

# COVID-19: beta-thalassemia subjects immunised?

Edouard Lansiaux<sup>1</sup>, Philippe Pierre Pébay<sup>2</sup>, Jean-Laurent Picard<sup>3</sup>, Joachim Son-Forget<sup>4</sup>

<sup>1</sup> Med student, Henry Warembourg Medecine School, Lille University, 59000, Lille, France

<sup>2</sup> Distinguished scientist, NexGen Analytics, 82801, Sheridan, Wyoming, United States of America

<sup>3</sup> University Lecturer, Conservatoire National des Arts et Métiers, 75141, Paris, France

<sup>4</sup> 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](mailto: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, SARSCoV-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, SARSCoV-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<sup>5,6,12</sup>. Thrombotic complications are also commonly reported such as acute pulmonary embolism<sup>14,23</sup>. This pneumonia was first discovered in December 2019 in the South China Seafood Market Hubei Province, China<sup>11</sup>. There is a high contagiousity 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<sup>8</sup> 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<sup>1,3,16,19</sup>.

The incidence and distribution of B-thalassemia syndromes in Italy and in Sardinia have been previously and broadly documented<sup>2,4,9,10,15,20,21</sup>. 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-thalassaemic 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 thalassaemia 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 gathered 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<sup>7</sup>. A parallel multiple group analysis was performed, and demographic data for each group was collected with reference date 01/01/2020<sup>22</sup>.

## 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 ( $x_0$ ), and its growth rate (k). The abscissa of the curve's midpoint indicates when the peak of the epidemics will occur / occurred.

$$\text{logistic function } f(t) = \frac{L}{1 - e^{-k(x-x_0)}}$$

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

### Populations

7,048,535 people were included in this study (population on 01/01/20) (Table 1)<sup>2,4,9,10,15,20,21,22</sup>.

According to MANOVA test with a risk of 5%, all the groups were well balanced except on the BMI, the diabetes and the thalassemia heterozygote prevalence.

Populations of those three regions have been studied during the period between 02/25/2020 and 04/10/2020.

### Outcomes

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-thalassaemic 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 thalassaemia 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-thalassaemic

heterozygotes (1/20)<sup>15</sup>, but still, they are part of an aphelion disease (world heterozygote prevalence: 1.5%)<sup>4</sup>. 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<sup>1,3,16,19</sup>, 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<sup>16</sup>, 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

1. Arbres décisionnels pour le diagnostic et la caractérisation moléculaire des hémoglobinopathies. *ABC*. 2020;(68):455–64.
2. Carcassi U. Aspetti medico sociali della microcitemia in Sardegna. *Atti Giornate Studio Microcitemia*. 1963 Oct;34–48.
3. Bonello-Palot N, Badens C. Bases moléculaires des syndromes thalassémiques et facteurs génétiques modulateurs de sévérité de la beta-thalassémie. *RMGH*. 2011;
4. Galanello R, Origa R. Beta-thalassaemia. *Orphanet Journal of Rare Diseases*. 2011, (5):11.
5. Wang D, Hu B, Hu C. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in Wuhan, China. *JAMA*. 2020;
6. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet*. 2020 Jan 24;
7. Csuthi E, Montuori M. Coronavirus, la situazione in Italia [Internet]. Gedi Visual. Available from: [lab.gedidigital.it/gedi-visual/2020/coronavirus-i-contagi-in-italia/?refresh\\_ce](http://lab.gedidigital.it/gedi-visual/2020/coronavirus-i-contagi-in-italia/?refresh_ce)
8. Wenzhong L, Hualan L. COVID-19: Attacks the 1-Beta Chain of Hemoglobin and Captures the Porphyrin to Inhibit Human Heme Metabolism. *ChemRxiv*. 2020 Apr 9;
9. Silvestroni E, Bianco I. Dati statistici sulla frequenza della microcitemia e del morbo di Cooley nella Sardegna meridionale. *Nuovi Annali d'Igiene e Microbiologia*. 1960;(11):339–48.
10. Silvestroni E, Bianco I. Diffusione e frequenza della microcitemia e della anemia microcitemiche nell'Italia continentale e in Sicilia. In: *Atti delle Giornate di Studio*. Roma: Istituto Italiano di Medicina Sociale; 1962.

11. Li Q, Guan X, Wu P. Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. *The New England Journal of Medicine*. 2020 Jan 29;
12. Chang D, Minggui L, Lai W. Epidemiologic and clinical characteristics of novel coronavirus infections involving 13 patients outside Wuhan, China. *JAMA*. 2020;
13. Diao K, Han P, Pang T, Li Y, Yang Z. HCTR Imaging Features in Representative Imported Cases of 2019 Novel Coronavirus Pneumonia. *Precision Clinical Medicine*. 2020;
14. Klok FA, Kruip MJHA, van der Meer NJM, et al. Incidence of thrombotic complications in critically ill ICU patients with COVID-19. *Thrombosis Research*. 2020;2020.
15. La talassemia in cifre [Internet]. Associazione Piera Cutino Guarire dalla Talassemia. Available from: <http://www.pieracutino.it/la-talassemia-in-cifre/>
16. Leroy-Viard K, Rouyer-Fessard P, Sauvage C, Scoot Mark, Beuzard Y, Modèles expérimentaux de la beta-thalassémie. *Synthèse*, 1992 Oct;(8):784-789.
17. Siniscalco M, Bernini L, Filippi G, Latte B, Meera Khan P, Piomelli S, et al. Population genetics of haemoglobin variants, thalassemia, and glucose-6-phosphate dehydrogenase deficiency with particular reference to the malaria hypothesis. *Bulletin of the World Health Organization*. 1966;(34):379–94.
18. Li X, Zai J, Wang X, Li Y. Potential of large ‘first generation’ human to human transmission of - - 2019 nCoV. *Journal of Medical Virology*. 2020 Jan 30;
19. Syndromes thalassémiques majeurs et intermédiaires. *Protocole national de diagnostic et de soins pour une maladie rare*. HAS. 2008 Jun;
20. Cao A, Galanello R, Furbetta M, Muroli PP, Garbato L, Rosatelli C, et al. Thalassemia types and their incidences in Sardinia. *Journal of Medical Genetics*. 1979 Jan;(15):443–7.

21. Ceci A, Mangiarini L, Felisi M, Bartoloni F, Cianco A, Capra M, et al. The Management of Iron Chelation Therapy: Preliminary Data from a National Registry of Thalassaemic Patients. *Anemia*. 2011 Apr 11;2011.
22. Population and household data [Internet]. Istituto Nazionale di Statistica. Available from: [istat.it](http://istat.it)
23. Rotzinger DC, Beigelman-Aubry C, von Garnier C, Qanadli SD. Pulmonary embolism in patients with COVID-19: Time to change the paradigm of computed tomography. *Thrombosis Research*. 2020 Apr 8; 2020.

## TABLES

		Puglia	Sardinia	Sicilia
Inhabitants (number)		4,029,053	1,639,591	4,999,891
Middle age (years)		44.7	47.2	44.3
Sex ratio (male/female)		0.95	0.97	0.95
Male life expectancy (years)		81.1	80.4	79.9
Female life expectancy (years)		85.2	85.9	84
Birth rate		7.1	5.7	7.8
Death rate		10	10.3	10.8
BMI (on subjects over 18 years)	<18	2	3,1	3
	18-25	46,9	54,7	47,7
	25-30	37,7	32,3	37,4
	>30	13,5	9,9	11,9
Diabetes prevalence (%)		7,1	4,7	6,5
Thalassemic heterozygotes prevalence (%)		6,5	12,9	7,5

Table 1. Demographic data on 01/01/20

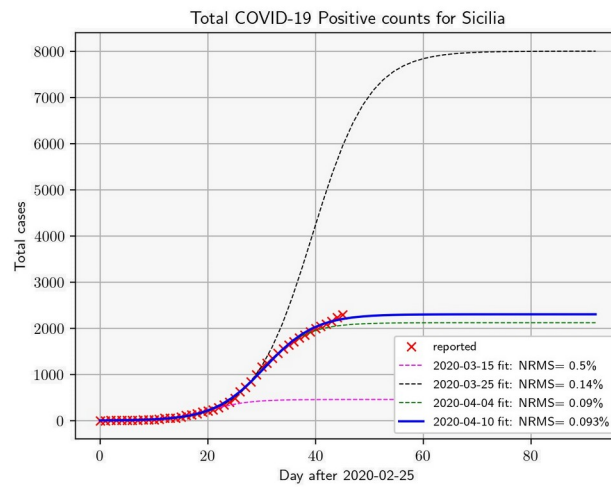
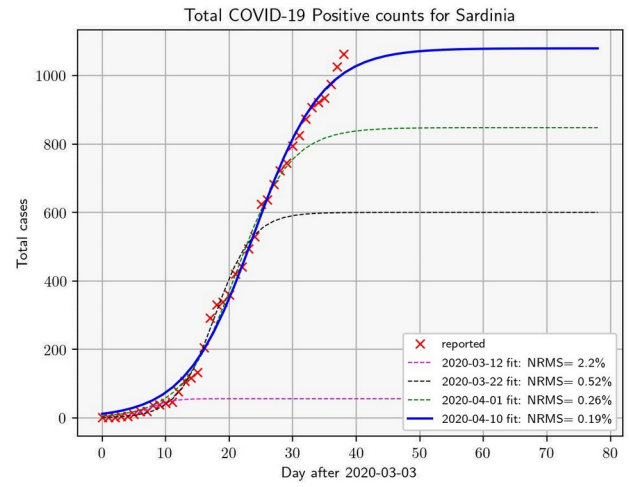
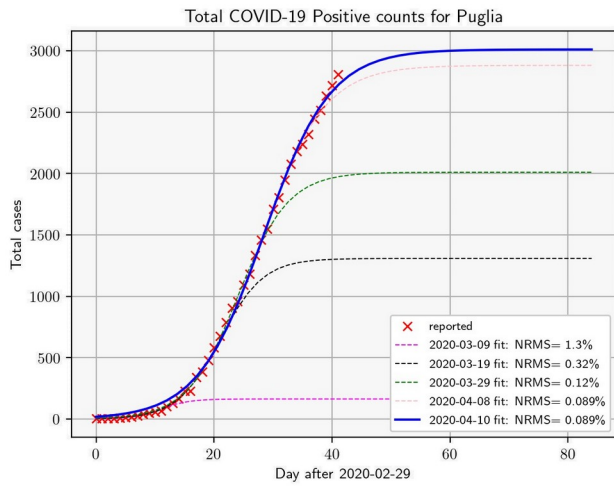


Table 2. Plots with logistic function applied upon accumulated positive confirmed counts for each region (Puglia, Sardinia, Sicilia).

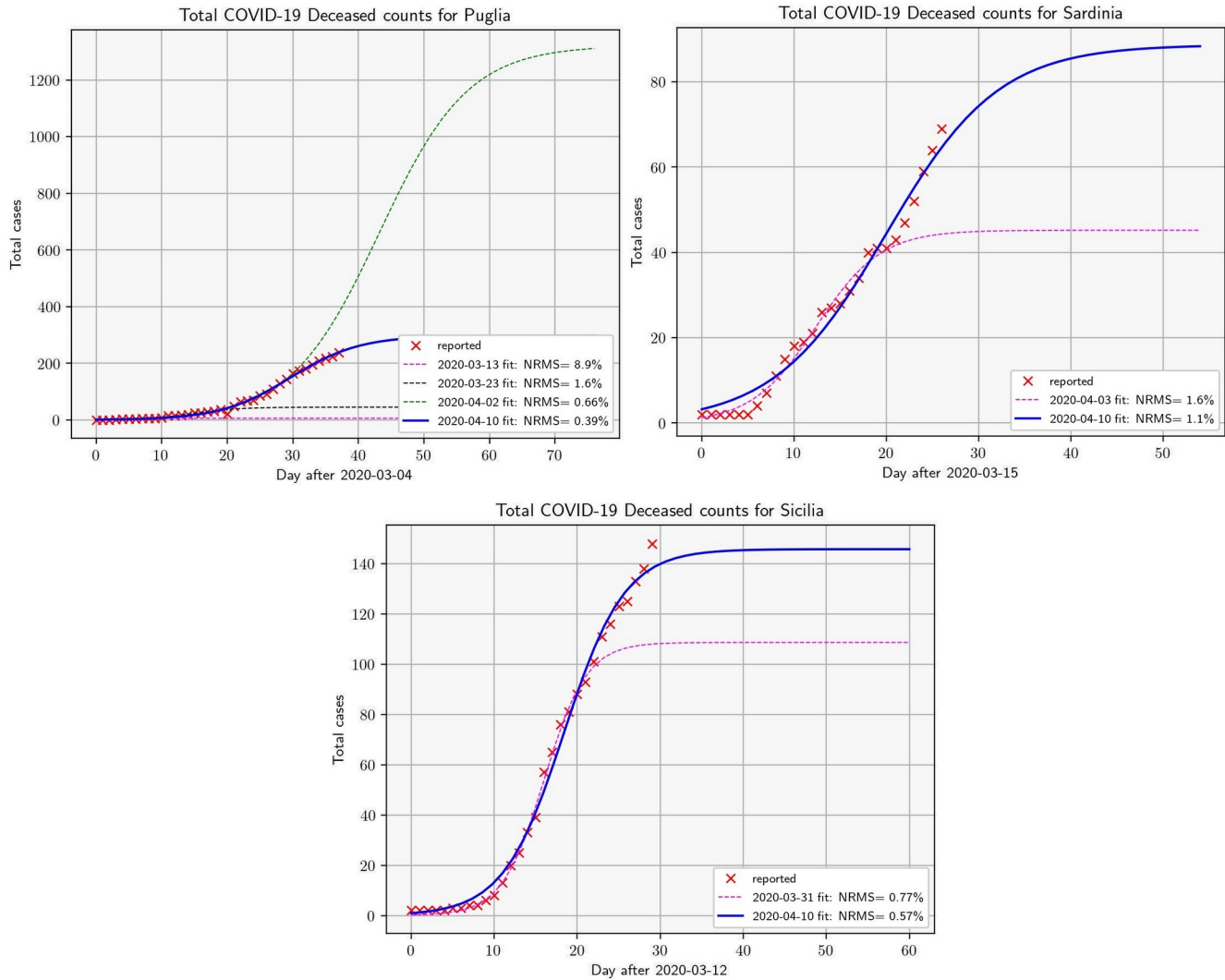


Table 3. Plots with logistic function applied upon deceased counts for each region (Puglia, Sardinia, Sicilia).

### Accrued Healed to Dead Ratio

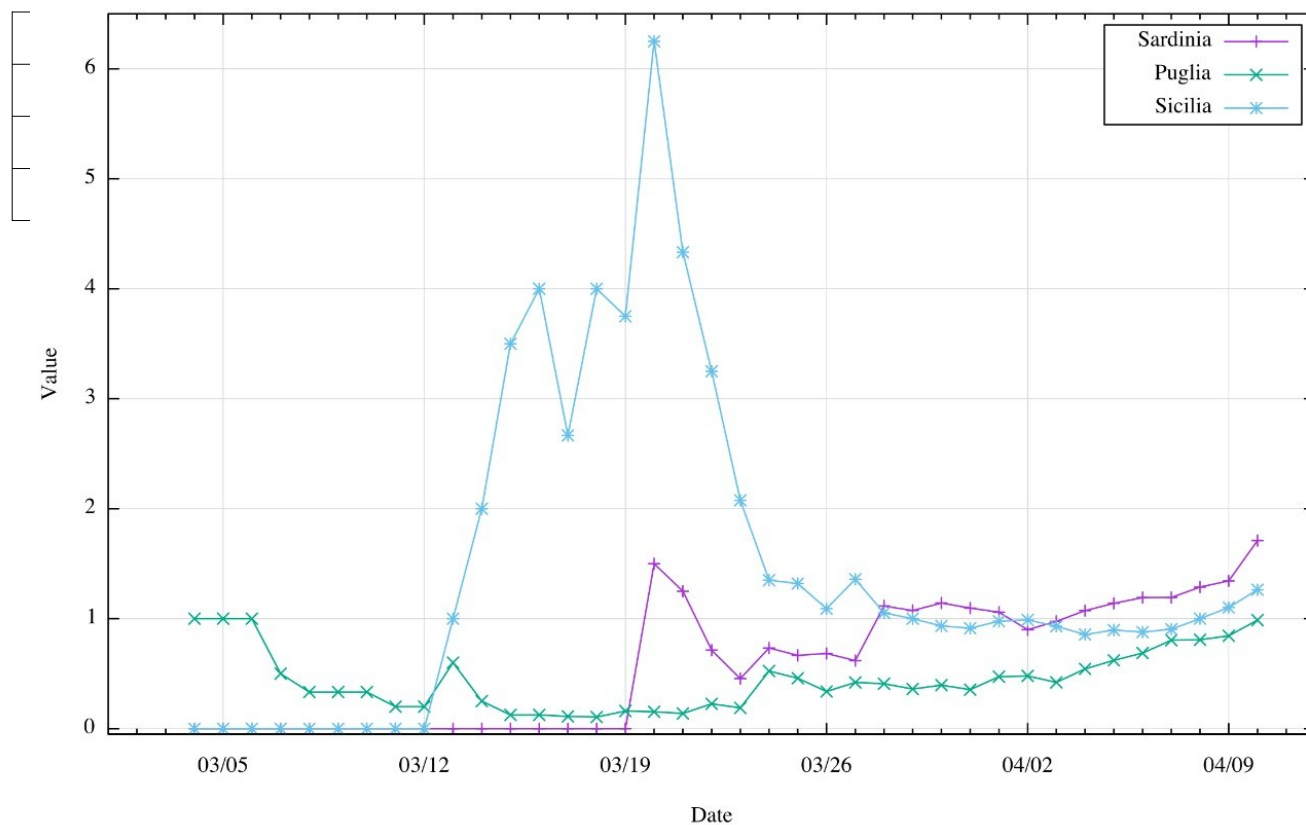


Table 4. Accrued healed to dead ratio by region



$$Y=A*x+PA$$

$$= (A_{puglia}, A_{sardinia}, A_{sicilia}) * X + (P_{puglia}, P_{sardinia}, P_{sicilia})$$

$$= (0.0187, 0.0751, 0.0714) * X + (0.065, 0.129, 0.075)$$

$A_i$  the linear coefficient from *Table 4*

$P_i$  the thalassemic prevalence from *Table 1*

Table 5. Matrix for the multiple linear regression

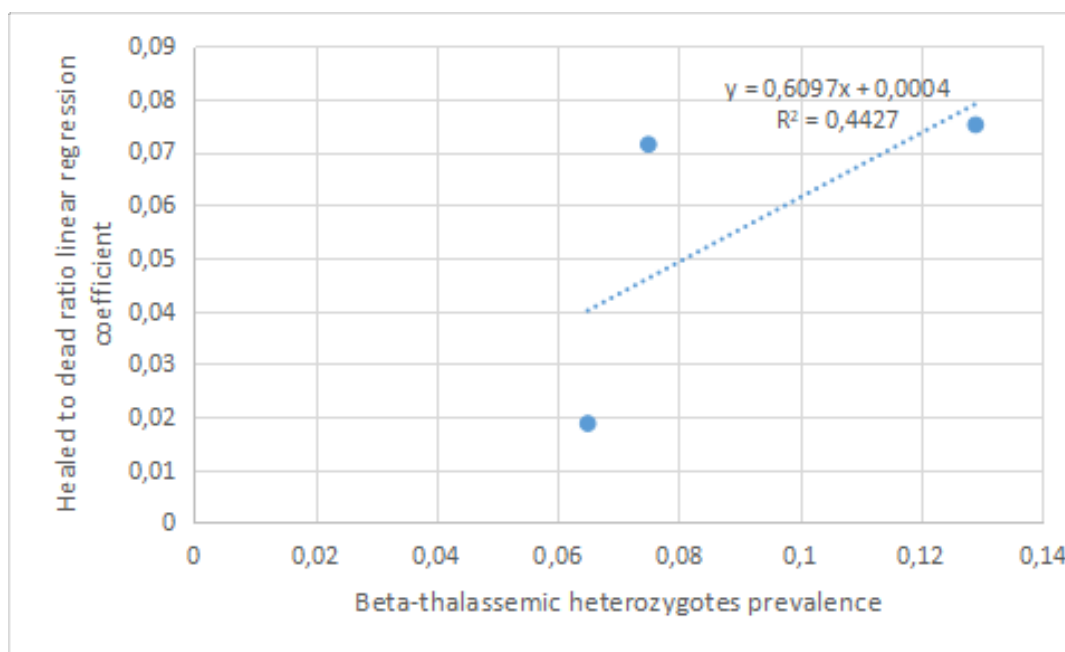


Table 6. Multiple least square adjustment