Socioeconomic Inequality and COVID-19 prevalence across municipalities in Catalonia, Spain

Ryohei Mogi^{1*}, Gento Kato^{2*}, Susumu Annaka³

Last Updated: May 20, 2020

Main text: 2562 words

Display items (figures, tables): 5 figures

¹ Centre d'Estudis Demogràfics, Universitat Autònoma de Barcelona. Carrer de Ca n'Altayó, Edifici E2. Universitat Autònoma de Barcelona. 08193 Bellaterra / Barcelona

² Department of Political Science, University of California, Davis. One Shields Avenue, Davis, CA, USA 95616.

³ Waseda Institute for Advanced Study, Waseda University, Tokyo, Nishiwaseda, Shinjuku-ku, Tokyo 169-0051, Japan

* Corresponding authors. Email: rmogi@ced.uab.es (R.M.); gkato@ucdavis.edu (K.G.)

Abstract

This study provides preliminary evidence regarding associations between socioeconomic inequalities and variations in the number of COVID-19 confirmed cases across 923 municipalities in Catalonia, Spain, as of the 14th of May, 2020. We consider three types of inequalities at municipality-level: 1) economic development, i.e., unemployment rate, average income, immigrants proportion, and the prevalence of small residence; 2) health vulnerability, i.e., crude death rate and the proportion of elderly (aged 65 +) population; and 3) information communication, i.e., the proportion of people with tertiary education. In addition to the static analysis with the total sum of COVID-19 cases, the dynamic analysis with daily moving weekly sum of cases is conducted. The result draws a rather complex picture of relationships between contextual socioeconomic inequalities and the spread of COVID-19. Many indicators of economic inequalities imply the opposite relationship as intuitively suggested: economically disadvantaged municipalities tend to have less cases of confirmed infection than economically advantaged counterparts. The implications from health inequality indicators show mixed patterns: crude death rate is positively associated, but elderly population is negatively associated, with the number of confirmed cases. The indicator of information inequality shows a consistent tendency, i.e., municipalities with more university educated have less confirmed cases, but this tendency transforms across time: the negative association is particularly strong during the first month of Spanish “state of alarm” measure (mid-March to mid-April). Our evidence suggests the need for more careful consideration regarding the association between socioeconomic inequalities and the regional progression of COVID-19 pandemic.

COVID-19, the disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), is now responsible for the worst global pandemic since Spanish flu in the early 20th century (1). As of May 14th, 2020, Spain is one of the most severely hit countries in the current pandemic: the second most confirmed cases and the fourth most deaths in Europe (2). Within Spain, however, not all regions are equally affected by COVID-19. Confirmed cases of COVID-19 are concentrated in two autonomous communities, Madrid and Catalonia, as these communities comprise 52.8% (28.6% in Madrid, 24.2% in Catalonia) of total confirmed cases in the country (229,540) on the 14th of May. Furthermore, within these communities, there are huge variations at the municipality level. Some municipalities experience significantly severer COVID-19 outbreaks than other municipalities. Focusing especially on Catalonia, this study examines how the spread of COVID-19 is associated with municipality-level social and economic inequalities.

The advancement of COVID-19 pandemic in Catalonia

The first positive case of COVID-19 in Catalonia was confirmed on the 27th of February, 2020 at Barcelona.¹ Since then, the number of confirmed cases in Catalonia increased exponentially, reaching 55,482 at the time of writing (the 14th of May, 2020). During this period, the Spanish and the Catalanian governments implemented several measures to prevent the spread of COVID-19. Namely, the Spanish government announced "the state of alarm" on the 14th of March (expected to expire on the 24th of May at the time of writing), preceded by the series of recommendations published on the 9th of March. These measures order people to stay home and

¹ The first PCR test in Catalonia was conducted on the 25th of February, 2020.

Figure 1: The cumulative number of confirmed positive cases of COVID-19 per 10,000 population across municipalities in Catalonia on the 14th of May, 2020

Note: Municipalities with less than 10 tests are greyed out.

limit travels on public roads and streets except for the essential activities such as purchasing essential items (e.g., foods, pharmaceuticals), attending health centers, going to work, returning to the usual residence, and caring for vulnerable families (3). The Catalan government also implemented measures to lockdown municipalities, starting with four municipalities (Igualada, Odena, Vilanova del Cami, and Santa Margarida de Montbui) on the 12th of March. Our analysis covers the entire period from the 27th of February to the 14th of May, but pays special attention

Figure 2: Time trends in the cumulative number of confirmed COVID-19 positive cases across municipalities in Catalonia from the 25th of February to the 14th of May

Note: It shows 50 municipalities that have the highest number of total COVID-19 cases excluding Barcelona, since Barcelona is an outlier with the excessive number of cases (given the excessive number of residents).

to the two months period following the week of the 9th, 12th, and 14th of March, when governments started to implement a series of preventive measures.

To grasp regional variations in the spread of COVID-19 in Catalonia, Figure 1 shows the municipality-level confirmed cases per 10,000 population. It shows that there are significant variations in the prevalence of COVID-19 cases across municipalities.² Then, Figure 2 shows time trends in the number of confirmed cases of COVID-19 in Catalonia from the 25th of February to the 14th of May among the most severely hit municipalities. The important observation here is that there are variations in when the acceleration of spreads occurs. Some

² Appendix C also shows the geographical distribution of PCR tests and positive rate of tests across municipalities in Catalonia for reference (not used in the main analysis).

municipalities experience the boost in confirmed cases in the early stage while others show the significant increase in relatively later days.

How can inequality matter in pandemic

Compared to known diseases, novel diseases like pandemic flu were often thought to “hit the rich and the poor alike.” However, the spread of novel infectious diseases is not equally distributed by individual’s or region’s socioeconomic characteristics (4, 5). Early research on Spanish flu in 1918, which is the worst pandemic the world has ever experienced, argued that the poor are affected more severely than the rich both in terms of morbidity and mortality (6–8). More rigorous evidence from recent studies suggests the mortality rate of Spanish flu is higher among socially-disadvantaged minority ethnic groups (9), individuals residing in smaller apartments (10), residents of districts with higher illiteracy rate (11), and low-skilled workers (12). Note that compared to mortality, evidence on morbidity is rather mixed. For example, research based on the survey conducted in Bergen, Norway between 1918 and 1919 reports that poorer (measured by smaller apartment size) experienced a higher rate of morbidity during the first wave of Spanish flu (July-September 1918) but a lower rate of morbidity in the second wave (October-December 1918) (13).

Evidence from 2009 H1N1 pandemic flu similarly suggests the socioeconomic disparity plays an important role in explaining the contraction and fatality of the illness. In the US, rates of both hospitalization and death from pandemic flu were reported to be higher among ethnic minorities, especially Hispanics (14, 15). Survey-based US study additionally finds that the lack of workplace policies such as paid sick leave is positively correlated with the contraction of

influenza-like illness during the 2009 pandemic (16). In England, the mortality rate of this pandemic flu was three times higher for residents of the most socioeconomically deprived quintile than those living in the least deprived quintile (17). Also, patient data from Spain indicates that the pandemic flu-related hospitalization rate is higher among non-Caucasian ethnicities and lower among those with secondary or higher levels of education (18).

Incoming preliminary evidence on COVID-19 raises similar concerns regarding the social determinants of the outbreak severity. The evidence to date provides three major implications. First, in the US and England, the ecological evidence emphasizes the higher rate of infection, positive test, and death among the economically disadvantaged population and socially-disadvantaged ethnic minorities (19, 20). While without sophisticated quantitative analysis, similar journalistic suggestions are also made in Spain (21). Second, the series of clinical and contextual evidence strongly suggests the elderly and those with underlying conditions are especially vulnerable to the infection and mortality of COVID-19 (22–24). Third, public opinion surveys on the knowledge and acts of preventive behavior in various countries (e.g., the US, China, Egypt, Paraguay, Peru, Jordan, Saudi Arabia, Kuwait, among others) suggest that educational attainment has a significant positive correlation with knowledge of COVID-19 characteristics and this knowledge plays a key role in agreeing with and following properly the country's measurements (25–30).

Given the above discussion, we measure socioeconomic inequalities across municipalities in Catalonia from three aspects: 1) Economic development: the proportion of unemployment among people aged 15-65, average taxable base income, the proportion of immigrants outside of the European Union, and the proportion of small size houses; 2) Health vulnerability: the crude

death rate in 2018 and the proportion of the elderly population (aged 65 +); 3) Information communication: the proportion of people with tertiary education. Population, population density and the average number of household members are included in the model to control the possible confoundings in associations.³

Results

Figure 3 presents the cross-sectional evidence using the raw cumulative number of confirmed cases in 923 counties of Catalonia on the 14th of May. We apply the hurdle model to estimate associations given the nature of data (see Data and Method section): the left panel shows the result for a hurdle component (having at least one positive case or not) and the right panel corresponds to the result for a positive count component (positive number of confirmed cases). Contrary to previous findings, our results caution against drawing the “simple picture”— the socially and economically disadvantaged areas face more severe effects of pandemic virus— and implies more complex relationships between municipality-level contexts and the spread of COVID-19.⁴

As most of our predictors do not have strong associations with the hurdle component (i.e., whether there are one or more confirmed cases or not), the following discussions focus on associations observed in the positive count component. Many of our economic inequality indicators show results contradicting with the expectation that poor areas are affected more than rich areas. Municipalities with a higher proportion of unemployed, a higher average income of residents, and a lower proportion of residents immigrated from outside of the EU have more

³ Appendix A discusses relationships among socioeconomic inequality indicators.

⁴ Results below are robust to alternative specifications. See Appendix D for robustness checks.

Figure 3: Associations between socioeconomic inequalities and total confirmed positive cases of COVID-19 in municipalities across Catalonia on the 14th of May

Note: Associations are estimated through the hurdle model. Each independent variable is standardized. The dot shows the odds / incidence-rate ratio estimate of each variable. The odds ratio (left panel) illustrates how many times the odds of non-zero count is more likely than zero count, and incidence-rate ratio (right panel) indicates how many times the odds of one more positive count is more likely than no more positive count. The gothic and light bars correspond to their 90% and 95% confidence intervals, respectively. The shape of dots (and colours of bars) indicates whether or not estimated odds/incidence-rate ratios are statistically significant at 5 or 10% levels. N=923.

confirmed cases, holding all other factors constant. These patterns imply, in contrast to a conventional expectation, economically advanced municipalities may have more confirmed cases than their economically backward counterparts. Note that the proportion of small size houses show a positive association (statistically significant at 10% level) with confirmed cases: the pattern consistent with a conventional expectation. Similarly, two indexes supposing to capture the health vulnerability have mixed results. The crude death rate in 2018 has a positive association, but the proportion of the elderly population has a negative association, with a

Figure 4: Associations between socioeconomic inequalities and positive cases of COVID-19 in municipalities across Catalonia on the 14th of May, controlling for test prevalence

Note: Associations are estimated through negative binomial regression. The dot shows the coefficient estimate of each variable. Each independent variable is standardized, thus, a incidence-rate ratio illustrates how many times the odds of one more positive count is more likely than no more positive count given 1SD increase in a independent variable. The gothic and light bars correspond to their 90% and 95% confidence intervals, respectively. The shape of dots (and colours of bars) indicates whether or not coefficients ratios are statistically significant at 5 or 10% levels. Municipalities with no tests are dropped from the analysis. N=807.

positive number of cases of COVID-19. The only evidence that consistently supports the prior findings is regarding the informational inequality: municipalities that have higher proportions of tertiary educated people tend to have less confirmed cases.

Recent studies criticize the use of the number of positive cases as the proxy for the spread of the virus. They argue that the number of confirmed cases can be an artifact of selection bias caused by differential test accessibility across regions (31). Here, it might be the case that more

economically advanced municipalities confirm a higher number of COVID-19 cases because they make tests more accessible (thus have higher ability to identify those infected), not because they are experiencing an outbreak of higher severity.⁵ To partially address this issue, we construct an additional model including the prevalence of tests (i.e., the number of tests per 100,000 population) as a control variable. If observed associations in Figure 3 are driven by the increase in test accessibility (and not the spread of the virus), the inclusion of test prevalence as a control should weaken those associations. Due to the fact that no positive case is confirmed if no test is conducted, the model is estimated only for municipalities with one or more tests conducted (N=807). Then, since the relevant dataset no longer has an excessive number of zero, the model is estimated by negative binomial regression.⁶

Figure 4 illustrates the results for the additional analysis. As expected, test prevalence (the last row) has a strong positive association with the number of positive cases. In comparison with the right panel of Figure 3, it shows the persistent patterns of associations for economic and informational inequalities, both in terms of direction and magnitude. Therefore, we find no evidence that supports the claim that the variation in test accessibility drives associations between economic/informational inequalities and confirmed COVID-19 cases. On the other hand, associations for health vulnerability indicators are significantly weakened after controlling for the test prevalence. Here we find the evidence consistent with the claim that associations between municipality-level health vulnerability and confirmed COVID-19 cases are driven at least partially by the accessibility to tests.

⁵ In New York, for example, it is found that poor neighborhoods are negatively associated with the number of tests, but positively associated with the likelihood that the test was positive .

⁶ Given the difference in methodology and datasets, Figure 3 and Figure 4 are not fully comparable while comparison provides some important implications. Further inquiry is required in the future studies.

Figure 5: Trends in associations between socioeconomic inequalities and daily moving weekly sum of COVID-19 positive cases in Catalonia between the 14th of March and the 14th of May (positive count component)

Note: On each day, associations are estimated by the hurdle model (N=923). This figure presents associations in positive count component (see Appendix Figure SE1 for associations in hurdle component). Each predictor is standardized, thus incidence-rate ratios correspond to its 1 standard deviation increase.

The above cross-sectional result reveals the non-simple connections between municipality-level socioeconomic inequality measures and the spread of COVID-19. On the other hand, the analysis was a static snapshot taken on the 14th of May and misses the dynamic aspect of the spreading process of COVID-19. Therefore, we calculate the daily moving weekly sum of confirmed cases (i.e., the sum of cases confirmed within one week prior to the given date) and analyze the dynamic trend in the relationship between regional socioeconomic

characteristics and COVID-19 cases by estimating the hurdle model on each date in the period between the 14th of March and the 14th of May. Figure 5 shows the results regarding the positive count component of the hurdle model. The relationship on the hurdle component of the hurdle model can be found in Figure SE1 in the supplementary material.

The general pattern of associations shown in Figure 5 coincides with Figure 3, but it provides more detailed information regarding the change in associations over time. The negative association of unemployment is slightly increasing in time. Contrary, the positive association of average income is weakening after the mid-April, then becomes negative in recent days. The proportion of small house and out-of-EU immigrants show a strong association in the early period while weakening in recent days. Lastly, the proportion of university-educated residents shows an interesting and potentially important trend. Their associations are weak (and even positive) in early and recent periods, but the strongest in the middle, the one month following the “state of alarm” announcement.

Discussion

This study examined how socioeconomic inequality is associated with the COVID-19 positive cases in 923 municipalities in Catalonia, Spain. We conducted analyses not only of the cumulative number of confirmed cases on the current date, but also at each date during the observed period, illustrating the dynamic change in their associations. Municipality-level socioeconomic inequalities are measured from three aspects: 1) economic inequality, 2) health vulnerability, and 3) informational inequality. Existing studies on COVID-19 argue for the “simple picture,” that disadvantaged areas unilaterally experience severer effects of COVID-19.

However, our preliminary results imply the evidence is at best mixed when comparing municipalities across Catalonia. These results persist even after considering the potential selection bias caused by municipalities' test accessibility.

Our results reveal two potentially important patterns that are worth discussion. First, in our data, the evidence on economic inequalities and health vulnerabilities is at best mixed. Many of our economic inequality indicators and the proportion of elderly even show opposite associations as expected: economically disadvantaged and elderly-populated municipalities tend to have *fewer* cases of COVID-19. Here, it is easy to conclude that ecological associations do not necessarily coincide with individual-level associations. On the other hand, this finding may imply the important role of community-level socioeconomic structure in the spread of COVID-19. For example, the positive association of income may suggest that municipalities in which richer people reside involve more frequent physical movements in and out of municipalities (e.g., commuting from homes in the suburbs to offices in the city). Those movements may have brought⁷ and spread the virus. If interpreted this way, the negative association of unemployment rate and the positive association of immigrants' proportion may also be reasonable: People with lower socioeconomic status are known to make physical movements within a smaller radius (32), thus expected to contribute less to the spread of the virus through movements in and out of municipalities.

Second, if the spread of COVID-19 is analyzed as dynamic rather than a static phenomenon, patterns of associations between municipality-level characteristics and the spread of COVID-19 transform across time. Especially, we found that the negative association between

⁷ By "bring," see also the left panel of Figure 3. The average income has a positive and statistically significant (at 5% level) association in the hurdle component of the hurdle model (i.e., increases the non-zero confirmation of COVID-19 cases).

informational advantage (measured by the high proportion of university graduates) and the COVID-19 confirmed cases is particularly strong (and otherwise weaker) during the one month immediately following the “state of alarm” intervention by the Spanish government. This pattern may imply that informational advantage may help municipalities to quickly make appropriate responses to the preventive measure (through higher levels of correcting information and of understanding the situation among residents). This advantage, on the other hand, tends to disappear in the recent period as the situation progresses and intentions of preventive measures penetrate to a larger population.

At least two caveats remain in the current analysis. Firstly, since the unit of analysis in our dataset is a municipality, it is risky to connect our findings directly to patterns at individual-level. However, municipality-level analysis has its own advantages. Since preventive measures in the pandemic can only be enforced at the aggregate level, there is an independent advantage to explore the patterns of the spread of the virus through aggregate data. Secondly, our current analysis only looks at the number of positive cases. Here, patterns in the severity of a disease, such as rates of hospitalization and mortality, represent another critical aspect of the current pandemic. We have not yet been able to obtain such data for municipalities of Catalonia, but if data become available, future research must explore the implication of municipality-level socioeconomic inequality for aspects of the current pandemic beyond the number of positive cases.

Data and Method

We obtain the daily data on the spread of COVID-19 in Catalonia from *dades obertes Catalunya* (<https://analisi.transparenciacatalunya.cat>) and other municipality-level statistics from the Statistical Institute of Catalonia (<https://www.idescat.cat>) and the 2011 Spanish Census (https://www.ine.es/censos2011_datos/cen11_datos_resultados_seccen.htm). Out of 947 municipalities in Catalonia, the main analysis includes all municipalities that have relevant statistics available (N = 923).

To estimate the association between COVID-19 cases and social and economic inequality, we use the hurdle model in the main analysis (33).⁸ Given that a significant fraction of Catalanian municipalities has zero cases of COVID-19, the hurdle model explains the number of COVID-19 confirmed cases in two stages. First, the hurdle component model explains if there is a positive count or not (estimated by binomial logit). Second, the positive count model explains the variation in the non-zero count of confirmed cases (estimated by negative binomial regression). This way, the hurdle model copes with the issue of heteroskedasticity caused by the excess number of zeros in the dependent variables. In the first analysis shown in Figure 3, we use cumulative confirmed cases of COVID-19 until the 14th of May, 2020 per municipality as a dependent variable. In the second analysis (Figure 5), we calculate the moving weekly sum of confirmed cases (i.e., the sum of cases confirmed in a week prior to the given date) and estimated its association with regional characteristics daily to assess changes in patterns during the period from the 14th of March to the 14th of May, 2020.

⁸ The analysis in Figure 4 uses negative binomial regression instead of the hurdle model due to reasons described in the Results section.

References

1. WHO, WHO Director-General's remarks at the media briefing on 2019-nCoV on 11 February 2020 (2020) (May 6, 2020).
2. Reuters, Tracking the spread of the novel coronavirus. *Reuters* (May 6, 2020).
3. M. Z. Alonso Artur Galocha, Antonio, Spain's state of alarm: the key measures that are now in place. *EL PAÍS* (2020) (May 13, 2020).
4. M. Marmot, Social determinants of health inequalities. *The Lancet* **365**, 1099–1104 (2005).
5. J. Sommerfeld, "Social Dimensions of Infectious Diseases" in *International Encyclopedia of Public Health (Second Edition)*, S. R. Quah, Ed. (Academic Press, 2017), pp. 569–573.
6. E. Sydenstricker, The Incidence of Influenza among Persons of Different Economic Status during the Epidemic of 1918. *Public Health Rep. 1896-1970* **46**, 154–170 (1931).
7. S. D. Collins, Age and Sex Incidence of Influenza and Pneumonia Morbidity and Mortality in the Epidemic of 1928-29 with Comparative Data for the Epidemic of 1918-19: Based on Surveys of Families in Certain Localities in the United States following the Epidemics. *Public Health Rep. 1896-1970* **46**, 1909–1937 (1931).
8. R. H. Britten, The incidence of epidemic influenza, 1918–19. *Public Health Rep.* **47**, 303–375 (1932).
9. S.-E. Mamelund, Spanish Influenza Mortality of Ethnic Minorities in Norway 1918–1919. *Eur. J. Popul. Rev. Eur. Démographie* **19**, 83–102 (2003).
10. S.-E. Mamelund, A socially neutral disease? Individual social class, household wealth and mortality from Spanish influenza in two socially contrasting parishes in Kristiania 1918–19. *Soc. Sci. Med.* **62**, 923–940 (2006).
11. K. H. Grantz, *et al.*, Disparities in influenza mortality and transmission related to sociodemographic factors within Chicago in the pandemic of 1918. *Proc. Natl. Acad. Sci.* **113**, 13839 (2016).
12. T. Bengtsson, M. Dribe, B. Eriksson, Social Class and Excess Mortality in Sweden During the 1918 Influenza Pandemic. *Am. J. Epidemiol.* **187**, 2568–2576 (2018).
13. S.-E. Mamelund, 1918 pandemic morbidity: The first wave hits the poor, the second wave hits the rich. *Influenza Other Respir. Viruses* **12**, 307–313 (2018).
14. D. L. Dee, *et al.*, Racial and Ethnic Disparities in Hospitalizations and Deaths Associated with 2009 Pandemic Influenza A (H1N1) Virus Infections in the United States. *Ann. Epidemiol.* **21**, 623–630 (2011).
15. D. L. Thompson, *et al.*, Risk Factors for 2009 Pandemic Influenza A (H1N1)–Related Hospitalization and Death Among Racial/Ethnic Groups in New Mexico. *Am. J. Public Health* **101**, 1776–1784 (2011).
16. S. Kumar, S. C. Quinn, K. H. Kim, L. H. Daniel, V. S. Freimuth, The Impact of Workplace Policies and Other Social Factors on Self-Reported Influenza-Like Illness Incidence During the 2009 H1N1 Pandemic. *Am. J. Public Health* **102**, 134–140 (2011).
17. P. D. Rutter, O. T. Mytton, M. Mak, L. J. Donaldson, Socio-economic disparities in mortality due to pandemic influenza in England. *Int. J. Public Health* **57**, 745–750 (2012).
18. J. M. Mayoral, *et al.*, Social factors related to the clinical severity of influenza cases in Spain during the A (H1N1) 2009 virus pandemic. *BMC Public Health* **13**, 118 (2013).
19. A. Y. Li, *et al.*, Multivariate Analysis of Factors Affecting COVID-19 Case and Death Rate in U.S. Counties: The Significant Effects of Black Race and Temperature. *medRxiv*,

- 2020.04.17.20069708 (2020).
20. T. C. Rose, *et al.*, Inequalities in COVID19 mortality related to ethnicity and socioeconomic deprivation. *medRxiv*, 2020.04.25.20079491 (2020).
 21. The Gurdian, Poor and vulnerable hardest hit by pandemic in Spain. *The Guardian* (2020) (May 14, 2020).
 22. J. B. Dowd, *et al.*, Demographic science aids in understanding the spread and fatality rates of COVID-19. *Proc. Natl. Acad. Sci.* **117**, 9696–9698 (2020).
 23. S. Sannigrahi, F. Pilla, B. Basu, A. S. Basu, The overall mortality caused by COVID-19 in the European region is highly associated with demographic composition: A spatial regression-based approach. *ArXiv200504029 Q-Bio* (2020) (May 15, 2020).
 24. P. Sun, X. Lu, C. Xu, W. Sun, B. Pan, Understanding of COVID-19 based on current evidence. *J. Med. Virol.* **92**, 548–551 (2020).
 25. A. S. Abdelhafiz, *et al.*, Knowledge, Perceptions, and Attitude of Egyptians Towards the Novel Coronavirus Disease (COVID-19). *J. Community Health* (2020) <https://doi.org/10.1007/s10900-020-00827-7> (May 5, 2020).
 26. J. M. Clements, Knowledge and behaviors toward COVID-19 among U.S. residents during the early days of the pandemic. *medRxiv*, 2020.03.31.20048967 (2020).
 27. A. Y. Naser, *et al.*, Knowledge and practices towards COVID-19 during its outbreak: a multinational cross-sectional study. *medRxiv*, 2020.04.13.20063560 (2020).
 28. C. M. Rios-González, Knowledge, attitudes and practices towards COVID-19 in Paraguayans during outbreaks: a quick online survey (2020) <https://doi.org/10.1590/SciELOPreprints.149> (May 5, 2020).
 29. J. Zegarra-Valdivia, B. N. C. Vilca, R. J. A. Guerrero, Knowledge, perception and attitudes in Regard to COVID-19 Pandemic in Peruvian Population (2020) <https://doi.org/10.31234/osf.io/kr9ya> (May 5, 2020).
 30. B.-L. Zhong, *et al.*, Knowledge, attitudes, and practices towards COVID-19 among Chinese residents during the rapid rise period of the COVID-19 outbreak: a quick online cross-sectional survey. *Int. J. Biol. Sci.* **16**, 1745–1752 (2020).
 31. G. J. Borjas, “Demographic Determinants of Testing Incidence and COVID-19 Infections in New York City Neighborhoods” (National Bureau of Economic Research, 2020) <https://doi.org/10.3386/w26952> (May 5, 2020).
 32. V. Frias-Martinez, J. Virseda-Jerez, E. Frias-Martinez, On the relation between socio-economic status and physical mobility. *Inf. Technol. Dev.* **18**, 91–106 (2012).
 33. C. E. Rose, S. W. Martin, K. A. Wannemuehler, B. D. Plikaytis, On the Use of Zero-Inflated and Hurdle Models for Modeling Vaccine Adverse Event Count Data. *J. Biopharm. Stat.* **16**, 463–481 (2006).

Data availability

All data analysed for this study are publicly available. The details are in the Data and Method section.

Code availability

The programming code to reproduce the results of the paper is available from G.K's GitHub page (<https://github.com/gentok/covid19spain>).

Author contributions

G.K was responsible for the data handling. S.A. did the literature review. G.K and R.M did the data analysis. All authors designed the study and R.M wrote the draft manuscript with assistance in the interpretation of the results from G.K. and S.A.

Competing interest declaration The authors declare no competing interests.

Appendix A: Relationships among Socioeconomic Inequality Indicators

Figure SA1: Correlations among socioeconomic inequality indicators

Note: For those variables with missing values, pairwise correlations are calculated.

Appendix B: Geographical Distribution of Socioeconomic Inequality Indicators

Figure SB1: Registered unemployment rate across municipalities in Catalonia (2019)

Figure SB2: Unemployment rate (alternative measure) across municipalities in Catalonia (2011)

Note: Municipalities with missing measurement are greyed out.

Figure SB3: Average taxable base income (in 10,000 euro) across municipalities in Catalonia (2017)

Figure SB4: Proportion of small residence (90 square meter or smaller) across municipalities in Catalonia (2011)

Note: Municipalities with missing measurement are greyed out.

Figure SB5: Proportion of Immigrants from outside of EU across municipalities in Catalonia (2018)

Figure SB6: Proportion of service sector workers in labour population across municipalities in Catalonia (2018)

Note: Municipalities with missing measurement are greyed out.

Figure SB7: Crude death rate across municipalities in Catalonia (2018)

Figure SB8: Proportion of residents who are 65 years old or over across municipalities in Catalonia (2019)

Figure SB9: Proportion of university educated across municipalities in Catalonia (2011)
Note: Municipalities with missing measurement are greyed out.

Figure SB10: Population across municipalities in Catalonia (2019)

Figure SB11: Population density across municipalities in Catalonia (2019)

Figure SB12: Household size across municipalities in Catalonia (2011)

Note: Municipalities with missing measurement are greyed out. In calculating the average household size, those who reside in the household of size 6 or more are counted as 7.

Appendix C: Geographical Distribution of Test Cases and Positive Rates

Figure SC1: The number of cumulative tested cases of COVID-19 per 10,000 population across municipalities in Catalonia on the 14th of May, 2020

Note: Municipalities with less than 10 cumulative tests are greyed out.

Figure SC2: The positive rate of COVID-19 tests across municipalities in Catalonia on the 14th of May, 2020

Note: Municipalities with less than 10 cumulative tests are greyed out.

Appendix D: Robustness Checks for Main Cumulative Results

Figure SD1: Associations between socioeconomic inequalities and total confirmed positive cases of COVID-19 in municipalities across Catalonia on the 14th of May (Model 2)

Note: Associations are estimated through the hurdle model. Each independent variable is standardized. The dot shows the odds / incidence-rate ratio estimate of each variable. The odds ratio (left panel) illustrates how many times the odds of non-zero count is more likely than zero count, and incidence-rate ratio (right panel) indicates how many times the odds of one more positive count is more likely than no more positive count. The gothic and light bars correspond to their 90% and 95% confidence intervals, respectively. The shape of dots (and colours of bars) indicates whether or not estimated odds/incidence-rate ratios are statistically significant at 5 or 10% levels. N=893.

Figure SD2: Associations between socioeconomic inequalities and total confirmed positive cases of COVID-19 in municipalities across Catalonia on the 14th of May (Model 3)

Note: Associations are estimated through the hurdle model. Each independent variable is standardized. The dot shows the odds / incidence-rate ratio estimate of each variable. The odds ratio (left panel) illustrates how many times the odds of non-zero count is more likely than zero count, and incidence-rate ratio (right panel) indicates how many times the odds of one more positive count is more likely than no more positive count. The gothic and light bars correspond to their 90% and 95% confidence intervals, respectively. The shape of dots (and colours of bars) indicates whether or not estimated odds/incidence-rate ratios are statistically significant at 5 or 10% levels. N=464.

Figure SD3: Associations between socioeconomic inequalities and total confirmed positive cases of COVID-19 in municipalities across Catalonia on the 14th of May (Model 4)

Note: Associations are estimated through the hurdle model. Each independent variable is standardized. The dot shows the odds / incidence-rate ratio estimate of each variable. The odds ratio (left panel) illustrates how many times the odds of non-zero count is more likely than zero count, and incidence-rate ratio (right panel) indicates how many times the odds of one more positive count is more likely than no more positive count. The gothic and light bars correspond to their 90% and 95% confidence intervals, respectively. The shape of dots (and colours of bars) indicates whether or not estimated odds/incidence-rate ratios are statistically significant at 5 or 10% levels. N=923.

Appendix E: Dynamic Analysis of Daily Moving Weekly Sum of Positive Cases (Hurdle Component)

Figure SE1: Trends in associations between socioeconomic inequalities and daily moving weekly sum of COVID-19 positive cases in Catalonia between the 14th of March and the 14th of May (hurdle component)

Note: On each day, associations are estimated by the hurdle model (N=923). This figure presents associations in hurdle component (see Figure 5 for associations in positive count component). Each predictor is standardized, thus odds ratios correspond to its 1 standard deviation increase.

Appendix F: Dynamic Analysis of Daily Cumulative Sum of Positive Cases

Figure SF1: Trends in associations between socioeconomic inequalities and daily cumulative sum of COVID-19 positive cases in Catalonia between the 14th of March and the 14th of May (positive count component)

Note: On each day, associations are estimated by the hurdle model (N=923). This figure presents associations in positive count component (see Appendix Figure SE1 for associations in hurdle component). Each predictor is standardized, thus incidence-rate ratios correspond to its 1 standard deviation increase.

Figure SF2: Trends in associations between socioeconomic inequalities and daily cumulative sum of COVID-19 positive cases in Catalonia between the 14th of March and the 14th of May (hurdle component)

Note: On each day, associations are estimated by the hurdle model (N=923). This figure presents associations in hurdle component (see Figure 4 for associations in positive count component). Each predictor is standardized, thus odds ratios correspond to its 1 standard deviation increase.