

Cite as: G. E. Mena *et al.*, *Science*
10.1126/science.abg5298 (2021).

Socioeconomic status determines COVID-19 incidence and related mortality in Santiago, Chile

Gonzalo E. Mena^{1†*}, Pamela P. Martinez^{2,3,4,5†*}, Ayesha S. Mahmud^{2,6}, Pablo A. Marquet^{7,8,9,10,11},
Caroline O. Buckee^{2‡}, Mauricio Santillana^{2,12,13‡}

¹Department of Statistics, University of Oxford, Oxford, UK. ²Center for Communicable Disease Dynamics, Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA. ³Department of Microbiology, University of Illinois at Urbana-Champaign, Urbana, IL, USA. ⁴Department of Statistics, University of Illinois at Urbana-Champaign, Champaign, IL, USA. ⁵Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL, USA. ⁶Department of Demography, University of California, Berkeley, CA, USA. ⁷Departamento de Ecología, Facultad de Ciencias Biológicas, Pontificia Universidad Católica de Chile, Santiago, Chile. ⁸Instituto de Ecología y Biodiversidad (IEB), Santiago, Chile. ⁹The Santa Fe Institute, Santa Fe, NM, USA. ¹⁰Instituto de Sistema Complejos de Valparaíso (ISCV), Valparaíso, Chile. ¹¹Centro de Cambio Global UC, Pontificia Universidad Católica de Chile, Santiago, Chile. ¹²Computational Health Informatics Program, Boston Children's Hospital, Boston, MA, USA. ¹³Department of Pediatrics, Harvard Medical School, Boston, MA, USA.

†These authors contributed equally to this work. ‡These authors contributed equally to this work.

*Corresponding author. Email: gonzalo.mena@stats.ox.ac.uk (G.E.M.); pamelapm@illinois.edu (P.P.M.)

The current COVID-19 pandemic has impacted cities particularly hard. Here, we provide an in-depth characterization of disease incidence and mortality, and their dependence on demographic and socioeconomic strata in Santiago, a highly segregated city and the capital of Chile. Our analyses show a strong association between socioeconomic status and both COVID-19 outcomes and public health capacity. People living in municipalities with low socioeconomic status did not reduce their mobility during lockdowns as much as those in more affluent municipalities. Testing volumes may have been insufficient early in the pandemic in those places, and both test positivity rates and testing delays were much higher. We find a strong association between socioeconomic status and mortality, measured either by COVID-19 attributed deaths or excess deaths. Finally, we show that infection fatality rates in young people are higher in low-income municipalities. Together, these results highlight the critical consequences of socioeconomic inequalities on health outcomes.

The coronavirus disease 2019 (COVID-19) pandemic is an ongoing public health crisis. While many studies have described the transmission of SARS-CoV-2 –the virus that causes COVID-19– in North America, Europe, and parts of Asia (1–5), the characterization of the pandemic in South America has received less attention, despite the severe impact in many countries during the Southern Hemisphere winter season. While confirmed COVID-19 cases are an important public health measure to estimate the level of spread of infections caused by SARS-CoV-2, they may not be a reliable indicator of incidence because of biases due to population-level health-seeking behavior, surveillance capacities, and the presence of asymptomatic individuals across regions (6). Analyses of COVID-19-related deaths as well as excess mortality provide an alternative and potentially less biased picture of epidemic intensity (7, 8). This is in part because ascertainment biases may be less pronounced for deaths than for confirmed cases, as people dying from COVID-19 are more likely to have experienced severe symptoms and thus, more likely to have been documented as COVID-19 positive cases by health surveillance systems. Age specific death data may also help explain the heterogeneity in COVID-19 burden and COVID-19 attributable deaths in different countries (9). However, the role of other factors, such as socioeconomic status – which is correlated with

health care access– on fatality and disease burden, remains a particularly important open question (10) for cities with significant economic disparities.

Here, we analyzed incidence and mortality attributed to SARS-CoV-2 infection and its association with demographic and socioeconomic status across the urban metropolitan area of the capital of Chile, known as ‘Greater Santiago’. Unlike many other countries, Chile set up a remarkably thorough reporting system and made many key data sets publicly available. To understand spatial variations in disease burden, we estimated excess deaths and infection fatality rates across this urban area. To understand disparities in the health care system, we analyzed testing capacity and delays across municipalities. We then demonstrate strong associations of these health indicators with demographic and socioeconomic factors. Together, our results show that socioeconomic disparities explain a large part of the variation in COVID-19 deaths and under-reporting, and that those inequalities disproportionately affected younger people.

Association between socioeconomic status and disease dynamics

The Greater Santiago area is composed of 34 municipalities

–defined as having more than 95% of its area urbanized– and is home to almost 7 million people. While this urban center accounts for 36% of the country’s population, it has reported 55% of the confirmed COVID-19 cases and 65% of the COVID-19 attributed deaths prior to epidemiological week 36 (end of August 2020). Socioeconomic status (SES) in the municipalities varies widely, with Vitacura having the highest value (SES = 93.7) and La Pintana the lowest one (SES = 17.0; Fig. 1A), and this difference is reflected in the impact of the pandemic during the Southern Hemisphere winter of 2020. The maximum incidence in Vitacura was 22.6 weekly cases per 10,000 individuals during the middle of May, while La Pintana reported a maximum of 76.4 weekly cases per 10,000 individuals during the first week of June (Fig. 1B). As shown in Fig. 1C and fig. S1, the attributed COVID-19 deaths follow a similar (yet lagged) temporal pattern to the number of reported COVID-19 cases. For instance, the highest rate of 4.4 weekly deaths per 10,000 individuals is observed in San Ramón, a municipality with a SES of 19.7, while Vitacura reported a maximum of 1.6 weekly deaths per 10,000 in June. These social inequalities impact the overall COVID-19 mortality rates as shown in Fig. 1D.

Changes in human mobility –a proxy for physical distancing– during lockdown periods follow a similar trend. Using human mobility indicators, inferred from anonymized mobile phone data obtained from the Facebook Data for Good Initiative, we show that the two municipalities with highest socioeconomic status exhibited a reduction in mobility by up to 61% during the full lockdown (dark green, Fig. 1E), compared to the ones with lowest SES, which, on average, reduced their mobility to 40% during the this period (dark pink, Fig. 1E). This relationship between reductions in mobility and SES was present during all time-periods considered for this study (Fig. 1F) and supports the hypothesis that people in poorer regions cannot afford to stay at home during lockdowns. Our result is consistent with analyses of New York City neighborhoods (11) and with findings from other studies conducted in Santiago that used different socioeconomic and mobility metrics (12–14).

Epidemic reconstruction reveals early transmission dynamics

In order to examine the possible bias present in the incidence data, we reconstructed SARS-CoV-2 infections over time by implementing a method called regularized mortality MAP (RmMAP). RmMAP back calculates the most likely infection numbers given the temporal sequence of deaths, the onset-to-death distribution, and the demography-adjusted infection fatality rate (IFR). Figure 2A shows the outcomes of this inference process, where the reconstructions from our approach and other methods are able to capture the main peak observed in May and June, with an estimate of the number

of infected individuals that is 5 to 10 times larger than the reported values.

The reconstructions also reveal important differences in the inferred number of infections during March of 2020, the month in which the virus was introduced in Chile by travelers from affluent municipalities. We analyzed the number of tests performed between March 8th and April 9th, and find a significantly higher number of tests performed in municipalities with high SES (Fig. 2B), especially during the first two weeks of March (Fig. 2D). In addition, an early peak of reported cases was only observed in high SES municipalities during middle March (Fig. 2C), despite the fact that several COVID-19 deaths, which are lagged with respect to infection by up to several weeks, were reported in low SES municipalities during the same period. These findings suggest that an early first wave of infections occurred during March and quickly spread through the rest of the city without being captured by the official counts. Our RmMAP estimates at the municipality level support this claim, as they capture a high volume of early infections in most municipalities (Fig. 2E), a scenario that largely deviates from the official tallies (Fig. 2C).

To further validate the hypothesis of an early under-reporting in low SES municipalities, and to rule that these early activity estimates are not an artifact of our method, we performed experiments on a synthetic elementary model of two peaks of different sizes separated in time (supplementary materials). These experiments confirm that RmMAP is capable of recovering this bi-modal phenomena, while other methods fail to do so; they over-smooth the true signal and the earlier peak is typically not recovered. This early under-reporting signal suggests that the patterns of mortality and testing observed across the Greater Santiago are partially explained by an early failure of healthcare systems in informing the population with sufficient situational awareness about the real magnitude of the threat (15).

Excess deaths match COVID-19 attributed deaths

Excess deaths –defined as the difference between observed and expected deaths– can provide a measure of the actual impact of the pandemic in mortality by quantifying direct and indirect deaths related to COVID-19 (7, 8, 16). We estimated the expected deaths for 2020 by fitting a Gaussian process model (17) to historical mortality data from the past twenty years, and used them to identify the increased mortality during to the pandemic, controlling for population growth and seasonality. As shown in Fig. 3A, the number of deaths observed between May and July 2020 is more than 1.73 [1.68, 1.79] times the expected value, with a peak surpassing 2110 death counts in epidemiological week 24 (first week of June, 2020) compared to an expected value of 802 deaths and an average number of deaths of 798 between 2015 and 2019.

When comparing the number of deaths by age in the year 2020 with our model's predictions we observe striking patterns. Although people younger than 40 years old have an overall lower mortality rate than those from older age groups as expected, they still exhibit a nearly two-fold increase in the total deaths with a peak in the observed deaths occurring 2 weeks earlier than for those older than 60 years old (Fig. 3B). For the age groups 40-60, 60-80, and older than 80, the observed deaths are 2.8, 3.2, and 2.4 times higher than expected, respectively. Even though the age group 80+ exhibits the highest expected mortality values for 2020, the group that contains people between 60 and 80 years old displays the highest weekly count (936 during epidemiological week 24), the biggest deviation from the predicted values, and the highest values of excess deaths (645 more deaths than expected, Fig. 3B).

COVID-19 attributed deaths for the entire Greater Santiago area fall within the credible intervals of excess deaths until late June, when the attributed deaths increase to rates that are even higher than the excess deaths, suggesting that under-reporting in COVID-19 attributed deaths is unlikely (Fig. 3C). COVID-19 confirmed deaths—those with a PCR-confirmed SARS-CoV-2 test—follow a similar temporal pattern, and the difference between confirmed and COVID-19 attributed deaths gets smaller toward the end of August, indicative of an improved testing capacity. This pattern is consistent when compared to normalized deaths by population size for each municipality (Fig. 3D), which also shows COVID-19 attributed deaths higher than the excess deaths in most of the cases. The anomalies in the observed versus predicted deaths for 2020 across different age groups also display a significant negative association with socioeconomic status, except for the 80+ group (Fig. 3E), suggesting a higher death burden in lower SES municipalities, independent of their age composition. Furthermore, the two municipalities with SES higher than 80 (Las Condes and Vitacura) had z-scores of much smaller magnitude (with the exception of the oldest age group) indicating that their patterns of mortality did not deviate much from what would have been expected on a normal year in people younger than 80 years old.

Although the observation that COVID-19 attributed deaths are greater than the estimated excess deaths might be counterintuitive (Fig. 3D), it may indicate the presence of changes in overall mortality patterns due to other causes, including a lower number of deaths due to reduction in the mobility. In addition, lower numbers of deaths were reported for respiratory infectious diseases such as influenza and pneumonia, and cancer during July and August of 2020 compared to the period 2015-2019 (Fig. 3F). Changes in mortality from respiratory diseases can be explained by a mild influenza season in the Southern Hemisphere during the winter of 2020 (18), which is consistent with our observation that much

fewer cases of respiratory viruses have been detected in Chile during the 2020 season (supplementary materials). A decrease in the number of cancer attributed deaths can be explained by mortality displacement (19, 20), but additional analyses need to be conducted to establish this hypothesis. Alternative explanations for changes in all-cause mortality should also consider possible changes in external and behavioral causes of mortality. We do not observe a substantial contribution from these causes (see supplementary materials, along with additional detailed analyses).

More testing with lower waiting times in wealthy areas

To further understand the consequences of insufficient early testing, we conducted a deeper analysis of different testing metrics at the municipality level. We first looked at testing capacity measured as weekly positivity rates, the fraction of tests that are positive for SARS-CoV-2. Our results show that the positivity signal tracked the course of the epidemic, peaking at times of highest incidence between May and July, and suggesting a highly saturated health-care system during this period across the entire city (Fig. 4A). A strong negative association between positivity and SES (Fig. 4B) further denotes difficulties in access to health care that is even more pronounced in lower SES municipalities. Despite changes in positivity rates over time, this testing metric also significantly correlated with number of cases (Fig. 4C) and number deaths (Fig. 4D).

Our findings on the number of tests conducted show a rather paradoxical association with SES and mortality. Many months into the epidemic, the early positive association between tests per capita and SES (Fig. 2B) reversed (Fig. 4E), indicative of an improvement in testing capacity over time, so that more tests were performed in the most affected areas. Similarly, the number of tests started to positively correlate with deaths (Fig. 4F), suggesting that the number of tests are strong predictors of mortality.

We also analyzed testing capacity by estimating the delays in obtaining test results. We inferred the distribution of the delay between onset of symptoms and report of the results, from which we obtained the proportion of cases that are publicly reported within one week since the onset of symptoms or *timeliness* (21). As shown in Fig. 4G, timeliness follows a similar temporal course as test positivity during May and part of June, but in the opposite direction. This metric is associated with SES, suggesting that municipalities with low SES, on average, get their test results later than the ones with high SES (Fig. 4H). Timeliness also negatively correlates with number of cases (Fig. 4I), total number of deaths (Fig. 4J), and with positivity (Fig. 4K). When looking at tests per death, a metric that can be used as a faithful proxy of testing capacity (22), we observe a positive correlation with socioeconomic status (Fig. 4L), indicating that testing disparities persisted

during the epidemic, with low SES areas being affected the most. In the supplementary materials we further discuss the associations between our metrics and case counts.

Infection fatality rate depends on socioeconomic status

In the absence of serological surveys, a direct inference of an infection fatality rate (IFR) is challenging. The degree of ascertainment depends on many factors, including testing capabilities and the likelihood of having symptomatic infections. Also, unlike deaths, age information of reported cases is not available at the municipality level, making this inference more challenging. To address these hurdles and to have estimates of the IFR, we implemented a hierarchical Bayesian model that considers the relationship between deaths, observed cases, and true infections across location, time, and age group. We first estimated the case fatality rate (CFR) by assigning total cases into age groups in a simple way that projects the overall age-distribution of cases to particular municipality demographics (Fig. 5A, see supplementary materials for details). With the exception of the oldest age group, case fatality rate shows a negative association with socioeconomic status. Similarly, our resulting IFR estimates once corrected for under-ascertainment display a similar pattern (Fig. 5B) but on an order of magnitude lower than the CFR estimates. We then grouped the municipalities into four categories of similar sizes and label them as low, mid-low, mid-high, and high socioeconomic category. When comparing the IFR ratio between the low and the high SES categories, the results show significantly higher infection fatality rate in the low SES group in people younger than 80 years old (Fig. 5C). The age groups 60-80 and 40-60 exhibit an IFR that are 1.4 and 1.7 times higher respectively in the low SES category, compared to the high SES one. The difference is even more pronounced in the younger age group (0-40 years old), which shows values of IFR that is 3.1 times higher for the municipalities with the lowest socioeconomic status. Altogether, these results are in line with the analyses of excess deaths presented in Fig. 3E. The lack of association between IFR and SES in the oldest age group can be attributed to a lower life expectancy (23), which in fact is factored in the estimation of SES (see methods for details) and that elderly people might be, in general, healthier enough to survive until that age.

Discussion

In order to understand the true burden of COVID-19, it is critical to consider demographic and socioeconomic factors and their consequences for the public health response. Here, we analyzed data from the capital of Chile, a highly segregated city. Our results align with the recent literature on uneven health risks globally, which has highlighted how socially and economically deprived populations are more vulnerable to

the burden of epidemics (24, 25). Mounting evidence suggests that such differences have also manifested in the context of the COVID-19 pandemic (26, 27). Since the pathways modulating these differential outcomes are not well-understood, comprehensive accounts are urgently needed (28), so that more resilient and socially-aware public-health strategies can be planned in advance of future pandemics. In Chile, recent studies have suggested a link between SES and effectiveness of non-pharmaceutical interventions such as stay-at-home orders (12, 13, 29). Our work further explores this topic by providing an holistic perspective about how the interplay between behavioral, social, economic, and public-health factors, modulates the observed heterogeneity in infection incidence and mortality. Along with the main findings, we also introduced several methodological innovations. Our Bayesian method for joint inference of infection fatality rates and under-reporting is a new contribution in this field. We show that it may not be necessary to have complete epidemiological data sets (here, age) to draw valid inferences, as long as the solution space is constrained enough by meaningful priors and demographic structure.

Our results show a strong link between socioeconomic and demographic factors with COVID-19 outcomes and testing capacity of COVID-19 in Santiago. This association is manifested as a reinforcing feedback loop, as highlighted by our findings. First, our analysis of human mobility indicates that municipalities with lower socioeconomic status were less compliant with stay-at-home orders, possibly because people from lower SES areas are unable to work from home, leaving them at a higher disease risk. Second, our analyses revealed an under-reporting of infections in low income areas at the start of the outbreak. Since public-health measures were taken in response to nominal case counts, these places were under prepared, with a poor health-care response that resulted in higher death counts. Third, anomalies in the overall excess deaths are higher in low SES areas, particularly in people younger than 80 years old, suggesting that more vulnerable municipalities were hit the hardest. Fourth, the analyses of test positivity rates, timeliness, and tests per death indicate an insufficient deployment of resources for epidemiological surveillance. Higher positivity rates in health care centers suggest the need for greater testing and detection. At the same time, slower turnaround in test results can lead to greater potential for transmission, since even small delays between onset of symptom, testing, and final isolation, significantly hinder the capability of public health systems to contain the epidemic (30). Finally, infection fatality rates were higher in lower SES municipalities, especially among younger people.

We propose two complementary explanations for the association between infection fatality rate and socioeconomic status. First, a higher IFR may reflect limited access to health

services during the pandemic, and the strong association between the number of tests per death and SES supports this claim. We also show in supplementary materials that the South and West zones (based on health coverage division) have 4 times fewer beds per 10,000 people and 4 times lower proportion enrolled in the private health system than the East zone, which contains all the municipalities with an SES of 60 or higher. Strikingly, more than 90% of the COVID-19 attributed deaths in the South and West zone occurred in places other than healthcare facilities, compared to 55% in the East. Second, more vulnerable communities may experience a higher prevalence of the comorbidities (31) that are associated with more severe presentations of COVID-19. People in low SES municipalities are more likely to be overweight and to live in overcrowded conditions (supplementary materials), factors that ultimately can put these populations at higher disease risk. The interaction of these two explanations can lead to a high disparity among different socioeconomic groups.

Our findings need to be considered in light of the following limitations. Mobility data from mobile phones are likely to be biased due to differential mobile phone ownership in different demographic groups. While Facebook mobility data can be biased in this way, our results are consistent with other studies in Santiago that used different socioeconomic and movement measurements [see (12–14) and supplementary materials]. Our methods depend on several assumptions. The back-calculated RmMAP estimates rely on a choice of the infection-to-death distribution and assume that the IFR do not change over time, and the excess mortality estimates depend on the choice of a kernel. Our IFR estimates are derived from a complex Bayesian model and are based on assumptions regarding reporting rates and age distribution of infections. Extensive sensitivity analyses suggest that our results are stable to deviations from these assumptions (supplementary materials).

To conclude, this study highlights major consequences of healthcare disparities in a highly segregated city, and provides new methodologies, that account for incomplete data, for studying infectious disease burden and mortality in other contexts.

Materials and methods

Data

Socioeconomic status

We define the Socioeconomic status index (SES) as $SES = 100 - SPI$, where SPI is the *Social Priority Index* (or ‘Indice de Prioridad Social’ in Spanish) estimated for 2019. The SPI index varies between 0 and 100, and has been reported yearly since 1995 by the Chilean Ministry of Social Development and Family. The SPI value denotes the priority of each municipality for the social programs of the regional government, and

thus, municipalities with lower SES have higher social priority. The SPI index equally weights three components: (i) income and poverty, (ii) access and quality of education, and (iii) health factors such as access to healthcare and life expectancy. For each component, the values are standardized on a common scale from 0 to 100, where the value 100 represents the worst relative situation (highest priority) and 0 the best situation (least priority).

COVID-19

At the end of January 2020, the Chilean government determined that all suspected cases of COVID-19 must be notified in a mandatory and immediate manner to the respective Health Epidemiology Unit and the Ministry of Health, through the specific form on the EPIVIGILA platform. In addition to the suspected cases that are identified in healthcare facilities, the government also implemented an active testing surveillance program to identify asymptomatic and pre-symptomatic cases. The criteria for the active testing are: i) people who have not been identified yet as confirmed or suspected COVID-19 or (ii) living in vulnerable areas, and (iii) individuals who live in institutions for a long time such as jails, nursing homes, the National Service for Minors, among others. Symptoms onset dates are reported by the patient to a physician, in the case that the person attended a health institution, or by the volunteers that are conducting surveillance in the community through a survey.

The Chilean Ministry of Science, Technology, Knowledge, and Innovation has made possible the access to aggregated data collected through the EPIVIGILA platform, which are available in the format of multiple reports. These reports also contain data on population projections for 2020, testing, positivity, and other metrics used in the study. One of the reports tracks the number of cases whose onset of symptoms started at a given epidemiological week, for each municipality. Given that they are published twice a week (typically Monday and Friday), we were able to analyze the history of such reports to estimate the delays. Timeliness is thus defined as the probability of getting a retrospective delay smaller than 7 days, based on the Monday’s reports. More details can be found in the supplementary materials.

Mortality

The Vital Statistics System in Chile is continuous, mandatory, and centralized. It is composed of the Civil Registry and Identification Service (CRIS), the National Institute of Statistics, and the Ministry of Health through the Department of Health Statistics and Information (DHSI). When a person dies, a medical death certificate is generated by the CRIS and distributed to health institutions. The mortality database is built with the death certificates, which are subjected to a rigorous validation process, to guarantee the reliability and validity of

the information. The DHSI standardizes the clinical terms in the format of the International Statistical Classification of Diseases (ICD-10). Since March 2020, the DHSI has implemented the recommendations of the WHO for coding the deaths resulting from COVID-19. In this study, confirmed COVID-19 deaths correspond to deaths in which the virus has been identified with a positive PCR test and have been coded as U07.1. Similarly, attributed COVID-19 deaths correspond to deaths in which the virus was not identified but clinically diagnosed as probable or suspected COVID-19 case, and have been coded as U07.2.

Human movement

Facebook's Data for Good has provided access to their Geoin-sights portal in response to COVID-19 crisis, from where it is possible to obtain aggregated data of their users (32). These data sets are anonymized and contain information of Facebook users that have a smartphone with the location services enabled. The movement vector from tile i to j (with $i \neq j$) at time t is defined as the transition from the modal location i at the preceding 8-hour bin to the modal location j in the current 8-hour bin. Facebook also provides a baseline value, defined as the average number of users who transit from tile i to j at a given day of week and time of day during a baseline period. The baseline period corresponds to the 45 days prior to the initiation of the movement data for that particular location (for Chile, the data collection was initialized on 03/25/2020). Using this data set, we calculated the percentage change compared to baseline for each i to j transition at a given 8-hour period, and then estimated the average percentage change for each municipality and epidemiological week. We only used the starting location (municipality) for the average percentage change estimation. The size of the side of the tile is approx. 2.4 km.

Models

Inference of SARS-CoV-2 infections with RmMAP

We aim to estimate the number of infected individuals over time I_s given a series of observed COVID-19 attributed deaths D_t and a known onset-to-death distribution T . We use a Poisson deconvolution model for deaths given I and T :

$$D_t | T, I \sim \text{Poisson}\left(\sum_s T_{t-s} I_s\right) \quad (1)$$

where $T_s = P(T = s)$ is the probability that the onset-to-death equals s days. Estimates of I maximizing Eq. 1 can be obtained with an expectation maximization algorithm (6, 33–35), but the outcome is typically unstable (36). RmMAP overcomes this issue by adding a quadratic penalty to the log-likelihood. The iterations of RmMAP write as

$$\hat{I}^{new} = \frac{1}{4\lambda} \sqrt{1 + 8\lambda I^{old}} - 1 \quad (2)$$

$$I_s^{new} = \hat{I}_s^{new} \frac{1}{\sum_t T_{t-s}} \sum_t \frac{D_t T_{t-s}}{\sum_{s'} T_{t-s'} \hat{I}_{s'}^{new}} \quad (3)$$

By scaling the final series I^{new} by the inverse IFR we obtain the inferred values of infected individuals over time. A detailed discussion of this method along with sensitivity analysis and comparison with existing methodology are presented in the in the supplementary materials.

Estimation of excess deaths

We used Gaussian Processes (GP) regression (17) to estimate excess deaths for 2020. GPs can be understood as an infinite dimensional Bayesian regression: in the finite dimensional case one fits $y_i = \sum_j w_j x_i + \epsilon_i$ where ϵ_i are Gaussian independent identically distributed errors, x_i are covariates and w_i coefficients sampled from a prior $p(w)$. Likewise, with GPs we fit $y_i = f(x_i) + \epsilon_i$ where f is a function sampled from a prior over functions $p(f)$. GPs are appealing because the level of complexity is automatically adjusted by the complexity of data, and because they are computationally tractable.

Priors over f are specified through a kernel K , which encodes the correlational structure of data so that $K(x, x')$ is simply the *prior* covariance between $f(x)$ and $f(x')$. K depends on a finite number of unknowns θ (so $K = K_\theta$) that have to be inferred as well.

We used a GP to account for both long-term trends in mortality as well as seasonality. As in (17), we consider kernels of the form

$$K_\theta = K_\theta^1 + K_\theta^2 \quad (4)$$

where K_θ^1 is an exponential kernel representing the long-term variation, and is given by

$$K_\theta^1(x, x') = \theta_1^2 \exp\left(-\frac{(x-x')^2}{2\theta_1^2}\right) \quad (5)$$

and K_θ^2 is a periodic times exponential kernel representing seasonal variation

$$K_\theta^2(x, x') = \theta_3^2 \exp\left(-\frac{(x-x')^2}{2\theta_4^2} - \frac{2\sin^2(\pi(x-x'))}{\theta_5^2}\right) \quad (6)$$

We considered an additional source of unstructured randomness through the term $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$. We performed Bayesian inference (MCMC) over the joint distribution parameters (θ, σ^2) and death counts for each time period of the 2020 year, based on 2000-2019 all-cause mortality data and suitable priors for the parameters. In the supplementary materials we comment on more specific aspects, and provide an extensive evaluation of our model.

Infection fatality rates

We deployed a hierarchical Bayesian joint model for reporting rates (and hence, IFR) per age group (a taking values 0 – 40, 40 – 60, 60 – 80 and 80 +) and municipality m , collapsing over the temporal dimension. We infer the number of infected individuals (and hence, IFR) based on reported cases C , positivity rates over time (t , month), and municipality, and total and COVID-19 attributed deaths D . The main appeal of this framework, is that although most of the components are not identifiable (e.g., if reporting rates and true cases are both unknown, the same observed case counts can be achieved by multiplying both by the same factor) (37), we can borrow from better known quantities (e.g., rough estimates of prevalence, reporting, etc) to enhance identification while propagating the appropriate levels of uncertainty over the parameters.

Specifically, the reporting rate $r_{m,t}$ links to the observed positivity rates $pos_{m,t}$ (in log-scale) through a logistic-linear relation (with parameters β), and we have include random effects $\epsilon_{m,t}$ to represent unobserved causes of reporting

$$\text{logit}(r_{m,t}) = \beta_0 + \beta_1 \times pos_{m,t} + \epsilon_{m,t} \quad (7)$$

Total infections by municipality and age $I_{m,a}$ are a fraction p_m of the total population $P_{m,a}$, i.e.

$$I_{m,a} \sim \text{Binomial}(P_{m,a}, p_m) \quad (8)$$

An implicit assumption in Eq. 8 is the existence of an underlying municipality-specific proportion infected p_m so that on each age group, the number of infected people is (on average) $p_m \times P_{m,a}$. We also assumed the following relation for P_m

$$\text{logit}(p_m) = p_0 + \mu_m \quad (9)$$

where p_0 represents a baseline of the proportion infected and μ_m is a municipality-specific random effect.

We use parameters $\gamma_{m,t} \in [0,1]$ to represent the temporal spread of infections; $\sum_t \gamma_{m,t} = 1$ so that $I_{m,a,t} = \gamma_{m,t} I_{m,a}$. Infections, cases, attributed deaths and age-stratified population sizes are linked through a cascade of binomial models. We relate infections, cases and reporting rates through

$$C_{m,t} \sim \text{Binomial}(I_{m,t}, r_{m,t}) \quad (10)$$

Infection fatality rates $IFR_{m,a}$ relate to infections and deaths through another binomial model

$$D_{m,a} \sim \text{Binomial}(I_{m,a}, IFR_{m,a}) \quad (11)$$

where the IFR s follow a stratified logistic-linear relation with socioeconomic status (SES) and age mediated by parameters α , η , δ :

$$\text{logit}(IFR_{m,a}) = \alpha_0 + (\alpha_1 + \eta_a) \times SES_m + \delta_a \quad (12)$$

A comprehensive explanation of this hierarchical Bayesian methodology, including a discussion of its assumptions and several sensitivity analysis and robustness check to misspecification of our assumptions appear in the supplementary materials.

REFERENCES AND NOTES

1. S. Chang, E. Pierson, P. W. Koh, J. Gerardin, B. Redbird, D. Grusky, J. Leskovec, Mobility network models of COVID-19 explain inequities and inform reopening. *Nature* **589**, 82–87 (2021). [doi:10.1038/s41586-020-2923-3](https://doi.org/10.1038/s41586-020-2923-3) [Medline](#)
2. M. U. G. Kraemer, C.-H. Yang, B. Gutierrez, C.-H. Wu, B. Klein, D. M. Pigott, L. du Plessis, N. R. Faria, R. Li, W. P. Hanage, J. S. Brownstein, M. Layan, A. Vespignani, H. Tian, C. Dye, O. G. Pybus, S. V. Scarpino, Open COVID-19 Data Working Group, The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science* **368**, 493–497 (2020). [doi:10.1126/science.abb4218](https://doi.org/10.1126/science.abb4218) [Medline](#)
3. G. Pullano, E. Valdano, N. Scarpa, S. Rubrichi, V. Colizza, Evaluating the effect of demographic factors, socioeconomic factors, and risk aversion on mobility during the COVID-19 epidemic in France under lockdown: A population-based study. *The Lancet Digital Health* **2**, e638–e649 (2020). [doi:10.1016/S2589-7500\(20\)30243-0](https://doi.org/10.1016/S2589-7500(20)30243-0)
4. M. Chinazzi, J. T. Davis, M. Ajelli, C. Gioannini, M. Litvinova, S. Merler, A. Pastore Y Piontti, K. Mu, L. Rossi, K. Sun, C. Viboud, X. Xiong, H. Yu, M. E. Halloran, I. M. Longini Jr., A. Vespignani, The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science* **368**, 395–400 (2020). [doi:10.1126/science.aba9757](https://doi.org/10.1126/science.aba9757) [Medline](#)
5. S. Lai, N. W. Ruktanonchai, L. Zhou, O. Prosper, W. Luo, J. R. Floyd, A. Wesolowski, M. Santillana, C. Zhang, X. Du, H. Yu, A. J. Tatem, Effect of non-pharmaceutical interventions to contain COVID-19 in China. *Nature* **585**, 410–413 (2020). [doi:10.1038/s41586-020-2293-x](https://doi.org/10.1038/s41586-020-2293-x) [Medline](#)
6. F. S. Lu, A. T. Nguyen, N. B. Link, J. T. Davis, M. Chinazzi, X. Xiong, A. Vespignani, M. Lipsitch, M. Santillana, Estimating the cumulative incidence of COVID-19 in the United States using four complementary approaches. medRxiv 2020.04.18.20070821 [Preprint]. 7 August 2020; <https://doi.org/10.1101/2020.04.18.20070821>
7. D. M. Weinberger, J. Chen, T. Cohen, F. W. Crawford, F. Mostashari, D. Olson, V. E. Pitzer, N. G. Reich, M. Russi, L. Simonsen, A. Watkins, C. Viboud, Estimation of excess deaths associated with the COVID-19 pandemic in the United States, March to May 2020. *JAMA Intern. Med.* **180**, 1336–1344 (2020). [doi:10.1001/jamainternmed.2020.3391](https://doi.org/10.1001/jamainternmed.2020.3391) [Medline](#)
8. R. J. Acosta, R. A. Irizarry, Monitoring health systems by estimating excess mortality. medRxiv 2020.06.06.20120857 [Preprint]. 9 June 2020; <https://doi.org/10.1101/2020.06.06.20120857>
9. M. O'Driscoll, G. Ribeiro Dos Santos, L. Wang, D. A. T. Cummings, A. S. Azman, J. Paireau, A. Fontanet, S. Cauchemez, H. Salje, Age-specific mortality and immunity patterns of SARS-CoV-2. *Nature* **590**, 140–145 (2021). [doi:10.1038/s41586-020-2918-0](https://doi.org/10.1038/s41586-020-2918-0) [Medline](#)
10. J. M. Feldman, M. T. Bassett, The relationship between neighborhood poverty and COVID-19 mortality within racial/ethnic groups (Cook County, Illinois). medRxiv 2020.10.04.20206318 [Preprint]. 6 October 2020; <https://doi.org/10.1101/2020.10.04.20206318>
11. S. M. Kissler, N. Kishore, M. Prabhu, D. Goffman, Y. Beilin, R. Landau, C. Gyamfi-Bannerman, B. T. Bateman, J. Snyder, A. S. Razavi, D. Katz, J. Gal, A. Bianco, J. Stone, D. Larremore, C. O. Buckee, Y. H. Grad, Reductions in commuting mobility correlate with geographic differences in SARS-CoV-2 prevalence in New York City. *Nat. Commun.* **11**, 4674 (2020). [doi:10.1038/s41467-020-18271-5](https://doi.org/10.1038/s41467-020-18271-5) [Medline](#)
12. A. Carranza, M. Goic, E. Lara, M. Olivares, G. Y. Weintraub, J. Covarrubia, C. Escobedo, N. Jara, L. J. Basso, The social divide of social distancing: Shelter-in-place behavior in Santiago during the COVID-19 pandemic. SSRN 3691373 [Preprint]. 12 September 2020; <http://dx.doi.org/10.2139/ssrn.3691373>
13. M. Bennett, All things equal? Heterogeneity in policy effectiveness against COVID-19 spread in Chile. *World Dev.* **137**, 105208 (2021). [doi:10.1016/j.worlddev.2020.105208](https://doi.org/10.1016/j.worlddev.2020.105208) [Medline](#)

14. N. Gozzi, M. Tizzoni, M. Chinazzi, L. Ferres, A. Vespignani, N. Perra, Estimating the effect of social inequalities in the mitigation of COVID-19 across communities in Santiago de Chile. *Nat. Commun.* **12**, 2429 (2021). [doi:10.1038/s41467-021-22601-6](https://doi.org/10.1038/s41467-021-22601-6)
15. M. Lipsitch, M. Santillana, "Enhancing situational awareness to prevent infectious disease outbreaks from becoming catastrophic" in *Global Catastrophic Biological Risks*, T. V. Inglesby, A. A. Adalja, Eds. (Current Topics in Microbiology and Immunology, vol. 424, Springer, 2019), pp. 59–74.
16. M. V. Kiang, R. A. Irizarry, C. O. Buckee, S. Balsari, Every body counts: Measuring Mortality from the COVID-19 pandemic. *Ann. Intern. Med.* **173**, 1004–1007 (2020). [doi:10.7326/M20-3100](https://doi.org/10.7326/M20-3100) [Medline](#)
17. C. K. Williams, C. E. Rasmussen, *Gaussian Processes for Machine Learning* (MIT Press, 2006), vol. 2.
18. S. J. Olsen, E. Azziz-Baumgartner, A. P. Budd, L. Brammer, S. Sullivan, R. F. Pineda, C. Cohen, A. M. Fry, Decreased influenza activity during the COVID-19 pandemic—United States, Australia, Chile, and South Africa, 2020. *Am. J. Transplant.* **20**, 3681–3685 (2020). [doi:10.1111/ajt.16381](https://doi.org/10.1111/ajt.16381) [Medline](#)
19. A. Banerjee, L. Pasea, S. Harris, A. Gonzalez-Izquierdo, A. Torralbo, L. Shallcross, M. Noursadeghi, D. Pillay, N. Sebire, C. Holmes, C. Pagel, W. K. Wong, C. Langenberg, B. Williams, S. Denaxas, H. Hemingway, Estimating excess 1-year mortality associated with the COVID-19 pandemic according to underlying conditions and age: A population-based cohort study. *Lancet* **395**, 1715–1725 (2020). [doi:10.1016/S0140-6736\(20\)30854-0](https://doi.org/10.1016/S0140-6736(20)30854-0) [Medline](#)
20. J. M. Aburto, R. Kashyap, J. Schöley, C. Angus, J. Ermisch, M. C. Mills, J. B. Dowd, Estimating the burden of the COVID-19 pandemic on mortality, life expectancy and lifespan inequality in England and Wales: a population-level analysis. *J. Epidemiol. Community Health* **10.1136/jech-2020-215505** (2021). [doi:10.1136/jech-2020-215505](https://doi.org/10.1136/jech-2020-215505)
21. O. Stoner, T. Economou, G. Drummond Marques da Silva, A hierarchical framework for correcting under-reporting in count data. *J. Am. Stat. Assoc.* **114**, 1481–1492 (2019). [doi:10.1080/01621459.2019.1573732](https://doi.org/10.1080/01621459.2019.1573732)
22. D. K. Kennedy, A. Seale, D. Bausch, "How experts use data to identify emerging COVID-19 success stories," *Our World in Data*, 30 June 2020; <https://ourworldindata.org/identify-covid-exemplars>.
23. U. Bilal, M. Alazraqi, W. T. Caiiffa, N. Lopez-Olmedo, K. Martinez-Folgar, J. J. Miranda, D. A. Rodriguez, A. Vives, A. V. Diez-Roux, Inequalities in life expectancy in six large Latin American cities from the SALURBAL study: An ecological analysis. *Lancet Planet. Health* **3**, e503–e510 (2019). [doi:10.1016/S2542-5196\(19\)30235-9](https://doi.org/10.1016/S2542-5196(19)30235-9) [Medline](#)
24. O. P. Ottersen, J. Dasgupta, C. Blouin, P. Buss, V. Chongsuvivatwong, J. Frenk, S. Fukuda-Parr, B. P. Gawanas, R. Giacaman, J. Gyapong, J. Leaning, M. Marmot, D. McNeill, G. I. Mongella, N. Moyo, S. Møgedal, A. Ntsaluba, G. Ooms, E. Bjertness, A. L. Lie, S. Moon, S. Roalkvam, K. I. Sandberg, I. B. Scheel, The political origins of health inequity: Prospects for change. *Lancet* **383**, 630–667 (2014). [doi:10.1016/S0140-6736\(13\)62407-1](https://doi.org/10.1016/S0140-6736(13)62407-1) [Medline](#)
25. E. R. Meara, S. Richards, D. M. Cutler, The gap gets bigger: Changes in mortality and life expectancy, by education, 1981–2000. *Health Aff. (Millwood)* **27**, 350–360 (2008). [doi:10.1377/hlthaff.27.2.350](https://doi.org/10.1377/hlthaff.27.2.350) [Medline](#)
26. B. Burström, W. Tao, Social determinants of health and inequalities in COVID-19. *Eur. J. Public Health* **30**, 617–618 (2020). [doi:10.1093/eurpub/ckaa095](https://doi.org/10.1093/eurpub/ckaa095) [Medline](#)
27. R. Laxminarayan, B. Wahl, S. R. Dudala, K. Gopal, C. Mohan B, S. Neelima, K. S. Jawahar Reddy, J. Radhakrishnan, J. A. Lewnard, Epidemiology and transmission dynamics of COVID-19 in two Indian states. *Science* **370**, 691–697 (2020). [doi:10.1126/science.abd7672](https://doi.org/10.1126/science.abd7672) [Medline](#)
28. Z. Wang, K. Tang, Combating COVID-19: Health equity matters. *Nat. Med.* **26**, 458 (2020). [doi:10.1038/s41591-020-0823-6](https://doi.org/10.1038/s41591-020-0823-6) [Medline](#)
29. Y. Li, E. A. Undurraga, J. R. Zubizarreta, Effectiveness of localized lockdowns in the COVID-19 pandemic. medRxiv 2020.08.25.20182071 [Preprint]. 30 March 2021; <https://doi.org/10.1101/2020.08.25.20182071>.
30. M. E. Kretzschmar, G. Rozhnova, M. C. J. Bootsma, M. van Boven, J. H. H. M. van de Wijert, M. J. M. Bonten, Impact of delays on effectiveness of contact tracing strategies for COVID-19: A modelling study. *Lancet Public Health* **5**, e452–e459 (2020). [doi:10.1016/S2468-2667\(20\)30157-2](https://doi.org/10.1016/S2468-2667(20)30157-2) [Medline](#)
31. S. V. Subramanian, I. Delgado, L. Jadue, J. Vega, I. Kawachi, Income inequality and health: Multilevel analysis of Chilean communities. *J. Epidemiol. Community Health* **57**, 844–848 (2003). [doi:10.1136/jech.57.11.844](https://doi.org/10.1136/jech.57.11.844) [Medline](#)
32. P. Maas et al., in *Proceedings of the 16th International Conference on Information Systems for Crisis Response and Management*, València, Spain, May 19–22, 2019 (ISCRAM Association, 2019), vol. 19, p. 3173.
33. L. A. Shepp, Y. Vardi, Maximum likelihood reconstruction for emission tomography. *IEEE Trans. Med. Imaging* **1**, 113–122 (1982). [doi:10.1109/TMI.1982.4307558](https://doi.org/10.1109/TMI.1982.4307558) [Medline](#)
34. Y. Vardi, L. A. Shepp, L. Kaufman, a statistical model for positron emission tomography. *J. Am. Stat. Assoc.* **80**, 8–20 (1985). [doi:10.1080/01621459.1985.10477119](https://doi.org/10.1080/01621459.1985.10477119)
35. E. Goldstein, J. Dushoff, J. Ma, J. B. Plotkin, D. J. D. Earn, M. Lipsitch, Reconstructing influenza incidence by deconvolution of daily mortality time series. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 21825–21829 (2009). [doi:10.1073/pnas.0902958106](https://doi.org/10.1073/pnas.0902958106) [Medline](#)
36. P. P. B. Eggermont, V. N. LaRiccia, V. LaRiccia, *Maximum Penalized Likelihood Estimation* (Springer, 2001), vol. 1.
37. G. Papadopoulos, J. Santos Silva, Identification issues in some double-index models for non-negative data. *Econ. Lett.* **117**, 365–367 (2012). [doi:10.1016/j.econlet.2012.06.001](https://doi.org/10.1016/j.econlet.2012.06.001)
38. N. M. Linton, T. Kobayashi, Y. Yang, K. Hayashi, A. R. Akhmetzhanov, S. M. Jung, B. Yuan, R. Kinoshita, H. Nishiura, Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: A statistical analysis of publicly available case data. *J. Clin. Med.* **9**, 538 (2020). [doi:10.3390/jcm9020538](https://doi.org/10.3390/jcm9020538) [Medline](#)
39. T. W. Russell, J. Hellewell, C. I. Jarvis, K. van Zandvoort, S. Abbott, R. Ratnayake, S. Flasche, R. M. Eggo, W. J. Edmunds, A. J. Kucharski, CMMID COVID-19 Working Group, Estimating the infection and case fatality ratio for coronavirus disease (COVID-19) using age-adjusted data from the outbreak on the Diamond Princess cruise ship, February 2020. *Euro Surveill.* **25**, 2000256 (2020). [doi:10.2807/1560-7917.ES.2020.25.12.2000256](https://doi.org/10.2807/1560-7917.ES.2020.25.12.2000256) [Medline](#)
40. R. Pastor-Barriuso, B. Pérez-Gómez, M. A. Hernán, M. Pérez-Olmeda, R. Yotti, J. Oteo-Iglesias, J. L. Sanmartín, I. León-Gómez, A. Fernández-García, P. Fernández-Navarro, I. Cruz, M. Martín, C. Delgado-Sanz, N. Fernández de Larrea, J. León Paniagua, J. F. Muñoz-Montalvo, F. Blanco, A. Larrauri, M. Pollán, ENE-COVID Study Group, Infection fatality risk for SARS-CoV-2 in community dwelling population of Spain: Nationwide seroepidemiological study. *BMJ* **371**, m4509 (2020). [doi:10.1136/bmj.m4509](https://doi.org/10.1136/bmj.m4509) [Medline](#)
41. M. H. Chitwood, M. Russi, K. Gunasekera, J. Havumaki, V. E. Pitzer, J. A. Salomon, N. Swartwood, J. L. Warren, D. M. Weinberger, T. Cohen, N. A. Menzies, Reconstructing the course of the COVID-19 epidemic over 2020 for US states and counties: results of a Bayesian evidence synthesis model. medRxiv 2020.06.17.20133983 [Preprint]. 6 April 2021; <https://doi.org/10.1101/2020.06.17.20133983>.
42. T. W. Russell, N. Golding, J. Hellewell, S. Abbott, L. Wright, C. A. B. Pearson, K. van Zandvoort, C. I. Jarvis, H. Gibbs, Y. Liu, R. M. Eggo, W. J. Edmunds, A. J. Kucharski, CMMID COVID-19 working group, Reconstructing the early global dynamics of under-ascertained COVID-19 cases and infections. *BMC Med.* **18**, 332 (2020). [doi:10.1186/s12916-020-01790-9](https://doi.org/10.1186/s12916-020-01790-9) [Medline](#)
43. G. Mena, P. Martinez, A. Mahmud, P. Marquet, C. Buckee, M. Santillana, Socioeconomic status determines COVID-19 incidence and related mortality in Santiago, Chile. Zenodo (2021); <https://doi.org/10.5281/zenodo.4699403>.
44. N. G. Becker, L. F. Watson, J. B. Carlin, A method of non-parametric back-projection and its application to AIDS data. *Stat. Med.* **10**, 1527–1542 (1991). [doi:10.1002/sim.4780101005](https://doi.org/10.1002/sim.4780101005) [Medline](#)
45. E. Levitan, G. T. Herman, A maximum a posteriori probability expectation maximization algorithm for image reconstruction in emission tomography. *IEEE Trans. Med. Imaging* **6**, 185–192 (1987). [doi:10.1109/TMI.1987.4307826](https://doi.org/10.1109/TMI.1987.4307826) [Medline](#)
46. R. L. White, in *Instrumentation in Astronomy VIII*, D. L. Crawford, E. R. Craine, Eds. (International Society for Optics and Photonics, 1994), vol. 2198, pp. 1342–1348.
47. P. J. Green, Bayesian reconstructions from emission tomography data using a modified EM algorithm. *IEEE Trans. Med. Imaging* **9**, 84–93 (1990). [doi:10.1109/42.52985](https://doi.org/10.1109/42.52985) [Medline](#)
48. A. C. Miller, L. Hannah, J. Futoma, N. J. Foti, E. B. Fox, A. D'Amour, M. Sandler, R. A. Saurous, J. A. Lewnard, Statistical deconvolution for inference of infection time series. medRxiv 2020.10.16.20212753 [Preprint]. 20 October 2020;

(CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>. This license does not apply to figures/photos/artwork or other content included in the article that is credited to a third party; obtain authorization from the rights holder before using such material.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/cgi/content/full/science.abg5298/DC1

Materials and Methods
Supplementary Text
Figs. S1 to S59
Tables S1 to S4
References (44–62)
MDAR Reproducibility Checklist

12 January 2021; accepted 22 April 2021
Published online 27 April 2021
10.1126/science.abg5298

- <https://doi.org/10.1101/2020.10.16.20212753>
49. B. W. Silverman, M. C. Jones, J. D. Wilson, D. W. Nychka, A smoothed Em approach to indirect estimation problems, with particular reference to stereology and emission tomography. *J. R. Stat. Soc. B* **52**, 271–303 (1990). [doi:10.1111/j.2517-6161.1990.tb01788.x](https://doi.org/10.1111/j.2517-6161.1990.tb01788.x)
 50. M. Titsias, in *Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics*, D. van Dyk, M. Welling, Eds., vol. 5 of *Proceedings of Machine Learning Research* (PMLR, 2009), pp. 567–574.
 51. A. Gelman, X.-L. Meng, H. Stern, Posterior predictive assessment of model fitness via realized discrepancies. *Stat. Sin.* **6**, 733–807 (1996).
 52. R. E. Barlow, Ed., *Statistical Inference Under Order Restrictions: The Theory and Application of Isotonic Regression* (Wiley Series in Probability and Mathematical Statistics, no. 8, Wiley, 1972).
 53. P. Gustafson, *Bayesian Inference for Partially Identified Models: Exploring the Limits of Limited Data* (CRC Press, 2015), vol. 140.
 54. D. Fisman, S. J. Drews, A. Tuite, S. O'Brien, Age-specific SARS-CoV-2 infection fatality and case identification fraction in Ontario, Canada. medRxiv 2020.11.09.20223396 [Preprint]. 12 November 2020; <https://doi.org/10.1101/2020.11.09.20223396>.
 55. P. de Valpine, D. Turek, C. J. Paciorek, C. Anderson-Bergman, D. T. Lang, R. Bodik, Programming with models: Writing statistical algorithms for general model structures with NIMBLE. *J. Comput. Graph. Stat.* **26**, 403–413 (2017). [doi:10.1080/10618600.2016.1172487](https://doi.org/10.1080/10618600.2016.1172487)
 56. M. Plummer et al., in *Proceedings of the 3rd International Workshop on Distributed Statistical Computing* (Vienna, Austria, 2003), vol. 124, pp. 1–10.
 57. J. A. Hoeting, D. Madigan, A. E. Raftery, C. T. Volinsky, Bayesian model averaging: A tutorial (with comments by M. Clyde, David Draper and E. I. George, and a rejoinder by the authors). *Stat. Sci.* **14**, 382 (1999). [doi:10.1214/ss/1009212519](https://doi.org/10.1214/ss/1009212519)
 58. Y. Yao, A. Vehtari, D. Simpson, A. Gelman, Using stacking to average Bayesian predictive distributions (with discussion). *Bayesian Anal.* **13**, 917 (2018). [doi:10.1214/17-BA1091](https://doi.org/10.1214/17-BA1091)
 59. A. Gelman, J. Hwang, A. Vehtari, Understanding predictive information criteria for Bayesian models. *Stat. Comput.* **24**, 997–1016 (2014). [doi:10.1007/s11222-013-9416-2](https://doi.org/10.1007/s11222-013-9416-2)
 60. A. Wilson, R. Adams, in *Proceedings of the 30th International Conference on Machine Learning* (PMLR, 2013), pp. 1067–1075.
 61. L. Wasserman, Bayesian model selection and model averaging. *J. Math. Psychol.* **44**, 92–107 (2000). [doi:10.1006/jmps.1999.1278](https://doi.org/10.1006/jmps.1999.1278) [Medline](#)
 62. V. Kontis, J. E. Bennett, T. Rashid, R. M. Parks, J. Pearson-Stuttard, M. Guillot, P. Asaria, B. Zhou, M. Battaglini, G. Corsetti, M. McKee, M. Di Cesare, C. D. Mathers, M. Ezzati, Magnitude, demographics and dynamics of the effect of the first wave of the COVID-19 pandemic on all-cause mortality in 21 industrialized countries. *Nat. Med.* **26**, 1919–1928 (2020). [doi:10.1038/s41591-020-1112-0](https://doi.org/10.1038/s41591-020-1112-0) [Medline](#)

ACKNOWLEDGMENTS

We would like to thank Marc Lipsitch, Oliver Stoner, Nick Link, Patricio Vargas, Jorge Perez, and Alonso Silva for insightful comments. We thank the Chilean Society of Intensive Medicine (SOCHIMI) and the School of Medicine of the Finis Terrae University (Chile) for making it possible to obtain the data on deaths recorded in health facilities. **Funding:** G.E.M. was partially supported by a Harvard Data Science Initiative fellowship, P.P.M. by U54GM088558, C.O.B. by Schmidt futures award, and M.S. by R01GM130668. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. **Author contributions:** G.E.M. and P.P.M. designed the study, conducted the analyses, and wrote the manuscript. A.S.M. and P.A.M. assisted with the analyses and conceptualization. C.O.B. and M.S. designed and oversaw the study. All authors edited the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Socioeconomic status data are available at www.desarrollosocialyfamilia.gob.cl/documentos/metropolitana. COVID-19 data are available at <https://github.com/MinCiencia/Datos-COVID19> in the format of multiple reports. Mortality data are available at <https://deis.minsal.cl>. Mobility data are available upon request at <https://dataforgood.fb.com>. All code and data to reproduce the figures in the main text are available at Zenodo (43). This work is licensed under a Creative Commons Attribution 4.0 International

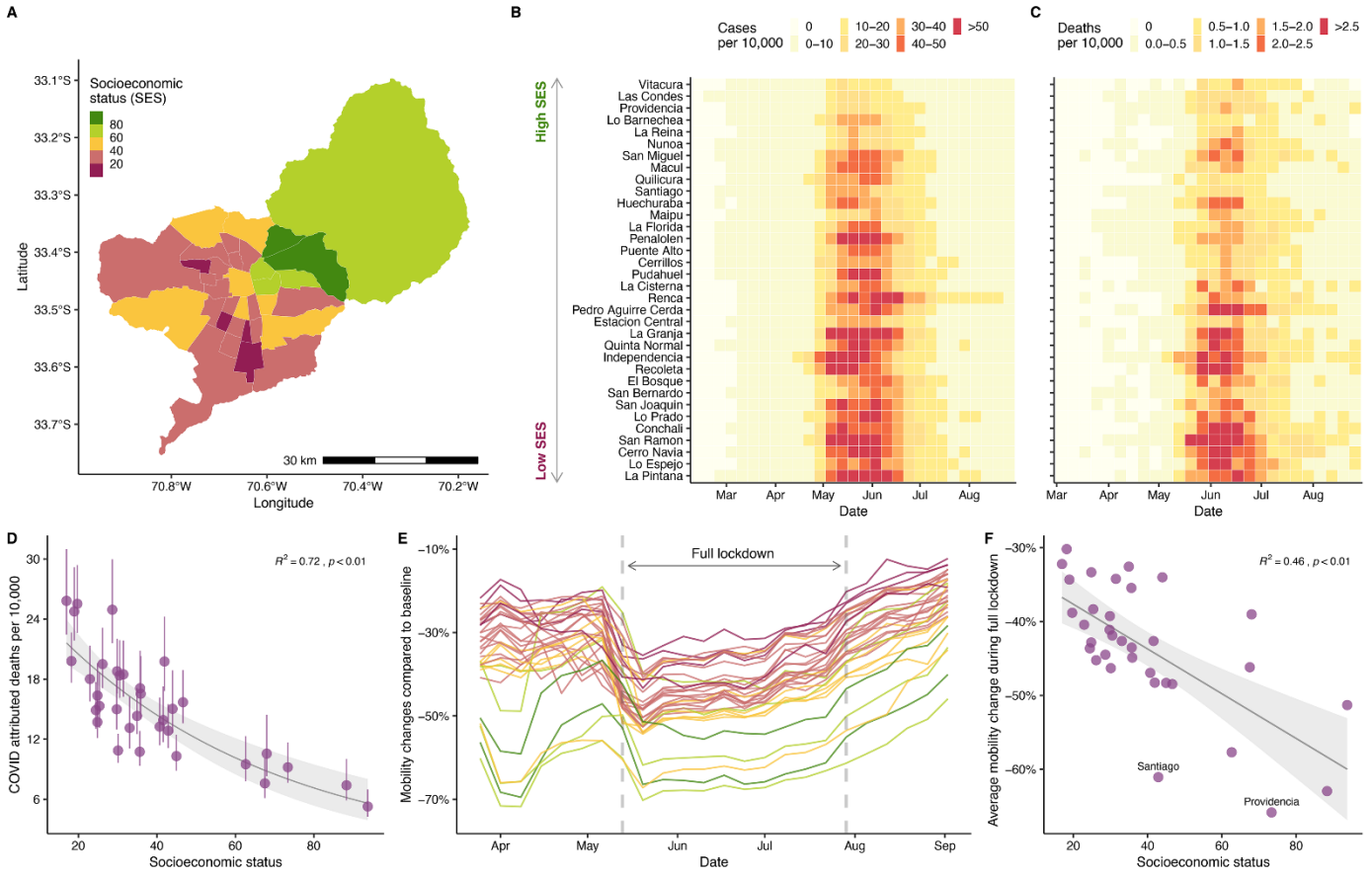


Fig. 1. Socioeconomic status, COVID-19 cases and deaths, and mobility data in Greater Santiago. (A) Municipalities that are part of the Greater Santiago are colored according to their socioeconomic status (SES), where a lower score is indicative of a lower SES. (B) COVID-19 cases normalized by population size per municipality. Municipalities are sorted by SES starting with the one that has the highest SES at the top. (C) COVID-19 attributed deaths normalized by population size per municipality. (D) Age-adjusted adjusted COVID-19 attributed death rates and its association with SES. The dots and the whiskers represent the median and the 95% confidence intervals respectively, reflecting uncertainty on the standard population used for weighting. (E) Daily reduction in mobility by municipality colored by its SES value. (F) Average reduction in mobility during the full lockdown period and its association with SES. The urban and the business centers, Santiago and Providencia respectively, experienced a greater reduction in mobility than expected based just on their socioeconomic profile.

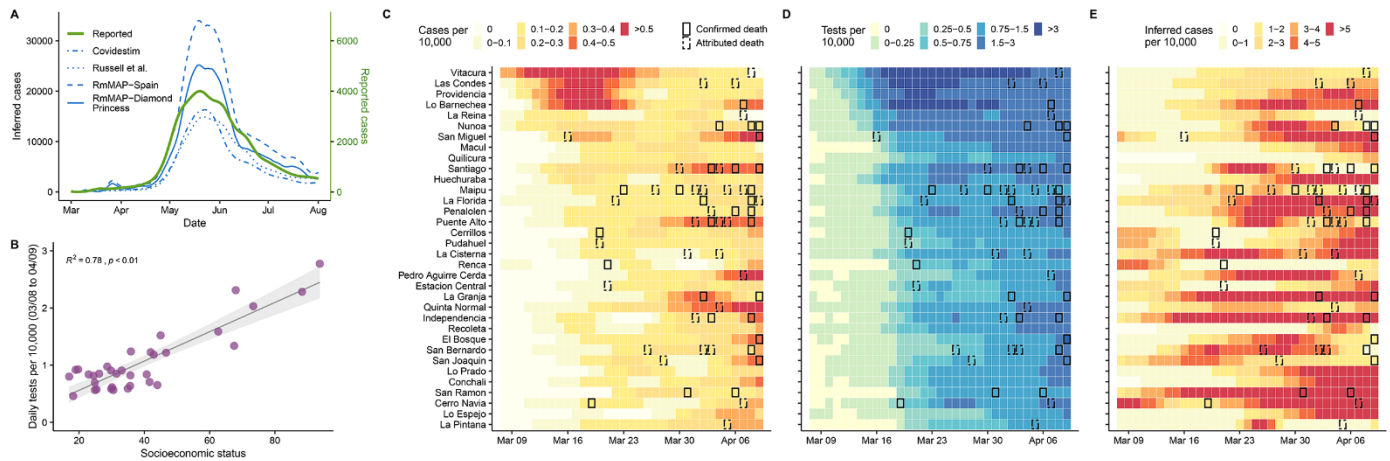


Fig. 2. Inferred cases and reported tests conducted for the Greater Santiago area. (A) Inferred and reported cases over time. For our RmMAP reconstructions, we considered the log-normal onset-to-death distribution described in (38) and two age-stratified IFR estimates, one from Diamond Princess cruise ship (39) and another from a seroprevalence study in Spain (40). For comparison, we also present reconstructions based on the Covidestim method (41), and by the re-scaling of case counts by the under-reporting estimates obtained with the method of Russell et al. (42). (B) Association between average daily test and SES during the early peak. The early peak is defined as those cases reported between 03/08 and 04/09. (C) Reported cases per 10,000 by municipality during the early peak. (D) Test per 10,000 by municipality during the early peak. (E) Inferred cases obtained from the RmMAP-Spain model per 10,000 by municipality during the early peak. For panels C-E, the record of at least one COVID-19 confirmed or attributed death for that particular week is highlighted with solid and dashed boxes respectively.

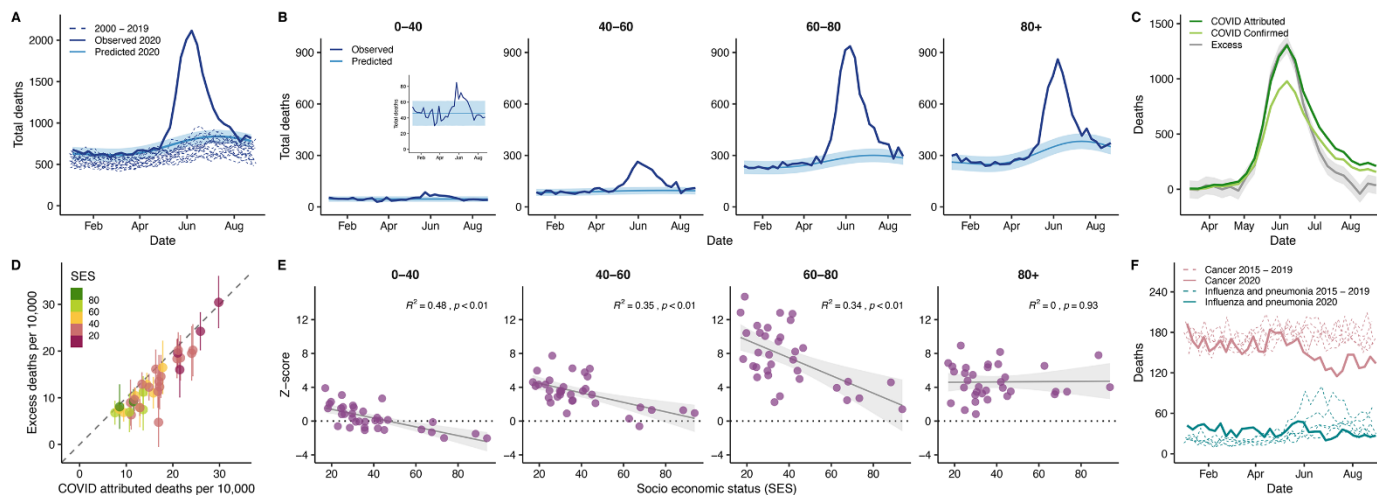


Fig. 3. Excess deaths and its association with demographic and socioeconomic factors.

(A) Observed deaths (solid dark blue line) in Greater Santiago compared to predicted deaths for 2020 (solid light blue line and its confident intervals shaded in lighter color), using a Gaussian process regression model built with historical mortality data from 2000 to 2019 (dashed blue lines). The values contain all the possible causes of deaths. (B) Age-specific trends of the observed deaths compared to the predicted deaths for 2020. (C) COVID-19 deaths versus excess deaths. COVID-19 confirmed deaths are shown in light green, while COVID-19 attributed deaths are shown in dark green. Excess deaths correspond to difference between observed and predicted deaths. (D) Comparison of excess deaths and COVID-19 attributed deaths per municipality colored by SES, and normalized by population size. (E) Monthly average of Z-scores of observed deaths between April and July by age group. The Z-scores correspond to the standard deviations over expected values. (F) Historical deaths due to influenza and pneumonia (teal dashed lines), and cancer (pink dashed lines) compared to the observed deaths during 2020 (solid lines).

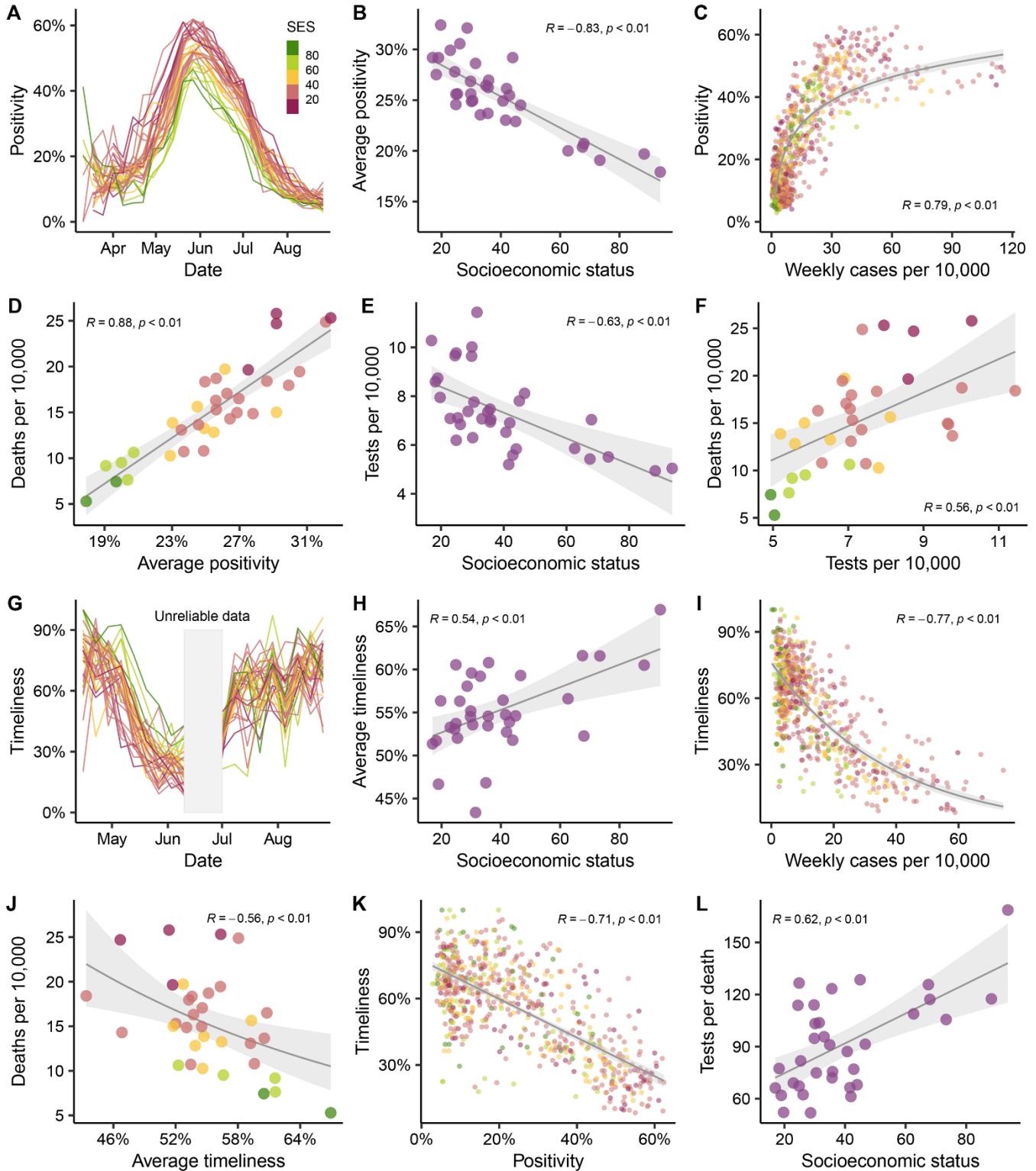


Fig. 4. Testing capacity and waiting times. (A) Positivity over time. Positivity is defined as the proportion of PCR tests that are positive on a given week. (B) Association between average positivity and SES. (C) Association between positivity and weekly number of cases per 10,000. (D) Association between the overall age-adjusted number of deaths per 10,000 and the average positivity over the same period. (E) Association between average daily tests per 10,000 and SES. (F) Association between tests per 10,000 and deaths per 10,000. (G) Timeliness over time. Timeliness is defined as the proportion of PCR tests that appear in the public records within one week from the onset of symptoms. Two weeks in June (shaded in gray) were excluded from the analysis due to inconsistencies in data, leading to unreliable delay estimates. (H) Association between average timeliness and SES. (I) Association between timeliness and weekly number of cases per 10,000. (J) Association between the overall age-adjusted number of deaths per 10,000 and the average timeliness. (K) Association between timeliness and positivity. Dots are representative of weekly data per municipality. (L) Association between tests per death (age-adjusted) and SES. Figures with different dot colors illustrate the SES value according to the reference presented in (A).

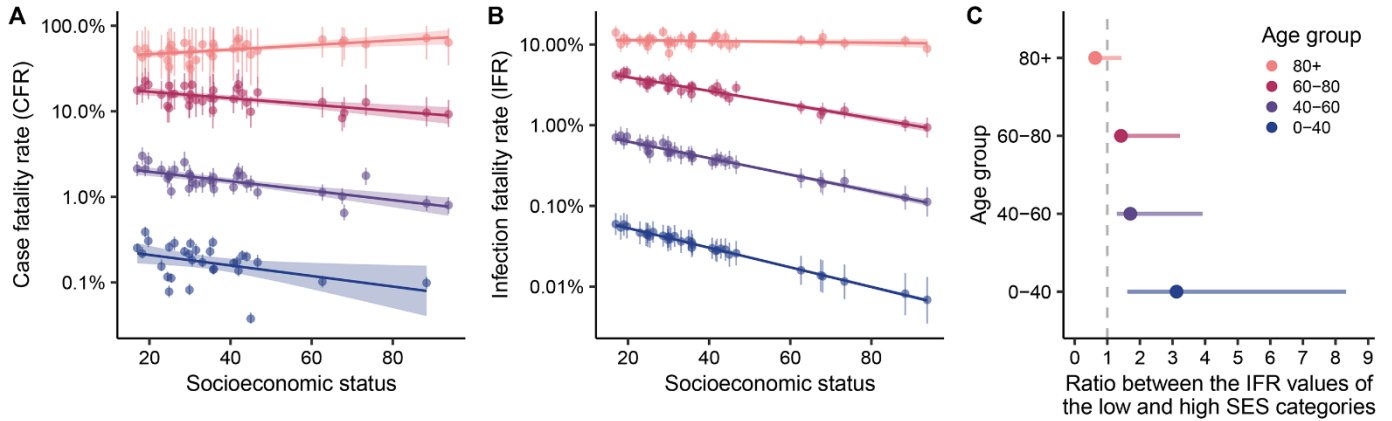


Fig. 5. Inference of case and infection fatality rates by age and socioeconomic status. (A) Estimates of case fatality rate (CFR) by age and SES based on a simple assignment of cases to age groups. Our estimates of CFR have been validated by the official ICOVID platform (www.icovidchile.cl/) which confirmed that 119 out of the 136 observed CFRs fall within our confidence intervals. CIs are derived from a bootstrap procedure described in the supplementary materials. (B) Inferred infection fatality rate (IFR) by age and SES using our ensemble of hierarchical Bayesian models, along with associated credible intervals. (C) IFR ratio between the low and high socioeconomic category by age group. Four socioeconomic categories were defined based on SES quantiles: low, mid-low, mid-high, and high.

Socioeconomic status determines COVID-19 incidence and related mortality in Santiago, Chile

Gonzalo E. Mena, Pamela P. Martinez, Ayesha S. Mahmud, Pablo A. Marquet, Caroline O. Buckee and Mauricio Santillana

published online April 27, 2021

ARTICLE TOOLS

<http://science.sciencemag.org/content/early/2021/04/26/science.abg5298>

SUPPLEMENTARY MATERIALS

<http://science.sciencemag.org/content/suppl/2021/04/26/science.abg5298.DC1>

REFERENCES

This article cites 49 articles, 15 of which you can access for free
<http://science.sciencemag.org/content/early/2021/04/26/science.abg5298#BIBL>

PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title *Science* is a registered trademark of AAAS.

Copyright © 2021 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution License 4.0 (CC BY).