Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Response to Naudet et al.

Édouard Lansiaux\textsuperscript{a,}\textsuperscript{*}, Philippe P. Pébay\textsuperscript{b}, Jean-Laurent Picard\textsuperscript{c}, Joachim Forget\textsuperscript{d}

\textsuperscript{a}Henri Waremboeuf School of Medicine, Lille University, 59000 Lille, France
\textsuperscript{b}NeoGen Analytics, Sheridan, WY 82801, USA
\textsuperscript{c}Conservatoire National des Arts et Métiers, 75141 Paris, France
\textsuperscript{d}Assemblée Nationale, 75355 Paris, France

\textbf{A R T I C L E  I N F O}

Article history:
Received 17 October 2020
Revised 5 November 2020
Accepted 6 November 2020
Available online 2 December 2020

Obviously, correlation doesn't mean causality. However, in sciences and especially in medicine, when a physio-pathological context is added, evidences were reinforced. Legitimate reviews were done on our previous modest study, we will try to reply here.

1. General considerations

The correlation between vitamin D and sunlight exposure has been broadly documented and known for a long time (Holick, 2016). Our hypothesis was initially based upon COVID-19 outcomes differences between BAME (Black, Asian, Middle East) and White people living at the same latitudes, that could be explained by vitamin D deficiency in the BAME population (Harris, 2006). Links between vitamin D deficiency and COVID-19 fatal outcomes were widely documented since (Mitchell, 2020), it was not the case at the time of submission date. Furthermore, to assess this, the French National Academy of Medicine has recommended the rapid serum vitamin D (i.e. 25 OHD) testing in people over 60 years of age with Covid-19, and a loading dose of 50,000 to 100,000 IU in case of deficiency, which could help limit respiratory complications (Académie Nationale de Médecine, 2020).

Our manuscript (Lansiaux et al., 2020) had as objective to prospect around this possible link between COVID-19, sunlight exposure and vitamin D. To this aim, in view of our team heterogeneity, we used open public data (because we didn't have access to others) to illustrate this link.

The studied COVID-19 outcome was the «mortality rate». Indeed, it was the number of COVID-19 deaths divided by the number of COVID-19 confirmed cases, related to the prognosis of COVID-19.

2. Specific considerations

The sunlight exposure average has been measured since 2018 (not the same data as correspondents have used (Naudet et al., 2020)). Weather data are actually unreachable due to a recent website update (Meteo France, 2020). In more, COVID-19 outcomes data were extracted after the daily update so we have different numbers (as presented on our first manuscript), so the 25/05/2020 data were included in our manuscript and not in their letter and they were published online on the 26/05/2020 by Santé Publique France.

Firstly, concerning the so-called mistake of «assumption of binormal distribution» (Naudet et al., 2020), we are sure that it did not escape the authors that we have used a Shapiro–Wilk test, as it is assessed in our previous manuscript and described in Table 1. As a reminder, it is a test of normality in frequentist statistics which test the null hypothesis that a sample came from a normally distributed population. Monte Carlo simulation has found that Shapiro–Wilk test has the best power for a given significance (Razali and Wah, 2011). As the p-value for the sunlight exposure is 0.3697 (which is higher than the traditional 0.05 p-value and especially higher than the 0.001 p-value that we have fixed as the limit), we can't reject the null hypothesis. In this way, as we have assessed the normal distribution of each variable (including sunlight exposure), we don't violate any Pearson's correlation law because:

1) we have established the link between the sunlight exposure and the COVID-19 mortality (by r computation: \( R = -0.635688603 \)).
2) We have obtained the regression line \( y = -55.459x + 2888.5 \).
According to our correspondents, if we use their analysis upon the 12 regions as N, we obtain $t = 2.604$ and a p-value of 0.0263 (the correlation stays significant with this p-value).

In this way, the use of Spearman’s correlation would be inappropriate here.

In a second hand, we formally disagree on the follow statement «at least for sunlight exposure, since two neighboring regions have more similar climates than two distant regions» (Naudet et al., 2020). Indeed, for example in the «Provence-Alpes-Côte d’Azur» region, we can observe a diversity of climate (mountain, mediterranean); in more, «Auvergne-Rhônes-Alpes» region has mountain, but borderline regions have different climates (for instance, semi-oceanic and oceanic for «Nouvelle-Aquitaine» region) (Météo France, 2020). To prove this statement, firstly, we invite our opponents to use of basic knowledges which assess that regions, with an average size of $44,605.083$ km$^2$ (Table 2), have climate more independent than neighboring towns. This is all the more reinforced as we reason on sunlight exposure average by region. Then, to use the Shapiro-Wilk test (as we have done in our manuscript (Lansiaux et al., 2020)), we have to prove the pseudo-independence (indeed, the perfect independence can’t be prove) of a same variable different values between them. In order to do this, we have used the «turning point test» (published by Irénée-Jules Bienaymé in 1874 (Seneta, 2001)), the time on the abscissa axis was replaced by French regions (therefore it’s not a time-series data as usual but a geographical series data). This method «is reasonable for a test against cyclicity but poor as a test against trend». As we wanted to study the inner-independence of variables (and especially for the sunlight exposure), we had to break free from cyclicity thanks to the turning points test. Our null hypothesis was that they are independently, identically distributed random variables. As we have 128 (or 140 if Corsica is included) turning points (Table 3), we obtained a z of 0.205 with 12 regions (without Corsica) and a z of 0.393 with Corsica inclusion (Table 4). As the observed number of turning points is in the 95% trust interval, we can’t reject the null hypothesis; we may have used, in our manuscript, inner-independent identically distributed random variables.

To respond to the reductio ad absurdum, we are extremely disappointed that our correspondents can’t read and can’t interpret correctly a Pearson/Spearman coefficient. In fact, a negative coefficient indicates a negative correlation and not a positive one as our correspondents assess (Naudet et al., 2020): «sunlight exposure makes people building nursing homes». With a Pearson/Spearman coefficient of −0.69, the correct statement would be «sunlight ex-
Thanks to this «negators» (according to the antique definition), even if we are wrong we are right. Therefore, we thank them for the fact that they, thus, ensure the validity of our hypothesis, despite the mistakes they think have discovered.

### 3. Conclusion

We are extremely enthusiastic about debating in a relaxing manner. Although the heart has its reasons that reason ignores, we must show restraint and rigor, especially in these sanitary hard times. This formal correspondence was very instructive for us (on modesty, respect…), we hope that it was and will for our correspondents.

### References

Académie Nationale de Médecine. 2020. Communiqué de l’Académie nationale de Médecine : Vitamine D et Covid-19 [Internet]. Académie Nationale de Médecine [cited 2020 Aug 25]. Available from: http://www.academie-medecine.fr/communique-de-lacademie-nationale-de-medecine-vitamine-d-et-covid-19/.

Holick, M.F., 2016. Biological effects of sunlight, ultraviolet radiation, visible light, infrared radiation and vitamin D for health. Anticancer Res. 36 (3), 1345–1356.

Harris, S. 2006. Vitamin D and African Americans. J. Nutr. 136, 1126–1129.

Lansiaux, E., Pébay, P., Picard, J.-L., Son-Forget, J. 2020. Covid-19 and Vit-D: disease mortality negatively correlates with sunlight exposure. Spat. Spatio-temporal Epidemiol. 35, 100362.

Meteo France, 2020. Bienvenue sur le Nouveau site de Météo-France ! [Internet]. Meteo France. Available from: http://meteofrance.com/actualites-et-dossiers/actualites/bienvenue-sur-le-nouveau-site-de-meteo-france/.

Météo France, 2020. Le climat en France métropolitaine [Internet]. Météo France [cited 2020 Aug 25]. Available from: http://meteofrance.com/comprendre-climat/france/le-climat-en-france-metropolitaine/.

Mitchell, F. 2020. Vitamin-D and COVID-19: do deficient risk a poorer outcome? Lancet Diabetes Endocrinol. 8.

Naudet, F., Locher, C., Braillon, A., Gillibert, A., 2020. Letter to the editor: Covid-19 and Vit-D: disease mortality negatively correlates with sunlight exposure. Spat. Spatio-temporal Epidemiol. 35, 100373.

Razali, N., Wah, Y.B., 2011. Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. J. Stat. Model. Anal. 2 (1), 21–33.

Remy G. Les 13 nouvelles régions françaises [Internet]. https://www.data.gouv.fr/, 2006 [cited 2020 Aug 25]. Available from: https://www.data.gouv.fr/fr/reuses/les-13-nouvelles-regions-francaises/.

Seneta, E., 2001. Irennée-Jules Bienaymé. In: Heyde, C.C., Seneta, E., Crépél, P., Fienberg, S.E., Gani, J. (Eds.), Statisticians of the Centuries. Springer, New York, NY. https://doi.org/10.1007/978-1-4613-0179-0_27.

### Table 4

| Number of Regions | Points | Turning points | 2  | Expected number of turning points | 95% Trust Interval |
|-------------------|--------|----------------|----|----------------------------------|-------------------|
| 12                | 192    | 128            | 0.20,489,932 | 128 | 116,5,489,471 | 139,4,310,529 |
| 13                | 208    | 140            | 0.393,606,616 | 138,6,666,667 | 126,7,480,329 | 150,5,833,004 |

### Table 5

Healthcare access data in Corse.

| Healthcare access | Corse |
|-------------------|-------|
| Elderly equipment rate (%) | 72    |
| Medical doctor density (‰) | 366   |
| Places in short hospitalisation service (1,163) | (1,163) |
| Hospitalization bed density (‰) | 321   |

### Table 6

Healthcare access statistical parameters with the inclusion of the Corse region.

| Healthcare access | Average | Variance | SD | CI 95% |
|-------------------|---------|----------|----|--------|
| Elderly equipment rate (%) | 144,5 | 707,103 | 26,591 | 130,045 147,303 |
| Medical doctor density (‰) | 326,7 | 1769,244 | 42,062 | 303,835 330,226 |
| Hospitalization bed density (‰) | 387,5 | 933,923 | 30,56 370,887 390,505 |

Exposure makes people building less nursing homes», if we interpret correctly the negative coefficient.

Concerning the exclusion of the Corsican region, we have done it after the data extraction from the different public health already quoted in our manuscript (Lansiaux et al., 2020). In order to justify, we used two indicators: the elderly equipment rate (in view of the large aged people infected by the COVID-19) and the hospitalization bed density (indeed, infected seniors require an hospitalization due to theirs comorbidities and not only a lockdown). Two of three trackers (the last one, the Corsican medical doctor density, had no significant difference with others regions) had significant differences with the others French regions (Tables 5 and 6) (Lansiaux et al., 2020). Therefore, they have conducted us to exclude the Corsican region before the sunlight exposure data extraction (so it was an a priori choice), and not in order to «hack» the p-value what we have been accused of.

Finally, if our correspondents persist in their thesis of p-hacking, we will oppose them that their computed p-value 0,03 (Naudet et al., 2020). Using a ratio absolute, this one stays above the usual medical significance threshold of 5%. In fact,