COVID-19: beta-thalassemia subjects immunised?

Edouard Lansiaux¹, Philippe Pierre Pébaÿ², Jean-Laurent Picard³, Joachim Son-Forget⁴

¹ Med student, Henry Warembourg Medecine School, Lille University, 59000, Lille, France
² Distinguished scientist, NexGen Analytics, 82801, Sheridan, Wyoming, United States of America
³ University Lecturer, Conservatoire National des Arts et Métiers, 75141, Paris, France
⁴ MD, PhD and Member of Parliament, Assemblée Nationale, 75355, Paris, France

Corresponding author: Edouard Lansiaux

26, rue de la Presqu’île, 59870, Marchiennes, France
+33781112194
edouard.lansiaux.etu@univ-lille.fr

The authors declare no conflicts of interest.

Keywords: Novel Coronavirus; respiratory distress; Favipiravir; statistics; correlation; beta thalassemia; immunisation; Italy; Sardinia; regression; heme
ABSTRACT

The novel coronavirus pneumonia (COVID-19) is a contagious acute respiratory infectious disease whose causative agent has been demonstrated to be a novel virus of the coronavirus family, SARS-CoV-2. A recent pre-print study has showed a heme attack on the 1-beta chain of hemoglobin by COVID19. Beta-thalassemia results of a default in the hemoglobin beta-chain synthesis. 1.5% global population are heterozygotes for this disease. In this study, by a multiple linear regression, we have analyzed the evolution of COVID-19 infection in three Italian regions (Puglia, Sardinia, Sicilia) with different beta-thalassemic prevalences, in order to search a link. The results have showed that betathalassemic heterozygote population prevalence is correlated to immunity against COVID-19, by a regression. This paper is only for academic discussion, the hypotheses and conclusions needs to be confirmed by further research.
BACKGROUND
The novel coronavirus pneumonia (COVID-19) is a contagious acute respiratory infectious disease whose causative agent has been demonstrated to be a novel virus of the coronavirus family, SARS-CoV-2. Patients with the coronavirus pneumonia have a fever, and the temperature above 38 degrees with symptoms such as dry cough, fatigue, dyspnea, difficulty breathing, and diarrhea. Thrombotic complications are also commonly reported such as acute pulmonary embolism. This pneumonia was first discovered in December 2019 in the South China Seafood Market Hubei Province, China. There is a high contagiosity for this disease. This pneumonia has now turned into a pandemic: Tens of thousands of people are infected worldwide.

HYPOTHESIS
A recent pre-print study shows that ORF8 and surface glycoproteins of the novel coronavirus could combine to the porphyrin to form a complex. Meanwhile, orf1ab, ORF10, and ORF3a proteins could coordinate attack the heme (porphyrin), formed into the mitochondria, on the 1-beta chain of hemoglobin to dissociate the iron ions from the heme. Furthermore, it revealed that Favipiravir could inhibit the envelope protein and ORF7a protein bind to porphyrin, thus preventing the virus from entering host cells, and catching free porphyrins. Explain here also quickly potential mechanisms of chloroquine if possible.

In addition, B-thalassemia has an interesting physiopathology in this context: genetic mutations induce decreased or even absent beta chains synthesis whereas alpha chains synthesis remains normal. This results in the excess of free alpha chains and less hemoglobin A. Either free alpha chains associate themselves to form unstable 4-alpha tetramers, which subsequently oxidize...
and precipitate themselves in erythroblast cytoplasm, (inducing oxygenated radicals liberation and then surface glycoproteins oxidation, and then dyserythropoiesis and hemolysis) or they are proteolyzed\textsuperscript{1,3,16,19}.

The incidence and distribution of B-thalassemia syndromes in Italy and in Sardinia have been previously and broadly documented\textsuperscript{2,4,9,10,15,20,21}. Prevalence of B-thalassemia heterozygotes in Sardinia (12.9\%) is higher than in Sicilia, which is itself higher than in Puglia.

In this context, we hypothesized that beta-thalassemic patients, most concentrated in Sardinia, could develop an immunity to SARS-CoV-2 infectious consequences on haemoglobin, as the beta chain, potential target of the virus, could be either absent or less prominent in the blood.

**EVALUATION OF THE HYPOTHESIS**

**Study and participants**

We conducted a multi-group descriptive observational transversal study in order to define a hypothetical relationship between beta thalassemia and SARS-CoV-2 immunization. The source and targeted populations are the whole humanity in view of the ongoing COVID-19 pandemic. The eligible population is constituted by all Italians.

The study was directed by a consortium of two data analysts, a MD-PhD specialized in radiology and brain research, and a medical student in clinical years. NexGen Analytics had no role in designing the study, or making the decision to submit manuscript to the publication, nor did receive any fee or compensation in the context of this work. The first author vouches for the data and analyses, as well as for the fidelity of this report to the study protocol.
Enrollment

Population included and studied gather people of three southern Italian regions: Puglia, Sardinia and Sicilia. There were no exclusion criteria. This choice was made according to differences upon beta-thalassemia heterozygote prevalence between all Italian regions and upon the geographical localization (which allowed us to eliminate environmental variables interference).

We gathered COVID-19-related data from various public health and social sources. A parallel multiple group analysis was performed, and demographic data for each group was collected with reference date 01/01/2020.

Outcome measures

Three primary variables were taken into account: accrued confirmed cases of COVID-19; accrued deaths causatively associated to COVID-19 infection; and accrued healed to death ratio.

The primary and second study outcomes were respectively the accumulated confirmed cases of COVID-19 and the accumulated deaths due to COVID-19. These were useful in order to assess the evolution of pandemic. The secondary study outcome was the accumulated deaths due to COVID-19 infection. To the immune status of the population, we used the daily ratio healed upon dead by day as the third outcome.

Statistical analysis

We found that the logistic function can be used to modelize various kind of epidemiological data such as the number of confirmed cases or the number of deaths. After performing regressions, we found that the calculated functions surprisingly fit many reported data, with NRMSE (Normalized Root Mean Square Error) usually below 0.1%. This excellent fit can be observed at different geographic scales (we studied country-level data as well as smaller geographical subdivisions such as provinces or cities). When the model is relevant to the
reported data, we observe the following as we gather more data over time: 1/decreasing NRMSE, 2/ the 3 degrees of freedom of the logistic curve should also converge. Regressions will provide the values of the 3 parameters / degrees of freedom of the logistic curve: its maximum (L), the abscissa of its midpoint (x₀), and its growth rate (k). The abscissa of the curve's midpoint indicates when the peak of the epidemics will occur / occurred.

\[
logistic\ function\ f(t) = \frac{L}{1-e^{-k(x-x_0)}},
\]

For each cluster, a linear regression was established to show contrasts. To assess these differences, a comparison of the linear regression coefficient (A) was done.

EMPIRICAL DATA

Concerning the primary outcome (accumulated confirmed positive counts), Sicilia is substantially less impacted than Puglia for a population almost similar (respectively 4,999,891 and 4,029,053). Two regions have nearly reached their stage phase, except Sardinia. The confirmed prevalence of COVID-19 on 04/10/2020 is: 0.070% for Puglia, 0.046% for Sicilia, 0.065% for Sardinia (Table 2).

Concerning the secondary outcome (accumulated deceased counts), Sicily is significantly less impacted than Puglia for a population almost similar (respectively 4,999,891 and 4,029,053 inhabitants). Two regions have nearly reached their stage
phase, except Sardinia. The death rate due to COVID-19 (number of deaths due to COVID-19 upon number of infected people by COVID-19) on 04/10/2020 is: 8.47% for Puglia, 6.43% for Sicilia, 6.49% for Sardinia (Table 3).

The third outcome is the accrued healed to dead ratio of COVID-19 infected patients (Table 4). This tracker has allowed us to define the immunization of population. Different variations could be observed: a Sardinian and Sicilian peak around 03/04/20 and 03/05/20. On 04/10/20, this ratio is 0.987 for Puglia, 1.735 for Sardinia, 1.263 for Sicilia.

The relationship of this last plot (Table 4) with the beta-thalassemic prevalence has been drawn through multiple regressions (Tables 4, 5, 6). As a first step, we have ordered data with beginning on the first day where the healed to dead ratio is different from 0 (for Puglia on 03/04/20; for Sardinia on 03/20/20; for Sicilia on 03/12/20). Then, a linear regression was done. This one has allowed to show, by the linear coefficient (A), that ratio increased faster in Sardinia ($a=0.0751$) than in Sicilia ($A=0.0714$) and faster in Sicilia than in Puglia ($A=0.0187$).

Finally, we had to calculate the regression between each regional prevalence and the regional ratio linear regression through a matrix (Table 5).

**DISCUSSION**

Using a multiple linear regression, we have shown that thalassemia heterozygote population prevalence is significantly correlated to the hypothesized immunity against COVID-19 (Table 6).

This study is of course limited by different important factors. Firstly, the short studying period (1.5 month) didn’t allow to predict perfectly the next evolution of this correlation. That’s a crucial point in this pandemic context. The second limit of our study concerned the population. Indeed, Italy has an important prevalence of B-thalassemic
heterozygotes (1/20)\textsuperscript{15}, but still, they are part of an aphelion disease (world heterozygote prevalence: 1.5\%)\textsuperscript{4}. The third limit is the immunity tracker; as we didn’t have data concerning serology tests on the Italian population, we have simulated this immunity tracker (accrued healed to dead ratio). Last but not least, there could be a major concern with the influence of well-known but hereby undocumented co-factors that are associated with bad prognosis in Intensive Care Unit population infected by COVID-19, such as BMI or diabetes (Table 1), and other unknown population co-factors, ranging from genetics to political management (PCR tests, face masks, containment of population, social distancing, treatment at early stages with antiviral drugs).

Nevertheless, our regression, linked with the hypothesized physiopathology\textsuperscript{1,3,16,19}, suggests a first order effect at least. In this global context of medical research concerning the COVID-19 pandemic, our study and its results must be taken into account in order to open new therapeutical and diagnostic perspectives.

Our hypothesis could be confirmed by screening the prevalence of COVID-19 infected among beta-thalassemic patients. In addition, in vitro cell studies and animal models, as the thalassemic erythrocyte or the beta-thalassemic mice\textsuperscript{16}, could be of interest to test our statistical correlation. Further studies could be conducted to identify other correlations between COVID-19 and other blood pathologies such as drepanocytosis, G6PD deficiency or malaria.
ACKNOWLEDGEMENTS

The authors declare no conflicts of interest.
REFERENCES


22. Population and household data [Internet]. Instituto Nazionale di Statistica. Available from: istat.it

### TABLES

<table>
<thead>
<tr>
<th></th>
<th>Puglia</th>
<th>Sardinia</th>
<th>Sicilia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inhabitants (number)</td>
<td>4,029,053</td>
<td>1,639,591</td>
<td>4,999,891</td>
</tr>
<tr>
<td>Middle age (years)</td>
<td>44.7</td>
<td>47.2</td>
<td>44.3</td>
</tr>
<tr>
<td>Sex ratio (male/female)</td>
<td>0.95</td>
<td>0.97</td>
<td>0.95</td>
</tr>
<tr>
<td>Male life expectancy (years)</td>
<td>81.1</td>
<td>80.4</td>
<td>79.9</td>
</tr>
<tr>
<td>Female life expectancy (years)</td>
<td>85.2</td>
<td>85.9</td>
<td>84</td>
</tr>
<tr>
<td>Birth rate</td>
<td>7.1</td>
<td>5.7</td>
<td>7.8</td>
</tr>
<tr>
<td>Death rate</td>
<td>10</td>
<td>10.3</td>
<td>10.8</td>
</tr>
<tr>
<td>BMI (on subjects over 18 years)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;18</td>
<td>2</td>
<td>3.1</td>
<td>3</td>
</tr>
<tr>
<td>18-25</td>
<td>46.9</td>
<td>54.7</td>
<td>47.7</td>
</tr>
<tr>
<td>25-30</td>
<td>37.7</td>
<td>32.3</td>
<td>37.4</td>
</tr>
<tr>
<td>&gt;30</td>
<td>13.5</td>
<td>9.9</td>
<td>11.9</td>
</tr>
<tr>
<td>Diabetes prevalence (%)</td>
<td>7.1</td>
<td>4.7</td>
<td>6.5</td>
</tr>
<tr>
<td>Thalassemic heterozygotes</td>
<td>6.5</td>
<td>12.9</td>
<td>7.5</td>
</tr>
<tr>
<td>prevalence (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Demographic data on 01/01/20
Table 2. Plots with logistic function applied upon accumulated positive confirmed counts for each region (Puglia, Sardinia, Sicilia).
Table 3. Plots with logistic function applied upon deceased counts for each region (Puglia, Sardinia, Sicilia).
Table 4. Accrued healed to dead ratio by region
\[ Y = A^*x + PA \]

\[ = (A_{\text{puglia}}, A_{\text{sardinia}}, A_{\text{sicilia}})^*x + (P_{\text{puglia}}, P_{\text{sardinia}}, P_{\text{sicilia}}) \]

\[ = (0.0187, 0.0751, 0.0714)^*x + (0.065, 0.129, 0.075) \]

A, the linear coefficient from Table 4

P, the thalassemic prevalence from Table 1

Table 5. Matrix for the multiple linear regression
Table 6. Multiple least square adjustment