Demographic determinants of testing incidence and COVID-19 infections in Catalonia

Regional Quantitative Analysis Research Group (AQR–UB)

AQR COVID-19 / #11

Barcelona, June 11th 2020

BACKGROUND AND OBJECTIVES

The growing availability of statistical information on the spread and incidence of COVID-19 with territorial detail is making possible to monitor the spatial distribution of the epidemic and to link its incidence with various characteristics of the territories. In the case of Catalonia, the Catalan Agency for Health Quality and Assessment (AQuAS) has been providing, since the end of March, information on the daily number of positive cases of COVID-19 by Basic Health Areas (BHA) and by municipalities.

In previous short letters we have shown evidence of the association between the geographical distribution of COVID-19 and some factors of interest, such as population density, meteorological factors or the socioeconomic conditions of the population in the different territories, among other factors. These previous analyses, as mostly done in the existing literature, focused on the incidence of covid diffusion measured as the fraction of persons who tested positive as percentage of population. Borjas (2020) argues that analyses based on a single number might face an issue. In fact, the rate of infection in the population is the result of the product of two different factors: the frequency of tests in a particular area and the fraction of positive tests among those tested. Considering the case of New York, Borjas (2020) argued that the incidence of testing in the city was not random and analyzed how socioeconomic characteristics correlate with each of the two determinants of rate infection. Results have shown that some socio-economic factors were correlated in opposite directions with each of the two determinants of rate infections therefore yielding a “zero” correlation with the infection rates over population. As a consequence, the effect that these variables had on covid spread remain hidden. More in detail, he showed that the incidence of testing was positively correlated with household income in certain neighborhoods, thus pointing out that richer areas were more likely to be tested. Nonetheless, once tests were administered, the likelihood of being infected was inversely related to income, poorer household were more likely to be infected. Also, overall his results have shown that the conditional probability of a positive test result was far greater for persons living in poor neighborhoods, in neighborhoods where large numbers of people reside together and in neighborhoods with a large black or immigrant population. At the same time, however, persons residing in poor or immigrant neighborhoods were less likely to be tested. These results might depend on the characteristics of the US health system, whereas health insurance (or sufficient resources) is needed in order to get health assistance since most health care facilities are owned and operate by private businesses.

The health system in (most) European countries is different from the one of US, since it is mostly public and (almost) totally financed by the State. Therefore, getting access to medical assistance in Europe is not conditioned by having health insurance. A natural question therefore arises: how do socio-economic factors correlate with each of the two determinants of the infection rates in Europe compared with the US? This will help to understand whether differences in health systems might play a role in how the health system has fought the covid...
diffusion. In this note, we consider the European region of Catalonia (Spain), and test the role of various socio-economic factors on infection rates and on its two determinants, using data defined at Basic Health Area (BHA) level.

The empirical analysis uses cross-section data for April 8th, which is the last day that information on the number of tests carried out was provided by the Catalan health authorities. We analyze the association between rate of infections over population, rate of tests over population and the rate of infections over tests, with a set of data related to demographic and socio-economic indicators of Catalonia population. As for descriptive evidence, for the total region of Catalonia at the date of April the 8th, there have been carried out 65,394 tests, which account for less than 1% of population. Also, around 43% of tests gave positive results, while the number of positive cases per 10,000 inhabitants was 36.44. Moreover, the correlation between positive cases over population and the number of tests over population is 0.89, while the correlation between the number of infections over population and number of infections over tests is 0.67. Finally, the correlation between the number of tests over population and the number of detected infections over tests is 0.34.

We focus our interest in the following explanatory variables: density, percentage of people aged over 65, mortality rate, percentage of dependent population aged 15 and over, an index of socio-deprivation, and a set of variables related to the foreign presence in the population such as number of foreign individuals also distinguished by regions of origin such as Developing Countries, China, Italy and the EU.

DRIVERS OF INFECTION RATES, TESTS RATE AND RATE OF INFECTION OVER TESTS

We run a group logit regression using weighted least squares for the sample of 371 BHA at the date of 8th of April 2020. All regressions control for meteorological factors (using average temperature and humidity between two and three weeks before the date of reference), pollution (using the average over the last years of NO2 and PM10), the day since the first test (or case) and its squared value, and a dummy variable identifying an area (Conca d’Odena) where there has been a significant outbreak in Catalonia. Since our independent variables are mostly highly correlated across them, we will insert them separately in a series of regressions. Nonetheless, we also run the regressions with all the variables, to check individual findings. Table 1 summarizes the results obtained in terms of signs and significance of each variable.

From Table 1 it is possible to see that the rate of infections over population is positively associated with many of the variables taken into account such as the percentage of population aged over 65, the percentage of population with dependency aged over 15, the density of economic activity and the index of socio-deprivation. On the other hand, we find out a negative association with the relative presence in BHAs of immigrants, in particular when coming from the European Union and, to a lesser extent, Italy. We do not find any significant association with BHAs with relatively high presence of immigrants coming from China and from the developing countries, even if the association with China is positive but slightly imprecisely estimated. Further, when tested jointly these patterns remain generally confirmed, even though there are significant correlations across pairs of variables.

Table 1. Explaining the infection rate, the tests rate and the rate of infection over tests: Comparison of sign and significance of the parameters estimates for key variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Sign for Infection Rates</th>
<th>Sign for Tests Rate</th>
<th>Sign for Rate of Infection over Tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>Share population &gt;65</td>
<td>**</td>
<td>**</td>
<td>+</td>
</tr>
<tr>
<td>Share of dependent pop.</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Mortality rate</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Density of population</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Socio-economic deprivation</td>
<td>+</td>
<td>**</td>
<td>+</td>
</tr>
<tr>
<td>Foreign population</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Pop. Developing Countries</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Chinese population</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Italian population</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>EU population</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
</tbody>
</table>

Source: Own elaboration. Significance at 1% ***, 5% ** and 10% *.

We now look at the determinants of the infection rate over population: the incidence of tests over population and the incidence of positive tests over total tests. We analyze each of the variables of interest separately.

As for what concern the age structure and morbidity of the population, it is possible to observe that a relatively higher number of tests per capita has been carried out in BHAs characterized by a higher percentage of elderly population (>65) and mortality rate. This might have been justified by the need of “covering” relatively more
this segment of the population, more likely to suffer of serious form of the illness. Also, it could be in line with the policy followed in Catalonia where only people with severe symptoms have been tested for COVID, and these likely were elderly people. Nonetheless, keeping fixed the number of tests, there is no a significant association with the incidence of infections (as in Borjas, 2020). Therefore, the higher number of infections over population associated with these two variables is driven by the relatively higher number of tests, suggesting no higher “predisposition” in getting sick for this segment of the population. On the other hand, when considering the BHAs characterized by a relatively higher number of people with a condition of dependency, higher infections rates have been found for this segment of the population, suggesting higher likelihood of being infected for those persons with morbidity conditions. Nonetheless, there has not been put more attention on testing this segment of the population with respect to the others, since the association with the ratio of tests over population is insignificant.

As for density, it is well known the positive association with the infection rate over the population, and this is confirmed also in the case of Catalonia1. Also, both the rate of tests over population and the rate of positive tests over total tests show a positive association, with higher magnitude for the first. These remain confirmed also when testing the variable jointly, especially for what concern the tests rate. On the opposite, the relation gets widely insignificant when considering the age and morbidity structure of the population in the case of rate of infections over tests.

Considering more deprived areas we do also find a positive association with infections rates, driven by both the incidence of tests over population and the incidence of positive cases over tests. In this case, the magnitude of the effect is higher for the number of positive cases. Hence, BHAs characterized by higher deprivation were more likely to be tested and positive results were found relatively more with respect to less deprived BHAs. These results hold also when testing jointly all variables of interest2.4.

Finally, considering the foreign/native structure of the population, other interesting findings arise. In fact, all BHAs with a relative high number of foreign population have been relatively less tested than other BHAs, with the exception (although individually slightly not precisely estimated) of those characterized by a relatively higher percentage of Chinese population.

Nonetheless, considering immigrants in general, the association with positive cases is insignificant. Hence, the negative association with the infection rate over population is driven by the lesser extent to which these areas have been tested, rather than by lower infection rates over tests.

We now consider the specific ethnicities. As for what concern the Chinese population, results of the infection rate over population have been driven by a relatively higher attention to make tests in areas characterized by high presence of this ethnicity, than by higher number of infection rates with respect to other areas. These results hold also when testing variables jointly.

Taking into account BHAs with high percentage of population coming from the developing countries the association with the infection rate in principle is not significantly different from 0. However, this is the result of the balancing of associations of inverse sign with respect to each of the two determinants. In fact, these areas have been relatively less likely to be tested. Nonetheless, given the number of tests, higher infections rates have been found. The latter finding might be related to deprivation, which increases the likelihood of being infected, since higher deprivation might be associated to areas with a high percentage of population coming from the developing counties. However, this does not seem to be the case since when testing jointly variables of interest, given the same level of deprivation, the shown correlations do not significantly change. Hence, less tests have been carried out also in more deprived areas characterized by a higher percentage of population coming from the developing countries as compared to more deprived areas characterized by a higher percentage of native population, even if BHAs with high incidence of population coming from the developing countries seem to be relatively more affected.

As for migrants from Italy and the EU in general, the infection rate over population is driven by both lower test incidence and lower infection rate over tests, pointing out lower attention towards these areas that showed less infections.

**CONCLUSION**

Overall, these results show that, compared to the case of US, Catalonia seems to have behaved differently. While the US has concentrated more attention on testing individuals in richer areas, although higher infections were then found over the relatively poorer segments of the population, in Catalonia more attention has been given to areas in principle more likely to suffer severe consequences of the illness such as those characterized by higher deprivation and the one with higher presence of elderly persons (in the latter case, as in the US). Nonetheless, as in the US, Catalonia put less attention to test areas with a higher percentage of foreign population. This might be justified in the case of “richer” immigrants such as from Italy and the EU (whereas the incidence of infection rate has been lower), but not when considering foreign population coming from developing countries,
whereas the incidence of infection rate has been higher. This picture suggests that, at least to some extent, the Catalan health system (that shares several features with the rest of European public health systems) has given more assistance to poorer segments of the population with respect to the US health system to handle the COVID-19 outbreak. At the same time, it is not clear whether this assistance has privileged relatively more the native population with respect to the weaker segment of foreign population. More analyses will be needed to better clarify the latter point.

Another important outcome of the overall analysis is the relevance of making publicly available the data on the number of tests conducted because, as we have shown, they are able to highlight several aspects of the handling of the infection that is not possible to gather without access to these data. Therefore, it is questionable the understanding of why the Catalan health authorities (as likely the Spanish ones) have decided to make no longer available these data (at least at the moment of writing this research note).

---

**Regional Quantitative Analysis Research Group (AQR)**

University of Barcelona
Faculty of Economics and Business
Department of Econometrics, Statistics and Applied Economics
Av. Diagonal 690, 08034 Barcelona
Contact: aqr@ub.edu
www.ub.edu/aqr

---

1. It is also possible that persons residing in the wealthier neighborhoods might have had better information networks or could more easily afford to obtain tests elsewhere in the city (cit. Borjas, 2020).
2. See the methodological note for the sources of data: https://www.ub.edu/aqr/covid19/docs/AQR_Covid19_NotaMetodologica_cast.pdf
3. It refers to people that need some form of assistance to carry out basic daily activities due to health conditions.
5. Data for the developing countries come from “Community Health Indicators” database from the Generalitat de Catalunya.
6. It is important to stress that some of these variables might in principle be positively correlated with the likelihood of both being tested and of getting the infection such as for instance the number of people aged over 65. In this case, a positive correlation with high testing incidence would measure a response to the outbreak. On the contrary, for other variables such as the index of socio-deprivation or the percentage of foreign population, this endogeneity question is likely to be avoided, i.e. it is not expected a positive correlation with both the determinants of total infection rates because of specific outbreaks according to these dimensions.
7. See the methodological note for the results of the regressions: https://www.ub.edu/aqr/covid19/docs/AQR_Covid19_NotaMetodologica_cast.pdf
8. See a more detailed analysis of these factors in previous research notes (Doc AQR-COVID/#01 and #03).
9. These results are also in general confirmed when testing jointly the variables, even if the number of infections over tests in some cases is positively correlated with the percentage of people aged 65 and over, while the mortality rate negatively, but this is likely due to the high correlation across these two variables (73%).
12. It would be even better, as recently done in Italy, to give information on data on tests and “new” tests, when persons are tested for the first time. In fact, the data on the number of tests sum up new tests with tests to covid patients to check that the infection is gone.