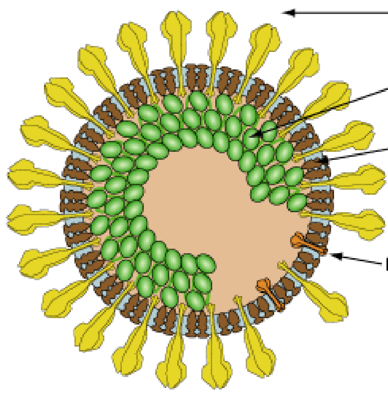
Versus-Virus Hackathon: COVID-19 and Air Pollution

Re-evaluation of a potential relationship between Particulate Matter and COVID-19 prevalence and severity in Switzerland.

Luca Mehl – International School of Geneva
Anurag Amol Sawarkar – International School of Geneva
Ahnaf Rahman – International School of Geneva

Introduction

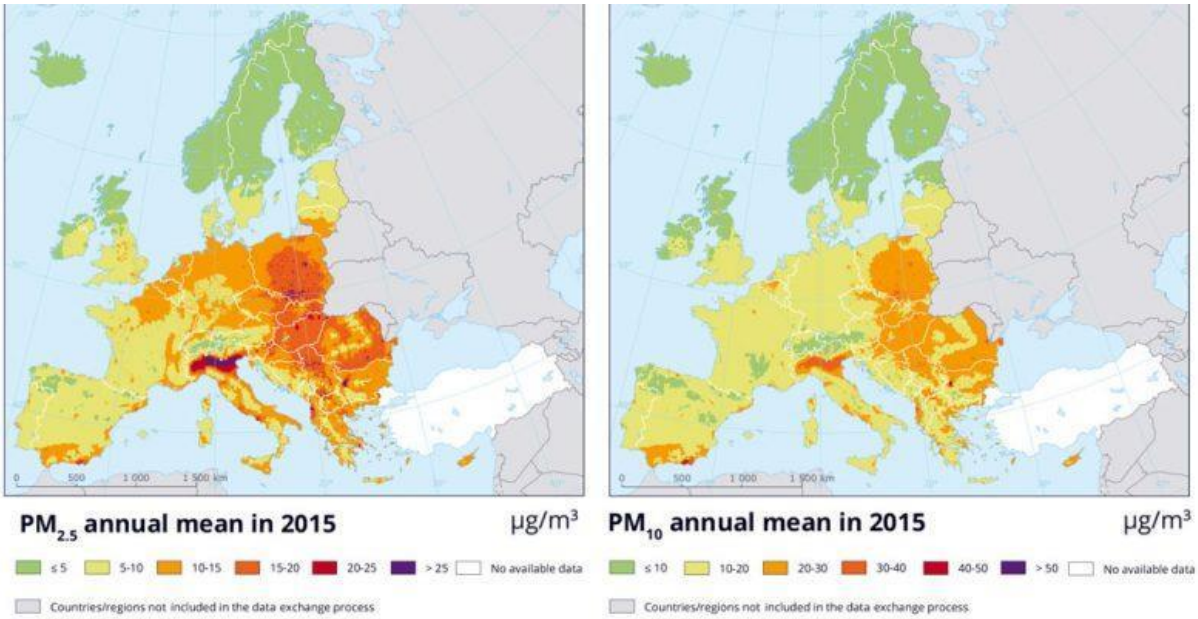


This paper aims to investigate the relationship between levels of PM2.5 & PM10 pollution vs. COVID-19 prevalence in Switzerland. Specifically, we aim to reproduce the findings of the investigation by Leonardo Setti et al, from a Swiss perspective where air pollution levels are comparatively low. Hence, we hypothesize a moderate relationship between levels of airborne particulate matter (PM) and prevalence of COVID-19. COVID-19, as shown by the viral glycoprotein in Figure 1, is known to demonstrate chemical polarity [1] allowing them to bond well with PM2.5 and PM10 ions such as sulfates and nitrates [2]. From this knowledge we can advance several hypotheses:

1) The first is that regions of Switzerland exhibiting relatively elevated PM concentration during the 2 months since the first Swiss COVID-19 case will tend to contain the highest COVID-19 cases:
    a. High PM concentrations can be measured as both average concentrations and as number of days where the concentration surpassed a 'public safety' threshold over a given time period, both of which we expect should show the expected positive correlation with COVID-19 cases.
2) Secondly, regions with consistently elevated PM concentrations over the long term (annual scale) are liable to health complications among the resident population. For this reason, we expect a more severe reception of the virus in said regions, as manifested by heightened cases, hospitalizations, and fatalities.

3) Lastly, knowing that SARS-CoV-2 can be carried by particulate matter and that highly elevated PM concentrations can cause an increased rate of diffusion, we expect “spikes” in COVID-19 rate of dissemination (incidence) to follow shortly after ‘spikes’ in PM concentrations, with lag times matching the 2-14 day incubation periods published by the US CDC [3] and WHO [4].

The results of this investigation will help determine whether Switzerland was made especially vulnerable, as the country with the most COVID-19 cases per million in the world, due to pollutant concentrations. They may also serve to guide the Swiss government as to whether reducing PM emissions should be made an absolute priority in this time of crisis.



Data sources & Data used

COVID-19 data

Data concerning the diffusion of SARS-CoV-2 cases and severity of the ensuring COVID-19 pandemic were supplied from The Statistical Office of the Canton of Zurich, providing cantonal data across Switzerland including daily counts of cases, hospitalizations, and fatalities since the first recorded case in Switzerland. [6]

NABEL data

Air pollution data was sourced from The National Air Pollution Monitoring Network (NABEL), comprising of air pollution levels at 16 distinct locations over in Switzerland. [7]

Population data

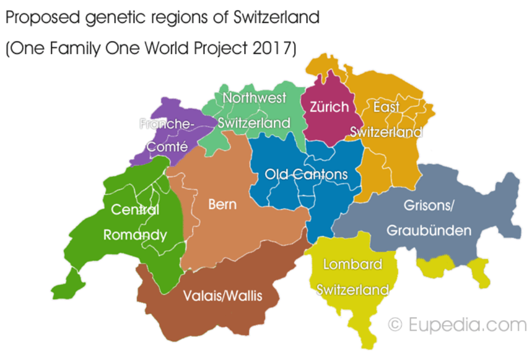
Data containing the population of Switzerland by cantons was supplied by the Federal Statistical Office (FSO). [8]

To evaluate whether air pollution influenced the propagation of the virus, historical data was identified and found for both COVID-19 cases per canton and average air pollution levels by canton. The COVID-19 data used was composed of cumulative cases, number of fatalities and hospitalization since the first date of data collection: 25/02/2020. Average air pollution was composed of data pertaining to the PM10 and PM2.5 levels across cantons, measured in units of $\mu g/m^3$. The data for PM10 and PM2.5 both contained a daily average in the range 14/02/2020 to 03/04/2020, and an annual average for PM10 from 2000 to 2018, respectively. Because the PM2.5 dataset lacked annual average data predating 2016, it was excluded from consideration for assessing the hypothesis relating long-term elevated PM concentrations to more severe COVID-19 occurrences. Furthermore, to ensure that population did not play a role in determining the correlation between the air pollution and COVID-19 Cases, it was factored in during data processing.

Overview of methodologies

Region-Splitting

Because the NABEL data included only 16 weather stations in their monitoring network, the 26 Swiss cantons had to be split into groups, with a roughly equal distribution of weather stations among them. 10 regions were created according to the regions proposed by the One Family One World Project (Eupedia) [9], which fit this requirement:



The 16 locations from which data is derived can be seen in figure 4. Hence, the cantons were split into regions seen in the table below*:*

Region	Canton(s) contained
Central Romandy	Geneva **(GE)**, Vaud **(VD)**
Valais	Valais **(VS)**
Lombard Switzerland	Ticino **(TI)**
Bern	Bern **(BE)**, Fribourg **(FR)**
Pranche	Neuchâtel **(NE)**, Jura **(JU)**
Northwest Switzerland	Aargau **(AG)**, Basel-Stadt **(BS)**, Basel-Landschaft **(BL)**, Solothurn **(SO)**
Graubünden	Graubünden **(GR)**
East Switzerland	Glarus **(GL)**, St. Gallen **(SG)**, Thurgau **(TG)**, Schaffhausen **(SH)**, Appenzell-Ausserrhoden **(AR)**, Appenzell-Innerrhoden **(AI)**
Zurich	Zurich **(ZH)**
Old Cantons	Lucerne **(LU)**, Zug **(ZG)**, Schwyz **(SZ)**, Uri **(UR)**, Nidwalden **(NW)**, Obwalden **(OW)**

Data-Grouping

To evaluate the validity of the three hypotheses given in the introduction section of this paper, the data was split into three data forms:

1. Current number of cases by canton weighted by the population of that canton plotted against the number days that PM10 levels exceeded a certain threshold (elaborated further)
2. Current number of cases, fatalities and hospitalizations by canton weighted by the population of that canton plotted against the PM10 annual average in the last 2, 5, and 10 years, and since first collection of data.
3. Historical data for the percentage of new cases per day plotted against the daily PM10 and PM2.5 levels since 10/02/2020

Data Processing

Cases by canton weighted

To ensure that population did not play a role in this correlation, it was decided to use the Cases Per Million(CPM) of each region to validly compare figures from each of the 10 regions. This number was calculated by finding the sum of all cases in a region, divided by the total population of that region, multiplied by a million. (Equation (1)):

$$CPM = \frac{c_1 + c_2 + \cdots + c_n}{P_1 + P_2 + \cdots + P_n} \times 10^6 \qquad (1)$$

Where CPM is the Cases Per Million of a *region*, c_n is the number of cases of within a single canton of a *region*, and P_n is the population of a single canton within a *region*. Similarly, for the Fatalities Per Million (FPM) and Hospitalizations Per Million (HPM), equation (1) was re-written with f_n and h_n representing the fatalities and hospitalization in a canton respectively.

PM level exceedances.

The number of days for which PM10 and PM2.5 concentrations exceeded a certain threshold value was used as a proxy for the pollution. While the WHO and EU maintain a general guideline threshold for PM10 and PM2.5 exposure, there also exist more/less stringent thresholds. For this reason, we chose to examine 3 threshold values to see if lowering the threshold value would have any effect on the correlation. This relates to the lower pollution ratings seen in Switzerland, as aforementioned, meaning that the strict limit of $50\mu g/m^3$ is rarely exceeded in most cantons. Below are the three threshold values we chose for both PM10 and PM2.5:

PM-10	$50\mu g/m^3$	$40\mu g/m^3$	$30\mu g/m^3$
PM-2.5	$25\mu g/m^3$	$20\mu g/m^3$	$15\mu g/m^3$

PM annual average

The annual average was considered to find if there were any long-term effects of the presence of PM10 in the atmosphere on the amount of hospitalizations, deaths, and cases. Only PM10 was considered because PM2.5 lacked enough datapoints. The annual averages were split into four timespans to vary the duration of the 'long-term effects': average of 2 years, 5 years, 10 years, and since collection of data.

Cases Growth factor

To find the factor by which a certain day's cases grew relative to the previous day(F_g), equation (2) was used, wherein the ratio between the number of cases on day 2, d_2 to day 1, d_1 was found:

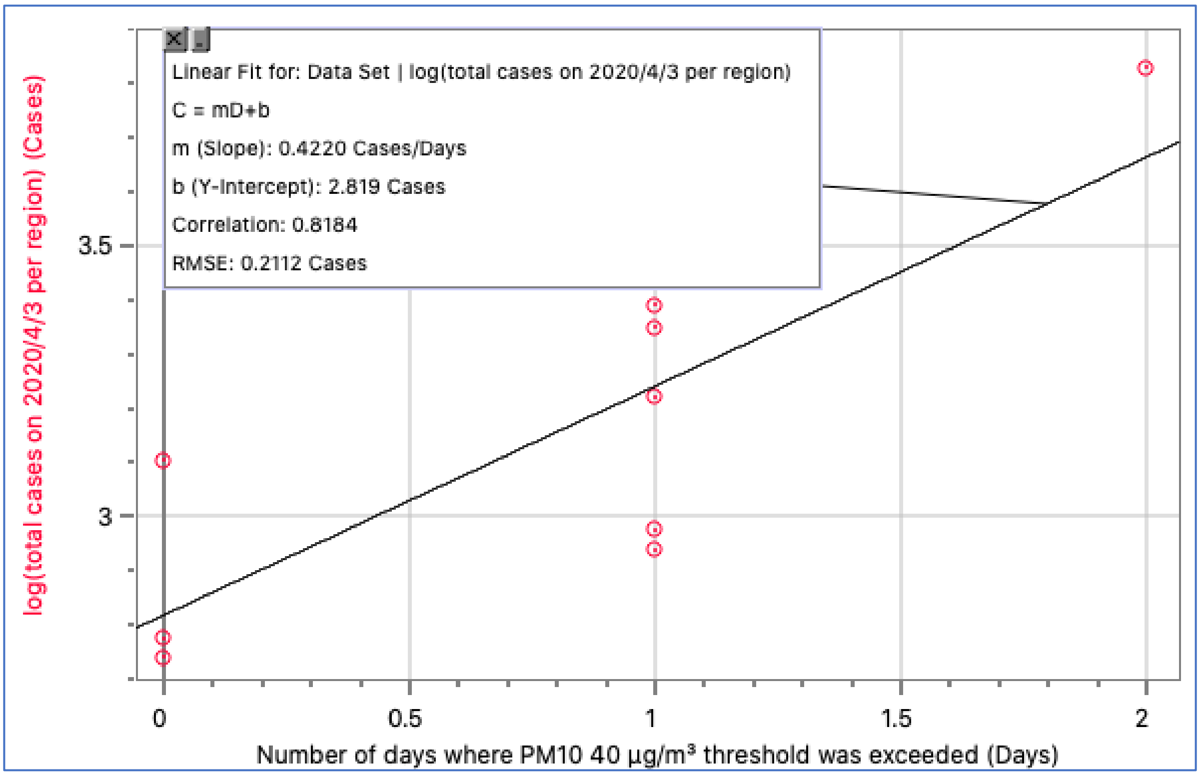
$$F_g = \frac{d_2}{d_1} \quad (2)$$

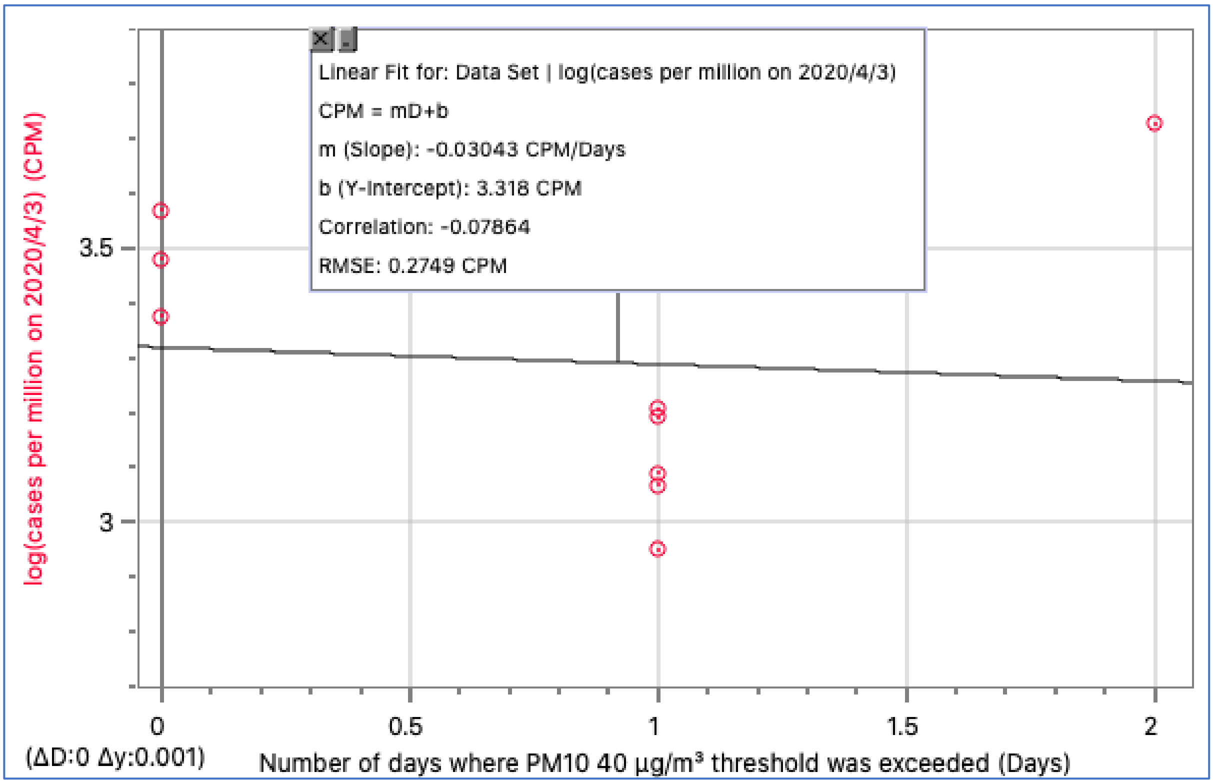
Data Compiling and Analysis

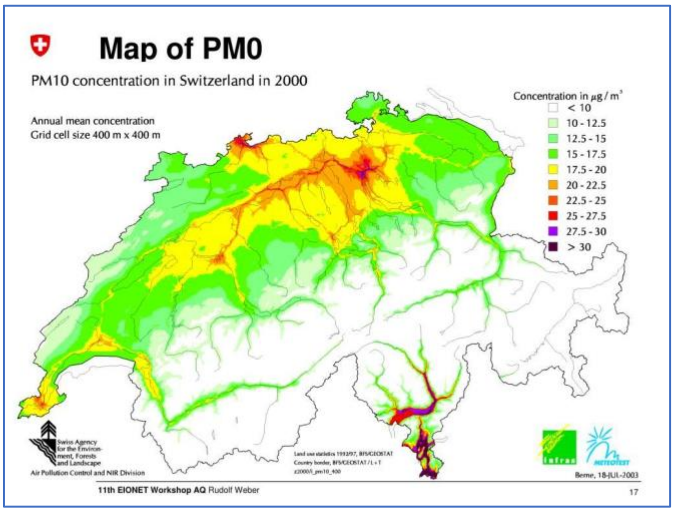
The Pearson correlation coefficient *r* was computationally calculated for the comparison of each pollution indicator (the avg. daily concentration across the two months for PM10 & PM2.5) and proxy indicator (PM10 & PM2.5 exceedance number at 3 thresholds each). This was repeated with and without the inclusion of the "Bern" region, which proved a frequent outlier in the data—possibly as a result of the region comprising both a weather station in the center of the city of Bern, and a weather station on the border with the more rural Fribourg

Region	Total cases by region on 2020/4/3	log (total)	Total CPM by region on 2020/4/3	log (total CPM)	Number of Days Exceeding threshold						Average daily concentration	
					PM10			PM2.5				
					TH50	TH40	TH30	TH25	TH20	TH15	PM10	PM2.5
Central Romandy	6894	3.838	5308.691886	3.725	0	2	3	2	4	7	11.5	7.8
Valais	1273	3.105	3701.065546	3.568	0	0	0	0	0	0	7.6	3.6
Lombard CH	2377	3.376	6727.174445	3.828	6	8	10	9	11	22	21.0	15.6
Bern	1661	3.220	1227.015619	3.089	0	1	3	2	5	8	12.5	8.2
Pranche	593	2.773	2369.450471	3.375	0	0	1	0	0	0	7.2	0.0
Northwest CH	2236	3.349	1551.456885	3.191	1	1	4	1	3	10	12.3	8.3
Graubuden	598	2.777	3014.431971	3.479	0	0	1	0	0	0	4.5	0.0
East Switzerland	871	2.940	890.6458665	2.950	1	1	1	1	2	7	9.6	6.7
Zurich	2452	3.390	1612.131222	3.207	1	1	4	1	3	9	11.0	7.0
Old Cantons	948	2.977	1165.971348	3.067	0	1	1	1	2	7	8.1	6.0
					Correlation coefficient R							
			log(total)	w/ outlier	0.248	0.419	0.520	0.403	0.552	0.520	0.595	0.644
				wo/ outlier	0.163	0.817	0.698	0.728	0.722	0.599	0.749	0.725
			log(total CPM)	w/ outlier	0.443	0.532	0.446	0.504	0.369	0.203	0.340	0.206
				wo/ outlier	-0.519	-0.079	-0.128	-0.195	-0.271	-0.524	-0.270	-0.363

The data shows that the highest correlations are present in the “log(total) wo/ outlier” row.

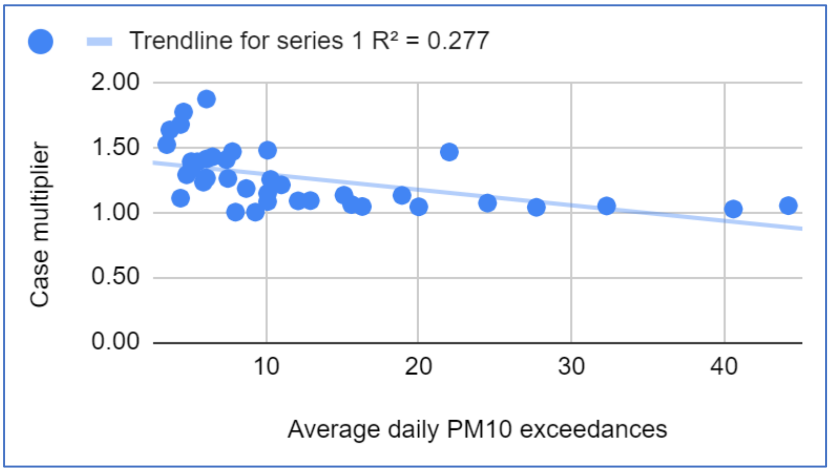



Linear Fit for: Data Set | log(cases per million on 2020/4/3)
CPM = mD+b
m (Slope): -0.03043 CPM/Days
b (Y-Intercept): 3.318 CPM
Correlation: -0.07864
RMSE: 0.2749 CPM
log(cases per million on 2020/4/3) (CPM)
3.5
3
0
0.5
1
1.5
2
(ΔD:0 Δy:0.001)
Number of days where PM10 40 µg/m³ threshold was exceeded (Days)


Map of PM0
PM10 concentration in Switzerland in 2000
Annual mean concentration
Grid cell size 400 m x 400 m
Concentration in µg / m³
< 10
10 - 12.5
12.5 - 15
15 - 17.5
17.5 - 20
20 - 22.5
22.5 - 25
25 - 27.5
27.5 - 30
> 30
11th EIONET Workshop AQ Rudolf Weber

Region	R Value
Central Romandy	-0.526
Valais	-0.326
Lombard Switzerland	-0.458
Bern	-0.383
Pranche	-0.197
Northwest Switzerland	-0.373
Graubuden	-0.192
East Switzerland	-0.253
Zurich	-0.303
Old Cantons	-0.235

As shown in table 3, the peak r values for each distinct Swiss geographical region have been calculated by calibrating the case multiplier with the most likely date of first case as shown in an example for Central Romandy in fig. 7.



However as shown by this example calculation, we can see that any correlation that exists is invalid due to it being negative and frankly rather small in most cases.

Limitations

With the assumption that the data used in this investigation is accurate, we must consider the limitations in its analysis. Firstly, we must note that the Swiss investigation is inherently different from the Italian one due to significantly lower levels of PM pollution in Switzerland [5]. Hence regional variations in PM pollution levels may not directly correlate with prevalence of COVID-19. With the weak correlation found in this investigation, we need to keep in mind that correlation does not necessarily equal causation. Even with the hypothesized bonding of the viral glycoprotein with ionic particulate matter, there just may not be enough pollution to act as carriers.

Also, the paper was not able to control for variations in regional population densities across Switzerland. Intuitively, we can conclude that higher population densities will result in more frequent contact-driven transmission of COVID-19. Similarly, the Italian research team does not account for population densities but as shown in fig. 6. Even with measurements of COVID-19 prevalence with the population per million parameter, we see that the correlation is so weak that there is a negative trend. This suggests that PM pollution is not at all a major a factor in the proliferation of COVID-19.

Lastly, variations in testing criteria among Swiss and Italian medical institutions has the potential to vastly impact the number of reported cases and the number of actual cases. To some extent, this can be solved by using the SIR model to gain a theoretical prediction of the virus' behavior using data from regions with similar testing criteria. However, due to its novelty, the current coronavirus infection is rather difficult to model with accuracy and hence such models will quite likely not to prove to be much more useful than the real-time data. These limitations are what results in small Pearson product-moment correlation coefficient values (r) when comparing COVID-19 prevalence and PM emissions of both PM2.5 and PM10 types.

Conclusions and Suggestions

In conclusion, this investigation finds no statistically significant correlation between air pollution and cases of COVID-19 prevalence in Switzerland. Thus, it is only appropriate to state that the dissemination and severity of the COVID-19 pandemic in Switzerland is quite likely resulting from other factors such as contact transmission and contamination. Regional factors such as population densities and PM emissions should be considered to allow for any corrective measures to be taken. For example, Switzerland ranks among the least polluted countries in the world in terms of PM emissions. This may explain the lack of correlation between minute changes in already low pollution levels and COVID-19 prevalence.

Further study should be conducted with data from countries with more PM emission levels. While Switzerland may be enjoying no correlation, it is quite possible that states plagued with damaging levels of PM emissions will suffer an increased prevalence of COVID-19 as hinted by the research of Leonardo Setti et al.

Also, it may be wise to divide regions according to the concentration of heavy industry instead of just looking at historically divided regions. This will allow us to filter the polluted regions with more precision and give us clearer data if a link between PM concentration and COVID-19 prevalence does exist. For the time being, Switzerland must focus on other methods of prevention such as reducing contact transmission by limiting non-essential movement among the general population.

References

[1] "Glycoprotein - Chemistry Encyclopedia - structure, proteins." http://www.chemistryexplained.com/Ge-Hy/Glycoprotein.html (accessed Apr. 05, 2020).

https://annals.org/aim/fullarticle/2762808/incubation-period-coronavirus-disease-2019-COVID-19-19-from-publicly-reported

https://www.eupedia.com/genetics/regional_dna_project_switzerland.shtml

https://energyindustryreview.com/analysis/pollution-in-eu-wreaking-havoc-on-human-health/