COVID19 outcome prediction: differences between raw laboratory and human pre-processed data

**Authors:** Silvia Gómez-Zorrilla1,2,3, Inmaculada López-Montesinos1,2,3, Eva Pérez4, Pau Pomés5, Maria Luisa Sorli1,2,3, Robert Güerri-Fernández 1,2,3, Santiago Grau6, Albert Márquez7, Jordi Martínez-Roldán8, Alessandro Principe2,4,5\*, Juan Pablo Horcajada1,2,3

**Affiliations:**

1 Infectious Diseases Department, Hospital del Mar, Infectious Pathology and Antimicrobials Research Group (IPAR)

2 Institut Hospital del Mar d’Investigacions Mèdiques (IMIM), Universitat Autònoma de Barcelona (UAB)

3 CEXS-Universitat Pompeu Fabra, Barcelona, Spain

4 Epilepsy Unit, Neurology Department, Hospital del Mar

5 Bioenginery-Universitat Pompeu Fabra (UPF), Barcelona, Spain

6 Pharmacy Department, Infectious Pathology and Antimicrobials Research Group (IPAR)

7 Information and Comunication System department, Hospital el Mar, Barcelona, Spain.

8 Innovation and Digital Transformation Manager, Hospital del Mar, Barcelona, Spain.

\*Corresponding author. Email: alessandro.principe@upf.edu.

Supplementary Text:

# Results

Towards a clinical trial

Since the first wave, the clinical follow-up of COVID19 patients changed and most of subjects with mild symptoms are now monitored from home. Home hospitalization can provide patients with oxymeters and even low-grade ventilation devices. For this reason, we tried to build early discharge models trying to predict the first grade of oxygen support intensification, which is ventilation beyond nasal cannulae. Only confirmed SARS-CoV2 patients were used. Moreover, instead of 6 RDF models, 10 were employed in order to provide a human readable percentage increase of deterioration risk, since each model would account for a 10% of the risk increase. The anticipation of EHR and LR models were of 3±10 and 5±21 days, respectively, the sensibility was 100% in both cases. However, the specificity of EHR models overcame that of LR models (96.1% vs. 95.2%), even though EHR models accounted for more subjects. ROC-AUCs were 98.0% and 97.6%, respectively, the same as when using the previous mecanical ventilation threshold. When predicting invasive mechanical ventilation and ICU translation with EHR, changing the number of RDFs to 10 and considering also non laboratory crosschecked SARS-CoV2 patients (2023 vs. 1685) did not significanlty change the ROC-AUC, which reached 90.8%, with 91.9% and 89.7% sensitivity and specificity, and anticipation rates of 5±13 days. The new repository is available at: [joricomico/COVID19\_trial: bundle to run clinical trials with COP applied on COVID19 (github.com)](https://github.com/joricomico/COVID19_trial.git).

Variable change through time and models (laboratory results)

The most important difference between LR and EHR sources is the number of variables and therefore their average weight in models, which was 3.2% for EHR and 0.8% for LR. Using EHR, 9 different variables out of 31 (29%) overcame the 3.2% threshold for the prediction of all goals, ED, OSi and ICU. Using LR on the other hand, 48 (40%) for ED, 50 (42%) for OSi and 53 (44%) for ICU overcame the 0.8% threshold. Therefore, more LR variables were needed to reach all objectives. Despite that, LR top variables were 5 times on average above their average weight, while EHR top variables were only 4 times above their threshold. These differences suggest that some LR key variables account for higher amounts of predictive power than EHR variables, which are more uniformly distributed. Interestingly, *Interleukin-6* and *Age* ranked similarly across sources, while *D-dimer* C-reactive protein ranked higher in EHR models. Moreover, white blood cell formula values (e.g., Neutrophils %) ranked higher than absolute values.

|  |  |  |
| --- | --- | --- |
| **Early Discharge** | **Intensification of OS** | **ICU admission** |
| **variable** | **weight** | **variable** | **weight** | **variable** | **weight** |
| Total CO2 (arterial) | 4.0%±0.6% | Age  | 4.1%±0.6% | Age  | 3.3%±0.7% |
| paCO2  | 3.9%±0.7% | paCO2 | 4.0%±0.6% | paCO2 | 3.2%±0.2% |
| Bicarbonate (arterial) | 3.9%±0.3% | Total CO2 (arterial) | 3.1%±0.4% | Interleukin-6  | 2.8%±0.6% |
| Age  | 3.5%±0.2% | Bicarbonate (arterial) | 2.9%±0.4% | Total CO2 (arterial) | 2.6%±0.3% |
| paO2  | 3.2%±0.6% | paO2  | 2.7%±0.5% | Urea | 2.4%±0.7% |
| pH (arterial) | 3.2%±0.3% | pH (arterial) | 2.6%±0.7% | Bicarbonate (arterial) | 2.4%±0.3% |
| Sat O2 (arterial) | 3.1%±0.3% | Sat O2 (arterial) | 2.2%±0.5% | BNP  | 2.2%±0.5% |
| Base Excess (arterial) | 2.0%±0.2% | Interleukin-6 | 2.2%±0.6% | Prealbumin | 2.2%±1.3% |
| D-Dimer  | 1.6%±0.2% | Neutrophils (%) | 1.9%±0.4% | D-Dimer | 2.1%±0.6% |
| BNP | 1.5%±0.4% | Lymphocytes (%) | 1.9%±0.7% | paO2  | 2.1%±0.3% |
| Days | 1.5%±0.1% | Urea  | 1.6%±0.6% | Sat O2 (arterial)  | 2.0%±0.2% |
| Urea | 1.4%±0.2% | BNP | 1.6%±0.6% | Magnesium | 1.9%±1.2% |
| Interleukin-6  | 1.4%±0.2% | Base Excess (arterial) | 1.6%±0.3% | Base Excess (arterial)  | 1.7%±0.3% |
| C-reactive protein | 1.3%±0.2% | RDW-SD | 1.6%±0.2% | LDH  | 1.5%±0.1% |
| Creatinine  | 1.3%±0.1% | Days | 1.5%±0.2% | pH (arterial) | 1.4%±0.2% |
| Neutrophils (%)  | 1.3%±0.2% | Calcium  | 1.5%±0.3% | RDW-CV | 1.4%±0.4% |
| Procalcitonin | 1.2%±0.2% | LDH  | 1.5%±0.4% | ESR | 1.4%±0.5% |
| Legionella urinary antigen test | 1.2%±0.3% | Procalcitonin  | 1.5%±0.6% | Neutrophils (%) | 1.4%±0.4% |
| Troponin T  | 1.2%±0.1% | D-Dimer  | 1.4%±0.5% | C-reactive protein | 1.3%±0.2% |
| LDH  | 1.1%±0.1% | Lymphocytes (103/µL) | 1.3%±0.2% | Days | 1.3%±0.2% |
| Base Excess (venous) | 1.1%±0.1% | Lymphocytes T CD4 (%) | 1.3%±0.4% | RDW-SD | 1.3%±0.3% |
| Calcium  | 1.1%±0.1% | Lymphocytes T CD4 (103/µL) | 1.3%±0.3% | Procalcitonin | 1.3%±0.2% |
| Lymphocytes (%) | 1.1%±0.1% | C-reactive protein | 1.3%±0.2% | Glucose  | 1.3%±0.3% |
| ALT  | 1.0%±0.1% | GGT  | 1.2%±0.4% | Monocytes (%)  | 1.2%±0.2% |
| Bicarbonate (venous) | 1.0%±0.1% | Neutrophils (103/µL) | 1.2%±0.2% | Creatine kinase | 1.2%±0.1% |
| ALP  | 1.0%±0.1% | Lymphocytes T CD8 (103/µL) | 1.1%±0.4% | Calcium | 1.2%±0.2% |
| RDW-SD | 1.0%±0.1% | ALP  | 1.1%±0.1% | Lymphocytes (%) | 1.1%±0.4% |
| Neutrophils (103/µL) | 1.0%±0.1% | RDW-CV | 1.1%±0.1% | Lymphocytes (103/µL) | 1.1%±0.2% |
| Lymphocytes (103/µL) | 0.9%±0.1% | Monocytes (%) | 1.1%±0.1% | MCH | 1.1%±0.1% |
| Monocytes (%) | 0.9%±0.1% | MCH | 1.0%±0.1% | Creatinine | 1.1%±0.2% |
| aPTT  | 0.9%±0.1% | Creatinine | 1.0%±0.2% | MCV | 1.0%±0.1% |
| Chloride | 0.9%±0.1% | MCV | 1.0%±0.1% | Platelets (103/µL) | 1.0%±0.1% |
| Platelets (103/µL) | 0.9%±0.0% | Glucose  | 1.0%±0.3% | pH (venous) | 1.0%±0.3% |
| RDW-CV | 0.9%±0.0% | Bilirubin | 1.0%±0.2% | ALT  | 1.0%±0.2% |
| MCH | 0.9%±0.0% | RBC | 1.0%±0.1% | Total cholesterol | 1.0%±0.4% |
| ESR | 0.9%±0.2% | Creatine kinase | 1.0%±0.1% | ALP  | 1.0%±0.1% |
| MCV | 0.9%±0.1% | CD4/CD8 ratio | 1.0%±0.2% | PDW | 1.0%±0.2% |
| Glucose | 0.9%±0.0% | Plasma Lactate | 1.0%±0.3% | Phosphate | 1.0%±0.4% |
| aPTT ratio | 0.9%±0.1% | Hematocrit | 1.0%±0.1% | RBC  | 1.0%±0.1% |
| Proteins | 0.9%±0.1% | Magnesium | 1.0%±0.6% | Monocytes (103/µL) | 0.9%±0.1% |
| Creatine kinase | 0.9%±0.1% | WBC | 0.9%±0.2% | GGT  | 0.9%±0.2% |
| AST  | 0.9%±0.0% | Lymphocytes T CD8 (%) | 0.9%±0.4% | Plasma Lactate | 0.9%±0.1% |
| Ferritin  | 0.9%±0.1% | Chloride | 0.9%±0.2% | Troponin T | 0.9%±0.2% |
| Bilirubin | 0.9%±0.1% | Platelets (103/µL) | 0.9%±0.1% | aPTT  | 0.9%±0.1% |
| RBC  | 0.9%±0.0% | ALT | 0.9%±0.1% | AST  | 0.9%±0.1% |
| Plasma Lactate | 0.9%±0.0% | PDW | 0.9%±0.1% | Chloride | 0.9%±0.1% |
| WBC | 0.9%±0.1% | AST | 0.9%±0.1% | aPTT ratio | 0.9%±0.1% |
| MCHC | 0.9%±0.1% | Hemoglobin  | 0.9%±0.1% | Hematocrit | 0.9%±0.1% |
| Prothrombin time (%) | 0.8%±0.0% | MCHC | 0.9%±0.1% | Neutrophils (103/µL) | 0.9%±0.1% |
| Hematocrit | 0.8%±0.0% | Monocytes  | 0.9%±0.1% | Hemoglobin  | 0.9%±0.1% |
| GGT  | 0.8%±0.1% | aPTT | 0.8%±0.1% | WBC | 0.9%±0.1% |
| Triglycerides | 0.8%±0.1% | aPTT ratio | 0.8%±0.1% | Potassium | 0.9%±0.1% |
| PDW | 0.8%±0.0% | Triglycerides | 0.8%±0.3% | MCHC | 0.9%±0.1% |
| Monocytes (103/µL) | 0.8%±0.0% | Prothrombin time (%) | 0.8%±0.1% | Prothrombin time (%) | 0.8%±0.1% |
| Lymphocytes T CD4 (103/µL) | 0.8%±0.1% | Potassium | 0.8%±0.1% | Lymphocytes T CD4 (%) | 0.8%±0.2% |
| Hemoglobin  | 0.8%±0.0% | Troponin T | 0.8%±0.3% | Triglycerides | 0.8%±0.2% |
| Eosinophils (%) | 0.7%±0.1% | Albumin | 0.7%±0.2% | Prothrombin time ratio | 0.8%±0.1% |
| Prothrombin time ratio | 0.7%±0.0% | ESR | 0.7%±0.2% | Albumin | 0.8%±0.1% |
| Potassium | 0.7%±0.0% | Sodium | 0.7%±0.1% | Sodium | 0.8%±0.1% |
| CD4/CD8 ratio | 0.7%±0.1% | Prothrombin time ratio | 0.7%±0.1% | Bilirubin | 0.8%±0.1% |
| Sodium | 0.7%±0.1% | Proteins | 0.6%±0.3% | Ferritin  | 0.7%±0.3% |
| Lymphocytes T CD4 (%) | 0.7%±0.1% | Ferritin  | 0.6%±0.1% | CD4/CD8 ratio | 0.7%±0.2% |
| Lymphocytes T CD8 (103/µL) | 0.7%±0.2% | INR | 0.6%±0.2% | INR | 0.7%±0.1% |
| Eosinophils (103/µL) | 0.7%±0.1% | Zinc | 0.6%±0.3% | Bicarbonate (venous) | 0.7%±0.1% |
| Albumin | 0.7%±0.1% | pH (venous)  | 0.6%±0.3% | Zinc | 0.7%±0.2% |
| Zinc | 0.7%±0.2% | Eosinophils (%) | 0.6%±0.1% | Vitamin D | 0.7%±0.1% |
| Total cholesterol | 0.7%±0.1% | Phosphate | 0.6%±0.2% | Base Excess (venous) | 0.6%±0.1% |
| Lymphocytes T CD8 (%) | 0.7%±0.1% | pvO2  | 0.5%±0.2% | Lymphocytes T CD8 (%) | 0.6%±0.1% |
| pvCO2  | 0.7%±0.1% | Eosinophils (103/µL) | 0.5%±0.1% | Lymphocytes T CD8 (103/µL) | 0.6%±0.1% |
| INR | 0.7%±0.1% | Sat O2 (venous) | 0.5%±0.2% | Lymphocytes T CD4 (103/µL) | 0.6%±0.1% |
| Basophils (%) | 0.6%±0.1% | Immunoglobulin A (Ig A) | 0.5%±0.1% | Eosinophils (%) | 0.6%±0.1% |
| pvO2  | 0.6%±0.1% | Immunoglobulin M (Ig M) | 0.5%±0.2% | Vitamin B12 | 0.6%±0.2% |
| Sat O2 (venous)  | 0.6%±0.1% | Vitamin D | 0.5%±0.1% | Eosinophils (103/µL) | 0.6%±0.1% |
| Lipase | 0.6%±0.1% | Basophils (%) | 0.4%±0.0% | pvCO2  | 0.6%±0.1% |
| pH (venous) | 0.6%±0.1% | Total cholesterol | 0.4%±0.1% | Proteins | 0.6%±0.2% |
| Plasma Fibrinogen | 0.5%±0.1% | Uric acid | 0.4%±0.1% | Basophils (%) | 0.5%±0.1% |
| Parathyroid Hormone  | 0.5%±0.2% | Base Excess (venous) | 0.4%±0.1% | Immunoglobulin A (Ig A) | 0.5%±0.1% |
| Basophils (103/µL) | 0.5%±0.0% | pvCO2 | 0.4%±0.1% | Uric acid | 0.5%±0.1% |
| Transferrin saturation | 0.5%±0.1% | Transferrin | 0.4%±0.3% | Reticulocytes (%) | 0.5%±0.2% |
| Phosphate | 0.4%±0.0% | Plasma Fibrinogen | 0.4%±0.1% | Reticulocyte Hb | 0.5%±0.2% |
| Magnesium | 0.4%±0.0% | Basophils | 0.4%±0.1% | Reticulocytes (103/µL) | 0.4%±0.1% |
| Uric acid | 0.4%±0.0% | Bicarbonate (venous) | 0.4%±0.1% | pvO2  | 0.4%±0.1% |
| Reticulocytes (103/µL) | 0.4%±0.1% | Vitamin B12 | 0.4%±0.1% | Sat O2 (venous) | 0.4%±0.1% |
| Reticulocyte Hemoglobin | 0.4%±0.0% | Lipase | 0.3%±0.1% | Basophils (103/µL) | 0.4%±0.1% |
| Reticulocytes (%) | 0.4%±0.0% | Reticulocytes (103/µL) | 0.3%±0.1% | Immunoglobulin G (Ig G) | 0.4%±0.1% |
| Sex | 0.4%±0.1% | Folate | 0.3%±0.1% | Plasma Fibrinogen | 0.4%±0.1% |
| Immunoglobulin A (Ig A) | 0.4%±0.1% | Reticulocytes (%) | 0.3%±0.1% | Immunoglobulin M (Ig M) | 0.4%±0.1% |
| Immunoglobulin M (Ig M) | 0.4%±0.1% | Immunoglobulin A (Ig A) | 0.3%±0.1% | Folate | 0.3%±0.1% |
| Vitamin D | 0.4%±0.0% | Ferrum | 0.3%±0.1% | Legionella urinary antigen test | 0.3%±0.1% |
| Transferrin | 0.4%±0.0% | Prealbumin | 0.3%±0.1% | Lipase | 0.3%±0.1% |
| Vitamin B12 | 0.4%±0.0% | Sex | 0.3%±0.1% | Ferrum | 0.2%±0.1% |
| Tirotropin (TSH) | 0.4%±0.0% | Reticulocyte Hemoglobin | 0.3%±0.1% | Sex | 0.2%±0.1% |
| Immunoglobulin G (Ig G) | 0.3%±0.0% | Tirotropin (TSH) | 0.2%±0.1% | HDL cholesterol | 0.2%±0.1% |
| HDL cholesterol  | 0.3%±0.1% | Legionella urinary antigen test | 0.2%±0.1% | Transferrin | 0.2%±0.1% |
| Ferrum | 0.3%±0.0% | LDL cholesterol | 0.1%±0.0% | LDL cholesterol | 0.2%±0.1% |
| LDL cholesterol | 0.3%±0.0% | Transferrin saturation | 0.1%±0.1% | Tirotropin (TSH) | 0.2%±0.1% |
| Folate | 0.3%±0.0% | HDL cholesterol | 0.1%±0.0% | Urinary sediment (Gram)  | 0.1%±0.0% |
| Urinary sediment (Gram) | 0.3%±0.0% | Urinary sediment (Gram)  | 0.1%±0.1% | Free thyroxine (T4) | 0.1%±0.1% |
| Urine culture | 0.3%±0.0% | Parathyroid Hormone | 0.1%±0.1% | Transferrin Saturation | 0.1%±0.1% |
| Albumin  | 0.2%±0.0% | Urine culture | 0.1%±0.0% | Urine culture | 0.1%±0.0% |
| Prealbumin | 0.2%±0.0% | Albumin | 0.1%±0.0% | Glycated hemoglobin (%) | 0.1%±0.0% |
| Glycated hemoglobin | 0.2%±0.0% | Free thyroxine (T4) | 0.1%±0.0% | Albumin (%) | 0.1%±0.0% |
| Free thyroxine (T4) | 0.1%±0.0% | Glycated hemoglobin | 0.1%±0.0% | Blood cultures (aerobes) | 0.1%±0.0% |
| Blood cultures (anaerobes) | 0.1%±0.0% | Blood cultures (anaerobes) | 0.0%±0.0% | Blood cultures (anaerobes)  | 0.1%±0.0% |
| Blood cultures (aerobes) | 0.1%±0.0% | Blood cultures (aerobes) | 0.0%±0.0% | Parathyroid Hormone | 0.0%±0.1% |
| Lipidemia | 0.0%±0.0% | Lipidemia | 0.0%±0.0% | Lipidemia | 0.0%±0.0% |
| Hemolysis | 0.0%±0.0% | Hemolysis | 0.0%±0.0% | Hemolysis | 0.0%±0.0% |
| SARS-CoV2 rt-PCR (E gene) | 0.0%±0.0% | SARS-CoV2 rt-PCR (E gene) | 0.0%±0.0% | SARS-CoV2 rt-PCR (E gene) | 0.0%±0.0% |
| SARS-CoV2 rt-PCR (RdRP & N) | 0.0%±0.0% | SARS-CoV2 rt-PCR (RdRP & N) | 0.0%±0.0% | SARS-CoV2 rt-PCR (RdRP & N) | 0.0%±0.0% |
| SARS-CoV2 total antibodies | 0.0%±0.0% | SARS-CoV2 total antibodies | 0.0%±0.0% | SARS-CoV2 total antibodies | 0.0%±0.0% |
| SARS-CoV2 Point-of-Care Test | 0.0%±0.0% | SARS-CoV2 Point-of-Care Test | 0.0%±0.0% | SARS-CoV2 Point-of-Care Test | 0.0%±0.0% |
| HBV surface antigen | 0.0%±0.0% | HBV surface antigen | 0.0%±0.0% | HBV surface antigen | 0.0%±0.0% |
| HCV antibodies | 0.0%±0.0% | HCV antibodies | 0.0%±0.0% | HCV Antibodies | 0.0%±0.0% |
| HIV antibodies | 0.0%±0.0% | HIV antibodies | 0.0%±0.0% | HIV Antibodies | 0.0%±0.0% |
| Pneumococcal urinary antigen test | 0.0%±0.0% | Pneumococcal urinary antigen test | 0.0%±0.0% | Pneumococcal urinary antigen test | 0.0%±0.0% |
| SARS-CoV2 IgG | 0.0%±0.0% | SARS-CoV2 IgG | 0.0%±0.0% | SARS-CoV2 IgG | 0.0%±0.0% |
| SARS-CoV2 IgM | 0.0%±0.0% | SARS-CoV2 IgM | 0.0%±0.0% | SARS-CoV2 IgM | 0.0%±0.0% |

**Table S1**. Variable weights in LR (laboratory results) models according to the different outcomes evaluated. Abbreviations: OS, oxygen support; ICU, Intensive Care Unit; paCO2, partial arterial pressure of carbon dioxide; paO2, partial arterial pressure of oxygen; satO2, oxygen saturation; pvCO2, partial venous pressure of carbon dioxide; pvO2, partial venous pressure of oxygen; BNP, brain natriuretic peptide; LDH, lactate dehydrogenase; ALT, alanine aminotransferase; ALP, alkaline phosphatase; AST, aspartate aminotransferase; GGT, gamma-glutamyl transferase; RBC, red blood cells; WBC, white blood cells; RDW-SD, *red cell distribution width-standard deviation;* RDW-CV, *red cell distribution width-corpuscular volume; MCH, m*ean corpuscular hemoglobin; *MCHC, mean corpuscular hemoglobin concentration; MCV, mean cell volume; ESR,* erythrocyte sedimentation rate; PDW, platelet distribution width; *aPTT, activated* partial thromboplastin time; INR, international normalized ratio; HDL, high-density lipoprotein; LDL, low-density lipoprotein; rt-PCR, reverse transcriptase-polymerase chain reaction; HBV, Hepatitis B virus; HCV, hepatitis C virus; HIV; human immunodeficiency virus.