

## Research Article

# Disentangling the Association of Hydroxychloroquine Treatment with Mortality in Covid-19 Hospitalized Patients through Hierarchical Clustering

Augusto Di Castelnuovo <sup>1</sup>, Alessandro Gialluisi,<sup>2</sup> Andrea Antinori,<sup>3</sup> Nausicaa Berselli,<sup>4</sup> Lorenzo Blandi,<sup>5</sup> Marialaura Bonaccio,<sup>2</sup> Raffaele Bruno,<sup>6,7</sup> Roberto Cauda,<sup>8,9</sup> Simona Costanzo,<sup>2</sup> Giovanni Guaraldi,<sup>10</sup> Lorenzo Menicanti,<sup>11</sup> Marco Mennuni,<sup>12</sup> Iliaria My,<sup>13</sup> Giustino Parruti,<sup>14</sup> Giuseppe Patti,<sup>12</sup> Stefano Perlini,<sup>15,16</sup> Francesca Santilli,<sup>17</sup> Carlo Signorelli,<sup>18</sup> Giulio Stefanini,<sup>13</sup> Alessandra Vergori,<sup>19</sup> Walter Ageno,<sup>20</sup> Antonella Agodi,<sup>21</sup> Piergiuseppe Agostoni,<sup>22,23</sup> Luca Aiello,<sup>24</sup> Samir Al Moghazi,<sup>25</sup> Rosa Arboretti,<sup>26</sup> Filippo Aucella,<sup>27</sup> Greta Barbieri,<sup>28</sup> Martina Barchitta,<sup>29</sup> Paolo Bonfanti,<sup>30,31</sup> Francesco Cacciatore,<sup>32</sup> Lucia Caiano,<sup>20</sup> Francesco Cannata,<sup>13</sup> Laura Carrozzi,<sup>33</sup> Antonio Cascio,<sup>34</sup> Giacomo Castiglione,<sup>35</sup> Arturo Cicullo,<sup>8</sup> Antonella Cingolani,<sup>8,9</sup> Francesco Cipollone,<sup>17</sup> Claudia Colomba,<sup>34</sup> Crizia Colombo,<sup>12</sup> Annalisa Crisetti,<sup>27</sup> Francesca Crosta,<sup>14</sup> Gian Battista Danzi,<sup>36</sup> Damiano D'Ardes,<sup>17</sup> Katleen de Gaetano Donati,<sup>8,9</sup> Francesco Di Gennaro,<sup>37</sup> Giuseppe Di Tano,<sup>36</sup> Gianpiero D'Offizi,<sup>38</sup> Francesco Maria Fusco,<sup>39</sup> Carlo Gaudiosi,<sup>40</sup> Ivan Gentile,<sup>41</sup> Francesco Gianfagnal,<sup>20</sup> Gabriele Giuliano,<sup>8</sup> Emauele Graziani,<sup>42</sup> Gabriella Guarnieri,<sup>43</sup> Valerio Langella,<sup>44</sup> Giovanni Larizza,<sup>45</sup> Armando Leone,<sup>46</sup> Gloria Maccagni,<sup>36</sup> Federica Magni,<sup>20</sup> Stefano Maitan,<sup>24</sup> Sandro Mancarella,<sup>47</sup> Rosa Manuele,<sup>48</sup> Massimo Mapelli,<sup>22,23</sup> Riccardo Maragna,<sup>22,23</sup> Rossella Marcucci,<sup>49</sup> Giulio Maresca,<sup>44</sup> Silvia Marongiu,<sup>50</sup> Claudia Marotta,<sup>37</sup> Lorenzo Marra,<sup>46</sup> Franco Mastroianni,<sup>45</sup> Alessandro Mengozzi,<sup>51</sup> Marianna Meschiari,<sup>10</sup> Jovana Milic,<sup>10</sup> Filippo Minutolo,<sup>52</sup> Roberta Mussinelli,<sup>16</sup> Cristina Mussini,<sup>10</sup> Maria Musso,<sup>53</sup> Anna Odone,<sup>5</sup> Marco Olivieri,<sup>54</sup> Antonella Palimodde,<sup>50</sup> Emanuela Pasi,<sup>42</sup> Raffaele Pesavento,<sup>55</sup> Francesco Petri,<sup>30</sup> Carlo A Pivato,<sup>13</sup> Venerino Poletti,<sup>56,57</sup> Claudia Ravaglia,<sup>56</sup> Giulia Righetti,<sup>45</sup> Andrea Rognoni,<sup>12</sup> Marco Rossato,<sup>55</sup> Iliaria Rossi,<sup>17</sup> Marianna Rossi,<sup>30</sup> Anna Sabena,<sup>15</sup> Francesco Salinaro,<sup>15</sup> Vincenzo Sangiovanni,<sup>39</sup> Carlo Sanrocco,<sup>14</sup> Nicola Schiano Moriello,<sup>41</sup> Laura Scorzolini,<sup>58</sup> Raffaella Sgariglia,<sup>47</sup> Paola Giustina Simeone,<sup>14</sup> Michele Spinicci,<sup>49</sup> Enrica Tamburrini,<sup>8</sup> Carlo Torti,<sup>59</sup> Enrico Maria Trecarichi,<sup>59</sup> Roberto Vettor,<sup>55</sup> Andrea Vianello,<sup>43</sup> Marco Vinceti,<sup>4,60</sup> Agostino Virdis,<sup>51</sup> Raffaele De Caterina,<sup>33</sup> and Licia Iacoviello <sup>2,20</sup>

<sup>1</sup>Mediterranea Cardiocentro, Napoli, Italy

<sup>2</sup>Department of Epidemiology and Prevention, IRCCS Neuromed, Pozzilli, Italy

<sup>3</sup>UOC Immunodeficienze Virali, National Institute for Infectious Diseases "L. Spallanzani" IRCCS, Rome, Italy

<sup>4</sup>Section of Public Health, Department of Biomedical, Metabolic and Neural Sciences, University of Modena and Reggio Emilia, Modena, Italy

<sup>5</sup>Università di Pavia, Pavia, Italy

<sup>6</sup>Division of Infectious Diseases I, Fondazione IRCCS Policlinico San Matteo, Pavia, Italy

<sup>7</sup>Department of Clinical, Surgical Diagnostic and Paediatric Sciences, University of Pavia, Pavia, Italy

<sup>8</sup>Fondazione Policlinico Universitario A. Gemelli IRCCS, Roma, Italy

- <sup>9</sup>Università Cattolica Del Sacro Cuore- Dipartimento di Sicurezza e Bioetica Sede di Roma, Roma, Italy
- <sup>10</sup>Infectious Disease Unit, Department of Surgical, Medical Dental and Morphological Sciences, University of Modena and Reggio Emilia, Modena, Italy
- <sup>11</sup>IRCCS Policlinico San Donato, San Donato Milanese, Milan, Italy
- <sup>12</sup>University of Eastern Piedmont, Maggiore Della Carità Hospital, Novara, Italy
- <sup>13</sup>Humanitas Clinical and Research Hospital IRCCS, Rozzano, Milano, Italy
- <sup>14</sup>Department of Infectious Disease, Azienda Sanitaria Locale (AUSL) di Pescara, Pescara, Italy
- <sup>15</sup>Emergency Department, IRCCS Policlinico San Matteo Foundation, Pavia, Italy
- <sup>16</sup>Department of Internal Medicine, University of Pavia, Pavia, Italy
- <sup>17</sup>Department of Medicine and Aging, Clinica Medica, "SS. Annunziata" Hospital and University of Chieti, Chieti, Italy
- <sup>18</sup>School of Medicine, Vita-Salute San Raffaele University, Milano, Italy
- <sup>19</sup>HIV/AIDS Department, National Institute for Infectious Diseases Lazzaro Spallanzani IRCCS, Roma, Italy
- <sup>20</sup>Department of Medicine and Surgery, University of Insubria, Varese, Italy
- <sup>21</sup>Department of Medical and Surgical Sciences and Advanced Technologies G.F. Ingrassia, University of Catania, AOU Policlinico G.Rodolico - San Marco, Catania, Italy
- <sup>22</sup>Centro Cardiologico Monzino IRCCS, Milano, Italy
- <sup>23</sup>Department of Clinical Sciences and Community Health, Cardiovascular Section, University of Milano, Milan, Italy
- <sup>24</sup>UOC. Anestesia e Rianimazione, Dipartimento di Chirurgia Generale Ospedale Morgagni-Pierantoni, Forlì, Italy
- <sup>25</sup>UOC Infezioni Sistemiche Dell'Immunodepresso, National Institute for Infectious Diseases L. Spallanzani IRCCS, Rome, Italy
- <sup>26</sup>Department of Civil Environmental and Architectural Engineering, University of Padova, Padova, Italy
- <sup>27</sup>Fondazione I.R.C.C.S Casa Sollievo Della Sofferenza, San Giovanni Rotondo, Foggia, Italy
- <sup>28</sup>Department of Surgical Medical and Molecular Medicine and Critical Care, Azienda Ospedaliera Universitaria Pisana and University of Pisa, Pisa, Italy
- <sup>29</sup>Department of Medical and Surgical Sciences and Advanced Technologies G.F. Ingrassia, University of Catania, Catania, Italy
- <sup>30</sup>UOC Malattie Infettive, Ospedale San Gerardo, ASST Monza, Monza, Italy
- <sup>31</sup>School of Medicine and Surgery, University of Milano-Bicocca, Milano, Italy
- <sup>32</sup>Department of Translational Medical Sciences, University of Naples, Federico II, Naples, Italy
- <sup>33</sup>Cardiovascular and Thoracic Department, Azienda Ospedaliero-Universitaria Pisana and University of Pisa, Pisa, Italy
- <sup>34</sup>Infectious and Tropical Diseases Unit- Department of Health Promotion, Mother and Child Care, Internal Medicine and Medical Specialties (PROMISE) - University of Palermo, Palermo, Italy
- <sup>35</sup>Servizio di Anestesia e Rianimazione II UO Rianimazione Ospedale San Marco, AOU Policlinico G. Rodolico, San Marco, Catania, Italy
- <sup>36</sup>Department of Cardiology, Ospedale di Cremona, Cremona, Italy
- <sup>37</sup>Medical Direction IRCCS Neuromed, Pozzilli, Italy
- <sup>38</sup>UOC Malattie Infettive-Epatologia, National Institute for Infectious Diseases L. Spallanzani IRCCS, Roma, Italy
- <sup>39</sup>UOC Infezioni Sistemiche e Dell'Immunodepresso, Azienda Ospedaliera Dei Colli Ospedale Cotugno, Napoli, Italy
- <sup>40</sup>ASL Napoli3 Sud COVID HOSPITAL, Boscorecase, Napoli, Italy
- <sup>41</sup>Department of Clinical Medicine and Surgery, University of Naples Federico II, Napoli, Italy
- <sup>42</sup>Medicina Interna, Ospedale di Ravenna, AUSL Della Romagna, Ravenna, Italy
- <sup>43</sup>Respiratory Pathophysiology Division, Department of Cardiology, Thoracic and Vascular Sciences, University of Padova, Padova, Italy
- <sup>44</sup>UOC Medicina COVID- PO S. Maria di Loreto Nuovo, ASL Na 1 Centro, Napoli, Italy
- <sup>45</sup>COVID-19 Unit. EE Ospedale Regionale F. Miulli, Acquaviva Delle Fonti, Bari, Italy
- <sup>46</sup>UOC di Pneumologia P.O. San Giuseppe Moscati, Taranto, Italy
- <sup>47</sup>ASST Milano Nord - Ospedale Edoardo Bassini, Cinisello Balsamo, Italy
- <sup>48</sup>U.O. C. Malattie Infettive e Tropicali, P.O. "San Marco" AOU Policlinico "G. Rodolico - San Marco", Catania, Italy
- <sup>49</sup>Department of Experimental and Clinical Medicine, University of Florence and Azienda Ospedaliero-Universitaria Careggi, Firenze, Italy
- <sup>50</sup>Santissima Trinità di Cagliari, Cagliari, Italy
- <sup>51</sup>Department of Clinical and Experimental Medicine, Azienda Ospedaliera Universitaria Pisana University of Pisa, Pisa, Italy
- <sup>52</sup>Dipartimento di Farmacia, Università di Pisa, Pisa, Italy
- <sup>53</sup>UOC Malattie Infettive-Apparato Respiratorio, National Institute for Infectious Diseases L. Spallanzani IRCCS, Rome, Italy
- <sup>54</sup>Computer Service, University of Molise, Campobasso, Italy
- <sup>55</sup>Clinica Medica 3. Department of Medicine - DIMED, University Hospital of Padova, Padova, Italy
- <sup>56</sup>UOC Pneumologia. Dipartimento di Malattie Apparato Respiratorio e Torace, Ospedale Morgagni-Pierantoni Forlì, Forlì, Italy
- <sup>57</sup>Department of Respiratory Diseases & Allergy Aarhus University Hospital, Aarhus, Denmark
- <sup>58</sup>UOC Malattie Infettive Ad Alta Intensità di Cura, National Institute for Infectious Diseases L. Spallanzani IRCCS, Rome, Italy
- <sup>59</sup>Infectious and Tropical Diseases Unit, Department of Medical and Surgical Sciences, Magna Graecia University, Catanzaro, Italy
- <sup>60</sup>Department of Epidemiology, Boston University School of Public Health, Boston, USA

Correspondence should be addressed to Licia Iacoviello; [licia.iacoviello@uninsubria.it](mailto:licia.iacoviello@uninsubria.it)

Received 19 February 2021; Revised 8 April 2021; Accepted 16 May 2021; Published 25 June 2021

Academic Editor: Matteo Russo

Copyright © 2021 Augusto Di Castelnuovo et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The efficacy of hydroxychloroquine (HCQ) in treating SARS-CoV-2 infection is harshly debated, with observational and experimental studies reporting contrasting results. To clarify the role of HCQ in Covid-19 patients, we carried out a retrospective observational study of 4,396 unselected patients hospitalized for Covid-19 in Italy (February–May 2020). Patients' characteristics were collected at entry, including age, sex, obesity, smoking status, blood parameters, history of diabetes, cancer, cardiovascular and chronic pulmonary diseases, and medications in use. These were used to identify subtypes of patients with similar characteristics through hierarchical clustering based on Gower distance. Using multivariable Cox regressions, these clusters were then tested for association with mortality and modification of effect by treatment with HCQ. We identified two clusters, one of 3,913 younger patients with lower circulating inflammation levels and better renal function, and one of 483 generally older and more comorbid subjects, more prevalently men and smokers. The latter group was at increased death risk adjusted by HCQ (HR [CI95%] = 3.80[3.08-4.67]), while HCQ showed an independent inverse association (0.51[0.43-0.61]), as well as a significant influence of cluster\*HCQ interaction ( $p < 0.001$ ). This was driven by a differential association of HCQ with mortality between the high (0.89[0.65-1.22]) and the low risk cluster (0.46[0.39-0.54]). These effects survived adjustments for additional medications in use and were concordant with associations with disease severity and outcome. These findings suggest a particularly beneficial effect of HCQ within low risk Covid-19 patients and may contribute to clarifying the current controversy on HCQ efficacy in Covid-19 treatment.

## 1. Introduction

Hydroxychloroquine (HCQ) is an antimalarial drug suggested to be effective in inhibiting Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-Cov-2) replication in vitro [1, 2]. Indeed, HCQ is characterized by antiviral, anti-inflammatory, and antithrombotic actions, contrasting the main disruptive effects of SARS-CoV-2 infection on the organism [3]. For this reason, it has been heavily used in treating patients affected by SARS-CoV-2 infection related disease (commonly known as Covid-19), especially in the first phases of the current pandemics, when Covid-19 was quite unknown [4].

Despite these elements and initial suggestive evidence of efficacy based on daily clinical practice, in the last months, the potential benefit of HCQ for Covid-19 patients has been harshly debated [3, 5]. In particular, evidence supporting protective effects from observational studies [6–13] was in contrast with that suggesting no effect at all by recent randomized clinical trials (RCTs) [14–18]. More recently, a meta-analysis combining both RCTs and observational studies over more than 44,000 patients supported a protective effect of HCQ, driven by the findings of observational studies [5]. A potential explanation for this discrepancy may be due to the usually high dosage administered in RCTs (800 mg/day), compared to lower dosages reported in observational studies supporting HCQ efficacy ( $\leq 400$  mg/day), as hypothesized elsewhere [5]. As an alternative explanation, it is likely that the efficacy of HCQ treatment for Covid-19 may vary across patients and is influenced by subtypes of the disease, which in turn is largely dependent on patients' characteristics and their nonlinear combinations [19]. In this

“personalized medicine” view, response to HCQ may not be the same across all patients of the same age, or with similar circulating inflammation levels. In order to identify these combinations, the use of big data like Electronic Health Records (EHRs) and of machine learning (ML) algorithms to interpret hidden HCQ response patterns is of fundamental importance. ML is an umbrella term covering different algorithms designed for the identification of hidden patterns, information mining, and variable classification/estimation through modelling complex (including nonlinear) functions, usually adopted in a big data setting. These algorithms can be generally classified into supervised and unsupervised approaches. The formers are designed to learn to predict specific outcomes after proper training of the algorithm in independent datasets, trying to model relationships and dependencies between the input variables (or features) and the target prediction output (or label). In unsupervised algorithms, the machine simply learns to identify hidden patterns across a high number of observations over many features, without the need for labels, and is more descriptive—rather than predictive—in nature. These algorithms have shown promising findings in public health, in the management of both chronic [20] and acute health conditions [21, 22], but also during the current pandemics. In particular, useful ML applications have been reported in the prediction of Covid-19 diagnosis and prognosis [23]. Notwithstanding this, to the best of our knowledge, only one study attempted so far to predict response to HCQ treatment within Covid-19 patients, through the application of a supervised ML technique (gradient boosting). Interestingly, the authors reported a reduction of in-hospital mortality within patients treated with HCQ, which was even more

pronounced within those patients predicted to benefit most from the drug, in line with expectations [19].

Here, we attempted a personalized Covid-19 patient characterization to better disentangle the beneficial effects of HCQ previously reported within the COVID-19 RISK and Treatments (CORIST) study, a large retrospective cohort of patients hospitalized for SARS-CoV-2 infection in Italy [13]. This approach consisted of i) identifying the existence of subtypes of Covid-19 patients through an unsupervised ML algorithm—hierarchical clustering—comparing their characteristics and their clinical risks, ii) testing the resulting patients' clusters for association with mortality and modification of effect by treatment with HCQ, and iii) analyzing potential interactions between clusters and HCQ use. This approach represents a prominent example of how personalized medicine may support clinicians in Covid-19 treatment.

## 2. Methods

**2.1. Analyzed Cohort.** The COVID-19 RISK and Treatments (CORIST) study includes 4,396 patients hospitalized for SARS-Cov-2 infection in 35 hospitals across Italy, between February 2020 and May 2020. Molecular diagnosis of SARS-CoV-2 infection was based on polymerase chain reaction (PCR) of viral DNA extracted and amplified from nasopharyngeal swabs. Within each participating hospital, clinical data were abstracted at one-time point from electronic medical records or charts and collected using either a centrally designed electronic worksheet or a centralized web-based database. Collected data included patients' demographics, laboratory tests, medications in use, history of disease, and prescribed pharmacological therapy for Covid-19 treatment. For each participant, the study index date was defined as the date of hospital admission, while the study end point was death. Follow-up time was computed as the time between the index date and death, or alternatively between the index date and the date of discharge, applying right-censoring. Further details on the study are reported elsewhere [13, 24, 25].

### 2.2. Statistical Analyses

**2.2.1. Cluster Analysis.** All analyses were carried out in R 4.0.2 (<https://www.r-project.org/>) [26]. We applied a hierarchical clustering analysis on Covid-19 patients using their main clinical, lifestyles, and sociodemographic characteristics, which were suggested as the most influential on mortality risk by previous studies in the field [25, 27, 28]. These included age (years), sex, glomerular filtration rate (eGFR, mL/min/1.73 m<sup>2</sup>) and high-sensitivity plasma C-reactive protein levels (mg/L) at in-hospital admission, obesity (body mass index (BMI)  $\geq 30$  kg/m<sup>2</sup>), hypertension (Yes/No), and smoking status (never/previous/current smoker), as well as history of myocardial infarction, heart failure, chronic pulmonary disease, cancer, and diabetes (Yes/No). Missing data were imputed through a k-Nearest Neighbor approach, implemented in the *knn()* function of the VIM package (version 6.1.0; <https://cran.r-project.org/>

[web/packages/VIM/index.html](https://cran.r-project.org/web/packages/VIM/index.html)), with  $k = 10$  [29]. CRP was transformed on the natural logarithm scale to reduce skewness, and all the continuous variables were normalized through the *normalize()* function in Keras v2.3.0 (<https://cran.r-project.org/web/packages/keras/index.html>).

Cluster analyses were then performed on the 4,396 patients, using the variables specified above, through the Cluster package v2.1.0 (<https://cran.rproject.org/web/packages/cluster/index.html>). First, we computed a dissimilarity matrix based on Gower pairwise distance (Figure S1), through the *daisy()* function. Gower distance is a parameter in the [0; 1] range representing the average of partial dissimilarities across individuals (the higher the distance, the more the dissimilarities for a given pair of subjects) [30]. Second, we performed hierarchical clustering through the *hclust()* function applied to the Gower distance matrix computed above, which separated subjects based on their degree of pairwise dissimilarity, both from lowest to highest (agglomerative clustering) and from highest to lowest (divisive clustering). Third, we determined the appropriate number of clusters for patient classification, based on the Average Silhouette method (Figure S2). This computes the number of clusters, which maximizes the average silhouette width, a measure of the quality of clustering indicating how well each object lies within its cluster [31]. This method, applied through the *fviz\_nbclust()* function of the Factoextra package (v1.0.7, <https://cran.rproject.org/web/packages/factoextra/index.html>), computed  $k = 2$  as the optimal number of clusters. Finally, each patient was assigned to one of the two clusters determined above, through application of the *cutree()* function (Cluster package) to the results of the cluster analysis previously carried out.

**2.2.2. Comparison of Clusters.** First, we compared the classifications made through agglomerative and hierarchical clustering, which revealed high consistency (Odds Ratio = 34.0 [26.7-43.7], Fisher Exact Test  $p$  value  $< 10^{-15}$ ). In light of this homogeneity of classification, and since divisive clustering has been reported to be more accurate and robust [32], all the subsequent analyses were performed on cluster classification identified through the latter approach.

The two clusters of patients identified were then compared for all anamnestic variables mentioned above, through Fisher's Exact Test (for binary variables), Chi-squared test (for nonbinary categorical variables), and through Student's  $t$ -test or Wilcoxon Rank Sum tests (for continuous variables meeting and not meeting parametric assumptions, respectively). Similarly, we compared Covid-19 disease severity, classified by recruiting centers in asymptomatic/mild, nonsevere pneumonia, severe pneumonia, and acute respiratory distress syndrome (ARDS). Moreover, we compared the use of six common drugs for Covid-19 treatment between the two clusters, including hydroxychloroquine, antihypertensive drugs, anti-interleukin-6 antibody, antivirals (Remdesivir, Lopinavir, Darunavir), and corticosteroids. These were reported as binary variables (Yes/No) and were, therefore, compared with clusters through Fisher Exact Tests.

**2.2.3. Survival Analyses.** Once the clusters were characterized, we modelled incident mortality risk as a function of patients clusters and use of HCQ through Cox Proportional Hazards (PH) models, using the *cox.zph()* function of the survival package (v3.2.7, see URLs: <https://cran.r-project.org/web/packages/survival/index.html>) [33]. Only patients with complete information needed in each model were included in the analysis (case-complete approach; see below). A preliminary check of the basic Cox PH assumptions revealed no influential observations based on *dfbeta* residuals (Figure S3a), while Schoenfeld residuals tests revealed a statistical violation of the proportionality of hazards assumption, although these did not show any evident trend at a visual inspection (Figure S3b). For this reason, we carried out Cox PH models both with and without including an interaction term with time-to-event, a strategy commonly used to overcome this violation [34]. Incrementally adjusted models were analyzed: i) a crude model including only patients' clusters (Model 1;  $N=4,319$ ); ii) a model testing additive influence of clusters and HCQ use (Model 2: Model 1 + HCQ;  $N=4,212$ ); and iii) a model testing both additive and synergistic influence of clusters and HCQ use (Model 3: Model 2 + clusters\*HCQ;  $N=4,212$ ). Additional sensitivity analyses were carried out to rule out potential confounding effects of additional drugs in use for Covid-19 treatment (Model 4: Model 3 + other drugs). Risk estimates were computed as hazard ratios (HR) with 95% confidence intervals (95% CI) of dying, and HR with  $p$  values below  $\alpha=0.05$  were considered significant. To quantify the potential for unmeasured confounding effects, the E-value was calculated for all the HRs observed in Model 3, as described in [35] (<https://www.evalue-calculator.com/>). This represents the minimum association required for a potential unmeasured confounder with both the exposure and the outcome to explain away the observed association. In other words, the higher the E-value, the harder it is to attribute an association to an unmeasured covariate [36].

**2.2.4. Associations with Additional Endpoints.** To better evaluate the associations of Covid-19 patients clusters and HCQ use with negative outcomes other than death, we built a composite endpoint based on the occurrence of at least one of the following outcomes: in-hospital death, access to intensive care unit during hospitalization, or severe disease manifestation (either severe pneumonia or ARDS). In this case, the resulting binary variable was assigned a value of 1. Conversely, the variable got "0" value if one of the following alternative conditions is applied: (i) none of the above mentioned outcomes was verified; (ii) a patient survived without recurring to intensive cares or (iii) without showing severe manifestations of the disease. Six patients with missing values on survival were removed. Then, we modelled the risk of manifesting a bad outcome through a logistic regression (*glm()* function in R), modelling both additive and interactive models of Covid-19 patients cluster and HCQ use, as above. This analysis was motivated by the fact that the course of disease often differs across patients, e.g., with some subjects with less severe forms suddenly

worsening their conditions until death and others having severe manifestations but still surviving, possibly thanks to intensive cares. Therefore, a composite outcome variable represented a robust way to measure potential risk/protective effects of patients' clusters and HCQ use.

### 3. Results

While both agglomerative and divisive clustering approaches were developed, all statistical analyses presented below are based on clusters identified through the latter approach, since this showed a high homogeneity with the results of agglomerative clustering (see Methods section), and divisive clustering has been reported to be more accurate and robust [32, 34].

**3.1. Characteristics of the Clusters.** We identified two clusters of Covid-19 patients (with  $N=3,913$  and  $483$ , Figure 1). A comparison of the continuous variables used for their determination is reported in Figures 2(a)–2(d). The larger cluster was younger (mean (SD) age:  $65.2$  ( $15.6$ ) vs  $77.9$  ( $9.2$ ) years;  $t$ -test =  $-25.9$ ,  $p < 10^{-15}$ ), with better kidney function (eGFR:  $77.9$  ( $26.9$ ) vs  $52.6$  ( $25.6$ ) mL/min/1.73 m<sup>2</sup>;  $t$ -test =  $20.3$ ,  $p < 10^{-15}$ ) and lower circulating inflammation levels (CRP:  $34.6$  ( $62.4$ ) vs  $36.5$  ( $58.6$ ) mg/L; Wilcoxon-test =  $847,660$ ,  $p = 2.5 \times 10^{-14}$ ). Conversely, BMI did not show strong differences between the two clusters ( $28.0$  ( $4.2$ ) vs  $27.5$  ( $4.2$ ) Kg/m<sup>2</sup>;  $t$ -test =  $2.3$ ,  $p = 0.02$ ). Moreover, patients belonging to the larger cluster were less frequently men ( $60\%$  vs  $75\%$ ) and smokers ( $11.5\%$  vs  $19.5\%$ ) and showed a lower prevalence of chronic health conditions like myocardial infarction, heart failure, diabetes, hypertension, cancer, and lung disease (all  $p < 0.0001$ ), while no significant difference was observed in the prevalence of obesity (Table 1).

Clusters were also associated with severe Covid-19 disease manifestations, with  $65.8\%$  of patients in the smaller cluster presenting with either severe pneumonia or ARDS, compared to  $45.9\%$  in the larger cluster (Chi-squared =  $76.4$ ,  $p < 10^{-15}$ ; Table S1). For the characteristics mentioned above, the large and small clusters will be hereafter named as "low risk" and "high risk" cluster. When we compared the use of specific drugs, in the high risk cluster, we observed a less frequent use of HCQ ( $p < 0.001$ ) and of Lopinavir/Darunavir ( $p < 0.05$ ) and a more frequent use of corticosteroids and antihypertensive medications ( $p < 0.001$ ), compared to the low risk cluster (Table S2).

**3.2. Combined Influence of Clusters and HCQ Use on Mortality Risk.** In Cox PH regressions modelling mortality risk as a function of clusters (Model 1), we analyzed  $4,319$  patients with a case-complete approach, with a total of  $799$  deaths and a total of  $73,924$  person-days follow-up (median  $13$  days). In this model, patients belonging to the high risk cluster showed a significant increase of incident mortality risk, compared to those of the low risk cluster (HR [CI] =  $3.81$  [ $3.12$ – $4.65$ ]; Table 2). This association remained stable in a Cox regression modelling additive effects of clusters and HCQ use (Model 2:

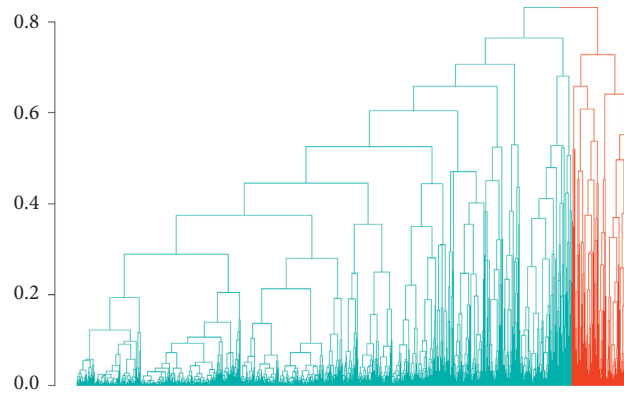


FIGURE 1: Hierarchical divisive clustering of Covid-19 hospitalized patients. Two main clusters of patients were identified, with  $N = 3,913$  (green) and 483 (red), respectively. Each line on the  $x$  axis represents a patient, while on the  $y$  axis the Gower distance between patients is reported. The higher the distance, the later the two patients join into a subcluster, and the more dissimilar they are.

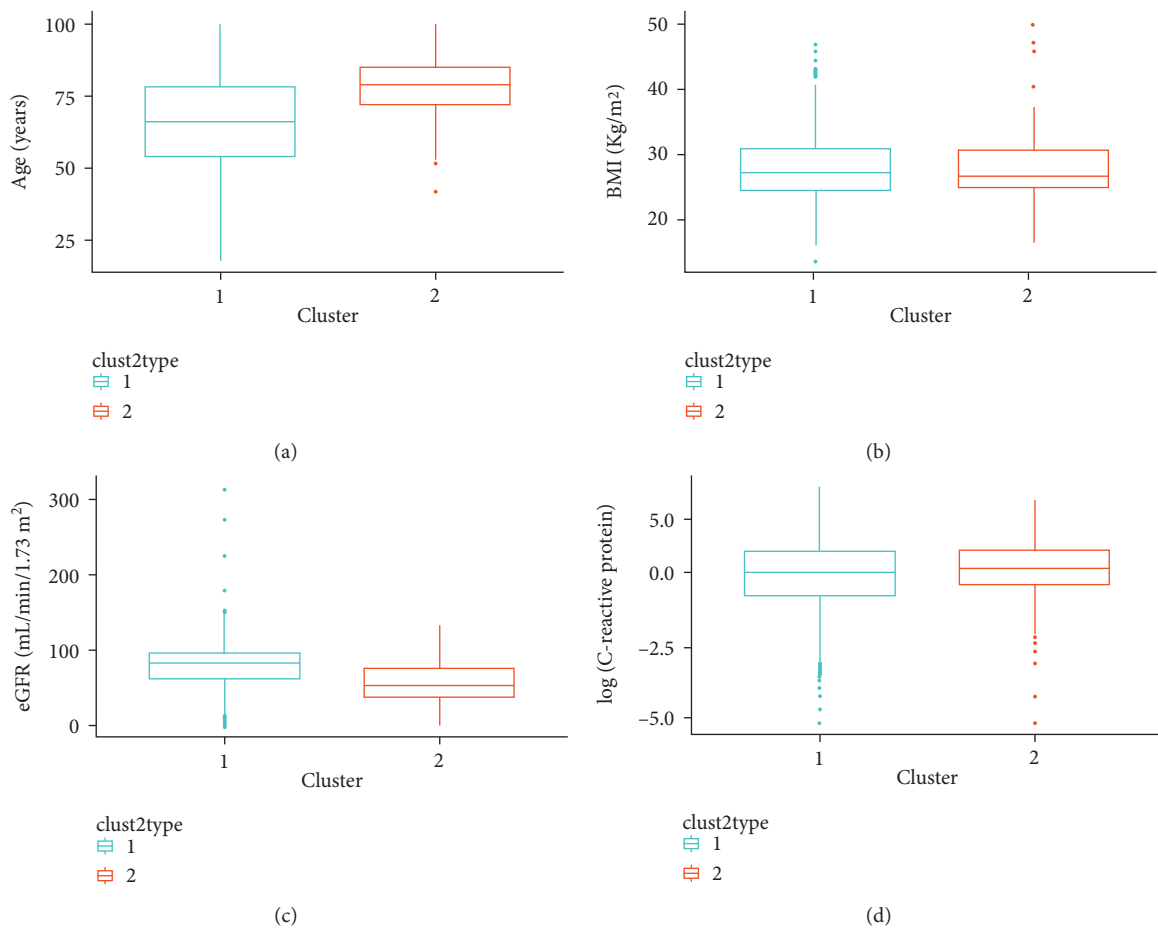


FIGURE 2: Characteristics of sample according to the two clusters identified. Comparison of the continuous variables used for hierarchical clustering—including (a) age (years), (b) BMI ( $\text{Kg}/\text{m}^2$ ), (c) eGFR ( $\text{mL}/\text{min}/1.73 \text{ m}^2$ ), and (d) C-reactive protein plasma levels ( $\text{mg}/\text{L}$ , log-scale) between the two clusters of Covid-19 patients identified, namely, the low (green) and the high risk (red) cluster. Here, these variables are represented through boxplots, with boxes showing the interquartile ranges ( $\text{IQR} = \text{Q1}-\text{Q3}$ ), continuous lines showing the whole distribution range from  $\text{Q1} - 1.5 \cdot \text{IQR}$  through  $\text{Q3} + 1.5 \cdot \text{IQR}$ , and dots showing more extreme values in the dataset.

$N = 4,212$ , 743 events, a total of 72,239 person-days follow-up, median 14 days). Indeed, the high risk cluster was associated with a significant increase of mortality (3.80 [3.08-4.67]), while

HCQ use was associated with a significant independent reduction (0.51 [0.43-0.61]). When we modelled additive and interactive associations of clusters and HCQ in a single model

TABLE 1: Comparison of main categorical variables between the two clusters identified.

Category (%)	Cluster 1 – low risk $N = 3,913$	Cluster 2 – high risk $N = 483$	p for difference
Men	2,346 (60.0%)	362 (74.9%)	$6 \times 10^{-11}$
Smoke			
Current smokers	450 (11.5%)	94 (19.5%)	$<10^{-15}$
Previous smokers	268 (6.8%)	94 (25.3%)	
Obesity ( $BMI \geq 30 \text{ Kg/m}^2$ )	546 (13.9%)	64 (13.3%)	0.73
Myocardial infarction	127 (3.2%)	335 (69.4%)	$<10^{-15}$
Heart failure	171 (4.4%)	315 (65.2%)	$<10^{-15}$
Diabetes	621 (15.9%)	276 (57.1%)	$<10^{-15}$
Hypertension	1,828 (46.7%)	453 (93.8%)	$<10^{-15}$
Cancer	392 (10.0%)	89 (18.4%)	$2 \times 10^{-07}$
Lung disease	415 (10.6%)	207 (42.8%)	$<10^{-15}$

P for difference resulting from comparison of the clusters—through Fisher’s Exact Test (for binary variables) or Chi-squared test (for nonbinary categorical variables, i.e., smoke)—are reported, along with absolute and % frequency of each condition within each cluster.

TABLE 2: Results of Cox PH regressions modelling incident mortality risk.

Model	N (deaths)	Cluster 2 vs 1	HCQ Yes vs no	Cluster*HCQ
Model 1: Death ~ cluster	4,319 (799)	<b>3.81 [3.12-4.65]</b> ( $<10^{-15}$ )		
Model 2: Death ~ cluster + HCQ	4,212 (743)	<b>3.80 [3.08-4.67]</b> ( $<10^{-15}$ )	<b>0.51 [0.43-0.61]</b> ( $8.8 \times 10^{-15}$ )	
Model 3: Death ~ cluster + HCQ + Cluster*HCQ	4,212 (743)	<b>2.45 [1.69-3.54]</b> ( $4.9 \times 10^{-4}$ )	<b>0.46 [0.38-0.55]</b> ( $<10^{-15}$ )	<b>1.90 [1.21-2.96]</b> ( $2.1 \times 10^{-4}$ )
Model 4: Death ~ cluster + HCQ + Cluster*HCQ + other drugs	3,736 (664)	<b>1.65 [1.20-2.26]</b> ( $2.2 \times 10^{-3}$ )	<b>0.52 [0.43-0.63]</b> ( $2.0 \times 10^{-11}$ )	<b>1.98 [1.36-2.89]</b> ( $4.0 \times 10^{-4}$ )

Associations between incident mortality risk, Covid-19 clusters identified, and use of Hydroxychloroquine (HCQ) were tested in the incremental models and in a sensitivity analysis including all the drugs used for Covid-19 treatment. No other covariates were included in the analysis. Hazard Ratios with 95% confidence intervals (HR [CI]) and relevant  $p$ -values (in brackets) are reported. Significant HRs ( $p < 0.05$ ) are highlighted in bold.

(Model 3), we observed a substantially stable protective association of HCQ (0.46 [0.38-0.55]), a reduced but still significant direct association of the “high risk” cluster (2.45 [1.69-3.54]), and a significant association of the cluster\*HCQ interaction term with incident mortality ( $p = 2.1 \times 10^{-4}$ ). This was driven by a differential association of HCQ use within the different clusters, since this was associated with a notable reduction of mortality risk in the low risk cluster (HR [CI] = 0.46 [0.39-0.54],  $p < 10^{-15}$ ) and with a milder nonsignificant reduction in the high risk cluster (0.89 [0.65-1.22],  $p = 0.47$ ). The abovementioned associations were quite robust against potential unmeasured confounding effects, with E-values of 3.1, 2.8, and 2.5 for the associations of mortality with the risk cluster, HCQ use, and cluster\*HCQ interaction in Model 3. Moreover, these associations remained substantially stable in Cox PH models including additional drugs in use (Table 2, Model 4), as well as in those including an interaction term with time (Table S3).

### 3.3. Associations with a Combined Covid-19 Outcome.

When we modelled the risk of bad clinical outcomes of the disease—i.e., severe Covid-19 manifestations, access to intensive care unit or death—as a function of clusters and HCQ use, we observed results in line with survival analyses, with increased risk for cluster 2 and decreased risk for HCQ users, both in the additive and in the interactive model

(Table 3). While the cluster\*HCQ interaction showed only a trend of significance ( $p = 0.08$ ), HCQ still presented a significant protective association in the low risk cluster (OR [CI] = 0.67 [0.56-0.79],  $p = 4.0 \times 10^{-6}$ ) and a substantially null association in the high risk cluster (OR [CI] = 0.98 [0.66-1.46],  $p = 0.92$ ).

## 4. Discussion

In the present work, we report differential influence of HCQ treatment on Covid-19 mortality through a hierarchical clustering analysis applied to patients hospitalized for SARS-CoV-2 infection. This revealed the existence of two separate clusters of Covid-19 patients, based on their clinical and sociodemographic characteristics: one of younger patients with less comorbidities, lower circulating inflammation, and better renal function (“low risk” cluster), and one of older and more comorbid patients, more prevalently men and smokers (“high risk” cluster). The former cluster showed a higher prevalence of severe manifestations of Covid-19, ranging from severe pneumonia to ARDS. Moreover, survival analyses showed an almost four-fold increase of incident in-hospital mortality for the high risk compared to the low risk cluster. Although a previous study attempted to identify subtypes of Covid-19 patients, associating them with disease severity [37], this represents the first attempt to use clustering in disentangling the effect of HCQ on different

TABLE 3: Results of logistic regressions modelling Covid-19 composite bad outcome risk.

Model	N	Cluster 2 vs 1	HCQ	Cluster*HCQ
Bad outcome ~ cluster	4,373	<b>2.53 [2.09-3.08]</b> ( $<10^{-15}$ )		
Bad outcome ~ cluster + HCQ	4,265	<b>2.53 [2.08-3.09]</b> ( $<10^{-15}$ )	<b>0.71 [0.61-0.83]</b> ( $2.1 \times 10^{-5}$ )	
Bad outcome ~ cluster + HCQ + Cluster*HCQ	4,265	<b>1.94 [1.35-2.79]</b> ( $3.8 \times 10^{-4}$ )	<b>0.67 [0.56-0.79]</b> ( $4.0 \times 10^{-6}$ )	1.46 [0.95-2.26] 0.08
Bad outcome ~ cluster + HCQ + Cluster*HCQ + other drugs	3,786	<b>1.80 [1.21-2.68]</b> ( $3.8 \times 10^{-3}$ )	<b>0.55 [0.45-0.66]</b> ( $9.0 \times 10^{-10}$ )	1.47 [0.92-2.36] (0.10)

The composite bad outcome was defined as one of the following: death, access to intensive care unit, or severe Covid-19 manifestation (either severe pneumonia or ARDS). Associations were tested in three incremental models and in a sensitivity analysis including all the drugs used for Covid-19 treatment, as for Cox PH regressions. Odds Ratios with 95% confidence intervals (OR [CI]) and relevant  $p$ -values (in brackets) are reported. Significant ORs ( $p < 0.05$ ) are highlighted in bold.

types of patients, by testing associations with incident in-hospital mortality risk. Specifically, we tested and observed both additive and interactive associations of HCQ and Covid-19 subtypes. Indeed, the high risk cluster was consistently associated with increased mortality across all models, while treatment with HCQ was generally associated with a halving of death risk, in line with previous evidence from both observational [6–13] and experimental studies [19]. While we already reported evidence suggesting a protective influence of HCQ against mortality in a largely overlapping sample [13], here, we have further deepened this relationship by testing and reporting a significant association between cluster-by-HCQ interaction and mortality, which was driven by a differential association within the two clusters. Indeed, the low risk cluster showed a significant “protective” influence of HCQ on in-hospital deaths, while the high risk cluster showed a concordant but nonsignificant association. This represents an element of novelty of the present study, since, in our previous work, we observed a “protective” association between HCQ and mortality within the totality of patients (about 75% of the current sample size), and when stratifying by age, sex, and other characteristics [13], but not within different subtypes of patients combining all these characteristics together in a nonlinear setting, as can be built through unsupervised ML algorithms. Moreover, here, our evidence is supported also by concordant associations with a composite and possibly more robust outcome of the disease, based on the occurrence of death, access to ICU, and severity of manifestations.

Recently, an approach based on the definition of subtypes of Covid-19 patients has been already proven to be successful in identifying which patients benefit most from HCQ treatment, in a multicenter trial involving six US hospitals and 290 patients hospitalized for Covid-19, the IDENTIFY study [19]. HCQ treatment was associated with higher survival in the treated harm, and especially within those patients that were predicted to benefit most based on a supervised ML algorithm applied to their characteristics, which included blood pressure, heart rate, temperature, respiratory rate, oxygen saturation, white blood cell and platelet count, lactate, blood urea nitrogen, creatinine, and bilirubin levels [19]. Interestingly, lactate and creatinine levels were the most important features in this algorithm

[19], the latter representing an index of renal function, which was also a characteristic feature of the low risk cluster in the present study, where HCQ was more effective. Moreover, patients eligible for HCQ treatment as derived by the algorithm of [19] were shown to be younger and less comorbid than the whole population studied, in line with the evidence reported in the present work, suggesting that HCQ treatment may be more effective for younger patients with better general health conditions.

**4.1. Strengths and Limitations.** Although, to the best of our knowledge, this study represents the largest and broadest cluster analysis on Covid-19 patients and a novel approach in analyzing the influence of a pharmacological treatment on Covid-19 mortality and outcomes, it also presents few limitations.

First, the observational retrospective design does not allow us to completely control for confounders and randomization of treatments across individuals. The former aspect is quite unlikely, since a potential residual confounder should be strongly associated with in-hospital mortality to take away observed associations in the interactive model, as suggested by the computed E-values [35, 36]. As for drug therapy, we cannot rule out that assignment to specific treatments was driven also by clinical conditions of the participants, as usually found in common clinical practice. For the same reason, the protective association observed for HCQ may be hypothesized to be driven by other coadministered medications. However, here, HCQ and patients’ cluster showed significant independent associations, which remained substantially stable across models and survived correction for other drugs in use for Covid-19 treatment. Lastly, our evidence is in contrast with RCTs published so far [14–18], which are commonly conceived as the gold standard for establishing drug efficacy and safety. While we generally agree with this view, we would like to underline that these studies did not randomize patients to treatment arms based on combinations of their features, but rather based on single characteristics such as age and sex. This may be the reason for this discrepancy, along with the hypothesis that the high dosage of HCQ administered in RCTs may be harmful for



patients, compared to lower dosages reported in observational studies supporting HCQ efficacy [5]. Of interest, a recent critical review underlined the aspect of suboptimal randomization methods of RCTs, which often do not take into account the whole patient profile and disease severity and may lead to misleading conclusions [38].

## 5. Conclusions

Overall, the evidence supported here and elsewhere [19] suggests that HCQ treatment may be more effective in specific subtypes of Covid-19 patients and indicates machine learning as a useful approach to identify the most “promising” patients in terms of success rate of this treatment. In the future, further studies on independent datasets are warranted, possibly using supervised ML techniques as in other clinical settings (e.g., [39, 40]), to validate this hypothesis and test the feasibility of predicting responsiveness to HCQ before intervention. Ideally, a trial administering low dosages of HCQ ( $\leq 400$  mg/day) and randomizing subjects based on their Covid-19 subtype profile rather than on single characteristics may be warranted to clarify the effects of HCQ on mortality risk in SARS-CoV-2 infection, especially within those patients with a “low risk” profile.

This may help solving current controversies on the use of HCQ as a medication for Covid-19 and maximize the efficacy of treatment strategies for this yet largely unknown disease, especially in low-income and developing countries with poorer national health systems.

## Data Availability

The data used to support the findings of this study may be released upon application to the CORIST collaboration board, who can be contacted through corresponding author.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Authors' Contributions

The raw data analyzed in the present work may be made available under approval by each local center involved in the study, in a way which does not affect patients' privacy. ADiC, RDC, and LI conceived the CORIST study. AG, ADiC, and LI conceptualized the present work. AG performed statistical analyses and wrote the first draft of the manuscript, under the supervision of ADiC and LI. All the co-authors contributed to collection, curation, and elaboration of the analyzed data, and/or to a critical review and editing of the manuscript. Di Castelnuovo and Gialluisi contributed equally to this manuscript.

## Acknowledgments

AG was supported by Fondazione Umberto Veronesi. The authors thank the participating clinical centers included in this cohort. This Article is dedicated to all the patients who suffered or died, often in solitude, due to COVID-19; their

tragic fate gave us moral strength to start and carry out this project. The authors are responsible for the views expressed in this article, which do not necessarily represent the views, decisions, or policies of the Institutions they are affiliated with. A preprint of this article is available at <https://www.medrxiv.org/content/10.1101/2021.01.27.21250238v1>.

## Supplementary Materials

Figure S1. Cross-patient dissimilarity matrix based on Gower distance. Figure S2. Optimal number of clusters (k) to apply in the hierarchical clustering analysis. Figure S3. Check for basic assumptions of Cox PH models. Table S1. Contingency table of Covid-19 cluster by disease severity. Table S2. Contingency tables of Covid-19 clusters by drugs used for treatment. Table S3. Results of Cox PH regressions modelling incident mortality risk, including interaction terms with time. (*Supplementary Materials*)

## References

- [1] X. Yao, F. Ye, M. Zhang et al., “In vitro antiviral activity and projection of optimized dosing design of hydroxychloroquine for the treatment of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2),” *Clinical Infectious Diseases*, vol. 71, no. 15, pp. 732–739, 2020.
- [2] C. Biot, W. Daher, N. Chavain et al., “Design and synthesis of hydroxyferroquine derivatives with antimalarial and antiviral activities,” *Journal of Medicinal Chemistry*, vol. 49, no. 9, pp. 2845–2849, 2006.
- [3] X. Li, Y. Wang, P. Agostinis et al., “Is hydroxychloroquine beneficial for COVID-19 patients?” *Cell Death & Disease*, vol. 11, 2020.
- [4] A. Cassone, L. Iacoviello, and R. Cauda, “Chloroquine/hydroxychloroquine and COVID-19: need to know more about,” *Future Microbiology*, vol. 15, pp. 13–16, 2020.
- [5] A. Di Castelnuovo, S. Costanzo, A. Cassone et al., “Low Dose Hydroxychloroquine Is Associated with Lower Mortality in COVID-19 a Meta-Analysis of 26 Studies and 44521 Patients, 2020, <https://medrxiv.org/cgi/content/short/2020.11.01.20223958>.
- [6] S. Arshad, P. Kilgore, Z. S. Chaudhry et al., “Treatment with hydroxychloroquine, azithromycin, and combination in patients hospitalized with COVID-19,” *International Journal of Infectious Diseases*, vol. 97, pp. 396–403, 2020.
- [7] L. Catteau, N. Dauby, M. Montourcy 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*, vol. 56, 2020.
- [8] L. Ayerbe, C. Risco-Risco, and S. Ayis, “The association of treatment with hydroxychloroquine and hospital mortality in COVID-19 patients,” *Internal and Emergency Medicine*, vol. 15, no. 8, pp. 1501–1506, 2020.
- [9] F. J. M. D. Novales, G. Ramírez-Olivencia, M. Estébanez et al., “Early Hydroxychloroquine Is Associated with an Increase of Survival in COVID-19 Patients: An Observational Study, 2020, <https://www.preprints.org/manuscript/202005.0057/v1>.
- [10] T. Mikami, H. Miyashita, T. Yamada et al., “Risk factors for mortality in patients with COVID-19 in New York city,” *Journal of General Internal Medicine*, vol. 36, pp. 1–10, 2020.

- [11] T. Sulaiman, A. Mohana, L. Alawdah et al., "The Effect of Early Hydroxychloroquine-Based Therapy in COVID-19 Patients in Ambulatory Care Settings: A Nationwide Prospective Cohort Study, 2020, <http://medrxiv.org/content/early/2020/09/13/2020.09.09.20184143.abstract>.
- [12] B. Yu, C. Li, P. Chen et al., "Low dose of hydroxychloroquine reduces fatality of critically ill patients with COVID-19," *Science China Life Sciences*, vol. 63, no. 10, pp. 1515–1521, 2020.
- [13] A. Di Castelnuovo, S. Costanzo, A. Antinori 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*, vol. 20, 2020.
- [14] G. The RECOVERY Collaborative, "Effect of hydroxychloroquine in hospitalized patients with covid-19," *The New England Journal of Medicine*, vol. 383, pp. 2030–2040, 2020.
- [15] A. B. Cavalcanti, F. G. Zampieri, R. G. Rosa et al., "Hydroxychloroquine with or without azithromycin in mild-to-moderate covid-19," *New England Journal of Medicine*, vol. 383, no. 21, pp. 2041–2052, 2020.
- [16] H. PanR. Peto et al., Repurposed antiviral drugs for covid-19 - interim WHO solidarity trial results WHO Solidarity Trial Consortium 2020, <http://www.ncbi.nlm.nih.gov/pubmed/33264556>.
- [17] C. P. Skipper, K. A. Pastick, N. W. Engen et al., "Hydroxychloroquine in nonhospitalized adults with early COVID-19," *Annals of Internal Medicine*, vol. 173, no. 8, pp. 623–631, 2020.
- [18] P. W. Horby and J. R. Emberson, "Hydroxychloroquine for COVID-19: balancing contrasting claims," *European Journal of Internal Medicine*, vol. 82, pp. 18–19, 2020.
- [19] H. Burdick, C. Lam, S. Mataraso et al., "Is machine learning a better way to identify COVID-19 patients who might benefit from hydroxychloroquine treatment?-the IDENTIFY trial," *Journal of Clinical Medicine*, vol. 9, no. 12, p. 3834, 2020.
- [20] Z. F. Khan and S. R. Alotaibi, "Applications of artificial intelligence and big data analytics in m-health: a healthcare system perspective," *Journal of Healthcare Engineering*, vol. 2020, Article ID 8894694, 15 pages, 2020.
- [21] Z. Zhang, E. P. Navarese, B. Zheng et al., "Analytics with artificial intelligence to advance the treatment of acute respiratory distress syndrome," *Journal of Evidence-Based Medicine*, vol. 13, no. 4, pp. 301–312, 2020.
- [22] M. Barchitta, A. Maugeri, G. Favara et al., "Cluster analysis identifies patients at risk of catheter-associated urinary tract infections in intensive care units: findings from the SPIN-UTI Network," *The Journal of Hospital Infection*, vol. 107, pp. 57–63, 2021.
- [23] L. Wynants, B. Van Calster, G. S. Collins et al., "Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal," *BMJ*, vol. 369, p. 26, 2020.
- [24] A. Di Castelnuovo, S. Costanzo, A. Antinori et al., "RAAS inhibitors are not associated with mortality in COVID-19 patients: findings from an observational multicenter study in Italy and a meta-analysis of 19 studies," *Vascul Pharmacol*, vol. 200, 2020.
- [25] A. Di Castelnuovo, M. Bonaccio, S. Costanzo et al., "Common cardiovascular risk factors and in-hospital mortality in 3,894 patients with COVID-19: survival analysis and machine learning-based findings from the multicentre Italian CORIST Study," *Nutrition, Metabolism and Cardiovascular Diseases*, vol. 30, 2020.
- [26] R Core Team, R: A Language and Environment for Statistical Computing, 2020, <https://www.r-project.org/>.
- [27] F. M. Noor and M. M. Islam, "Prevalence and associated risk factors of mortality among COVID-19 patients: a meta-analysis," *Journal of Community Health*, vol. 45, no. 6, pp. 1270–1282, 2020.
- [28] A. Maugeri, M. Barchitta, and A. Agodi, "A clustering approach to classify Italian regions and provinces based on prevalence and trend of sars-cov-2 cases," *International Journal of Environmental Research and Public Health*, vol. 17, pp. 1–14, 2020.
- [29] A. Kowarik and M. Templ, "Imputation with the R package VIM," *Journal of Statistical Software*, vol. 74, pp. 1–16, 2016.
- [30] J. C. Gower, "A general coefficient of similarity and some of its properties," *Biometrics*, vol. 27, no. 4, p. 857, 1971.
- [31] P. J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53–65, 1987.
- [32] A. Sharma, Y. López, and T. Tsunoda, "Divisive hierarchical maximum likelihood clustering," *BMC Bioinformatics*, vol. 18546 pages, 2017.
- [33] T. M. Therneau and P. M. Grambsch, *Modeling Survival Data: Extending the Cox Model*, Springer, New York, NY, USA, 2000, <https://cran.r-project.org/web/packages/survival/citation.html>.
- [34] M. Roux, "A comparative study of divisive and agglomerative hierarchical clustering algorithms," *Journal of Classification*, vol. 35, no. 2, pp. 345–366, 2018.
- [35] M. B. Mathur, P. Ding, C. A. Riddell, and T. J. VanderWeele, "Web site and R package for computing E-values," *Epidemiology*, vol. 29, no. 5, pp. e45–e47, 2018.
- [36] T. J. VanderWeele and P. Ding, "Sensitivity analysis in observational research: introducing the E-Value," *Annals of Internal Medicine*, vol. 167, no. 4, pp. 268–274, 2017.
- [37] W. Ye, W. Lu, Y. Tang et al., "Identification of COVID-19 clinical phenotypes by principal component analysis-based cluster Analysis," *Frontiers in Medicine*, vol. 7, Article ID 570614, 2020.
- [38] V. R. Emani, S. Goswami, D. Nandanor, S. R. Emani, N. K. Reddy, and R. Reddy, "Randomised controlled trials for COVID-19: evaluation of optimal randomisation methodologies—need for data validation of the completed trials and to improve ongoing and future randomised trial designs," *International Journal of Antimicrobial Agents*, vol. 57, Article ID 106222, 2020.
- [39] M. Barchitta, A. Maugeri, G. Favara et al., "Early prediction of seven-day mortality in intensive care unit using a machine learning model: results from the SPIN-uti project," *Journal of Clinical Medicine*, vol. 10, no. 5, p. 992, 2021.
- [40] M. Barchitta, A. Maugeri, G. Favara et al., "A machine learning approach to predict healthcare-associated infections at intensive care unit admission: findings from the SPIN-UTI project," *Journal of Hospital Infection*, vol. 112, 2021.