors and not enough information to recompute it was given.

Conclusions

Our data show the extent of between-patient variability in progression of the disease in terms of both length of hospital stay, duration of various types of breathing support and basic markers.

We provide very weak observational evidence against a substantial beneficiary effect of using azithromycin (both with or without hydroxychloroquine) and against substantial negative effect of using favipiravir in hospitalized Covid-19 patients. We also observed better outcomes associated with taking hydroxychloroquine, which is likely linked to substantial confounding by indication. Where our results contradict randomized trials, the most likely explanation is systematic bias in our dataset.

A lesson from our analysis is that the assessment of treatment efficacy from observational data is very sensitive to modelling assumptions while it is usually almost impossible to determine which of the models is more likely to reflect reality

medRxiv preprint doi: https://doi.org/10.1101/2020.12.03.20239863; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

(if any). We believe that using multiverse analysis is an appropriate way to explore data in such contexts as it lets us be transparent about this sensitivity. We further believe that using hidden Markov models is a promising complement to the standard Cox proportional hazards analysis when detailed information on disease progression is available, particularly because it lets us impose additional structure on the model and thus make inferences with more disease states than would be possible to handle in the Cox framework, making better use of the available data.

Additionally, our experience indicates that a substantial fraction of published prognostic models will perform much worse on new patients than on the datasets they were built for and that external validation is crucial. We suggest that comparing the prognostic models against simple baselines (e.g. decade of age as the single predictor) should be a first step in validation. Furthermore, some of the published scores lack enough information to let others implement the score in the same way.

Conflict of Interest

The authors declare that there is no conflict of interest.

Data Availability

The data that support the findings of this study are available on request from the corresponding author, if legal obligations under GDPR can be met. The data are not publicly available due to privacy restrictions. We will be happy to arrange to run any analytical code locally and share the results, provided the code and the results do not leak personal information.

Acknowledgements

This work was supported by ELIXIR CZ research infrastructure project (MEYS Grant No: LM2018131) including access to computing and storage facilities.

References

1. Sanders JM, Monogue ML, Jodlowski TZ, Cutrell JB. Pharmacologic Treatments for Coronavirus Disease 2019 (COVID-19): A Review. *JAMA*. 2020;323(18):1824-1836. doi:10.1001/jama.2020.6019

2. Tobaiqy M, Qashqary M, Al-Dahery S, et al. Therapeutic management of patients with COVID-19: a systematic review. *Infection Prevention in Practice*. 2020;2(3):100061. doi:10.1016/j.infpip.2020.100061

medRxiv preprint doi: https://doi.org/10.1101/2020.12.03.20239863; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

3. RECOVERY CG. Dexamethasone in Hospitalized Patients with Covid-19 — Preliminary Report. *New England Journal of Medicine*. Published online July 17, 2020. doi:10.1056/NEJMoa2021436

4. Sterne JAC, Murthy S, Diaz JV, et al. Association Between Administration of Systemic Corticosteroids and Mortality Among Critically Ill Patients With COVID-19: A Meta-analysis. *JAMA*. Published online September 2, 2020. doi:10.1001/jama.2020.17023

5. Beigel JH, Tomashek KM, Dodd LE, et al. Remdesivir for the Treatment of Covid-19 — Final Report. *New England Journal of Medicine*. 2020;0(0):null. doi:10.1056/NEJMoa2007764

6. WHO Solidarity trial consortium, Pan H, Peto R, et al. *Repurposed Antiviral Drugs for COVID-19 –Interim WHO SOLIDARITY Trial Results*. Infectious Diseases (except HIV/AIDS); 2020. Accessed November 25, 2020. http://medrxiv.org/lookup/doi/10.1101/2020.10.15.20209817

7. Horby P, Mafham M, Linsell L, et al. Effect of Hydroxychloroquine in Hospitalized Patients with COVID-19: Preliminary results from a multi-centre, randomized, controlled trial. *medRxiv*. Published online July 15, 2020:2020.07.15.20151852. doi:10.1101/2020.07.15.20151852

8. Skipper CP, Pastick KA, Engen NW, et al. Hydroxychloroquine in Nonhospitalized Adults With Early COVID-19. *Annals of Internal Medicine*. Published online July 16, 2020. doi:10.7326/M20-4207

9. Boulware DR, Pullen MF, Bangdiwala AS, et al. A Randomized Trial of Hydroxychloroquine as Postexposure Prophylaxis for Covid-19. *New England Journal of Medicine*. 2020;0(0):null. doi:10.1056/NEJMoa2016638

10. Mitja O, Ubals M, Corbacho M, et al. A Cluster-Randomized Trial of Hydroxychloroquine as Prevention of Covid-19 Transmission and Disease. *medRxiv*. Published online July 26, 2020:2020.07.20.20157651. doi:10.1101/2020.07.20.20157651

11. Rajasingham R, Bangdiwala AS, Nicol MR, et al. Hydroxychloroquine as pre-exposure prophylaxis for COVID-19 in healthcare workers: a randomized trial. *medRxiv*. Published online September 21, 2020:2020.09.18.20197327. doi:10.1101/2020.09.18.20197327

12. Furtado RHM, Berwanger O, Fonseca HA, et al. Azithromycin in addition to standard of care versus standard of care alone in the treatment of patients admitted to the hospital with severe COVID-19 in Brazil (COALITION II): a randomised clinical trial. *The Lancet*. 2020;0(0). doi:10.1016/S0140-6736(20)31862-6

13. RECOVERY CG. *Statement from the Chief Investigators of the Randomised Evaluation of COVid-19 thERapY (RECOVERY) Trial on Lopinavir-Ritonavir,*

medRxiv preprint doi: https://doi.org/10.1101/2020.12.03.20239863; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

29June2020.; 2020. Accessed September 12, 2020. https://www.recoverytrial.net/files/lopinavir-ritonavir-recovery-statement-29062020_final.pdf

14. Agarwal A, Mukherjee A, Kumar G, et al. Convalescent plasma in the management of moderate COVID-19 in India: An open-label parallel-arm phase II multicentre randomized controlled trial (PLACID Trial). *medRxiv*. Published online September 10, 2020:2020.09.03.20187252. doi:10.1101/2020.09.03.20187252

15. Li L, Zhang W, Hu Y, et al. Effect of Convalescent Plasma Therapy on Time to Clinical Improvement in Patients With Severe and Life-threatening COVID-19: A Randomized Clinical Trial. *JAMA*. 2020;324(5):460-470. doi:10.1001/jama.2020.10044

16. Catteau L, Dauby N, Montourcy M, et al. Low-dose hydroxychloroquine therapy and mortality in hospitalised patients with COVID-19: a nationwide observational study of 8075 participants. *International Journal of Antimicrobial Agents*. Published online August 24, 2020:106144. doi:10.1016/j.ijantimicag.2020.106144

17. Del Valle DM, Kim-Schulze S, Huang H-H, et al. An inflammatory cytokine signature predicts COVID-19 severity and survival. *Nature Medicine*. Published online August 24, 2020:1-8. doi:10.1038/s41591-020-1051-9

18. Castelnuovo AD, Costanzo S, Antinori A, et al. Use of hydroxychloroquine in hospitalised COVID-19 patients is associated with reduced mortality: Findings from the observational multicentre Italian CORIST study. *European Journal of Internal Medicine*. Published online August 25, 2020. doi:10.1016/j.ejim.2020.08.019

19. Chen C, Zhang Y, Huang J, et al. Favipiravir versus Arbidol for COVID-19: A Randomized Clinical Trial. *medRxiv*. Published online April 15, 2020:2020.03.17.20037432. doi:10.1101/2020.03.17.20037432

20. Wilkinson J, Dahly D. *Statistical Review of Favipiravir Versus Arbidol for COVID-19: A Randomized Clinical Trial*. Zenodo; 2020. Accessed October 26, 2020. https://zenodo.org/record/3734198#.X5amU-0o-Uk

21. Lou Y, Liu L, Qiu Y. Clinical Outcomes and Plasma Concentrations of Baloxavir Marboxil and Favipiravir in COVID-19 Patients: an Exploratory Randomized, Controlled Trial. *medRxiv*. Published online May 5, 2020:2020.04.29.20085761. doi:10.1101/2020.04.29.20085761

22. Dauby N, Van Praet S, Vanhomwegen C, Veliziotis I, Konopnicki D, Roman A. Tolerability of favipiravir therapy in critically ill patients with COVID-19: A report of four cases. *J Med Virol*. Published online September 29, 2020:jmv.26488. doi:10/ghhwf6

medRxiv preprint doi: https://doi.org/10.1101/2020.12.03.20239863; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

23. Takahashi H, Iwasaki Y, Watanabe T, et al. Case studies of SARS-CoV-2 treated with favipiravir among patients in critical or severe condition. *International Journal of Infectious Diseases*. 2020;100:283-285. doi:10.1016/j.ijid.2020.08.047

24. Shrestha DB, Budhathoki P, Khadka S, Shah PB, Pokharel N, Rashmi P. Favipiravir versus other antiviral or standard of care for COVID-19 treatment: a rapid systematic review and meta-analysis. *Virol J*. 2020;17(1):141. doi:10.1186/s12985-020-01412-z

25. Siemieniuk RA, Bartoszko JJ, Ge L, et al. Drug treatments for covid-19: living systematic review and network meta-analysis. *BMJ*. 2020;370. doi:10.1136/bmj.m2980

26. Maeda T, Obata R, Rizk DO D, Kuno T. The association of interleukin-6 value, interleukin inhibitors, and outcomes of patients with COVID-19 in New York City. *J Med Virol*. Published online August 25, 2020:jmv.26365. doi:10/gg6m3s

27. Brat GA, Weber GM, Gehlenborg N, et al. International electronic health record-derived COVID-19 clinical course profiles: the 4CE consortium. *npj Digital Medicine*. 2020;3(1):1-9. doi:10.1038/s41746-020-00308-0

28. Hood K, Dahly D, Wilkinson J. Statistical review of Efficacy and safety of lopinavir/ritonavir or arbidol in adult patients with mild/moderate COVID-19: an exploratory randomized controlled trial. Published online May 1, 2020. doi:10.5281/zenodo.3779933

29. Hood K, Goulao B, Dahly D, Yap C. *Statistical Review of Remdesivir in Adults with Severe COVID-19: A Randomised, Double-Blind, Placebo-Controlled, Multicentre Trial*. Zenodo; 2020. Accessed September 12, 2020. https://zenodo.org/record/3819778#.X1yBlotS-Uk

30. Wynants L, Calster BV, Collins GS, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. *BMJ*. 2020;369. doi:10.1136/bmj.m1328

31. Steegen S, Tuerlinckx F, Gelman A, Vanpaemel W. Increasing Transparency Through a Multiverse Analysis. *Perspect Psychol Sci*. 2016;11(5):702-712. doi:10.1177/1745691616658637

32. R Core Team. *R: A Language and Environment for Statistical Computing*.; 2020. https://www.R-project.org/

33. Wickham H, Averick M, Bryan J, et al. Welcome to the Tidyverse. *JOSS*. 2019;4(43):1686. doi:10.21105/joss.01686

34. Therneau TM, Grambsch PM. *Modeling Survival Data: Extending the Cox Model*. Springer; 2000.

medRxiv preprint doi: https://doi.org/10.1101/2020.12.03.20239863; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

35. Bürkner P-C. Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal*. 2018;10(1):395. doi:10.32614/RJ-2018-017

36. Williams JP, Storlie CB, Therneau TM, Jr CRJ, Hannig J. A Bayesian Approach to Multistate Hidden Markov Models: Application to Dementia Progression. *Journal of the American Statistical Association*. 2020;115(529):16-31. doi:10.1080/01621459.2019.1594831

37. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. 2011;12(1):77. doi:10.1186/1471-2105-12-77

38. Parackova Z, Zentsova I, Bloomfield M, et al. Disharmonic Inflammatory Signatures in COVID-19: Augmented Neutrophils' but Impaired Monocytes' and Dendritic Cells' Responsiveness. *Cells*. 2020;9(10):2206. doi:10.3390/cells9102206

39. Lu J, Hu S, Fan R, et al. ACP risk grade: a simple mortality index for patients with confirmed or suspected severe acute respiratory syndrome coronavirus 2 disease (COVID-19) during the early stage of outbreak in Wuhan, China. *medRxiv*. Published online February 23, 2020:2020.02.20.20025510. doi:10.1101/2020.02.20.20025510

40. Chen X, Liu Z. Early prediction of mortality risk among severe COVID-19 patients using machine learning. *medRxiv*. Published online April 19, 2020:2020.04.13.20064329. doi:10.1101/2020.04.13.20064329

41. Shi Y, Yu X, Zhao H, Wang H, Zhao R, Sheng J. Host susceptibility to severe COVID-19 and establishment of a host risk score: findings of 487 cases outside Wuhan. *Critical Care*. 2020;24(1):108. doi:10.1186/s13054-020-2833-7

42. Caramelo F, Ferreira N, Oliveiros B. Estimation of risk factors for COVID-19 mortality - preliminary results. *medRxiv*. Published online February 25, 2020:2020.02.24.20027268. doi:10.1101/2020.02.24.20027268

43. Bello-Chavolla OY, Bahena-López JP, Antonio-Villa NE, et al. Predicting Mortality Due to SARS-CoV-2: A Mechanistic Score Relating Obesity and Diabetes to COVID-19 Outcomes in Mexico. *J Clin Endocrinol Metab*. 2020;105(8):2752-2761. doi:10.1210/clinem/dgaa346