




Hypothesis-Agnostic Network-Based Analysis of Real-World Data Suggests Ondansetron is Associated with Lower COVID-19 Any Cause Mortality

Gregory M. Miller¹ · J. Austin Ellis² · Rangaprasad Sarangarajan¹ · Amay Parikh³ · Leonardo O. Rodrigues¹ · Can Bruce¹ · Nischal Mahaveer Chand¹ · Steven R. Smith³ · Kris Richardson¹ · Raymond Vazquez⁴ · Michael A. Kiebish¹ · Chandran Haneesh³ · Elder Granger¹ · Judy Holtz³ · Jacob Hinkle⁵ · Niven R. Narain¹ · Bret Goodpaster³ · Jeremy C. Smith^{6,7} · Daniel S. Lupu³ 

Accepted: 27 March 2022 / Published online: 9 July 2022
© The Author(s) 2022

Abstract

Background The COVID-19 pandemic generated a massive amount of clinical data, which potentially hold yet undiscovered answers related to COVID-19 morbidity, mortality, long-term effects, and therapeutic solutions.

Objectives The objectives of this study were (1) to identify novel predictors of COVID-19 any cause mortality by employing artificial intelligence analytics on real-world data through a hypothesis-agnostic approach and (2) to determine if these effects are maintained after adjusting for potential confounders and to what degree they are moderated by other variables.

Methods A Bayesian statistics-based artificial intelligence data analytics tool (bAIcis[®]) within the Interrogative Biology[®] platform was used for Bayesian network learning and hypothesis generation to analyze 16,277 PCR+ patients from a database of 279,281 inpatients and outpatients tested for SARS-CoV-2 infection by antigen, antibody, or PCR methods during the first pandemic year in Central Florida. This approach generated Bayesian networks that enabled unbiased identification of significant predictors of any cause mortality for specific COVID-19 patient populations. These findings were further analyzed by logistic regression, regression by least absolute shrinkage and selection operator, and bootstrapping.

Results We found that in the COVID-19 PCR+ patient cohort, early use of the antiemetic agent ondansetron was associated with decreased any cause mortality 30 days post-PCR+ testing in mechanically ventilated patients.

Conclusions The results demonstrate how a real-world COVID-19-focused data analysis using artificial intelligence can generate unexpected yet valid insights that could possibly support clinical decision making and minimize the future loss of lives and resources.

✉ Daniel S. Lupu
Daniel.Lupu@adventhealth.com

¹ BERG, Framingham, MA 01701, USA

² National Center for Computational Sciences, Oak Ridge National Laboratory, Oak Ridge, TN 37830, USA

³ Translational Research Institute, AdventHealth Research Institute, 301 E Princeton St, Orlando, FL 32804, USA

⁴ AdventHealth Information Technology, Orlando, FL 32804, USA

⁵ Computational Sciences and Engineering Division, Oak Ridge National Laboratory, Oak Ridge, TN 37830, USA

⁶ Department of Biochemistry and Cellular and Molecular Biology, University of Tennessee, Knoxville, TN 37996, USA

⁷ UT/ORNL Center for Molecular Biophysics, Oak Ridge National Laboratory, Oak Ridge, TN 37830, USA

Key Points

Electronic health records data collected from 16,277 patients with COVID-19 analyzed using artificial intelligence found 223/239 and 3/239 significant variables to be associated with increased and decreased any cause mortality at any timepoint post-COVID-19 PCR+ testing, respectively.

Two of three variables referred to the commonly prescribed antiemetic ondansetron, showing that in hospital ondansetron treatment either early (less than 7 days) or late (more than 28 days) after a COVID-19 PCR+ test is associated with decreased any cause mortality at any timepoint post-COVID-19 PCR+ testing in patients aged under 60 years.

Inpatient sub-group analyses that accounted for variables associated with any cause mortality 30 days post-PCR+ testing and their interactions found that ondansetron treatment early (less than 7 days) after COVID-19 PCR+ testing is associated with decreased any cause mortality 30 days post-PCR+ testing in patients who received mechanical ventilation.

1 Introduction

1.1 Background and Significance

As of February 2022, an estimated 383 million cases of COVID-19 and over 5.6 million deaths have been reported worldwide [1]. The USA has reported the largest proportion of COVID-19 cases estimated at 40 million with approximately 900,000 reported deaths [1]. Worldwide efforts are currently focused on the implementation of an aggressive vaccination program to control the pandemic. Despite the control strategies of limiting COVID-19 infections by physical measures (use of masks, isolation, social distancing) and vaccination, the emergence of new SARS-CoV-2 variants across the globe, the increased incidence of breakthrough infections, especially in the younger population, and the evolving understanding of infection cycles and re-emergence provide impetus for continued investigation of real-world data (RWD). This investigation can generate insights into disease susceptibility and long-term effects, and can provide potential therapeutic strategies.

The revolution in computational analytics, including the considerable progress achieved in the application of artificial intelligence (AI) and machine learning capabilities,

in tandem with access to high-density RWD and clinical evidence, provides a suitable environment to generate hypothesis-agnostic insights for the management of health and disease. Further, the availability of supercomputers and cloud-based high-performance computing capabilities significantly increases analytical depth and reduces the time required to perform higher order AI/machine learning analytics of large population-based datasets, thus permitting a better understanding of disease etiology and facilitating the identification of novel information pertinent to disease management.

Artificial intelligence has been extensively applied to analyze various COVID-19 data, including to aid diagnostics and in therapeutic design [2]. In RWD AI, machine learning has been used to predict the probability of acute respiratory distress syndrome based on the clinical characteristics of patients with COVID [3]. A further study, on 3194 COVID-19 cases in the Emory Healthcare Network, assessed whether the need for hospitalization in a patient with COVID-19 can be predicted at the time of their RT-PCR test using electronic medical record data prior to the test [4].

Although concern has been raised about the use of untested AI programs and small data sets in COVID-19 research [5], AI continues to play a major role in COVID-19 decision making. Of particular importance in the current pandemic is finding novel hypotheses for disease outcomes, and in this respect, Bayesian networks are ideal for taking an event that occurred and predicting the likelihood that any one of several possible causes was the main contributing factor. Bayesian networks create a network of dependency links among variables of interest [6]. Such analysis has the benefit of determining which independent variables are directly associated with a clinical outcome variable of interest (e.g., death, admission to intensive care unit), and which variables are located further upstream. A drawback of Bayesian networks is that their computational complexity is relatively high, but this can be overcome with sufficiently large computational power.

The bAIcis[®] algorithm generates a network of directed associations connecting variables present in the input dataset. As described previously [7], the algorithm first generates many families, each consisting of combinations of parent variables for each child variable. Here, directed association flows from the parents to the child. Next, the many families are combined to produce a final network. The resulting network topology may serve as a feature selection for subsequent multivariate modeling.

The current study combines a large amount of COVID-19-focused RWD from AdventHealth that were analyzed using a Bayesian statistics-driven platform on a supercomputer at the Oak Ridge National Laboratory. The objective of this study was to generate Bayesian network models to

identify factors associated with any cause mortality post-COVID-19 PCR+ testing, including drugs that have the potential to improve outcomes in patients with COVID-19. We proposed to employ the bAIcis algorithm [7] to develop Bayesian networks based on various patient subpopulations and patient features from different time windows, before and after COVID-19 diagnosis, in order to identify those variables having a likely association with any cause mortality post-COVID-19 PCR+ testing. Finally, we wanted to determine if these effects are maintained after adjusting for potential confounders and to what degree they are moderated by other variables.

2 Materials and Methods

2.1 Data Collection

AdventHealth, headquartered in Orlando, Florida, is one of the largest non-profit healthcare systems in the USA with over 50 hospitals in 12 states, and 5 million patient encounters annually (including inpatient, outpatient, and emergency visits) [8]. Early in the COVID-19 pandemic, AdventHealth established the Registry and Biorepository of COVID-19 (RECOVER-19), a registry of all patients tested for SARS-CoV-2 within the AdventHealth Enterprise. The registry comprised raw data extracted from the AdventHealth Cerner electronic health records system and was made available in the Data Lake powered by the Integrated Data for Enterprise Analytics platform. The registry collected associated data from inpatients and outpatients, structured

in nine data tables, using the approach described in Table 1 of the Electronic Supplementary Material (ESM): patient IDs, COVID-19 encounters, diagnoses, problems (patient personal medical history starting 2016), procedures related to the COVID-19 visit, clinical events, lab results, in-house medication administration records, and recorded home and prescription discharge medications. To facilitate a comprehensive selection and inclusion of data elements in potential studies using the registry data, the Clinical Classifications Software Refined was implemented [9]. This allowed the aggregation of diagnostic and procedure codes in a manageable number of clinical categories across clinical domains and body systems. This study was approved by the AdventHealth Institutional Review Board (IRBnet #1590483).

2.2 Cohort Selection and Subgroup Definitions

We stratified the RECOVER-19 registry patients based on SARS-CoV-2 test type, as well as inpatient vs outpatient status (Fig. 1). The RECOVER-19 registry included 279,281 inpatients and outpatients tested for SARS-CoV-2 infection by antigen, antibody, or PCR methods from January to December 2020. From the 35,504 positive patients thus found, we selected for this work a sub-cohort of PCR+ patients ($n = 16,277$), owing to the higher sensitivity and specificity of this diagnostic method (Fig. 1). This cohort was used for the initial AI analysis (bAIcis[®]), with patients stratified by age, race, or ethnicity to identify potential factors related to any cause mortality post-COVID-19 PCR+ testing (Fig. 2).

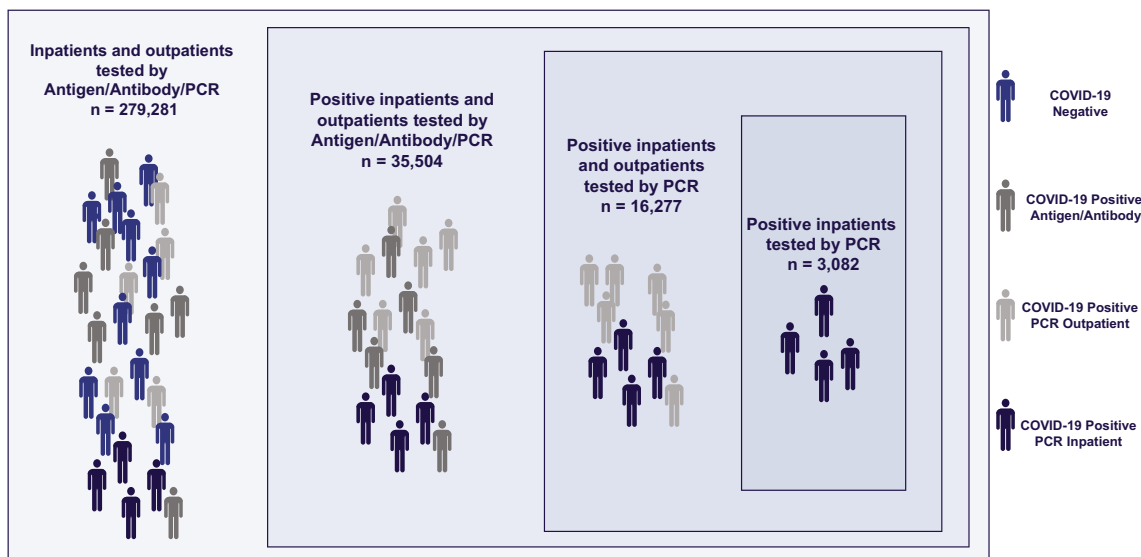


Fig. 1 RECOVER-19 registry included 279,281 inpatients and outpatients tested for SARS-CoV-2 infection by antigen, antibody, or PCR methods from January to December 2020. There were 16,277 PCR+ patients selected for this analysis

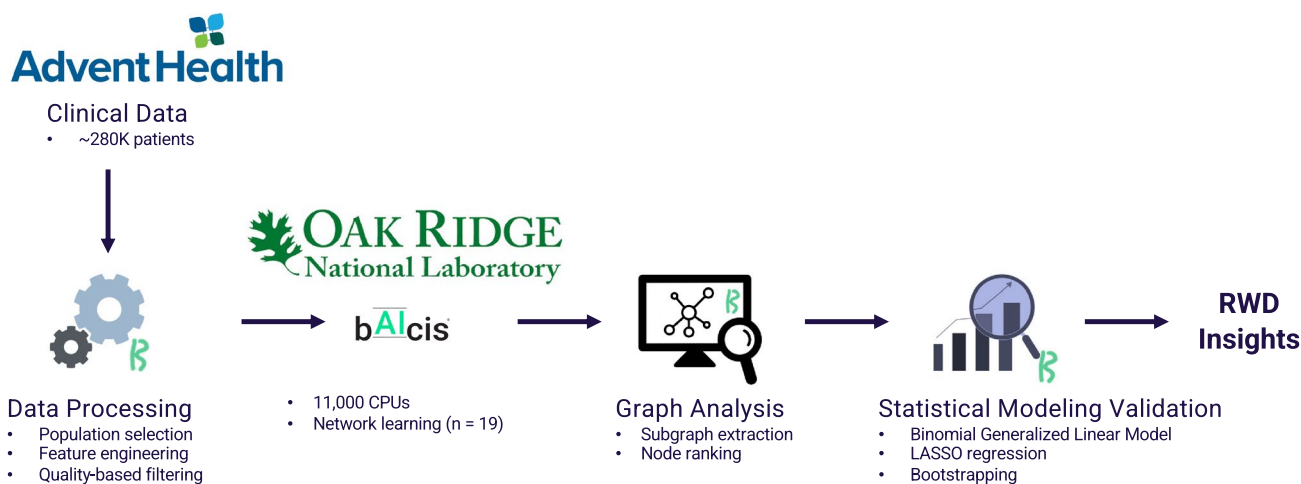


Fig. 2 Overview of the data processing workflow. *LASSO* least absolute shrinkage and selection operator, *RWD* real-world data

This cohort of 16,277 patients included a subset of individuals who were admitted to the inpatient setting ($n = 3082$) (Fig. 1, Table 1). As most any cause mortality events post-COVID-19 PCR+ testing occurred in the inpatients group, further statistical analyses of factors relating to this mortality were focused on this smaller cohort. The inpatients cohort was divided into subgroups according to age: 18–39 years ($n = 389$), 40–49 years ($n = 391$), 50–59 years ($n = 548$), 60–69 years ($n = 655$), and 70+ years ($n = 1099$). To allow a less granular approach to patient age, two additional subgroups of inpatients were created: age < 60 years ($n = 1328$) and age 60+ years ($n = 1754$).

2.3 Outcome Definition

The focus of this work was to identify factors associated with any cause of mortality in patients who received a positive RT-PCR test for COVID-19. During the initial discovery phase, patients with COVID-19 with any cause mortality were defined as having *deceased_flag* = 1 in the patients table at any timepoint, or having had an encounter with *discharge_disposition* = “Expired – 20.” To better understand the effect of ondansetron use in the inpatient setting, this definition was refined to be 30-day any cause COVID-19 mortality in the multivariate analysis portion of this work. Here, 30-day any cause of mortality following COVID-19 was defined as having had an encounter with *discharge_disposition* = “Expired – 20” within 30 days of a COVID-19 RT-PCR+ test.

2.4 Alignment of Patient Journeys

To enable Bayesian network inference, patients were aligned along their COVID-19 disease journey and features were defined in 11 time bins in relation to the time of a PCR+

specimen collection. The time windows utilized in the feature generation were: > 12, 9–12, 6–9, 3–6, 1–3, and < 1 month prior to and < 7, 7–14, 14–21, 21–28, and >28 days after the time of COVID-19 PCR+ test sample collection (Fig. 3). Using these time bins, features were derived from all data tables.

The rationale for selecting time bins was as follows. First, time windows preceding a COVID-19 PCR+ test were selected to allow intermediate (on order of months prior to infection) to longer term (more than a year prior to infection) co-morbidities or prior procedures to be captured. Conversely, time windows following COVID-19 infection were selected on a shorter time period (weekly) to capture potential treatments administered according to COVID-19 disease progression. Finally, the overall number of time bins ($n = 11$) was selected to keep the number of variables defined in the dataset manageable to reduce the computational complexity for bAIcis[®] learning.

2.5 Data Analysis Methodology Summary

Briefly, the analysis performed in this work involved the following steps:

- (1) To identify factors associated with any cause mortality after a COVID-19 PCR+ test in a hypothesis-free manner, the bAIcis algorithm was used to learn Bayesian networks, consisting of clinical variables (e.g., demographics, medications, procedures, and lab tests) recorded for 16,277 patients with a COVID-19 PCR+ test (inpatients and outpatients). Similarly, the bAIcis algorithm was also utilized to learn Bayesian networks generated from subsets of this patient cohort (e.g., patients with a COVID-19 PCR+ test who were admitted to an inpatient setting).

Table 1 General characteristics of the two populations studied

	COVID-19 PCR+ patients	COVID-19 PCR+ inpatients
Deceased ^a	373 (2.3%)	262 (8.5%)
Mechanical ventilation	444 (2.7%)	444 (14.4%)
Female	8441 (51.9%)	1473 (47.8%)
Male	7836 (48.1%)	1609 (52.2%)
White	7787 (47.8%)	1674 (54.3%)
African American	2417 (14.8%)	649 (21.1%)
Asian	166 (1%)	48 (1.6%)
Unknown race	2398 (14.7%)	48 (1.6%)
Hispanic	6072 (37.3%)	1231 (39.9%)
Not Hispanic	7047 (43.3%)	1737 (56.4%)
Unknown ethnicity	3158 (19.4%)	114 (3.7%)
Age, years		
18–39	7146 (43.9%)	389 (12.6%)
40–49	2731 (16.8%)	391 (12.7%)
50–59	2615 (16.1%)	548 (17.8%)
60–69	1883 (11.6%)	655 (21.3%)
70+	1902 (11.7%)	1099 (35.7%)
Heart failure	530 (3.3%)	452 (14.7%)
COPD	412 (2.5%)	330 (10.7%)
Asthma	1198 (7.4%)	348 (11.3%)
Kidney disease	765 (4.7%)	658 (21.3%)
Neoplastic disease	383 (2.4%)	186 (6%)
Ondansetron	1438 (8.8%)	924 (30%)
Dexamethasone	792 (4.9%)	540 (17.5%)
Tocilizumab	295 (1.8%)	295 (9.6%)
Convalescent plasma	321 (2%)	321 (10.4%)
Remdesivir	1158 (7.1%)	1158 (37.6%)
Azithromycin	1547 (9.5%)	1208 (39.2%)
All	16,277	3082

COPD chronic obstructive pulmonary disorder

^aDefinition for ‘deceased’ is detailed in Sect. 2.3. Briefly, for ‘all patients’, deceased means any cause mortality at any time post-COVID-19 PCR+ testing, and for all ‘inpatients’ deceased means any cause mortality at 30 days post-COVID-19 PCR+ testing

(2) Based on the results of the first step of the analysis, we focused on a subset of 3082 hospitalized patients and examined whether the effect of ondansetron is independent of 24 other variables known or suspected to influence any cause mortality at a 30-day post-COVID-19 PCR+ test. The presence of missing data in some of the variables of interest necessitated the use of imputation to create a complete data set amenable for logistic regression analysis. Because of the stochastic nature of the imputation method utilized, five versions of the imputed data were generated and analyzed independently.

- (3) The relationship between ondansetron and any cause mortality at 30 days after a COVID-19 PCR+ test was further investigated by accounting for possible interactions effects between the 25 variables and this mortality. We considered 333 variables, including the original 25 main-effect variables and their pairwise interaction terms. We used least absolute shrinkage and selection operator (LASSO) regression, a penalized regression approach, to identify the subset of these variables that are the most likely to explain any cause mortality at a 30-day post-COVID-19 PCR+ test.
- (4) Finally, we estimated the variability of our findings in step 3. This was done by (1) generating ten new independently imputed data sets, and (2) from each of these imputed data sets, generating 1000 data sets with similar patient distributions, and (3) applying LASSO regression as before. The results of this analysis allowed the estimation of the variance of the regression coefficients for all the variables of interest.

2.6 Bayesian Network Learning and Hypotheses Generation

BAIcis[®] learning was employed to generate hypotheses of factors related to any cause mortality at any time after a COVID-19 PCR+ test. In the generated Bayesian networks, nodes represent the analyzed features and edges represent the directed relationships, and where an upstream/parent node drives changes in downstream/child nodes. In this context, bAIcis[®] allows the pre-definition of network hierarchies, thus variables can be constrained to not have parent nodes (constrained as ‘top’) or child nodes (constrained as bottom), such as, for example, the top variable, ‘Age’ that is not driven by any other variable. In this regard, the top and bottom variables were selected in alignment with the data structure, and Bayesian networks were inferred containing the potential cause-and-effect relationships. Following bAIcis[®] learning, features related to any cause mortality after a COVID-19 PCR+ test were identified by the following approach. First, subgraphs were extracted from the original bAIcis[®] networks by removing infrequent edges (edges present in the ensemble model with a frequency $\leq 20\%$), then extracting the nodes connected to the any cause mortality post COVID-19 PCR+ test node by the first, second, or third degree (Table 2, column 4). All nodes selected by these criteria were then assessed by a univariate statistical analysis to have a significant relationship to any cause mortality after a COVID-19 PCR+ test in the respective patient cohort used for network learning. Features with a p -value ≥ 0.05 by the Fisher’s exact test were considered insignificant and those between a 0.67 and 1.5 odds ratio (OR) were considered of low effect size, and these were disregarded in a further

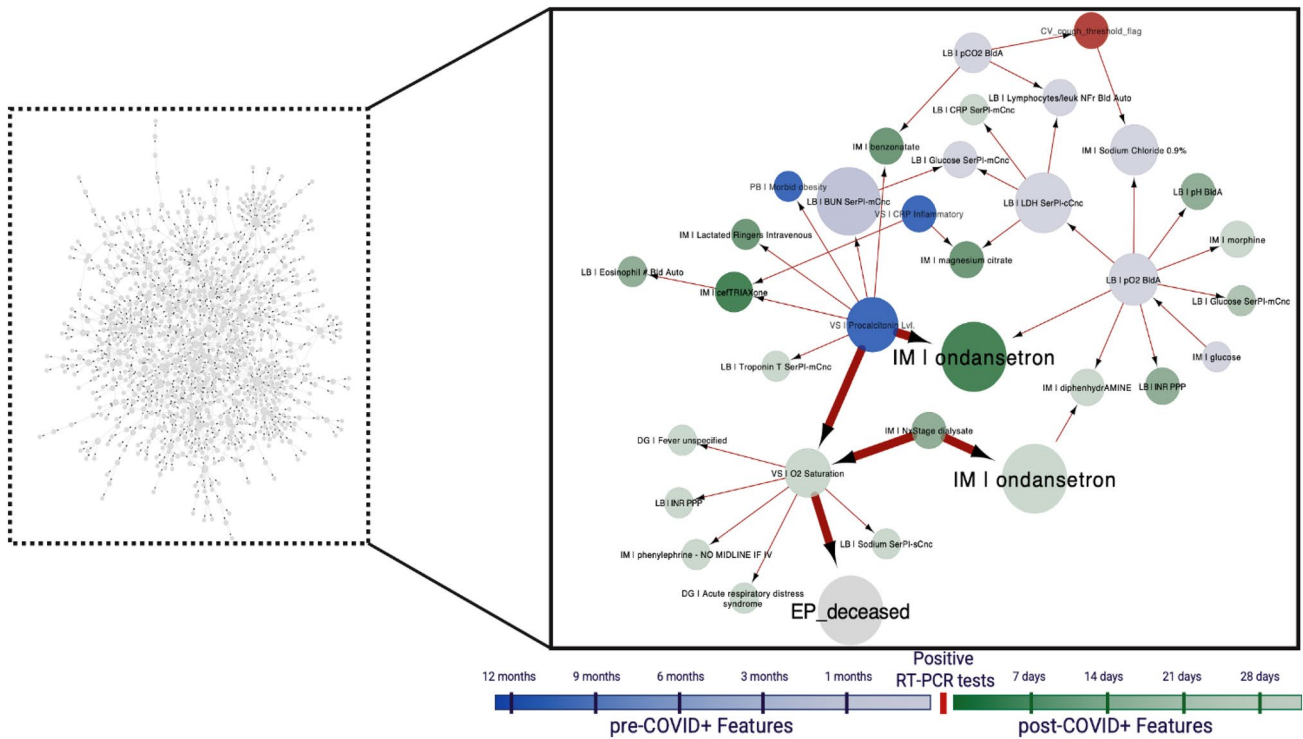


Fig. 3 Illustrative example of the analysis approach. Left: a visualization of the Bayesian artificial intelligence analytics (bAIcis[®]) network learned from the age < 60 years population admitted to the inpatient setting ($n = 1328$). Right: subgraph of the “Inpatients age < 60” bAIcis[®] network illustrating the linkage between ondansetron use and any cause mortality at any time after the COVID-19 PCR+ test. Ondansetron use within the first week and after 28 days after the COVID-19 PCR+ test was significantly associated with decreased

any cause mortality (highlighted nodes). Features were defined in 11 time bins in relation to the time of COVID-19 PCR+ specimen collection. The time windows utilized in feature generation were: > 12, 9–12, 6–9, 3–6, 1–3, and < 1 months prior to the time of COVID-19 PCR+ test sample collection and < 7, 7–14, 14–21, 21–28, and > 28 days after the time of COVID-19 PCR+ test sample collection. *CV* COVID-19 visit, *CRP* C-reactive protein, *DG* diagnosis, *EP* endpoint, *IM* in-house medication, *LB* lab results, *PC* procedures, *VS* vital signs

analysis. Multiple testing correction was not applied to allow for more features to be included.

2.7 Computational Resources for Bayesian Network Training

The deep Bayesian networks for this COVID-19 effort were trained using the Andes supercomputer at the Oak Ridge Leadership Computing Facility. Andes is a 704-node machine with two AMD EPYC 7302 CPUs per node and is primarily focused on large-scale scientific discovery via data processing and modeling.

2.8 Multivariable Regression Analysis for Assessment of Validity of the Effect of Ondansetron on COVID-19-Associated 30-Day All-Cause Mortality

To clarify the relationship between ondansetron use and any cause mortality after a COVID-19 PCR+ test discovered by bAIcis[®], a multivariate regression analysis focused on

inpatients was undertaken. In addition to ondansetron use, potential confounders (e.g., demographics, comorbidities, lab values, and medications) were included to fit a multivariable regression model with the outcome being 30-day any cause mortality following a COVID-19 PCR+ test. Specifically, the following groups, comprising 25 variables, were examined using multi-variable regression for their ability to predict mortality in a COVID-19+ patient cohort: demographics (age [approximate], sex, race, and ethnicity), medications (ondansetron, azithromycin, remdesivir, dexamethasone, tocilizumab, convalescent plasma), comorbidities (diabetes mellitus, chronic obstructive pulmonary disease, asthma, coronary artery disease, heart failure, neoplastic disease, kidney disease), laboratory analytes (C-reactive protein, D-dimer, alanine aminotransferase, ferritin, aspartate transaminase, blood urea nitrogen, lymphocytes), and ventilator status. The de-identified dataset had patient ages structured in bins, such as 18–39, 40–49, ..., 70+ years. For the purpose of regression modeling, these were converted to approximate ages as 35 for the 18–39 bin and 45 for the 40–49 bin. Out of the 3082 inpatients, 2259 had

Table 2 BERG's Bayesian artificial intelligence analytics (bAICis[®]) generated 19 networks that enabled unbiased identification of significant predictors of any cause mortality post-COVID-19 PCR+ testing for specific patient populations

Name	Number of patients	Number of features	Number of features connected to any cause mortality	Number of selected significant predictors of any cause mortality
All patients	16,277	2386	487	20
Inpatients	3082	1946	227	20
Inpatients post-COVID-19+ PCR test	3082	1285	100	20
Inpatients aged 60–69	655	796	32	20
Inpatients aged 70+	1099	1238	32	20
Inpatients aged 60+	1754	1570	77	20
Inpatients Hispanic	1231	1080	93	20
Inpatients white non-Hispanic	898	981	44	20
Inpatients aged < 60	1328	991	30	19
Inpatients pre-COVID-19+ PCR test	3082	678	23	17
Inpatients aged 40–49	391	448	26	10
Inpatients white non-Hispanic pre-COVID-19+ PCR test	898	315	7	7
Inpatients aged 60+ pre-COVID-19+ PCR test	1754	519	9	6
Inpatients Hispanic pre-COVID-19+ PCR test	1231	279	7	4
Inpatients aged 50–59	548	604	7	4
Inpatients age 18–39	389	386	7	3
Inpatients age <60 pre-COVID-19+ PCR test	1328	259	5	3
Inpatients African American Non-Hispanic	600	675	5	3
Inpatients African American Non-Hispanic pre-COVID-19+ PCR test	600	161	16	3

complete observations for the variables indicated above. To increase the statistical power, imputation of missing values was accomplished by a multiple imputation approach using the predictive mean matching method [10], as implemented by the R package mice [11]. In the least-squares regression analysis and LASSO regression analysis, five imputed data sets were generated after 25 iterations of the predictive mean matching (pmm) method; in the bootstrap analysis, ten imputed datasets were generated with 100 iterations of the pmm method, as the basis for each bootstrap sampling.

To study the ondansetron-associated effect of the above variables on any cause 30-day mortality, a logistic regression was performed in a generalized linear model for the binary outcome of deceased or survived, using the R package **glm**. A subsequent analysis focused on fitting the logistic regression model to the 25 clinical variables (described above), the corresponding 300 ($25 \times 24/2$) interaction terms (each equaling the pairwise product of their binary values), and the eight squared terms for the continuous terms, for a total of 333 terms. Because of the inherent limitation of logistic regression in fitting such a model to the available data, we used the LASSO regression method (as implemented by the R package **glmnet** (6), with the parameter *alpha* set to 1) to select those covariates most likely to have non-zero coefficients. The *lambda* parameter was optimized with a cross-validation approach using the *cv.glmnet* function. *Lambda*

was set conservatively to a value that is one standard deviation away from the minimum value determined from the cross-validation approach. In addition, ten versions of the dataset with imputed values were generated, and each of them in turn was used to generate a population of 1000 datasets with similar underlying distributions, by the bootstrap method. We ensured that each bootstrap sample was balanced in terms of the proportion of patient any cause 30-day mortality. The 10,000 samples were each analyzed by LASSO regression. Bootstrapping was performed by sampling with replacement of the dataset, while retaining the proportion of deceased to survived patients at 30 days after a COVID-19 PCR+ test in the bootstrap datasets.

3 Results

3.1 Studied Population Characteristics

The study population for the initial AI analysis comprised 16,277 patients who were COVID-19 PCR+ (Table 1). In this cohort, the majority were female at 8441 (51.9%), 7836 (48.1%) were male, 7787 (47.8%) were white, 6072 (37.3%) were Hispanic, 2417 (14.8%) were black, and 166 (1%) were Asian. This cohort included a subset of individuals who were admitted to the inpatient setting ($n = 3082$)

Table 3 Features with significant relationship to decreased any cause mortality post- COVID-19 PCR+ testing

Network name	Feature name	<i>P</i> -value	OR	95% CI
Inpatients post-COVID-19+ PCR test	DG_> 28d_postCOVID-19 Z20.828 contact with and (suspected) exposure to other viral communicable diseases	2.00E-07	0.16	0.06, 0.37
Inpatients age <60	IM_> 28d_postCOVID-19 ondansetron	0.001	0.079	0.0019, 0.52
Inpatients age <60	IM_upto7d_postCOVID-19 ondansetron	0.03	0.45	0.2, 0.93

CI confidence interval, DG_ diagnosis, IM_ in-house medication, OR odds ratio

and was used for all subsequent analyses characterizing the effect of ondansetron on any cause mortality at 30 days after a COVID-19 PCR+ test. In the inpatient population, the majority were male at 1609 (52.2%), 1473 (47.8%) were female, 1674 (54.3%) were white, 1231 (39.9%) were Hispanic, 649 (21.1%) were black, and 48 (1.6%) were Asian. In the inpatient cohort, 444 (14.4%) were mechanically ventilated and 262 (8.5%) were deceased. The comorbidity with the highest prevalence was kidney disease (21.3%), followed by heart failure (14.7%) and asthma (11.3%).

3.2 Features Associated with Increased COVID-19 Any Cause Mortality

Using data collected from 16,277 patients who were COVID-19+ by a PCR test, we generated 19 networks that enabled unbiased identification of significant predictors of any cause mortality after a SARS-CoV-2 PCR+ test at any time for specific patient populations (Table 2, Table 3 of the ESM). These networks included inpatient cohorts of multiple age groups, races, and ethnicities, before (pre-COVID-19) and after (post-COVID-19) their PCR+ test. The “All patients” Bayesian network includes the full patient cohort of 16,277 and exhibited 2386 features, 487 of which were connected to any cause mortality after a COVID-19 PCR+ test, with 20 being significant predictors of this mortality. Other networks with high numbers of significant features were “Inpatients age 60 to 69”, “Inpatients age 70+”, “Inpatients Hispanic”, and “Inpatients White Non-Hispanic”.

Of the 239 significant any cause mortality post-COVID-19 PCR+ test features identified in specific patient subpopulations, the majority (223/239) were found to be associated with increased any cause mortality at any time (Table 2 of the ESM). Thirteen of the identified significant any cause mortality features had more than two factor values, thus producing contingency tables larger than 2×2 (Table 4 of the ESM).

As expected, being placed on a ventilator or being admitted in the intensive care unit was found to be associated with increased any cause mortality consistently across 16/19 networks (Table 3 of the ESM) and 11/19 networks, respectively. Similarly, the length of stay was found to be associated with any cause mortality (with longer stays associated

with higher mortality) across 6/19 networks (Table 2). As expected, inpatient medications commonly administered in the intensive care unit setting, such as fentanyl, midazolam, and cisatracurium, were also found to be associated with increased any cause mortality across multiple networks.

3.3 Features Associated with Decreased Any Cause Mortality

Three of the 239 significant features were found to be associated with decreased any cause mortality post-COVID-19 PCR+ testing (Table 3), two of which referred to the use of ondansetron in patients aged younger than 60 years (Table 3 of the ESM). Within this patient population, in-hospital ondansetron treatment either early (< 7 days) or late (> 28 days) after COVID-19 PCR+ testing was found to decrease any cause mortality ($p = 0.03$, OR = 0.45, 95% confidence interval [CI] 0.2, 0.93; and $p = 0.001$, OR = 0.079, 95% CI 0.0019, 0.52, respectively) (Fig. 3). The third feature associated with decreased any cause mortality was the use of *International Statistical Classification of Diseases and Related Health Problems*, 10th Revision code Z20.828 (“Contact with and (suspected) exposure to other viral communicable diseases”).

3.4 Multivariate Analysis of Variables Associated with Any Cause Mortality at 30 Days

To account for potential confounders in the association of ondansetron with decreased any cause mortality post-COVID-19 PCR+ testing observed in the full clinical data set, we focused on a smaller subset of 3082 patients who were hospitalized with COVID-19. Together with ondansetron, we included 24 other variables of relevance to COVID-19 outcomes, including demographics, co-morbidities, and treatments received during hospitalization. Results from the logistic regression analysis performed to predict which variables are significantly associated with any cause mortality at 30 days post-COVID-19 PCR+ testing are summarized in Table 4. The 95% CIs for the coefficients obtained from

Table 4 Coefficient means and 95% CIs of logistic regression fitted to five versions of imputed data, and ranked by their *p*-values

Covariate	Coefficient (mean)	Coefficient (95 CI)	<i>P</i> -value
(Intercept)	-7.26	(- 7.31, -7.21)	< 0.001 ^a
Age	0.047	(0.046, 0.048)	< 0.001 ^a
Ferritin	4.93E-05	(4.39e-05, 5.47e-05)	< 0.001 ^a
BUN	0.018	(0.017, 0.018)	< 0.001 ^a
Mechanical_Ventilation	2.83	(2.81, 2.85)	< 0.001 ^a
Ondansetron	-0.63	(- 0.64, -0.62)	0.001 ^a
Ddimer	0.043	(0.036, 0.051)	0.001 ^a
Heart_Failure	0.47	(0.46, 0.49)	0.01 ^a
Remdesivir	0.4	(0.38, 0.42)	0.03 ^a
Neoplastic_Disease	-0.76	(- 0.78, - 0.73)	0.03 ^a
COPD	0.33	(0.32, 0.34)	0.13
Gender_male	0.24	(0.24, 0.25)	0.14
CRP	0.001	(0.001, 0.001)	0.15
Lymphocytes	-0.16	(- 0.16, - 0.15)	0.17
CAD	0.26	(0.25, 0.26)	0.18
Dexamethasone	-0.26	(- 0.27, - 0.25)	0.2
AST	0.002	(0.001, 0.002)	0.22
ALT	-0.002	(- 0.002, -0.002)	0.24
Race_white	-0.17	(-0.19, - 0.15)	0.26
Convalescent_plasma	0.2	(0.19, 0.21)	0.34
Azithromycin	-0.16	(- 0.17, - 0.15)	0.36
Ethnicity_Hispanic	0.068	(0.010, 0.126)	0.42
Diabetes	-0.059	(- 0.068, - 0.051)	0.71
Tocilizumab	-0.037	(- 0.052, - 0.022)	0.82
Asthma	-0.05	(- 0.064, - 0.036)	0.82
Kidney_disease	-0.02	(- 0.033, - 0.006)	0.87

ALT alanine aminotransferase, *AST* aspartate transaminase, *BUN* blood urea nitrogen, *CAD* coronary artery disease, *COPD* chronic obstructive pulmonary disease, *CRP* C-reactive protein

^aDenotes significance; negative sign means a significant association with decreased mortality

the five imputed data sets for the variables cover a relatively narrow range, and the *p*-values derived from each data set are essentially identical as presented in Table 4. We also performed the same analysis without using the data imputation method, on a set of 2259 patients who had complete data, and the results were similar (Table 5 of the ESM). In the dataset with imputed data, we set a logistic regression model on the effect of ondansetron on any cause mortality at 30 days post-COVID-19 PCR+ testing, adjusted for the possible confounding effect of the other 24 variables of interest. The direct effect of ondansetron on any cause 30-day mortality was significant ($p = 0.001$) with a coefficient of -0.63 (95% CI $-0.64, -0.62$). In this analysis, we effectively simplified a possibly complex set of causal interactions involving ondansetron and any cause 30-day mortality into one that assumes only independent effects of these variables on this mortality. A descriptive statistic of the ondansetron-treated vs non-treated cohort is presented in Table 5 of the ESM.

3.5 Ondansetron Use in Conjunction with Mechanical Ventilation is Associated with Decreased Any Cause 30-Day Mortality After Adjusting for Interactions

We used the elastic net regression method, a combination of the LASSO and ridge regression methods [12], to determine if previous findings related to any cause mortality at 30 days of COVID-19 PCR+ testing are maintained after adjusting for potential confounders. Using five versions of the dataset with imputed data resulted in the generation of five models (Table 5). The covariates *Age*, *Heart_Failure*, *D-dimer*, *Ferritin*, *BUN*, and *Mechanical_Ventilation* were identified by LASSO in all five versions of the imputed dataset; all with positive coefficients, indicative that they associated with increased any cause mortality at 30 days. The covariate *Ondansetron* was selected in two of the five datasets, with negative coefficients, again supporting the potential reduction in any cause mortality at 30 days. Least absolute shrinkage and selection operator identified additional covariates

Table 5 Covariates and their coefficients selected by least absolute shrinkage and selection operator regression on *five* versions of the datasets with imputed data

Covariate	Imputed dataset				
	1	2	3	4	5
(Intercept)	-4.70E+00	-4.78E+00	-4.39E+00	-4.77E+00	-5.51E+00
Age	1.62E-02	1.90E-02	1.43E-02	1.89E-02	2.74E-02
Age:BUN	1.94E-04				
Age:Ferritin	2.99E-07				
Age:Mechanical_Ventilation	1.19E-02				
Age:Remdesivir	1.14E-03				
BUN	1.39E-02	1.36E-02	1.26E-02	1.35E-02	1.45E-02
CAD					5.77E-02
CAD:AST	7.32E-05				
CAD:BUN	1.10E-03				
COPD:BUN	1.47E-03				
COPD:Mechanical_Ventilation	1.42E-01				
CRP					8.50E-05
CRP:BUN	6.12E-06				
Ddimer	1.29E-02	3.97E-03	4.67E-03	6.57E-03	2.84E-02
Ferritin	1.56E-05	1.03E-05	7.56E-06	1.23E-05	2.10E-05
Gender:Dimer	1.12E-02				
Gender:Mechanical_Ventilation	2.02E-01				
Heart Failure	1.73E-01	1.01E-01	2.47E-02	1.03E-01	2.35E-01
Heart_Failure:BUN	6.76E-04				
Heart_Failure:CRP	1.06E-03				
Heart_Failure:Dimer	1.21E-02				
Mechanical_Ventilation	2.09E+00	2.53E+00	2.46E+00	2.52E+00	2.63E+00
Ondansetron	-7.97E-02				-1.95E-01
Ondansetron:Mechanical_Ventilation	-8.09E-02				
Remdesivir					3.87E-03

Covariates are listed alphabetically. All covariates are significant

Age approximate age of patient (one of 35, 45,..., 75 years), AST aspartate transaminase level, BUN blood urea nitrogen, CAD coronary artery disease, COPD chronic obstructive pulmonary disease, CRP C-reactive protein, : indicates interaction between two covariates

Negative sign means association with decreased mortality (bold)

that were observed only once out of the five datasets, the interaction term *Ondansetron:Mechanical_Ventilation* (with a negative coefficient) being one of them. These results suggested that the stochastic aspects of the data imputation and model fitting result in minor variability in the composition of the final model.

Table 6 lists the percentage of the 10,000 bootstrap samples for the top covariates selected by LASSO, their median coefficient values, as well as their 95% and 99% CIs. The LASSO regression results on the bootstrap samples show *BUN* and *Mechanical_Ventilation* identified as main (linear) terms, while the remaining covariates are interaction terms. *Age* is a quadratic (squared) term, indicating that any cause mortality at 30 days post-COVID-19 PCR+ testing increases curvilinearly with age. The most frequently

identified covariates are *Mechanical_Ventilation* and *Age*² and both are positively associated with any cause mortality at 30 days. The interaction term *Ondansetron:Mechanical_Ventilation* is identified in 74.4% of the bootstrap sample and is negatively associated with any cause mortality at 30 days. None of the 95% confidence intervals, and except for *COPD:Mechanical_Ventilation*, none of the 99% confidence intervals of the coefficients include zero, suggesting that these coefficients are stable in their sign despite the variability in the sample sets.

From the regression analysis we conducted on the bootstrapped samples of the dataset, the median value of the regression coefficient for mechanically ventilated patients treated with ondansetron was -0.365. This means that when

Table 6 Covariates selected by logistic least absolute shrinkage and selection operator regressions in greater than 50% of 1000 bootstrap permutations of ten versions of the imputed dataset

Name	Percent selected	Coefficient		
		Median	99% CI	95% CI
(Intercept)	100	-5.10E+00	(-6.03, -3.92)	(-6.31, -3.57)
Tocilizumab:Mechanical_Ventilation	56.3	-5.00E-01	(-1.59, -0.0213)	(-1.94, -4.06e-03)
Gender:Neoplastic_Disease	61.7	-4.87E-01	(-1.69, -0.0274)	(-2.1, -4.98e-03)
Ondansetron:Mechanical_Ventilation	74.4	-3.65E-01	(-1.14, -0.0201)	(-1.38, -4.51e-03)
Age:Ferritin	55.3	3.47E-07	(2.60e-09, 1.22e-06)	(2.06e-08, 9.57e-07)
CRP:BUN	68	2.80E-05	(2.03e-07, 1.17e-04)	(1.42e-06, 9.15e-05)
Age^2	99.1	2.68E-04	(3.25e-05, 5.00e-04)	(7.95e-05, 4.42e-04)
Age:Remdesivir	75.9	3.67E-03	(3.75e-05, 1.68e-02)	(2.36e-04, 1.29e-02)
BUN	52.9	7.39E-03	(2.95e-05, 2.23e-02)	(2.27e-04, 1.84e-02)
Gender:Dimer	60.6	2.37E-02	(1.70e-04, 8.38e-02)	(1.09e-03, 6.79e-02)
Gender:Mechanical_Ventilation	61.4	2.30E-01	(1.34e-03, 9.10e-01)	(9.56e-03, 7.19e-01)
COPD:Mechanical_Ventilation	51.3	2.64E-01	(-1.36e-01, 1.25e+00)	(7.95e-03, 9.44e-01)
Ondansetron:Convalescent_plasma	52.9	4.56E-01	(2.94e-03, 1.69e+00)	(2.33e-02, 1.34e+00)
Mechanical_Ventilation	99.8	2.54E+00	(4.73e-01, 3.76e+00)	(1.07e+00, 3.41e+00)

Covariates are sorted by the median value of their coefficients, along with their 95% and 99% CIs

Age approximate age of patient (one of 35, 45, ..., 75 years), BUN blood urea nitrogen, CI confidence interval, COPD chronic obstructive pulmonary disease, CRP C-reactive protein, : indicates interaction between two covariates

Negative median sign means association with decreased mortality (bold)

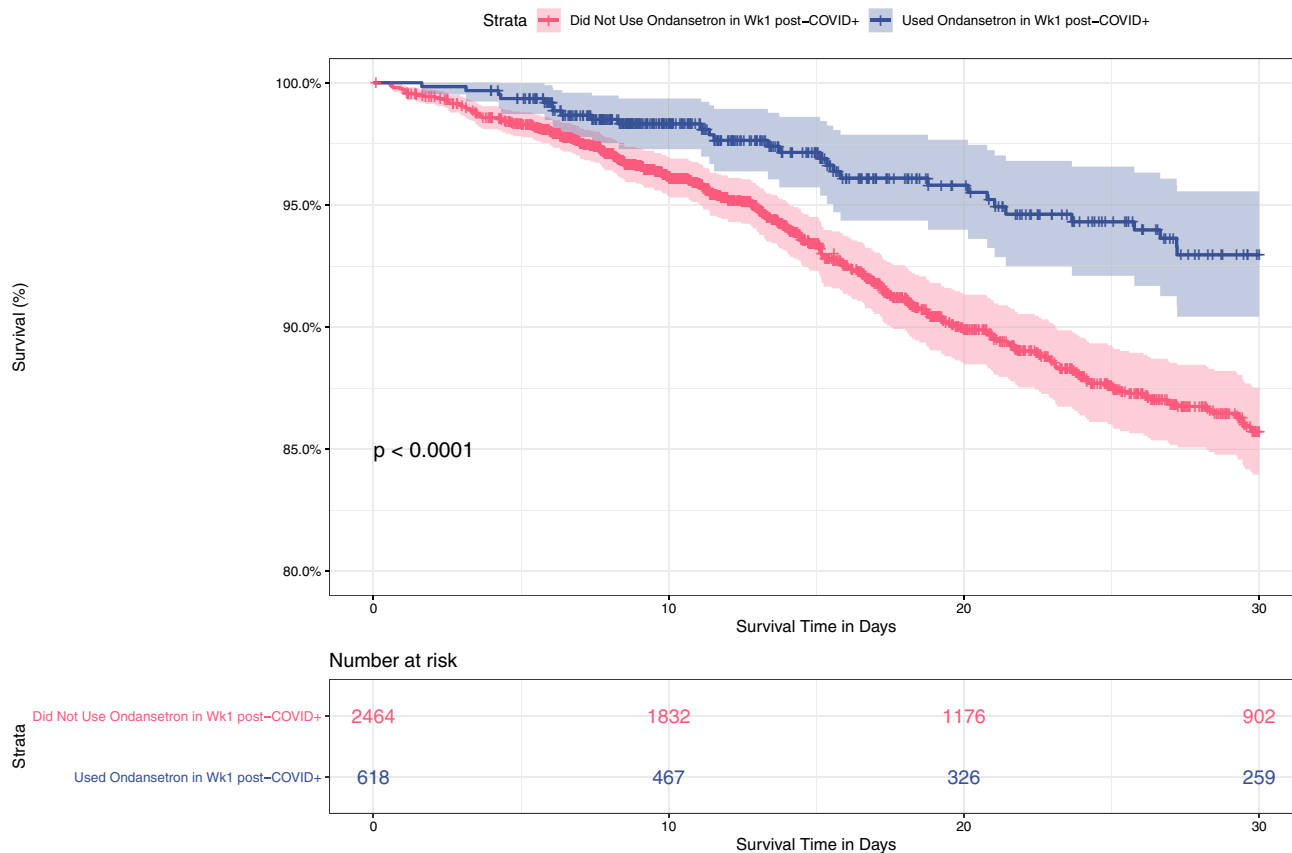


Fig. 4 Kaplan–Meier curve showing 30-day survival rates of hospitalized patients who received (blue) or did not receive (red) ondansetron in the first week (Wk) after a COVID-19 PCR+ test. Patients

who received ondansetron had improved 30-day survival compared with patients who did not ($p < 0.0001$, log-rank test)

this interaction term has a value of 1, the odds of any cause mortality at 30 days post-COVID-19 PCR+ testing are multiplied by $e^{(-0.365)} = 0.694$. We note that the coefficient for the effect of ondansetron on any cause mortality at 30 days is that of its total effect, while the coefficients for all the other variables (including the interaction terms that involve ondansetron) account for their direct effects on any cause 30-day mortality alone, disregarding any possible indirect effects.

3.6 Ondansetron Use Within 7 Days After a COVID-19 PCR+ Test is Associated with Improved 30-Day Survival

Regarding the timing of administration, of the 737 patients who received ondansetron at any time during the first 30 days, the majority (84%) received it within the first week after a COVID-19 PCR+ test. Patients who received ondansetron in the first week after a PCR+ test had improved 30-day survival compared with patients who did not ($p < 0.0001$, log-rank test) (Fig. 4).

4 Discussion

The goal of this study, representing a multi-institutional collaborative effort of collecting, structuring, and analyzing RWD through AI analytics, was to develop a model of COVID-19 any cause mortality-associated factors and identify potential new insights for therapeutic options for patients with COVID-19. The study involved two-stage data analytics: the primary discovery phase, involving a Bayesian statistics-based analysis to generate Bayesian networks of all patients to identify significant factors influencing any cause mortality at any time in the COVID-19+ cohort (Table 2), from a 12-month pre-PCR-based COVID-19 diagnosis to a 28-day post-PCR-based COVID-19 diagnosis; this was followed by an additional analysis incorporating potential confounders with the main findings on imputed and bootstrapped data, from a multivariable regression analysis to the LASSO logistic regression analysis. The Bayesian network findings were based on the analysis of demographic, clinical, and laboratory data from 16,277 patients with PCR-confirmed COVID-19 representing a subset of the 279,281 patients in the RECOVER-19 registry; the multivariate

logistic regression analyses were performed on data from the 3082 hospitalized patients.

The value for the coefficient of ondansetron obtained in the logistic regression analysis (Table 4) is the conditional total effect of ondansetron on the log odds of any cause mortality at 30 days after a COVID-19 PCR+ test, that is, the log OR for the total effect of ondansetron on this mortality at any given level of the 24 other variables that were considered in our model. It is important to note that the values of the other coefficients presented in Tables 4, 5, and 6 make a simplifying assumption that all the covariates other than ondansetron have only direct and independent effects on any cause 30-day mortality. Thus, the values given in these tables for variables other than ondansetron do not reflect their total effect on any cause 30-day mortality, i.e., their direct and indirect effects. Possible indirect effects include the effect of ondansetron on any cause 30-day mortality being mediated by another variable; or one of the variables affecting both ondansetron use and this mortality. The same caveats apply to the results of the more complex models presented in Tables 5 and 6.

We note that mechanical ventilation use is different from the other variables we considered in our multivariate analysis. Because it typically occurs later in time after hospitalization than other treatments, it can be seen as an intermediate variable in predicting any cause mortality at 30 days after a COVID-19 PCR+ test, that is, it could conceivably be an outcome of some of the earlier treatments. However, it is also a potential confounder, in that severity of the disease is likely to be associated with both mechanical ventilation and any cause 30-day mortality. In fact, in our multivariate models the variable for mechanical ventilation has a positive sign, giving the appearance that it is a risk factor of any cause 30-day mortality. This actually reflects the fact that sicker patients are more likely to be administered this treatment. Thus, ventilator use should be seen as a variable that modulates any cause 30-day mortality among patients with severe COVID-19. If we could have controlled for a confounder variable that reflected the severity of the disease, ventilator use would have had a negative association with any cause 30-day mortality. As with the other variables, controlling for ondansetron use did not cause ventilator use to acquire a negative coefficient. We conclude that ondansetron use does not predict disease severity. The beneficial effect of ondansetron is associated with its use during the first week of hospitalization, which is typically earlier than ventilator use, thus we considered the possibility that this effect is mediated by ventilator use. We found no evidence for this scenario, and conclude that ondansetron and ventilator use act independently in their effects on any cause 30-day mortality.

The LASSO regression analysis indicated that *Ondansetron* and *Mechanical_Ventilation* are interacting variables, that is, ondansetron is modulating the effect of mechanical

ventilation, which, as mentioned above, in turn modulates the effect of disease severity. It is possible that the beneficial effects of ondansetron last long enough that by the time mechanical ventilation is applied, the disease has become less lethal and thus ventilation becomes more effective in preventing 30-day mortality.

Our use of bootstrapping to generate variants of our patient population with similar proportions was performed by preserving the proportion of cases and controls (stratified bootstrapping). By forcing this requirement on the bootstrap samples, we are moving away from the desired aim to generate populations similar to what would have been drawn from the general population. However, this method ensures against the bootstrap sample having too few deceased patients (6% of the inpatients had an outcome of any cause 30-day mortality), which would lead to a loss of sensitivity in the characterization of this group. In a comparison of stratified bootstrapping with k-fold cross validation, the former appeared to be better in terms of bias and variance, when compared with regular cross-validation [13].

We identified ondansetron as the main factor associated with decreased any cause 30-day mortality in inpatients with COVID-19 who received mechanical ventilation. An initial unbiased search for predictors of any cause mortality after a COVID-19 PCR+ test at any time and within any patient population found ondansetron as the only medication associated with decreased any cause mortality (Table 3). This association was initially identified within a specific inpatient population (age < 60 years) and when ondansetron was administered at disparate times (up to 7 days and > 28 days post-COVID-19 diagnosis). To better quantify the relationship between ondansetron use and any cause mortality at the 30-day post-COVID-19 PCR+ test, multivariable logistic regression by LASSO showed significant effects for age and ondansetron use on this mortality. We found that the effect of ondansetron on any cause 30-day mortality is not mediated through the use of mechanical ventilation because ondansetron use does not predict mechanical ventilator use. In addition, the absence of a significant *Age:Ondansetron* interaction effect suggested that the effect of ondansetron applies to all age groups equally. However, it is the interaction term *Ondansetron:Mechanical_Ventilation* that is primarily selected by LASSO as a covariate of a non-zero coefficient, rather than the main term *Ondansetron*. This suggests that the beneficial effect of ondansetron is seen only in patients who received mechanical ventilation. This key finding complements a study by Bayat et al. that reported a reduction in 30-day all-cause mortality for all inpatients (including intensive care unit) with early administration of ondansetron after admission [14]. Further validation in an independent cohort may clarify the interaction between ondansetron use, mechanical ventilation, and age.

Ondansetron is a selective 5-HT₃ serotonin receptor antagonist used to prevent or treat nausea and vomiting through both central and peripheral mechanisms [15]. Ondansetron is not regularly used in intubated patients. In our cohort of ventilated patients who received ondansetron, 54% received it before ventilation and 46% received it after ventilation. This suggests that ondansetron is not a marker of surviving mechanical ventilation and it might have an effect on any cause mortality at different stages of disease progression.

It has been postulated that SARS-CoV-2 might have an indirect effect on enteroendocrine cells, triggering the release of neuroactive agents such as emesis-inducing serotonin [16]. Most studies showed that patients with COVID-19 have higher plasma serotonin levels and this correlates with increased interleukin-6 [17, 18], while others concluded they have decreased serotonin levels [19]. Considering the role of serotonin in regulating innate and adaptive immune responses [20], the observed beneficial effect of ondansetron might be due to the modulation of serotonin levels or could also be linked with a direct effect on the immune system [21] or on known COVID-19 comorbidities, such as liver and kidney disease or complications such as thrombosis [22–24]. There are also data suggesting that serotonin receptor signaling influences cellular activities that regulate the entry of diverse virus families [25].

It is interesting to note that while we do not find convalescent plasma to be a significant predictor of any cause mortality at the 30-day post-COVID-19 PCR+ test, patients receiving ondansetron and convalescent plasma were more likely to die (Table 6), suggesting a complex interaction between ondansetron, ventilator use, and convalescent plasma. It is now known that high-titer convalescent plasma does not improve COVID-19 survival or clinical outcomes when used in both inpatients [26, 27] and high-risk outpatients [28] and when a beneficial effect on the risk of death was observed it was not maintained for patients who had received mechanical ventilation [29]. As early in the pandemic convalescent plasma was usually reserved for patients with more severe COVID-19 pneumonia, this observed association might be explained by this confounding bias and not by a potential detrimental effect of convalescent plasma. In our cohort, tocilizumab had a similar effect on any cause mortality at 30 days after a COVID-19 PCR+ test to ondansetron for mechanically ventilated patients, in line with published evidence from large randomized controlled studies [30, 31].

Although being male is associated with COVID-19 mortality, we find that in the context of neoplastic disease this is reversed. Our results show a negative association between the interaction term *Gender:Neoplastic_Disease* and any cause mortality at the 30-day post-COVID-19 PCR+ test. This may be because of an indirect effect of ondansetron, as this is often prescribed to patients with cancer undergoing

chemotherapy, radiation therapy, and surgery. Thus, while being male is positively associated with any cause mortality at 30 days, in patients with cancer, this may be modulated by ondansetron use. Additionally, the association of COVID-19 mortality with cancer is not straightforward. The COVID-19 mortality of patients with cancer depends on the type of their cancer, with the main mortality drivers being age, sex, comorbidities, and hematological cancers [32–34].

In the present study, in addition to ondansetron and tocilizumab, other covariates interacting with ventilator use indicate that male individuals on a ventilator and patients with chronic obstructive pulmonary disease on a ventilator are more likely to have mortality from any cause at 30-days after a COVID-19 PCR+ test. The former of these two findings agrees with Nicholson et al., who showed that male patients with COVID-19 on a ventilator have a higher mortality rate than female patients (after correcting for comorbidities) [35]. Chronic obstructive pulmonary disease is also an already established comorbidity associated with an increased odds of hospitalization and death in patients with COVID-19 [36, 37].

Looking at interactions between other covariates, we found that the C-reactive protein and blood urea nitrogen dyad, laboratory biomarkers that are found in prognostic models for COVID-19 mortality, are also associated with any cause 30-day mortality after a COVID-19 PCR+ test in our cohort [38, 39]. Similarly, the previously observed mortality link of interacting factors ferritin and age [40] as well as higher D-dimer levels in male individuals were also confirmed by our analyses [41]. The association between any cause 30-day mortality with a combination of age and age-squared agrees with a previous finding that the infection fatality ratio has a log-linear increase by age among individuals aged older than 30 years [42].

Although a US Food and Drug Administration-approved drug for COVID-19, remdesivir was not found to increase survival in large randomized controlled trials [43–45]. We find that age and remdesivir use interact to increase any cause 30-day mortality. This has not been reported previously and may suggest that we see a similar confounding bias with convalescent plasma, as remdesivir was reserved for patients with more severe disease earlier on.

Diagnostic code Z20.828 (“Contact with and (suspected) exposure to other viral communicable diseases”) was one of the three features with a significant relationship with decreased any cause mortality at any time following a COVID-19 PCR+ test in our analysis. This code was used in 2020 when a clinician suspected exposure to SARS-CoV-2 without a test result available. In the RECOVER-19 registry, out of the approximately 200,000 unique patients seen with this diagnostic code in 2020, about 16,000 were found to be positive. We might see this mortality benefit because the patients who were COVID-19+ with the Z20.828 code

might have had a less severe form of COVID-19 (higher ambiguity without a positive test) or arrived earlier in the course of the disease and were designated patients under investigation, benefiting from early precautions. Generally, the patients admitted with a severe form of COVID-19 would have received another more definitive diagnostic code.

The results described in this work have limitations that should be acknowledged. First, the additional analyses incorporating clinical variables of interest used to quantify the relationship between ondansetron and any cause mortality found in our initial bAIcis[®] work were performed on a subset of the same patient cohort, albeit by different methodology (logistic regression and LASSO regression) and using data imputation methods. A more robust validation should be performed next, using an independent data set that had not been used during the initial discovery work.

Another limitation is related to the simplifying assumptions of the prediction model developed using the regression models. As stated previously, the models assume that the variables (or their interaction terms) other than ondansetron are only directly affecting any cause 30-day mortality and there are no indirect effects such as mediation or confounding effects. Therefore, the predictive models presented here should only be viewed as a first approximation of the likely complex set of causal interactions between ondansetron use and any cause mortality at a 30-day post-COVID-19 PCR+ test.

5 Conclusions

To our knowledge, this is the first use of a Bayesian network analysis of clinical data to report disease outcomes in patients with COVID-19. Using high-performance computer-driven Bayesian AI, we report here a negative association between any cause mortality at 30 days after a COVID-19 PCR+ test and ondansetron treatment for mechanically ventilated patients, as well as confirming the beneficial effects of tocilizumab and validating some of the already established factors associated with COVID-19 increased mortality, such as higher blood urea nitrogen, C-reactive protein, ferritin, and D-dimer levels. These results suggest that the bAIcis[®] platform can be used to generate hypotheses from RWD. Currently, there are no controlled trials examining the effect of ondansetron in patients with COVID-19. Our findings suggest that this Food and Drug Administration-approved drug should be investigated for its potential effectiveness against COVID-19.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s40801-022-00303-9>.

Acknowledgments We thank Lance McCain for his leadership and support in establishing the AdventHealth RECOVER-19 registry.

Declarations

Funding This research used resources of the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725. Research support was also provided by the AdventHealth Foundation.

Conflicts of interest/Competing interests GMM, RS, LOR, CB, NMC, KR, MAK, and NRN are employees of BERG, LLC and received support from BERG, LLC for the submitted work. GMM, RS, LOR, CB, NMC, KR, MAK, EG, and NRN have stock or stock options in BERG, LLC. GMM has received support for attending meetings and/or travel from BERG, LLC and patents planned, issued, or pending by BERG, LLC. LOR has other financial or non-financial interests in BERG, LLC. EG has a leadership or fiduciary role on the Cigna Board, BERG, LLC Board, Cerner Electronic Medical Record, and Defense Logistics Health Board. EG has stock or stock options in BERG, LLC, Cerner, and Defense Logistics Health, and Cigna. NRN is a co-founder of BERG, LLC. No other relationships or activities have influenced the submitted work.

Ethics approval This study has been approved by the AdventHealth Institutional Review Board (IRBnet #1590483).

Consent to participate A waiver of consent was granted by the AdventHealth Institutional Review Board.

Consent for publication Not applicable.

Availability of data and material Available upon request.

Code availability Not applicable.

Authors' contributions Concept and manuscript review (GM, DL, RS, NN, AP, JS), wrote the manuscript (GM, DL, CB, RS), network generation (AE, JH), data analysis (CB, GM, NMC, KR, LR, JH, DL), data architecture, annotation, and cleaning (RV, CH, JH), and manuscript review (EG, SRS, BG). All authors read and approved the final manuscript.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, which permits any non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

References

1. Johns Hopkins University (JHU), COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns

- Hopkins University (JHU). Baltimore, MD, United States, 2022. <https://coronavirus.jhu.edu/map.html>
2. DeGrave AJ, Janizek JD, Lee S-I. AI for radiographic COVID-19 detection selects shortcuts over signal. *Nat Mach Intell*. 2021;3(7):610–9.
 3. Xu W, Sun NN, Gao HN, Chen ZY, Yang Y, Ju B, et al. Risk factors analysis of COVID-19 patients with ARDS and prediction based on machine learning. *Sci Rep*. 2021;11(1):2933.
 4. Tariq A, Celi LA, Newsome JM, Purkayastha S, Bhatia NK, Trivedi H, et al. Patient-specific COVID-19 resource utilization prediction using fusion AI model. *NPJ Digit Med*. 2021;4(1):94.
 5. Health TLD. Artificial intelligence for COVID-19: saviour or saboteur? *Lancet Digit Health*. 2021;3(1): e1.
 6. Yu J, Smith VA, Wang PP, Hartemink AJ, Jarvis ED. Advances to Bayesian network inference for generating causal networks from observational biological data. *Bioinformatics*. 2004;20(18):3594–603.
 7. Zhang L, Rodrigues LO, Narain NR, Akmaev VR. bAICs: a novel Bayesian network structural learning algorithm and its comprehensive performance evaluation against open-source software. *J Comput Biol*. 2020;27(5):698–708.
 8. Becker's Healthcare, 100 of the largest hospitals and health systems in America 2020. United States, 2020. <https://www.beckershospitalreview.com/lists/100-of-the-largest-hospitals-and-health-systems-in-america-2020.html>
 9. Healthcare Cost and Utilization Project Clinical Classifications Software Refined (CCSR). Rockville, MD, United States, 2021. https://www.hcup-us.ahrq.gov/toolssoftware/ccsr/ccs_refined.jsp
 10. Rubin DB. Statistical matching using file concatenation with adjusted weights and multiple imputations. *J Business Econ Stat*. 1986;4(1):87–94.
 11. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. *J Stat Soft*. 2011;45(3):1–67. <https://doi.org/10.18637/jss.v045.i03>.
 12. Zou H, Hastie T. Regularization and variable selection via the elastic net. *J Royal Stat Soc Ser B (Stat Methodol)*. 2005;67(2):301–20.
 13. Kohavi R. A study of cross-validation and Bootstrap for accuracy estimation and model selection. Morgan Kaufmann Publishers Inc., Montreal, Canada, 1995.
 14. Bayat V, Ryono R, Phelps S, Geis E, Sedghi F, Etminani P, et al. Reduced mortality with ondansetron use in SARS-CoV-2-infected inpatients. *Open Forum Infect Dis*. 2021;8(7):ofa336.
 15. Griddine A, Bush JS. Ondansetron. [Updated 2021 Sep 29]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2022 Jan-. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK499839/>
 16. Andrews PLR, Cai W, Rudd JA, Sanger GJ. COVID-19, nausea, and vomiting. *J Gastroenterol Hepatol*. 2021;36(3):646–56.
 17. Ha S, Jin B, Clemmensen B, Park P, Mahboob S, et al. Serotonin is elevated in COVID-19-associated diarrhoea. *Gut*. 2021;70(10):2015–7.
 18. Zaid Y, Puhm F, Allaey I, Naya A, Oudghiri M, Khalki L, Limami Y, Zaid N, Sadki K, Haj RBE, Mahir W, Belayachi L, Belefquih B, Benouda A, Cheikh A, Langlois M-A, Cherrah Y, Flamand L, Guessous F, Boilard E. Platelets Can Associate With SARS-CoV-2 RNA and Are Hyperactivated in COVID-19. *Circ Res*. 2020;127(11):1404–18.
 19. Soria-Castro R, Meneses-Preza YG, Rodríguez-López GM, Romero-Ramírez S, Sosa-Hernández VA, Cervantes-Díaz R, Pérez-Fragoso A, Torres-Ruiz JJ, Gómez-Martín D, Campillo-Navarro M, Álvarez-Jiménez VD, Pérez-Tapia SM, Chávez-Blanco AD, Estrada-Parra S, Maravillas-Montero JL, Chacón-Salinas R. Severe COVID-19 is marked by dysregulated serum levels of carboxypeptidase A3 and serotonin. *J Leukocyte Biol*. 2021;110(3):425–31.
 20. Hamed MGM, Hagag RS. The possible immunoregulatory and anti-inflammatory effects of selective serotonin reuptake inhibitors in coronavirus disease patients. *Med Hypotheses*. 2020;144: 110140.
 21. Magrini E, Szabo I, Doni A, Cibella J, Viola A. Serotonin-mediated tuning of human helper T cell responsiveness to the chemokine CXCL12. *PLoS ONE*. 2011;6(8): e22482.
 22. Chen Y, Sun J, Fan X, Wang X, Zeng L, Zhang X, et al. Association of 5-hydroxytryptamine 3 receptor antagonists with the prognosis of liver failure. *Front Pharmacol*. 2021;12: 648736.
 23. Liu FC, Liou JT, Liao HR, Mao CC, Yang P, Day YJ. The anti-aggregation effects of ondansetron on platelets involve IP3 signaling and MAP kinase pathway, but not 5-HT3-dependent pathway. *Thromb Res*. 2012;130(3):e84–94.
 24. Datta A, Matlock MK, Le Dang N, Moulin T, Woeltje KF, Yanik EL, et al. “Black box” to “conversational” machine learning: ondansetron reduces risk of hospital-acquired venous thromboembolism. *IEEE J Biomed Health Inform*. 2021;25(6):2204–14.
 25. Mainou BA, Ashbrook AW, Smith EC, Dorset DC, Denison MR, Dermody TS. Serotonin receptor agonist 5-nonyloxytryptamine alters the kinetics of reovirus cell entry. *J Virol*. 2015;89(17):8701–12.
 26. Group RC. Convalescent plasma in patients admitted to hospital with COVID-19 (RECOVERY): a randomised controlled, open-label, platform trial. *Lancet*. 2021;397(10289):2049–59.
 27. Bégin, P., Callum, J., Jamula, E., Cook, R., Heddle, N. M., Tinnmouth, A., Zeller, M. P., Beaudoin-Bussières, G., Amorim, L., Bazin, R., Loftsgard, K. C., Carl, R., Chassé, M., Cushing, M. M., Daneman, N., Devine, D. V., Dumaresq, J., Fergusson, D. A., Gabe, C., Glesby, M. J., Li, N., Liu, Y., McGeer, A., Robitaille, N., Sachais, B. S., Scales, D. C., Schwartz, L., Shehata, N., Turgeon, A. F., Wood, H., Zarychanski, R., Finzi, A., Marceau, D., Huang, A., Carr, H., Lin, Y., Lall, R., Graham, C., Arsenault, C., Sales, V., Sidhu, D., Semret, M., Hamm, C., Arhanchiague, E., Solh, Z., Srouf, N., Soliman, K., Yee, C., Laroche, V., Nahirniak, S., Greenaway, C., Pai, M., Côté, A., Tsang, J. L. Y., Cserti-Gazdewich, C., Talbot, D., Poulin, S., Guimaraes, R., Rushton-Marovac, M., Langlois, A., Ning, S., Shih, A., Boileau, M., Singh, H., Ledingham, D., Ponnampalam, A., Yan, M., Prokopchuk-Gauk, O., Poirier, A., Girouard, G., Pavenski, K., Drouin, O., Harris, D., Durand, M., Rimmer, E., Ovakim, D., Ménard, F., Cuccarolo, G., Carruthers, J., Lucier, K., Arsenault, V., Auclair, M.-C., Avram, M., Brassard, M., Cerro, S., Martinez, V., Morin, J., Saint-Jacques, M., Veillette, M., Armali, C., Kron, A., Modi, D., Duncan, J., Justumus, P., John, M. S., St-Onge, G., Hadzi-Tosev, M., Dion, P.-M., McGillivray, L., de Moulliac, A. V., Nyman, S. A., Perilli, S., Van Vliet, P. J., Lane, S., Pavenski, K., Pereira, R., Siroitch, E., Abelson, J., Greene, S., Khandelwal, A., Thakar, S., Longo, S., Anand, S. P., Benlarbi, M., Bourassa, C., Boutin, M., Descôteaux-Dinelle, J., Gendron-Lepage, G., Goyette, G., Laumaea, A., Medjahed, H., Prévost, J., Richard, J., Kaufmann, D., Brunet-Ratnasingham, E., Chaumont, N., Drebot, M., Robinson, A., Mendoza, E., Dimitrova, K., Manguiat, K., Phillipson, C., Chan, M., Evans, D., Lin, J., Boyer, L., Cloutier, M., Drouin, M., Ducas, É., Dussault, N., Fournier, M.-J., Landy, P., Nolin, M.-È., Perreault, J., Tremblay, T., Nazy, I., Xie, F., Liu, D., Wong, M., Silverio, G., Walkus, K., Barton, M., Haveman, K., Mueller, D., Scott, A., Moher, M., Wood, G., Roarty, T., Auld, F., Carney, G., Thomson, V., Onell, R., Walley, K., Donohoe, K., Brunk, C., Hernandez, G., Jacobucci, T., Lazosky, L., Mann, P., Raval, G., Zampieri, L. A., Sekhon, M., Wright, A., James, N., Chang, G., Chen, R., Deol, K., Gantioqui, J., Larsen, E., Ramdin, N., Roche, M., Rosinski, K., Sham, L., Storms, M., Gillrie, M., Mahe, E., Suryanarayan, D., Ugarte-Torres, A., Robinson, T., Gibbs, M., Hewsgirard, J., Holmes, M., McCarthy, J., Ody, M., Doucette, K., Sligl, W., Sonpar, A., Robertson, K., Narayan, J., Ravindran,

- L., Stewart, B., Zapernick, L., Lee, S., Sy, E., Wong, A., Gryzb, K., Craddock, S., Fuchs, D., Myrah, D., Sunny, S., Harding, S. R., Kogilwaimath, S., Hodgson, N., Johnson, D., Meier, S., Thomson, K., Heendeniya, A., Houston, B., Kenyan, Y., Lothar, S., Olafson, K., Rush, B., Wuerz, T., Solvason, D., Albensi, L., Alias, S., Choi, N., Curtis, L., Hutmacher, M., Kashani, H., Lane, D., Marten, N., Pronyk-Ward, T., Rigaux, L., Silva, R., Tays, Q., Naidu, R., Mathews, J., Mai, M., Miceli, V., Molson, L., Radhakrishnan, G., Schaefer, L., Haddad, M., Landry, S., Chernish, R., Krui-selbrink, R., Liu, T., Jeromin, J., Siddiqui, A., Girolametto, C., Krokoszynski, K., Main, C., Fox-Robichaud, A., Rochweg, B., Kruja, E., Ellingham, D., Sampat, D., Tang, N., Leto, D., Karunakaran, M., Ricciuto, D., Fusco, K., Ghate, T., Robinson, H., Ball, I., Shalhoub, S., Slessarev, M., Silverman, M., Nano, E., Bentall, T., Campbell, E., Kinney, J., Parvathy, S., Fera, E., Delfa, A. L., Nadarajah, J., Solow, H., Mendoza, E., Engel, K., Monaco, D., Kononow, L., Suntharalingam, S., Fralick, M., Munshi, L., Saeed, S., Hajjaj, O., Hsu, E., Ali, K., Duan, E., Farjou, G., Jenson, L., Salib, M., Patterson, L., Anant, S., Ding, J., Jomy, J., Das, P., Geagea, A. and the C.-S. G. Convalescent plasma for hospitalized patients with COVID-19: an open-label, randomized controlled trial. *Nature Medicine*, 27, 11 (2021/11/01 2021), 2012–2024.
28. Korley, F. K., Durkalski-Mauldin, V., Yeatts, S. D., Schulman, K., Davenport, R. D., Dumont, L. J., El Kassab, N., Foster, L. D., Hah, J. M., Jaiswal, S., Kaplan, A., Lowell, E., McDyer, J. F., Quinn, J., Triulzi, D. J., Van Huysen, C., Stevenson, V. L. W., Yadav, K., Jones, C. W., Kea, B., Burnett, A., Reynolds, J. C., Greineder, C. F., Haas, N. L., Beiser, D. G., Silbergleit, R., Barsan, W. and Callaway, C. W. Early Convalescent Plasma for High-Risk Outpatients with Covid-19. *New England Journal of Medicine*, 385, 21 (2021), 1951–1960.
 29. Joyner MJ, Carter RE, Senefeld JW, Klassen SA, Mills JR, Johnson PW, et al. Convalescent plasma antibody levels and the risk of death from Covid-19. *N Engl J Med*. 2021;384(11):1015–27.
 30. Group RC. Tocilizumab in patients admitted to hospital with COVID-19 (RECOVERY): a randomised, controlled, open-label, platform trial. *Lancet*. 2021;397(10285):1637–45.
 31. Mariette X, Hermine O, Tharaux P-L, Resche-Rigon M, Steg PG, Porcher R, Ravaud P. Effectiveness of Tocilizumab in Patients Hospitalized With COVID-19: A Follow-up of the CORIMUNO-TOCI-1 Randomized Clinical Trial. *JAMA Int Med*. 2021;181(9):1241–3.
 32. Lee LYW, Cazier JB, Starkey T, Briggs SEW, Arnold R, Bisht V, et al. COVID-19 prevalence and mortality in patients with cancer and the effect of primary tumour subtype and patient demographics: a prospective cohort study. *Lancet Oncol*. 2020;21(10):1309–16.
 33. Lee LY, Cazier JB, Angelis V, Arnold R, Bisht V, Campton NA, et al. COVID-19 mortality in patients with cancer on chemotherapy or other anticancer treatments: a prospective cohort study. *Lancet*. 2020;395(10241):1919–26.
 34. Kuderer NM, Choueiri TK, Shah DP, Shyr Y, Rubinstein SM, Rivera DR, et al. Clinical impact of COVID-19 on patients with cancer (CCC19): a cohort study. *Lancet*. 2020;395(10241):1907–18.
 35. Nicholson CJ, Wooster L, Sigurslid HH, Li RH, Jiang W, Tian W, Lino Cardenas CL, Malhotra R. Estimating risk of mechanical ventilation and in-hospital mortality among adult COVID-19 patients admitted to Mass General Brigham: The VICE and DICE scores. *EClinicalMedicine*. 2021;33:100765.
 36. Lee SC, Son KJ, Han CH, Park SC, Jung JY. Impact of COPD on COVID-19 prognosis: a nationwide population-based study in South Korea. *Sci Rep*. 2021;11(1):3735.
 37. Gerayeli FV, Milne S, Cheung C, Li X, Yang C, Tam WT, et al. COPD and the risk of poor outcomes in COVID-19: a systematic review and meta-analysis. *EClinicalMedicine*. 2021;33: 100789.
 38. Altschul DJ, Unda SR, Benton J, de la Garza RR, Cezayirli P, Mehler M, et al. A novel severity score to predict inpatient mortality in COVID-19 patients. *Sci Rep*. 2020;10(1):16726.
 39. Booth AL, Abels E, McCaffrey P. Development of a prognostic model for mortality in COVID-19 infection using machine learning. *Mod Pathol*. 2021;34(3):522–31.
 40. Rasyid H, Sangkereng A, Harjianti T, Soetjipto AS. Impact of age to ferritin and neutrophil-lymphocyte ratio as biomarkers for intensive care requirement and mortality risk in COVID-19 patients in Makassar, Indonesia. *Physiol Rep*. 2021;9(10): e14876.
 41. He X, Yao F, Chen J, Wang Y, Fang X, Lin X, et al. The poor prognosis and influencing factors of high D-dimer levels for COVID-19 patients. *Sci Rep*. 2021;11(1):1830.
 42. O'Driscoll M, Ribeiro Dos Santos G, Wang L, Cummings DAT, Azman AS, Paireau J, et al. Age-specific mortality and immunity patterns of SARS-CoV-2. *Nature*. 2021;590(7844):140–5.
 43. Beigel JH, Tomashek KM, Dodd LE, Mehta AK, Zingman BS, Kalil AC, et al. Remdesivir for the treatment of Covid-19: final report. *N Engl J Med*. 2020;383(19):1813–26.
 44. WHO Solidarity Trial Consortium Repurposed Antiviral Drugs for Covid-19 — Interim WHO Solidarity Trial Results. *New Engl J Med*. 2020;384(6):497–511.
 45. Wang Y, Zhang D, Du G, Du R, Zhao J, Jin Y, et al. Remdesivir in adults with severe COVID-19: a randomised, double-blind, placebo-controlled, multicentre trial. *Lancet*. 2020;395(10236):1569–78.