Red Nacional de Investigadores en Economía

SARS-CoV-2 spread, detection, and dynamics in a megacity in Latin America*

Rachid Laajaj (University of Los Andes), ***Camilo De Los Rios*** (Inter American Development Bank), ***Ignacio Sarmiento-Barbieri*** (University of Los Andes), ***Danilo Aristizabal*** (University of Los Andes), ***Eduardo Behrentz*** (University of Los Andes), ***Raquel Bernal*** (University of Los Andes), ***Giancarlo Buitrago*** (Universidad Nacional de Colombia), ***Giancarlo Buitrago*** (Imperial College London), ***Fernando de la Hoz*** (Universidad Nacional de Colombia), ***Alejandro Gaviria*** (University of Los Andes), ***Luis Jorge Hernández*** (University of Los Andes), ***Leonardo León*** (University of Los Andes), ***Diane Moyano*** (Secretaria de Salud de Bogotá), ***Elkin Osorio*** (Secretaria de Salud de Bogotá), ***Andrea Ramírez Varela*** (University of Los Andes), ***Silvia Restrepo*** (University of Los Andes), ***Rodrigo Rodriguez*** (Secretaria de Salud de Bogotá), ***Norbert Schady*** (Inter American Development Bank), ***Martha Vives*** (University of Los Andes), ***Duncan Webb*** (Paris School of Economics)

DOCUMENTO DE TRABAJO N°78

Julio de 2021

*Publicado como Documento CEDE #18

Citar como:
Laajaj, Rachid, Camilo De Los Rios, Ignacio Sarmiento-Barbieri, Danilo Aristizabal, Eduardo Behrentz, Raquel Bernal, Giancarlo Buitrago, Zulma Cucunubá, Fernando de la Hoz, Alejandro Gaviria, Luis Jorge Hernández, Leonardo León, Diane Moyano, Elkin Osorio, Andrea Ramírez Varela, Silvia Restrepo, Rodrigo Rodriguez, Norbert Schady, Martha Vives, Duncan Webb (2021). SARS-CoV-2 spread, detection, and dynamics in a megacity in Latin America. *Documento de trabajo RedNIE N°78.*

SARS-CoV-2 spread, detection, and dynamics in a megacity in Latin America

Rachid Laajaj, Camilo De Los Rios, Ignacio Sarmiento-Barbieri, Danilo Aristizabal, Eduardo Behrentz, Raquel Bernal, Giancarlo Buitrago, Zulma Cucunubá, Fernando de la Hoz, Alejandro Gaviria, Luis Jorge Hernández, Leonardo León, Diane Moyano, Elkin Osorio, Andrea Ramírez Varela, Silvia Restrepo, Rodrigo Rodriguez, Norbert Schady, Martha Vives, Duncan Webb†*

Abstract

In many developing countries, the COVID-19 pandemic has spread much faster and wider than the number of detected cases implies. By combining data from 59,770 RT-PCR tests on mostly asymptomatic individuals with administrative data on all

†The authors acknowledge generous support from the Interamerican Development Bank, the Development Bank of Latin America (CAF), the University of Los Andes and the Universidad Nacional de Colombia. Duncan Webb gratefully acknowledges support from the ED 465 at University Paris 1, and the EUR project ANR-17-EURE-0001. The authors have no competing interests to declare. The authors declare no conflicts of interest. We thank members of the CoVIDA working group, Yessica Daniela Campaz Landazabal, Marylin Hidalgo, Juliana Quintero, Paola Betancourt, Pablo Rodríguez, Diana Sofia Rios Oliveros, Andrés Felipe Patiño, Jose David Pinzon Ortiz,Maribel Rincón, Alejandro Segura, Leonardo Salas Zapata.

*Laajaj:University of Los Andes, Bogotá, Colombia, r.laajaj@uniandes.edu.co; De Los Rios: Inter-American Development Bank, Washington D.C, United States, cdelosriosru@gmail.com; Sarmiento-Barbieri, University of Los Andes, Bogotá, Colombia, i.sarmiento@uniandes.edu.co; Aristizabal: University of Los Andes, Bogotá, Colombia, de.aristizabal411@uniandes.edu.co; Behrentz: University of Los Andes, Bogotá, Colombia, ebehrent@uniandes.edu.co; Bernal: University of Los Andes, Bogotá, Colombia, rbernal@uniandes.edu.co; Buitrago, Universidad Nacional de Colombia, Bogotá, Colombia, gbuitragog@unal.edu.co; Cucunubá, Imperial College London, London, UK, zulma.cucunuba@imperial.ac.uk; de la Hoz, Universidad Nacional de Colombia, Bogotá, Colombia, fpdelahozr@unal.edu.co; Gaviria: University of Los Andes, Bogotá, Colombia, agaviria@uniandes.edu.co; Hernández: University of Los Andes, Bogotá, Colombia, luishern@uniandes.edu.co; León: University of Los Andes, Bogotá, Colombia, lj.leon62@uniandes.edu.co; Moyano, Secretaria de Salud de Bogotá, Bogotá, Colombia, dmoyano@saludcapital.gov.co; Osorio, Secretaria de Salud de Bogotá, Bogotá, Colombia, ejosorio@saludcapital.gov.co; Ramírez Varela: University of Los Andes, Bogotá, Colombia, anrami2@uniandes.edu.co; Restrepo: University of Los Andes, Bogotá, Colombia, srestrep@uniandes.edu.co; Rodriguez, Secretaria de Salud de Bogotá, Bogotá, Colombia, rorofer21@gmail.com; Schady, Inter-American Development Bank, Washington D.C, United States, norberts@iadb.org; Vives: University of Los Andes, Bogotá, Colombia, mvives@uniandes.edu.co; Webb, Paris School of Economics, Paris, France, dmbwebb@gmail.com. Author contributions: R.L., G.B. and FdlH. conceived the study; R.L., CDLR., I.S. and D.W. analyzed the data; R.L, D.A, R.B, G.B, A.G, L. J. H., L.L, A.R.V, S.R, and M.V implemented the CoVIDA project, including tests and primary data collection; R.L., CDLR., I.S. and D.W. wrote the paper; D.M, E.O and R.R provided access to HSB data and contributed to the design of the CoVIDA project, all authors critically revised the various versions of the manuscript.

detected cases, we capture the spread and dynamics of the COVID-19 pandemic in Bogotá from June 2020 to early March 2021. Our data provide unusually broad and detailed information on mostly asymptomatic adults in Bogotá, allowing to describe various features of the pandemic that appear to be specific to a developing country context. We find that, by the end of March 2021, slightly more than half of the population in Bogotá has been infected, despite only a small fraction of this population being detected. In July 2020, after four months of generalized quarantine that mitigated the pandemic without curving it, the initial buildup of immunity contributed to the end of the first wave. We also show that the share of the population infected by February 2021 varies widely by occupation, socio-economic stratum, and location. This, in turn, has affected the dynamics of the spread: while the first wave of infections was driven by the lowest economic strata and highly-exposed occupations, the second peak affected the population more evenly. A better understanding of the spread and dynamics of the pandemic across different groups provides valuable guidance for efficient targeting of health policy measures and restrictions.

Key words: SARS-CoV-2, COVID-19, CoVIDA, Latin America
JEL Classification: I14, I15, I18, O54

Propagación, detección y dinámica del SARS-CoV-2 en una megaciudad de América Latina

Rachid Laajaj, Camilo De Los Rios, Ignacio Sarmiento-Barbieri, Danilo Aristizabal, Eduardo Behrentz, Raquel Bernal, Giancarlo Buitrago, Zulma Cucunubá, Fernando de la Hoz, Alejandro Gaviria, Luis Jorge Hernández, Leonardo León, Diane Moyano, Elkin Osorio, Andrea Ramírez Varela, Silvia Restrepo, Rodrigo Rodriguez, Norbert Schady, Martha Vives, Duncan Webb[†]*

Resumen

En muchos países en desarrollo, la pandemia de COVID-19 se ha propagado más rápida y ampliamente de lo muestran el número de casos detectados. Combinando datos de 59,770 pruebas de RT-PCR de personas mayoritariamente asintomáticas, con datos administrativos de todos los casos detectados, capturamos la propagación

[†]Agradecemos por la generosa financiación al Banco Interamericano de Desarrollo, CAF-Banco de Desarrollo de América Latina, a la Universidad de Los Andes y a la Universidad Nacional de Colombia. Duncan Web agradece la financiación del ED 465 en la Universidad Paris 1, y el projecto ANR-17-EURE-0001. Los autores no tienen conflictos de interes que declarar. Agradecemos tambien a los miembros del grupo de trabajo de CoVIDA, Yessica Daniela Campaz Landazabal, Marylin Hidalgo, Juliana Quintero, Paola Betancourt, Pablo Rodríguez, Diana Sofia Rios Oliveros, Andrés Felipe Patiño, Jose David Pinzon Ortiz,Maribel Rincón, Alejandro Segura, Leonardo Salas Zapata.

*Laajaj:University of Los Andes, Bogotá, Colombia, r.laajaj@uniandes.edu.co; De Los Rios: Inter-American Development Bank, Washington D.C, United States, cdelosriosru@gmail.com; Sarmiento-Barbieri, University of Los Andes, Bogotá, Colombia, i.sarmiento@uniandes.edu.co; Aristizabal: University of Los Andes, Bogotá, Colombia, de.aristizabal411@uniandes.edu.co; Behrentz: University of Los Andes, Bogotá, Colombia, ebehrent@uniandes.edu.co; Bernal: University of Los Andes, Bogotá, Colombia, rbernal@uniandes.edu.co; Buitrago, Universidad Nacional de Colombia, Bogotá, Colombia, gbuitragog@unal.edu.co; Cucunubá, Imperial College London, London, UK, zulma.cucunuba@imperial.ac.uk; de la Hoz, Universidad Nacional de Colombia, Bogotá, Colombia, fpdelahozr@unal.edu.co; Gaviria: University of Los Andes, Bogotá, Colombia, agaviria@uniandes.edu.co; Hernández: University of Los Andes, Bogotá, Colombia, luishern@uniandes.edu.co; León: University of Los Andes, Bogotá, Colombia, lj.leon62@uniandes.edu.co; Moyano, Secretaria de Salud de Bogotá, Bogotá, Colombia, dmoyano@saludcapital.gov.co; Osorio, Secretaria de Salud de Bogotá, Bogotá, Colombia, ejosorio@saludcapital.gov.co; Ramírez Varela: University of Los Andes, Bogotá, Colombia, anrami2@uniandes.edu.co; Restrepo: University of Los Andes, Bogotá, Colombia, srestrep@uniandes.edu.co; Rodriguez, Secretaria de Salud de Bogotá, Bogotá, Colombia, rorofer21@gmail.com; Schady, Inter-American Development Bank, Washington D.C, United States, norberts@iadb.org; Vives: University of Los Andes, Bogotá, Colombia, mvives@uniandes.edu.co; Webb, Paris School of Economics, Paris, France, dmbwebb@gmail.com. Contribuciones de autores: R.L., G.B. y FdlH. concibieron el estudio; R.L., CDLR., I.S. y D.W. analizaron los datos; R.L, D.A, R.B, G.B, A.G, L. J. H., L.L, A.R.V, S.R, y M.V implementaron el proyecto CoVIDA, incluyendo tests y recolección de datos primarios; R.L., CDLR., I.S. y D.W. escribieron el artículo; D.M, E.O y R.R dieron acceso a los datos de la SDS y contribuyeron al diseño del proyecto CoVIDA, todos los autores hicieron una revisión crítica de las diferentes versiones del artículo.

y la dinámica de la pandemia de COVID-19 en Bogotá desde junio de 2020 hasta principios de marzo de 2021. La riqueza de nuestros datos permite describir características propias de la pandemia en un país en desarrollo. Encontramos que, a fines de marzo de 2021, cerca de la mitad de la población de Bogotá ha sido infectada, a pesar de que una pequeña fracción de esta población fue detectada. También mostramos que la proporción de la población infectada varía ampliamente por ocupación, estrato socioeconómico, y localidad. Esto, a su vez, ha afectado la dinámica de la pandemia: mientras que la primera ola de infecciones fue impulsada por los estratos económicos más bajos y ocupaciones altamente expuestas, el segundo pico afectó a la población más uniformemente. Una mejor comprensión de la propagación y la dinámica de la pandemia entre diferentes grupos proporciona una guía valiosa para la focalización eficiente de las medidas y restricciones de las políticas de salud.

Palabras clave: SARS-CoV-2, COVID-19, CoVIDA, América Latina
Códigos JEL: I14, I15, I18, O54

1 Introduction

As of March 8th 2021, the World Health Organization (WHO) has declared 116 million detected cases of SARS-CoV-2 worldwide. 43% of these detected cases have occurred in developing countries (World Heath Organization, 2021), a group which comprises 81% of the global population. Taken at face value, these statistics would imply that SARS-CoV-2 is more prevalent in developed countries. But recent evidence points at high levels of infection in developing contexts (excluding those that were able to contain the pandemic) (Mattar et al., 2020, Uyoga et al., 2021), with some of the highest rates of infection expected to occur in Latin American countries (O'Driscoll et al., 2020). Moreover, high rates of infection in low-income areas may not be well reflected in recorded cases because the ability to detect cases may vary dramatically, for example ranging from an estimated one in four cases in Philadelphia, U.S.A., to one in 621 cases in Kenya (Flannery et al., 2020, Silveira et al., 2020, Stringhini et al., 2020, Kar et al., 2021, Uyoga et al., 2021).

Research documenting the differential impact of the COVID-19 pandemic across occupational or socioeconomic groups is largely limited to developed countries (Angelucci et al., 2020). However, developing economies have specific features that are likely to affect the transmission patterns of the virus. For example, workers may be less able to work remotely (Alfaro et al., 2020); rates of informal work may be higher, with less sick leave; institutions may have lower capacity to implement quarantine, isolation, testing and tracing measures; use of public transportation may be higher; and some cultures may have norms encouraging gathering and physical contact. It is therefore important to study the spread of the pandemic in a developing context in order to better inform local authorities' decisions for targeted interventions or restrictions.

2 Data and Estimation

Our primary data comes from the CoVIDA project, a sentinel community based surveillance led by the University of Los Andes that includes information on an average of 1,552 RT-PCR tests per week in Bogotá throughout a 9-month period (see Supplementary Materials SI.1.3 for a description of the test). We targeted and invited a mostly asymptomatic adult population. We over-sampled occupations that were expected to be most exposed, such as security guards, transportation workers and health workers. Half of the sample was randomly drawn from large lists of participants obtained through various agreements with partners. Many of the lists represented a large share of a given occupation, with the aim of obtaining a representative sample within each occupation. Another half was a convenience sample based on a public invitation for free testing. Results are qualitatively similar when restricted to the sample that is randomly drawn from lists (see Supplementary Table SI.1 and Figure SI.2). We use also administrative data collected by the Health Secretary of Bogotá (HSB, *Secretaria de Salud de Bogotá* in Spanish) that covers all the reported cases in Bogotá from the beginning of the pandemic on January 23rd, 2020 up until February 14, 2021.

In our main estimations, we re-weight observations by occupation in order to ensure our results are representative of the whole population of Bogotá. We exclude individuals with symptoms or known contacts with an infected person. This conservative assumption allows us to avoid the bias that would occur if these individuals (with typically higher prevalence rates) were more likely to seek testing or to accept it when invited (Vogel, 2020). This leaves us with 42,164 observations in the main estimations. In addition, for some estimations, we convert the positivity rate of CoVIDA tests into a number of daily new cases. To do this, we take into account the estimated sensitivity of the RT-PCR tests, which implies that individuals can be tested positive for a period of 17 days on average

(Miller et al., 2020) (for more details on the sample, calculation methods and robustness see Supplementary Methods, Sections SI.1.2 and SI.1.3).

3 From the tip of the iceberg to its actual size

We use the CoVIDA database to estimate total cases in Bogotá (the actual size of the iceberg) and compare it to the number of detected cases from the HSB (the tip of the iceberg). As shown in figure 1, our estimation of the cumulative number of cases per 100,000 inhabitants is highly consistent with the positivity rate obtained in a seroprevalence survey administered by the National Health Institute of Colombia (NHI) between October 26th and November 17th, 2020.

We draw multiple lessons from these figures. First, by March 3rd, 2021, 53% [95% CI: 45-62%] of the population in Bogotá had been infected. This is in line with other studies showing total incidence rates nearing 50% in Latin America (O'Driscoll et al., 2020, Mattar et al., 2020, Del Brutto et al., 2020), compared to rates of 1% to 15% in developed countries (Havers et al., 2020, Stringhini et al., 2020, Dopico et al., 2020). Second, only 8% of the population had been tested positive by the end of January, implying that approximately one in every 6.4 cases is detected (95% CI:5.4-7.5). This is a relatively high detection rate compared to other developing contexts (Flannery et al., 2020, Silveira et al., 2020, Kar et al., 2021, Uyoga et al., 2021). However, detection rates appear to vary significantly even within Bogotá, ranging from one in 10.1 for the lowest socioeconomic stratum to one in 5.9 for the highest strata (see Supplementary Table SI.2).

Third, other studies found that cases typically start to decrease between one and four weeks after the beginning of the quarantine in European countries, US, China, Iran and Turkey (but not in Russia, where it appeared to be ineffective) (Thu et al., 2020, Pan et al., 2020). By contrast, while the quarantine that started on March 24th, 2020 in

Bogotá helped to "flatten the curve" and avoid overcrowding of hospitals, the number of daily new cases continued to increase. We estimate that the reproduction number during the early phases of quarantine, when the fraction of the population that is susceptible (S) is assumed to be approximately 1, is $R_q = 1.22$ [95% CI: 1.17-1.27]. It appears to be stable during the first months of the quarantine (see Supplementary Material Section SI.1.4 or our companion paper (Laajaj et al., 2020)). Since the effective reproduction number R_e can be found by multiplying R_q by susceptible proportion S, it would require S to be below 0.82 for the R_e to fall below one. This is highly consistent with a downward trend of the first peak that started by the end of July, when the share of infected reached almost 20% of the population. In short, the quarantine alone was insufficient to curb the increase in daily new cases until a significant fraction of the population was immune.

Condensing the story of the pandemic in Bogotá into a simple framework, we attribute most variation in transmission to two dominant mechanisms: number of contacts and susceptibility. These are well reflected in the calculation of the effective reproduction number $R_e = S \times n \times SAR$, where n is the average number of contacts of an infected individual during their infectious period, SAR is the secondary attack rate (the probability of infection conditional on contact) and S is the share of the population that is susceptible to the disease (never infected, if infection is associated with immunity). Mobility levels in Bogotá went down considerably at the beginning of the quarantine, for example through less time commuting or at work, likely leading to large reductions in n (see Supplementary Figure SI.6). However, this mobility increased over time as a result of the growing need for people to resume their economic and social activities and a progressive loosening of the quarantine. Hence over time, exposure increased and with it the R_e. On the other hand, as a consequence of the growing share of the population having been infected shown in figure 1, the share of susceptible individuals went down over time, slowing down the expansion of the virus. After the end of the full quarantine, the two forces seemed to be balanced in the period from September to November. Then, the celebration of *novena de Navidad* (a festival celebrated on the nine days leading up Christmas), Christmas, and New Year led to an increase in exposure that was high enough to overcome the immunity effect of having more than one third of the population already infected before the beginning of the second peak. In February, once the celebrations were over, the decrease in n, combined with the growing immunity effect, both contributed to the fast decrease in the number of infections. If infections lead to durable immunity, then a third peak of a similar magnitude seems unlikely given the shrinking share of susceptible individuals.

4 An uneven distribution of infections

In Figure 2 we analyze heterogeneity in infections by occupation, socioeconomic strata, and locality. Figure 2a, shows that as expected, occupations that are most likely to be able to work remotely are among the least affected, while occupations that require physical interactions tend to be at the top of the distribution (Dingel and Neiman, 2020). Indeed, security guards, construction workers, shopkeepers, taxi drivers, public transportation workers, military and police were all (rightly) defined as priority populations for testing by the CoVIDA project because of their exposure to multiple contacts. Less anticipated were the high levels of infection among housewives and the unemployed, which may be explained by the low socioeconomic status of these groups. The figure also hints at quite different dynamics among the various groups. For example, security guards, transportation workers and tellers had already reached high levels of infection by the end of November. These occupations were considered "essential workers", exempt from the full quarantine. By contrast, most of the infections for shopkeepers, construction workers, babysitters and house workers occurred between November and February, which is consistent with the fact that they had stronger restrictions to commute and work on site early on in the pandemic (more details on quarantine restrictions by occupation in Supplementary Table SI.3).

Figure 2b displays a similar heterogeneity in infections by socioeconomic stratum, a classification used in Colombia to determine the costs of utilities and a proxy for the household's socioeconomic conditions. Neighborhoods are categorized into one of 6 levels, where 1 is the poorest and 6 is the wealthiest. To gain power, we pool together strata 1 and 2 and strata 5 and 6. The figure shows that the relationship between socioeconomic condition and infection is monotonic and that the poorest strata were four times more likely to have been infected than the wealthiest ones. Also, lower strata were hit particularly hard during the first period, whereas middle and higher strata had a large share of cases occurring during the second period. This is consistent with the observation that higher strata were more likely to be subject to the full quarantine in the earlier period, meaning that they started the second period with a high share of susceptible individuals. Figure 2c shows an even greater level of heterogeneity between districts (Bogotá includes 20 districts, which are the largest geographical division). The virus spread more widely and more quickly in districts that are poorer (see Supplementary Figure SI.9 for maps of infections and socioeconomic status).

5 Different dynamics

After finding that exposure varies by group and over time, we analyze how this translates into different dynamics of infections across the various groups. Figure 3a illustrates how much the dynamics can differ by occupational group. For example, the group of workers that are most able to work remotely (e.g. lawyers, engineers and scientist) was able to maintain low infection throughout the study period, tellers contributed mostly to the first peak, shopkeepers to the second peak, and security guards to both. (the dynamics of all groups appear in Supplementary Figures SI.8).

Figures 3b to 3d use three different sources to capture variations by stratum. The data from CoVIDA may suffer from limited statistical power due to lower sample size, especially at higher levels of disaggregation by group and time, whereas the data from the reported cases are subject to rates of detection that vary substantially by strata and over time and generate a systematic bias (Supplementary Table SI.2). Hence we also show daily COVID-19 reported fatalities per 100,000 inhabitants in Figure SI.7d which is less subject to both limitations.

Because the very first cases were imported by international travelers, higher case rates were initially observed in the wealthiest strata, but this pattern rapidly reversed. The data from the HSB thus show a very clear ordering during the first wave, with a strong decrease in infection rates as income increases (Figure 2b). Interestingly, this relationship is not observed at all in the second wave. Instead, all strata exhibit similar rates of daily new cases from COVID-19, and if anything, strata 5&6 have the highest rates of daily new cases reported at the beginning of the second wave. However, this is not reflected in the CoVIDA data nor in deaths associated to COVID-19 in official data (Figure 3d). This puzzle seems to be driven by the fact that, between July 2020 and January 2021, the share of cases that were detected grew from 5% to 52% in strata 5&6, while it remained relatively stable in the poorest strata. The differences in detection dynamics across strata may be a consequence of the government program PRASS (*"Programa de Pruebas, Rastreo y Aislamiento Selectivo Sostenible"*), the Colombian adaptation of the Test, Trace and Isolate strategy, which began in 08/12/2020 (Ministerio de Salud y Protección Social, 2020). The program decentralized the testing process so that health care providers were responsible for detecting the cases of their affiliated members. Although in Bogotá 97% of individuals are affiliated to a health care provider, the quality of service and appointment wait times vary greatly and is highly correlated with income levels. In keeping with this, the delegation of testing to providers appears to have dramatically increased the detection rate among the rich. This finding highlights a potential trade-off between a more centralized system that was on average less efficient but more equitable, and a decentralized system that led to improvements but only in the wealthiest strata.

By relating the recorded COVID-19 fatalities to the estimated cases from the CoVIDA data, we find an infection fatality ratio of 0.34%, a value that is plausible in a relatively young population (O'Driscoll et al., 2020). Interestingly, recorded new daily cases are greater in the first wave than in the second one, but this is not the case if we use the CoVIDA data or registered deaths. Hence, part of the increase in detected cases is driven by an increase in detection rate over time (Supplementary Table SI.2).

6 Discussion

A rich combination of primary and administrative data allows us to analyze the spread of COVID-19 in Bogotá. Caution should be used when interpreting the results, since we are unable to use a perfectly representative sample. However, the congruence between our results and the independent serology survey implemented in the middle of our study period lends credibility to our analysis. This is one of the most extensive studies of COVID-19 on asymptomatic participants, filling an important gap in our current knowledge. Moreover, the rich primary data provides a depth that is unique in such a context, in particular allowing us to identify potential biases in official data, and to analyze which occupational and socioeconomic groups are most affected by virus spread and how this varies over time.

Compared to Bogotá, the infection rates seen in the rest of Colombia have been even higher (estimated by the serology tests of the National Health Institute), with lower rate of detection (see Supplementary Figure SI.10). In general, other areas of Colombia are less socioeconomically developed and have poorer institutional and laboratory capacity. This finding is thus in line with our result within Bogotá showing that lower economic strata have higher rates of infection and lower detection rates and implies that a large part of Latin America is likely to have infections rates that are even higher despite their small number of detected cases.

We estimate that, as of March 3rd 2021, about 54% [95% CI: 46-63%] of the population has been infected. The implications of this result depend on the duration of the immunity that is gained from a prior infection, a question that is still the subject of some debate. Recent evidence indicates that durable immunity against secondary COVID-19 disease is likely for most individuals (Dan et al., 2021, Sasisekharan et al., 2021, Rodda et al., 2021). However, the spread of novel variants of SARS-CoV-19 has raised concerns that such variants may elude the immune response (Callaway, 2021) and may be related to the resurgence of COVID-19 in Manaus, Brazil in early 2021, despite high seroprevalence in late 2020 (Sabino et al., 2021). Protection from new variants and a better understanding of it become new priorities.

Under the assumption that infection does entail immunity for the majority of those infected, then our results show that Bogotá, Colombia more broadly, and perhaps a good part of the developing world may have reached immunity levels that will help to slow down the evolution of the pandemic considerably over the coming months. This diagnostic is particularly important at this point in time, when vaccines are slowly being rolled out in Colombia. Widespread immunity may contribute to lower case loads in the future. At the same time, the country is likely to reach its first two vaccination phases (targeting health staff and adults older than 60) only by June 2021. Without a clear diagnostic of the current situation, reductions in future cases may be misattributed to the vaccine roll out.

Our work shows that the high level of economic inequalities in Latin America translates into inequalities in infection among different groups.[1] We go beyond showing static differences between groups and describe how they vary over time. Understanding these

[1] In a companion paper, Laajaj et al. (2020), we attempt to understand the key channels that drive this inequality in infections between different socioeconomic groups.

dynamics is key for appropriate targeting of interventions. For example, in an active surveillance testing initiative like CoVIDA, targeting the groups with the highest infection rates will require targeting groups that are poor and have high occupational exposure during the first wave. But in the second wave, identifying populations with greater levels of exposure becomes more difficult, because the greater levels of exposure of some groups tends to be compensated by more immunity. Populations that have recently resumed their economic activity, such as teachers in the coming months, may combine low immunity and high exposure, and thus be at the greatest risk.

References

Alfaro, L., Becerra, O., and Eslava, M. (2020). Emes and covid-19: shutting down in a world of informal and tiny firms. Technical report, National Bureau of Economic Research.

Angelucci, M., Angrisani, M., Bennett, D. M., Kapteyn, A., and Schaner, S. G. (2020). Remote work and the heterogeneous impact of covid-19 on employment and health. Technical report, National Bureau of Economic Research.

Callaway, E. (2021). Fast-spreading covid variant can elude immune responses. *Nature*, 589(7843):500–501.

Challen, R., Brooks-Pollock, E., Tsaneva-Atanasova, K., and Danon, L. (2020). Meta-analysis of the sars-cov-2 serial interval and the impact of parameter uncertainty on the covid-19 reproduction number. *medRxiv*.

Dan, J. M., Mateus, J., Kato, Y., Hastie, K. M., Yu, E. D., Faliti, C. E., Grifoni, A., Ramirez, S. I., Haupt, S., Frazier, A., et al. (2021). Immunological memory to sars-cov-2 assessed for up to 8 months after infection. *Science*.

Del Brutto, O. H., Costa, A. F., Mera, R. M., Recalde, B. Y., Bustos, J. A., and García, H. H. (2020). Sars-cov-2 in rural latin america. a population-based study in coastal ecuador. *Clinical Infectious Diseases*.

Dingel, J. I. and Neiman, B. (2020). How many jobs can be done at home? *Journal of Public Economics*, 189:104235.

Dopico, X. C., Muschiol, S., Christian, M., Hanke, L., Sheward, D. J., Grinberg, N. F., Bogdanovic, G., Mcinerney, G. M., Allander, T., Wallace, C., et al. (2020). Seropositivity in blood donors and pregnant women during 9-months of sars-cov-2 transmission in stockholm, sweden. *medRxiv*.

Flannery, D. D., Gouma, S., Dhudasia, M. B., Mukhopadhyay, S., Pfeifer, M. R., Woodford, E. C., Gerber, J. S., Arevalo, C. P., Bolton, M. J., Weirick, M. E., et al. (2020). Sars-cov-2 seroprevalence among parturient women in philadelphia. *Science immunology*, 5(49).

Havers, F. P., Reed, C., Lim, T., Montgomery, J. M., Klena, J. D., Hall, A. J., Fry, A. M., Cannon, D. L., Chiang, C.-F., Gibbons, A., et al. (2020). Seroprevalence of antibodies to sars-cov-2 in 10 sites in the united states, march 23-may 12, 2020. *JAMA internal medicine*, 180(12):1576–1586.

He, X., Lau, E. H., Wu, P., Deng, X., Wang, J., Hao, X., Lau, Y. C., Wong, J. Y., Guan, Y., Tan, X., et al. (2020). Temporal dynamics in viral shedding and transmissibility of covid-19. *Nature medicine*, 26(5):672–675.

Hilton, J. and Keeling, M. J. (2020). Estimation of country-level basic reproductive ratios for novel coronavirus (sars-cov-2/covid-19) using synthetic contact matrices. *PLoS computational biology*, 16(7):e1008031.

Instituto Nacional de Salud (2020). Estudio nacional de seroprevalencia. *Estudio Nacional de Seroprevalencia*.

Kar, S. S., Sarkar, S., Murali, S., Dhodapkar, R., Joseph, N. M., and Aggarwal, R. (2021). Prevalence and time trend of sars-cov-2 infection in puducherry, india, august–october 2020. *Emerging Infectious Diseases*, 27(2):666.

Laajaj, R., Webb, D., Aristizabal, D., Behrentz, E., Bernal, R., Buitrago, G., Cucunubá, Z., de la Hoz, F., Delgado, G., Gaviria, A., Hernández, L. J., Keefer, P., León, E., Moyano, D., Osorio, E., Ramírez Varela, A., Restrepo, S., Rodríguez, R., Schady, N., and Vives, M. (2020). Understanding inequalities in sars-cov-2 infections across socioeconomic groups. Mimeo.

Mattar, S., Alvis-Guzman, N., Garay, E., Rivero, R., García, A., Botero, Y., Miranda, J., Galeano, K., de La Hoz, F., Martínez, C., Arrieta, G., Faccini-Martínez, A. A., Guzmán, C., Kerguelen, H., Moscote, M., Contreras, H., and Contreras, V. (2020). SARS-CoV-2 seroprevalence among adults in a tropical city of the Caribbean area, Colombia: are we much closer to herd immunity than developed countries? *Open Forum Infectious Diseases*. ofaa550.

McAloon, C., Collins, Á., Hunt, K., Barber, A., Byrne, A. W., Butler, F., Casey, M., Griffin, J., Lane, E., McEvoy, D., Wall, P., Green, M., O'Grady, L., and More, S. J. (2020). Incubation period of covid-19: a rapid systematic review and meta-analysis of observational research. *BMJ Open*, 10(8).

Miller, T. E., Garcia Beltran, W. F., Bard, A. Z., Gogakos, T., Anahtar, M. N., Astudillo, M. G., Yang, D., Thierauf, J., Fisch, A. S., Mahowald, G. K., et al. (2020). Clinical sensitivity and interpretation of pcr and serological covid-19 diagnostics for patients presenting to the hospital. *The FASEB Journal*, 34(10):13877–13884.

Ministerio de Salud y Protección Social (2020). Mediante decreto se reglamenta el programa prass. *Boletín de Prensa No 590 de 2020*.

O'Driscoll, M., Dos Santos, G. R., Wang, L., Cummings, D. A., Azman, A. S., Paireau, J., Fontanet, A., Cauchemez, S., and Salje, H. (2020). Age-specific mortality and immunity patterns of sars-cov-2. *Nature*, pages 1–6.

Pan, A., Liu, L., Wang, C., Guo, H., Hao, X., Wang, Q., Huang, J., He, N., Yu, H., Lin, X., et al. (2020). Association of public health interventions with the epidemiology of the covid-19 outbreak in wuhan, china. *Jama*, 323(19):1915–1923.

Rodda, L. B., Netland, J., Shehata, L., Pruner, K. B., Morawski, P. A., Thouvenel, C. D., Takehara, K. K., Eggenberger, J., Hemann, E. A., Waterman, H. R., et al. (2021). Functional sars-cov-2-specific immune memory persists after mild covid-19. *Cell*, 184(1):169–183.

Sabino, E. C., Buss, L. F., Carvalho, M. P., Prete, C. A., Crispim, M. A., Fraiji, N. A., Pereira, R. H., Parag, K. V., da Silva Peixoto, P., Kraemer, M. U., et al. (2021). Resurgence of covid-19 in manaus, brazil, despite high seroprevalence. *The Lancet*, 397(10273):452–455.

Sasisekharan, V., Pentakota, N., Jayaraman, A., Tharakaraman, K., Wogan, G. N., and Narayanasami, U. (2021). Orthogonal immunoassays for igg antibodies to sars-cov-2 antigens reveal that immune response lasts beyond 4 mo post illness onset. *Proceedings of the National Academy of Sciences*, 118(5).

Silveira, M. F., Barros, A. J., Horta, B. L., Pellanda, L. C., Victora, G. D., Dellagostin, O. A., Struchiner, C. J., Burattini, M. N., Valim, A. R., Berlezi, E. M., et al. (2020). Population-based surveys of antibodies against sars-cov-2 in southern brazil. *Nature Medicine*, 26(8):1196–1199.

Stringhini, S., Wisniak, A., Piumatti, G., Azman, A. S., Lauer, S. A., Baysson, H., De Ridder, D., Petrovic, D., Schrempft, S., Marcus, K., et al. (2020). Seroprevalence of anti-sars-cov-2 igg antibodies in geneva, switzerland (serocov-pop): a population-based study. *The Lancet*, 396(10247):313–319.

Thu, T. P. B., Ngoc, P. N. H., Hai, N. M., et al. (2020). Effect of the social distancing measures on the spread of covid-19 in 10 highly infected countries. *Science of The Total Environment*, 742:140430.

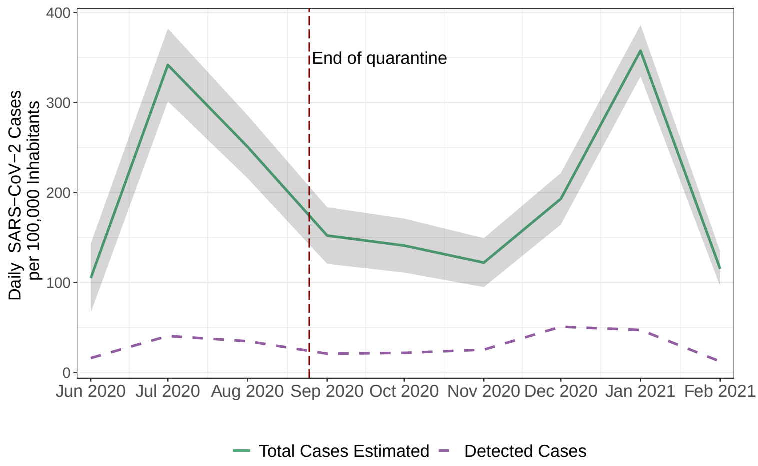
Uyoga, S., Adetifa, I. M., Karanja, H. K., Nyagwange, J., Tuju, J., Wanjiku, P., Aman, R., Mwangangi, M., Amoth, P., Kasera, K., et al. (2021). Seroprevalence of anti–sars-cov-2 igg antibodies in kenyan blood donors. *Science*, 371(6524):79–82.

Vogel, G. (2020). First antibody surveys draw fire for quality, bias.

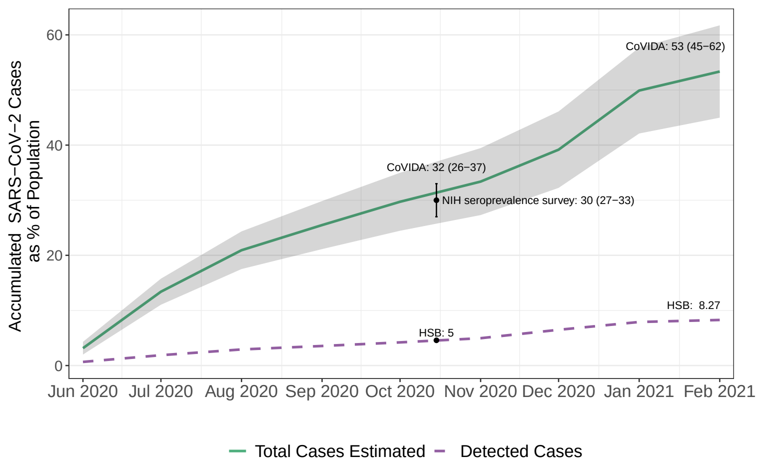
Wallinga, J. and Lipsitch, M. (2007). How generation intervals shape the relationship between growth rates and reproductive numbers. *Proceedings of the Royal Society B: Biological Sciences*, 274(1609):599–604.

World Heath Organization (2021). Who coronavirus (covid-19) dashboard. *WHO Coronavirus (COVID-19) Dashboard.*

7 Figures

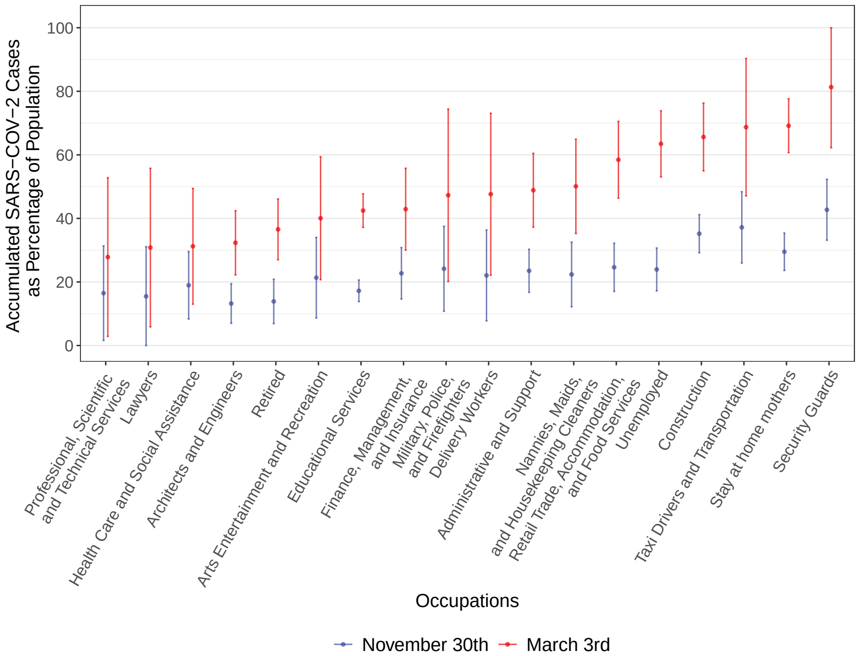


(a) Daily Cases per 100,000 Inhabitants

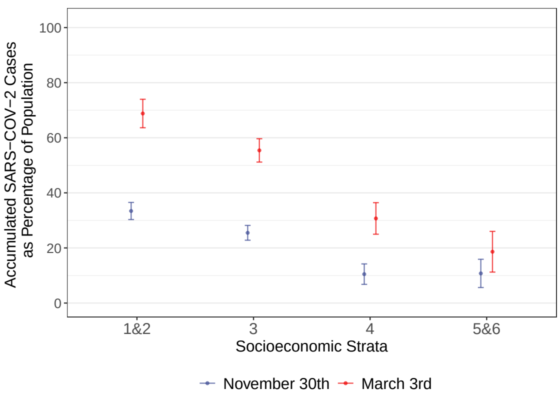


(b) Accumulated Cases as % of Population

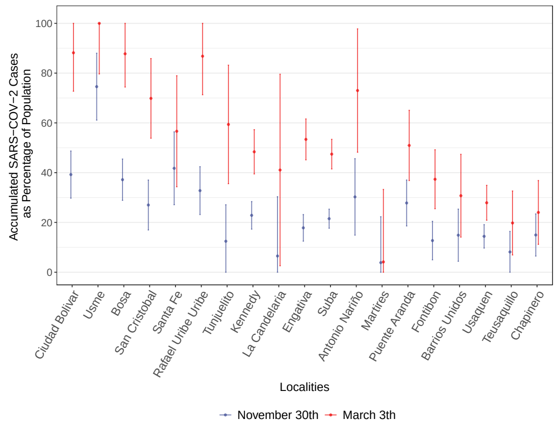
Figure 1. SARS-CoV-2 daily new cases per 100,000 inhabitant and accumulated cases as % of Population. Panel (a) shows total daily SARS-CoV-2 cases per 100,000 inhabitants based on CoVIDA data (solid line) and detected cases based on data from the Health Secretary of Bogotá (HSB) (dashed line). The vertical dashed line marks the end of quarantine on August 25, 2020. Panel (b) shows cumulative cases as % of Bogotá's population. It also shows in black the point estimate and 95% confidence interval of the test positivity rate from a seroprevalence survey run by the National Health Institute of Colombia (NHI) (Instituto Nacional de Salud, 2020). Estimated cases using CoVIDA data were calculated using a monthly weighted average and assuming a 17 day positivity window. Weights were calculated based on workers' occupation. Shaded regions denote 95% confidence intervals.



(a) Occupations

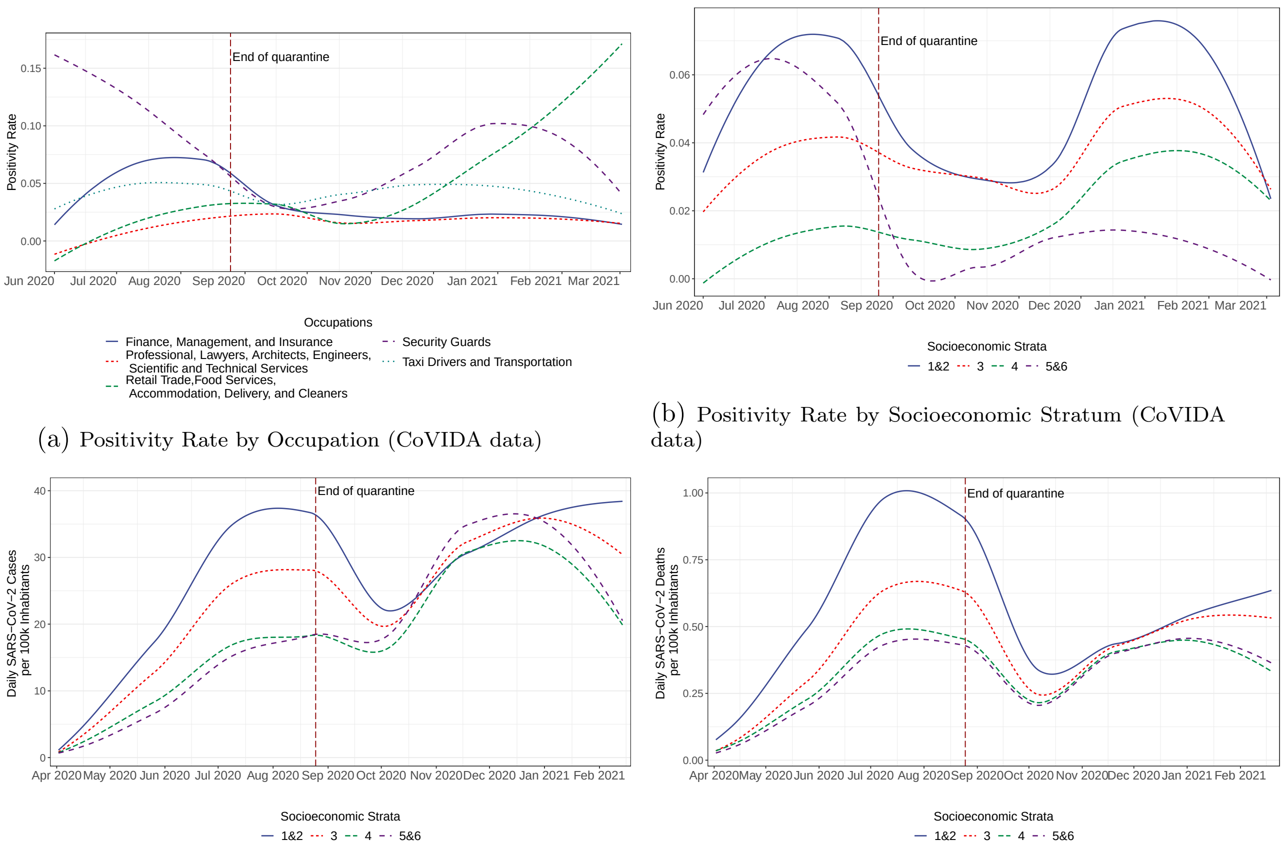


(b) Socio-economic Strata



(c) Localities

Figure 2. Estimated Accumulated SARS-CoV-2 Cases as a percentage of Population by Occupation, Socioeconomic Stratum, and District for two time periods: June 1st-November 30th and June 1st-March 3rd). Panel(a) shows estimated accumulated SARS-CoV-2 cases as percentage of the population of workers in each category. Panel (b) repeats the exercise using socioeconomic strata. Panel (c) runs the same estimation but grouping by Bogotá's districts, sorted by the mean stratum for individuals in a district. Observations are weighted by occupation to be representative of the Bogotá population. Error bars denote 95% confidence intervals.



(a) Positivity Rate by Occupation (CoVIDA data)

(b) Positivity Rate by Socioeconomic Stratum (CoVIDA data)

(c) Daily new cases by Socioeconomic Stratum (HSB data)

(d) Daily deaths by Socioeconomic Stratum (HSB data)

Figure 3. Daily Dynamics. The figure shows smoothed SARS-CoV-2 daily positivity rates (panels (a) and (b)) from CoVIDA data and smoothed SARS-CoV-2 daily cases per 100,000 inhabitants (panels (c) and (d)) from the Health Secretary of Bogotá (HSB). The vertical dashed line marks the end of quarantine on August 25, 2020. Positivity rates were calculated using worker's occupation weights. Daily rates and cases were smoothed using a local polynomial regression (loess) with a smoothing parameter of 0.7. See Supplementary Methods Figure SI.7 for same figures with 95% confidence intervals.

Supplementary Information

SI.1 Supplementary Methods

The Methods section is structured as follows: first we describe the RT-PCR test used and its reliability, then in 'Data Description' we describe the datasets used. In the 'Estimation' subsection we describe how we estimate positivity rates, sample construction and robustness, to then explain how we convert positivity rate into a number of daily new cases. Next, we proceed to describe how we calculate the basic reproduction number.

SI.1.1 The RT-PCR test used and its reliability

The tests to be used in this study are based on the reverse transcriptase technique - PCR or one-step RT-PCR (the U-TOPTM COVID-19 detection kit). This kit was validated in the facilities of the Public Health Laboratory of the Secretaria Distrital de Salud, finding that the kit allows detection of the SARS-CoV-2 virus, with results comparable to the Charité technique, Berlin (Rev, January 13, 2020).

SI.1.2 Data Description

SI.1.2.1 CoVIDA Data

Our primary data comes from the CoVIDA project led by the University of Los Andes. This community based sentinel surveillance initiative was integrated to the district's public health surveillance and organized by occupation group. The CoVIDA project was designed to help contain the spread of SARS-CoV-2 through active surveillance among mostly asymptotic individuals and to provide a range of information that differs from the self-selected symptomatic individuals tested in health facilities. The sample includes 59,770 RT-PCR tests of SARS-CoV-2 on 55,078 different individuals in Bogotá from the beginning of June 2020 to March 3rd, 2021. At the time of registration, individuals were surveyed to capture various characteristics, including occupation, socioeconomic strata, and address.

Two main strategies were employed to recruit participants, and about one half of the total sample comes from each strategy. First, through 74 agreements with institutions and companies, we obtained long-lists with that we used to contact and invite the participants. Most lists were specific to a given occupation through a large company, app or , and some lists of residents (i.e. beneficiaries of social programs), then randomly selected participants from the lists and contacted them to invite them to be tested for free. The total population of all lists covers 20% of the population in Bogotá, hence it is relatively close to a population-based sampling, but with an over-representation of some occupations that were prioritized in the CoVIDA project because they were expected to be more exposed (which is why we re-weight by occupation, as explained below).

The second source of participants' identification comes from public announcements made by the CoVIDA team through various communication channels to invite people to be tested, stating explicitly that the invitation is open to those that are asymptomatic.

SI.1.2.2 Health Secretary of Bogotá Data

Our second database comes from administrative records, collected by the Health Secretary of Bogotá (HSB, in Spanish the *Secretaria de Salud de Bogotá*), that cover the universe of cases of Bogotá residents that have been tested positive to SARS-CoV-2 by any laboratory using an RT-PCR test, starting from the beginning of the pandemic (January 23rd, 2020) until February 14th, 2021. All laboratories in Bogotá must report any positive test to the HSB, which in turn reports it to the National Health Institute that provides national statistics used by the World Health Organization. This administrative data also comes with basic socioeconomic characteristics from a form that is a mandatory part of the institution's report when recording a positive case to the HSB.

SI.1.3 Estimation

SI.1.3.1 Estimation of Positivity Rates

To obtain unbiased estimates of the true positivity rate, we need that, within each occupation, the likelihood of being tested is not systematically correlated with the likelihood of being positive within each occupation. This requires that people in the invited occupation lists are not significantly different from other non-invited lists. For example, when we have an agreement with one of the main taxi companies in Bogotá, who shared a list of all its affiliated workers, we need that those workers are not systematically different from other taxi drivers. This is plausible because inclusion on the list is not an individual decision of the person invited to be tested, so there is no reason to expect that workers from that company would have systematically more or less positivity than other taxi drivers.

Self-selection bias could also be generated if the participants' decision to be tested is correlated with their perceived likelihood of being positive. To account for this possible bias, the survey includes questions about COVID-related symptoms and any contact with a confirmed or probable case in the past 14 days. In our main results, we exclude those with a positive answer to any of the two questions. By doing so, we account for the most likely source of bias, in a conservative way, meaning that the results should be a lower bound of the actual positivity of the population. By excluding both symptomatic and known contacts of infected individuals, the sample is reduced to 42,164. Table SI.1 shows that positivity rates are lower when excluding those individuals (3.08%) compared to when calculated on the full sample (5.75%).

To account for the over-sampling of occupations, we weigh observations by the occupation population size in Bogotá divided by the number of individuals of that occupation in the sample. This reweights our sample so that it is representative of Bogotá as a whole. Our results, however, are not particularly sensitive to this reweighting exercise. In particular, the second row of Table SI.1 shows that for the main sample used in the paper, the estimated positivity rate remains almost unchanged when reweighting (moving from 3.08% to 3.06%.)

Since the sampling from lists is potentially less subject to bias than the public campaign, we reproduce basic statistics and the results of Figure 1, restricting our sample to this first group. In Table SI.1, we find a (weighted) positivity rate that is somewhat higher among invited participants from the lists (3.24) compared to the one among participants from the public campaign (2.99), but the difference between the two is not significant (p=0.14). As a result of the difference, the estimation of the share infected during the entire period reaches 63% when excluding participants from the public campaign an shown in Figure SI.2. In the main body of the paper we nevertheless present the main results including both groups, both because this assumption is more conservative and because the estimated positivity deviates further from the NHI seroprevalence study in October when using only the list-sampled group, indicating that the inclusion of both groups is a more accurate estimate.

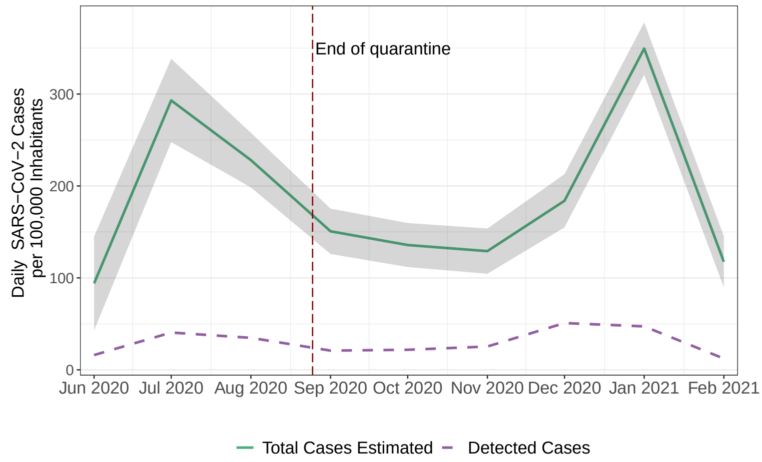
Also, running the estimation without weights don't affect the main conclusions (Supplementary Figure SI.1) hence the weights contribute to better representativity of the Bogotá population but results are robust to using crude positivity. By contrast, keeping all individuals who have symptoms or known contact with an infected person would increase the results substantially, probably as a result of the selection bias mentioned previously (Supplementary Table SI.1).

Self selection and acceptance will always make such data imperfect but they are unavoidable in any sample given that one cannot be forced to take the test, and the data still provides a unique opportunity in the region to obtain this coverage over time and different groups of the population. Tables SI.2, SI.3 and SI.5, show that a wide range of socio-economic strata, occupations and locations were reached through this approach.

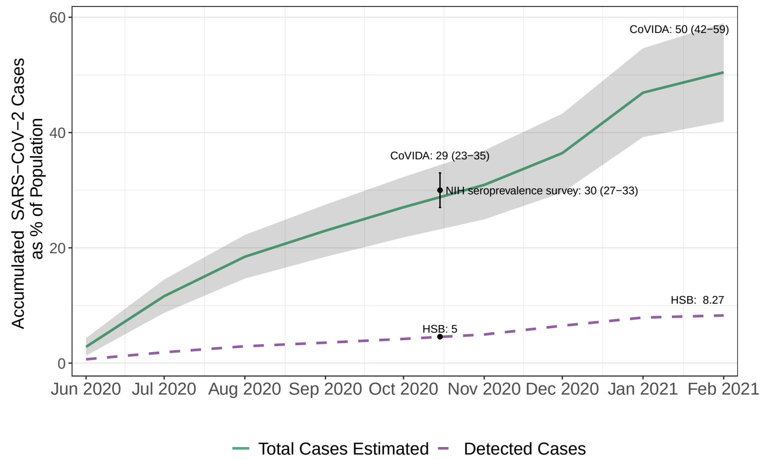
Table SI.1. Positivity rate using different CoVIDA subsamples

Sample	% Positivity Weighted	% Positivity Non-Weighted	Observations
Full sample	5.75 (5.57 - 5.94)	5.08 (4.91 - 5.26)	59,770
Excluding symptomatic and/or known contact*	3.08 (2.92 - 3.25)	3.06 (2.90 - 3.22)	42,164
Only symptomatic and/or known contact	12.00 (11.5 - 12.5)	9.93 (9.49 - 10.40)	17,606
Participants that we invited (from lists)	3.24 (3.00 - 3.49)	3.20 (2.96 - 3.44)	20,496
Participants from public campaign	2.99 (2.76 - 3.21)	2.93 (2.70 - 3.15)	21,668

Note: Estimated positivity using CoVIDA data were calculated using a an aggregated average and assuming a 17 day positivity window. Weights were calculated based on workers' occupation. Population of workers category was obtained from a review of official records from several sources. Analytical 95% Confidence Intervals in parentheses. Given an outbreak on a military battalion (69 positives out of 135) , all tested on the 2nd of July, all the samples exclude this battalion. * This is the main sample used for the estimations in the paper.

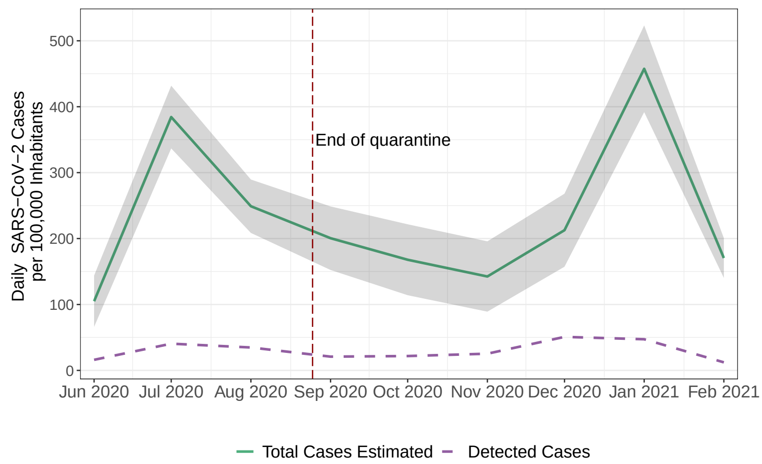


(a) Daily Cases per 100,000 Inhabitants



(b) Accumulated Cases as % of Population

Figure SI.1. Replicates Figure 1 unweighted, i.e. omits the use of occupations weights. It shows SARS-CoV-2 daily new cases per 100,000 inhabitant and accumulated Cases as % of the population: CoVIDA and Health Secretary of Bogotá (HSB). Panel (a) predicted daily SARS-CoV-2 cases per 100,000 inhabitants based on CoVIDA data (solid line) and Health Secretary of Bogotá (HSB) (dashed line). Panel (b) shows accumulated cases as % of Bogotá's population. It also shows in black point estimate and 95% confidence interval of positivity to a seroprevalence survey estimated by the National Health Institute of Colombia (NHI) (Instituto Nacional de Salud, 2020). Shaded regions denote 95% confidence intervals.



(a) Daily Cases per 100,000 Inhabitants

(b) Accumulated Cases as % of Population

Figure SI.2. Replicates Figure 1 but excludes participants from the public campaign. It shows SARS-CoV-2 daily new cases per 100,000 inhabitant and accumulated Cases as % of the population: CoVIDA and Health Secretary of Bogotá (HSB). Panel (a) predicted daily SARS-CoV-2 cases per 100,000 inhabitants based on CoVIDA data (solid line) and Health Secretary of Bogotá (HSB) (dashed line). Panel (b) shows accumulated cases as % of Bogotá's population. It also shows in black point estimate and 95% confidence interval of positivity to a seroprevalence survey estimated by the National Health Institute of Colombia (NHI) (Instituto Nacional de Salud, 2020). Occupation weights are included. Shaded regions denote 95% confidence intervals.

SI.1.3.2 From Positivity Rate to a number of infected individual per day

The key assumption for this conversion is the average number of days during which a person can be tested positive when infected. Using the estimations of (Miller et al., 2020), we estimate this number of days to be 17. Hence in order to obtain the number of cases per day and per inhabitant, one needs to divide the positivity rate by 17. Intuitively, on average, any person that tests positive was infected over the past 17 days .

The conversion from positivity to number of daily infections requires estimation expected number of days during which a person can be tested positive when infected. This expected duration is defined as the sum of the sensitivity over the entire period when the individual is infected. We estimated that there are, on average, 17 days in which a positive person can be tested positive by a PCR test (for example 2 days at 50% sensibility mean one day when the individual is detected in expected value). Hence we divided the positivity by 17 to obtain the number of daily cases. For illustration suppose that in a population of one hundred, one new person would get infected every day and the PCR test has a sensitivity of 100% during 17 days, then (after at least 17 days after the first infection can be detected) the test in that population should provide a positivity of 17%, capturing infections during the 17 days corresponding window.

Our estimation of the expected number of days during which one can be tested positive was based on Miller *et al.* (2020) (Miller et al., 2020) who estimated the sensibility of the PCR test day by day following the onset of the symptoms (see Figure SI.3). We used their results to sum the percentage of sensitivity of each day and added an extra 2 days to take into account the period prior the onset of the symptoms. We obtained a total of 17 days.

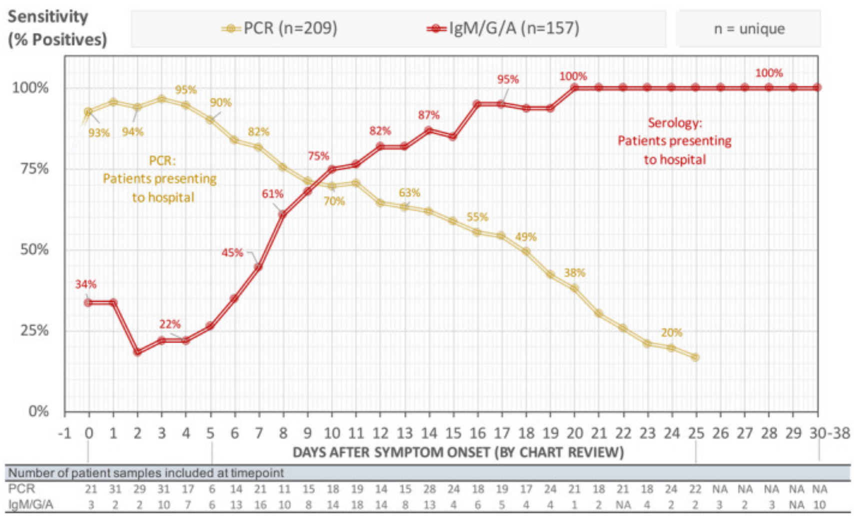


Figure SI.3. PCR sensitivity for SARS-CoV-2 taken from Miller et al. (2020) (Miller et al., 2020). It shows the clinical sensitivity to PCR tests after days of first symptoms appearance.

SI.1.4 Reproduction Number Calculation

Here we calculate the value of R_q, which we define as the average number of secondary infections generated by an infected individual at the start of the generalized quarantine period under the assumption that the proportion of susceptible individuals is 1. We treat this value as a constant in the early phase of the quarantine. To calculate the value of R_q, we use the Lotka-Euler equation (Wallinga and Lipsitch, 2007). This assumes exponential growth in new cases, assumes that all individuals in the population are susceptible ($S = 1$), and uses the rate of exponential growth in new cases r and the distribution of the generation interval $g(a)$ to calculate an estimate of R_q.

First, we calculate the rate of exponential growth in new confirmed cases per day by running an OLS regression with the natural log of daily confirmed cases as the outcome variable, and the date as the independent variable. We limit our sample to the early period of the epidemic April 1st 2020 to June 1st 2020, when the exponential growth curve fits the data well and when immunity is unlikely to play a role in case growth because $S \approx 1$. This yields an estimate of $r = 0.038$ (95% CI: 0.031, 0.046). Figure SI.4 displays the log daily confirmed new cases in Bogotá over time (the gray dots), and plots the line of best fit (in red) whose slope is equal to r.

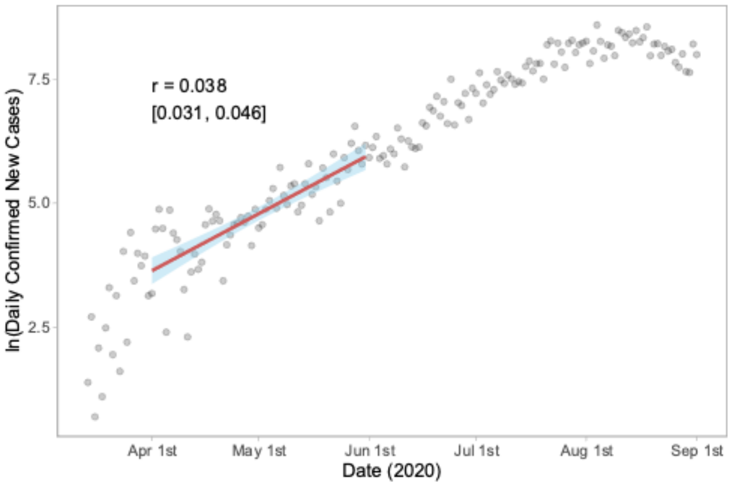


Figure SI.4. Estimate of the exponential rate of growth in new cases r

Second, we assume that the generation interval $g(a)$, where a is the number of days since infection, is described by a gamma distribution with a mean of 5.2 days and a shape parameter of 4.79. This yields a distribution with the shape seen in Figure SI.5. The choice of these parameters reflects an estimation process seen in a companion paper (Laajaj et al., 2020), in which we calibrate the generation interval based on data for the serial interval (He et al., 2020) and the incubation period (McAloon et al., 2020). Moreover, our mean of 5.2 days falls within the confidence range seen in a recent meta-analysis (Challen et al., 2020) , which estimated the mean generation interval to be 4.8 [95% CI 4.3-5.41] when using a fitted gamma distribution.

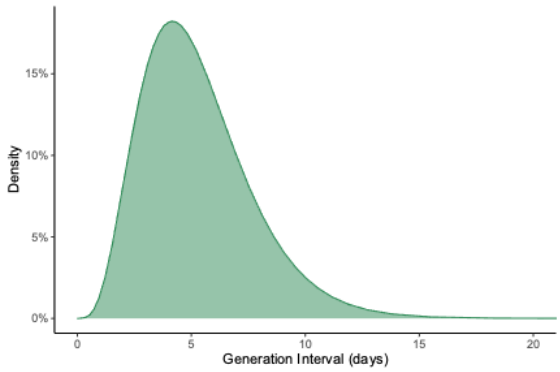


Figure SI.5. Assumed distribution of generation interval (gamma with mean = 5.2, shape = 4.79)

Using the estimated value of r, and the assumed distribution for $g(a)$, we calculate the initial value of R_q using the Lotka-Euler equation:

$$\frac{1}{R_q} = \int_{a=0}^{\infty} e^{(-ra)} g(a) da \tag{1}$$

Where $g(a)$ is the density of the generation interval as a function of the day since infection a, r is the rate of exponential growth in new cases.

Using the values of $r = 0.38$ and the $g(a)$ function from Figure 4, this yields an estimate of $R_q = 1.218$ [95% CI: 1.172, 1.266]. Note that this estimate comes during a period of strict lockdown in Bogotá, which explains why our estimate of R_q is significantly lower than the estimates of R_0 seen in the literature (which are typically calculated in conditions of full mobility (Hilton and Keeling, 2020)).

SI.2 Supplementary Tables

Table SI.2. One case detected out of...

Month	Stratum				Average
	1&2	3	4	5&6	
June	3.7 (0.5 - 6.8)	9.8 (4.0 - 15.6)	1.7 (0.0 - 6.1)	7.9 (0.0 - 24.4)	6.5 (4.1 - 8.9)
July	11.3 (8.4 - 14.1)	5.4 (3.3 - 7.4)	7.4 (2.8 - 12.0)	21.1 (9.7 - 32.6)	8.4 (7.4 - 9.4)
August	7.0 (5.6 - 8.4)	8.3 (6.5 - 10.1)	5.5 (1.8 - 9.3)	12.6 (4.1 - 21.1)	7.2 (6.2 - 8.2)
September	10.8 (8.6 - 13.1)	6.8 (5.0 - 8.5)	3.1 (1.0 - 5.2)	0.0 (0.0 - 0.0)	7.3 (5.8 - 8.8)
October	7.4 (5.3 - 9.6)	9.3 (7.4 - 11.1)	3.4 (1.9 - 5.0)	1.3 (0.0 - 2.6)	6.5 (5.1 - 7.8)
November	6.9 (4.8 - 9.1)	5.3 (4.0 - 6.6)	2.9 (1.7 - 4.2)	2.8 (1.3 - 4.4)	4.8 (3.7 - 5.9)
December	6.3 (4.6 - 8.1)	3.8 (2.9 - 4.7)	2.5 (1.5 - 3.5)	2.0 (0.9 - 3.1)	3.8 (3.2 - 4.4)
January	11.2 (9.0 - 13.3)	9.0 (7.5 - 10.5)	6.9 (4.9 - 9.0)	1.9 (0.5 - 3.3)	7.6 (7.0 - 8.2)
February	8.9 (5.5 - 12.2)	12.3 (8.5 - 16.0)	20.3 (12.7 - 27.9)	2.4 (0.0 - 5.1)	9.4 (7.8 - 11.0)
Average*	10.1 (7.8 - 12.5)	7.5 (5.8 - 9.3)	5.1 (2.7 - 7.5)	5.9 (2.4 - 9.4)	6.4 (5.4 - 7.5)

Note: the table shows the share of detection by the HSB. It is calculated as the total estimated cases using CoVIDA data divided by the total number of cases detected by the HSB the same period. Analytical Confidence Intervals in Parentheses. *For the stratums, average for the period July - January.

Table SI.3. Epidemiological week when occupation were able to resume work for the first time since the beginning of the pandemic (in week 12)

Occupation Group	Occupations Included	Positivity (%)	95% CI	Observations	Epidemiological Week* Allowed to Work
Administrative and Support	Secretary Staff (n=3,105); Call Center Employees (n=267); Insurance And Social Security Agent (n=47); Travel Agency Employees (n=35); Real Estate Agent (n=24)	3.09	(2.54, 3.66)	3,338	August 31st, 2020[1]
Architects and Engineers	Engineer (n=1,678); Web Designer (n=340); Computer Services Staff (n=254); Electrician (n=122); Mechanical Engineering Staff (n=43)	2.07	(1.51, 2.65)	2,318	August 31st, 2020[2]
Arts Entertainment and Recreation	Journalists And Writers (n=667); Cultural Activities Staff (n=613); Architects (n=362); Artists (n=196)	2.33	(1.62, 3.05)	1,719	*Always*[3]
Construction	Construction Workers (n=728); Seamstress And Related (n=87); Carpenters And Related (n=76); Shoemaker (n=3)	3.98	(2.81, 5.25)	880	April 20th, 2020[4]
Delivery Workers	Delivery Workers (n=2,101)	3.00	(2.32, 3.74)	2,066	*Always*
Educational Services	Students (n=3,213); College Professor (n=595); Secondary Teacher (n=74); Preschool Teacher (n=19); Primary Teacher (n=11)	2.60	(2.08, 3.13)	3,807	August 31st, 2020
Finance, Management, and Insurance	Personal Financial Services (n=2,484); Directors And Managers Of Companies (n=793)	2.96	(2.38, 3.58)	3,106	*Always*[5]
Health Care and Social Assistance	Nurse (n=2,271); Doctor (n=1,936); Physiotherapist (n=404); Dentists (n=373); Medical Assistants (n=236); Nutritionist (n=61); Optometrist (n=24); Hospital Admissions Staff (n=14)	2.09	(1.70, 2.49)	5,076	*Always*[6]
Lawyers	Lawyers (n=979)	1.91	(1.04, 2.84)	892	August 31st, 2020
Military, Police, and Firefighters	Police (n=646); Military (n=290); Firemen (n=32); Air Force Officers (n=6)	2.98	(1.95, 4.16)	972	*Always*
Nannies, Maids, and Housekeeping Cleaners	Personal Grooming (n=829); Trash Collectors (n=107); Babysitter (n=46); Car Washers (n=4)	3.58	(2.39, 4.82)	923	*Always*[7]
Professional, Scientific and Technical Services	Veterinarian (n=507); Psychologists (n=311); Biologist And Related (n=253); Economists (n=170); Sociologist, Anthropologist (n=44); Chemical (n=21); Geologist (n=13); Political Scientist And Related (n=9)	1.57	(0.95, 2.32)	1,270	*Always*[8]
Retail Trade, Accommodation, and Food Services	Street Vendor (n=1,406); Chefs (n=205); Pharmacist (n=199); Hairdressers And Related (n=117); Waiter (n=86); Baker And Related (n=33); Shop Seller (n=29)	6.31	(5.26, 7.40)	1,996	August 31st, 2020[9]
Retired	Retired (n=1043)	2.31	(1.39, 3.23)	994	August 31st, 2020
Security Guards	Security Guards (n=1,211)	5.02	(3.73, 6.23)	1,195	*Always*
Stay at home mothers	Stay at home mothers (n=1,332)	4.53	(3.39, 5.71)	1,279	August 31st, 2020
Taxi Drivers and Transportation	Taxi Drivers (n=3,312); Personal Transportation (n=423)	4.32	(3.67, 5.00)	3,684	*Always*
Unemployed	Unemployed (n=951)	4.20	(2.86, 5.57)	904	August 31st, 2020
Not Classified	-	2.71	(2.18, 3.27)	3,546	–

* Refers to the first day of the week in which the occupation with the most observations included in the occupation group was allowed to first work outside home since the beginning of the pandemic. 1. Call Centers were always open; 2. Electricians allowed to work outside home beginning on July 13th; 3. All but Journalists and Writers, allowed to work outside home beginning on August 31st; 4. All but Construction Workers, allowed to work outside home beginning on August 31st,2020; 5. CEOs were allowed to work outside home beginning on August 31st; 6. Dentists were allowed to work outside home beginning on August 31st; 7. Car Washers were allowed to work outside home beginning on August 31st; 8. All but Veterinarians and Psychologist, allowed to work outside home beginning on August 31st,2020; 9. Hairdressers allowed to work beginning on July 13th.

Table SI.4. Case Mortality Rate

Month	Avg. Daily Deaths	Avg. Daily Cases		Death Rate	
		HSB	CoVIDA	HSB	CoVIDA
June	16.5	1,290	9,591	1.28	0.17
July	77.6	3,262	24,686	2.38	0.31
August	83.3	2,798	20,062	2.99	0.42
September	40.7	1,680	11,954	2.42	0.34
October	28.5	1,754	11,454	1.63	0.25
November	29.7	2,035	10,636	1.46	0.28
December	44.7	4,008	15,667	1.12	0.29
January	117.1	4,009	28,670	2.92	0.23
Aggregated	55	2,605	16,086	2.11	0.34

Note: Data on deaths comes from HSB. The estimated cases using CoVIDA data were calculated using a monthly weighted average and assuming a 17 day positivity window. Weights were calculated based on workers' occupation. Population of workers category was obtained from a review of official records in several sources

SI.2.1 Populations

Table SI.5. Population by Locality

Locality	CoVIDA Data Population	CoVIDA Data %*	HSB Data Population	HSB Data %*	Official Population	Mean Stratum
Antonio Nariño	427	0.01	87,27	0.10	108,976	2.9
Barrios Unidos	1,007	0.01	13,101	0.16	276,453	3.4
Bosa	1,635	0.02	47,042	0.56	799,660	1.9
Chapinero	1,691	0.02	15,614	0.19	125,294	4.2
Ciudad Bolivar	1,229	0.01	38,909	0.46	776,351	1.4
Engativa	4,157	0.05	69,649	0.83	892,169	2.7
Fontibon	1,899	0.02	31,387	0.37	444,951	3.1
Kennedy	3,691	0.04	81,268	0.97	1,273,390	2.4
La Candelaria	165	0.00	2,937	0.04	21,830	2.4
Martires	346	0.00	8236	0.10	92,234	2.9
Puente Aranda	1,339	0.02	25,624	0.31	211,802	3.0
Rafael Uribe Uribe	1,128	0.01	29,841	0.36	341,886	2.3
San Cristobal	1,198	0.01	30,055	0.36	387,560	2.1
Santa Fe	547	0.01	10,758	0.13	91,111	2.2
Suba	7,690	0.09	91,344	1.09	1,381,597	2.8
Teusaquillo	1,552	0.02	13,990	0.17	139,369	3.9
Tunjuelito	550	0.01	15,278	0.18	183,067	2.3
Usaquen	5,751	0.07	43,175	0.52	476,931	3.8
Usme	622	0.01	22,536	0.27	348,332	1.5

Note: Sumapaz was excluded from all estimations; *Percentage of official population taken from the Colombian National Statistical System (DANE in Spanish)

Table SI.6. Population by Strata

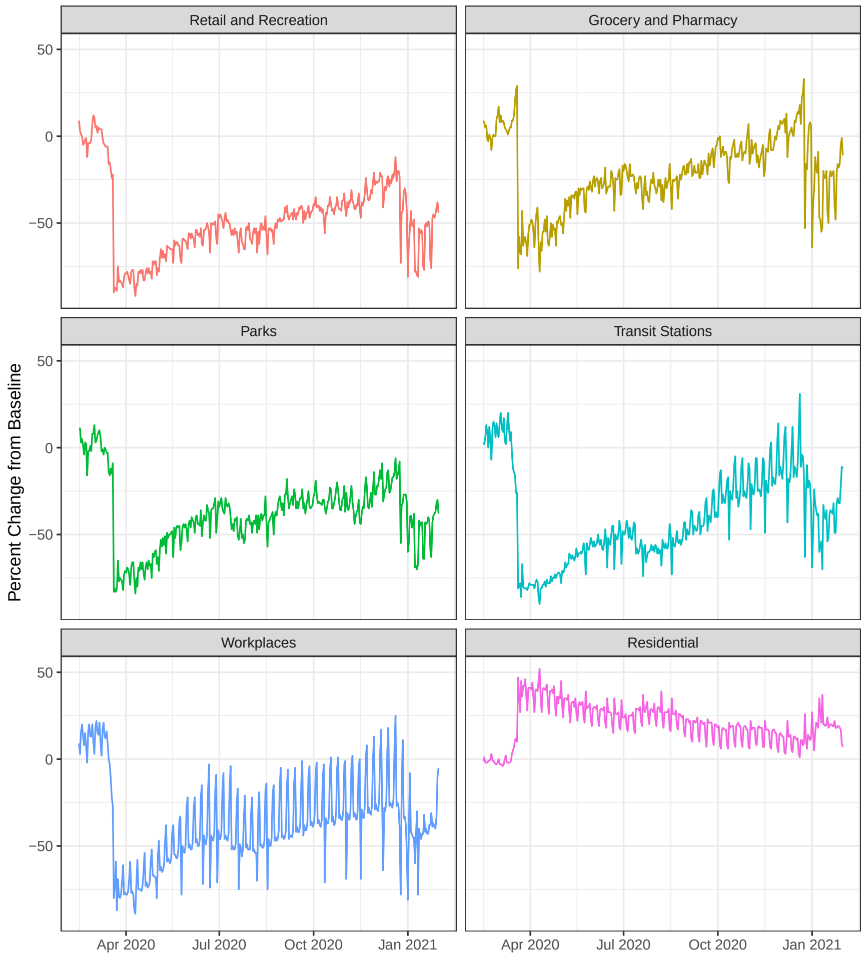
Stratum	CoVIDA Data Population	CoVIDA Data % *	HSB Data Population	HSB Data %*	Official Population
1 & 2	10,986	0.14	356,899	4.44	4,063,470
3	16,568	0.21	220,324	2.74	2,857,861
4	8,646	0.11	46,431	0.58	757,923
5 & 6	4,699	0.06	23,686	0.29	365,459

Note: *Percentage of official population taken from the Colombian National Statistical System (DANE in Spanish)

SI.3 Supplementary Figures

SI.3.1 Mobility Changes in Bogotá

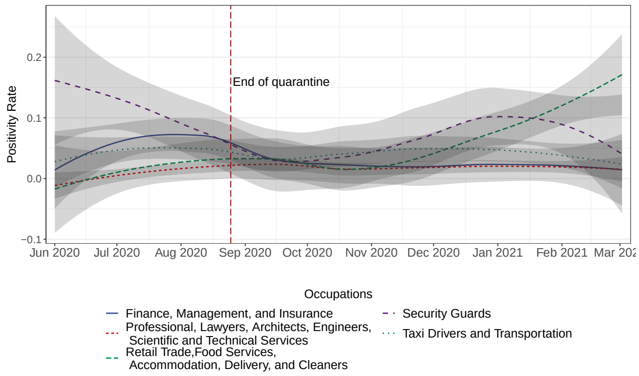
Figure SI.6



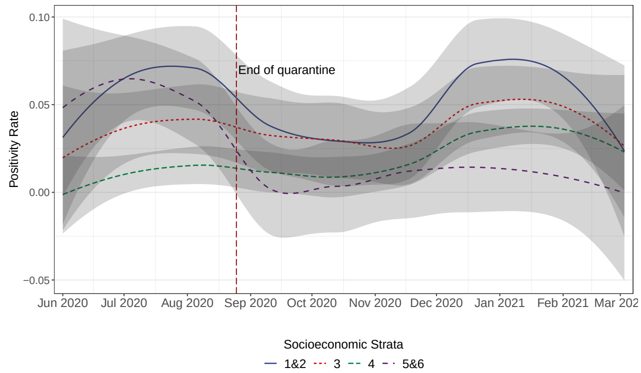
Mobility Changes by Location. Data comes from Google's COVID-19 Community Mobility. Report Baseline is the median value, for the corresponding day of the week, during the 5- week period Jan 3–Feb 6, 2020

SI.3.2 Dynamics by Occupations, Strata, and Geography

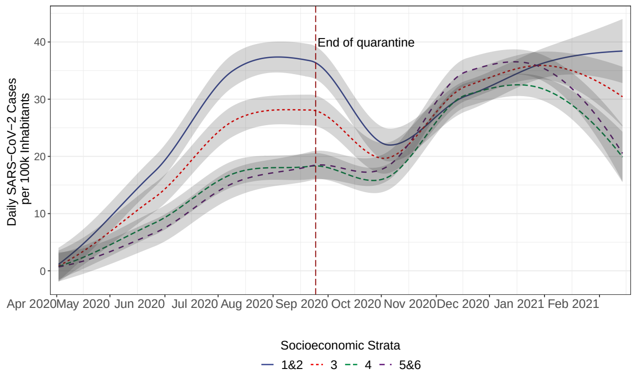
SI.3.2.1 Daily Occupations and Strata Dynamics



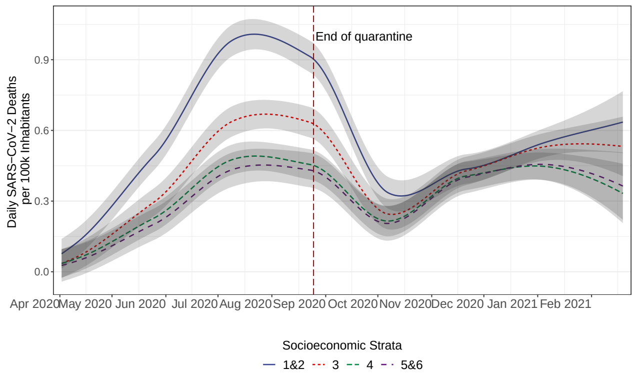
(a) Positivity by Occupation (CoVIDA data)



(b) Positivity by socio-economic Stratum (CoVIDA data)



(c) Daily new cases by socio-economic Stratum (HSB data)



(d) Daily deaths by socio-economic Stratum (HSB data)

Figure SI.7. Daily Dynamics. The Figure complements Figure SI.7 by adding 95% confidence intervals in gray. It shows smoothed SARS-CoV-2 daily positivy rates (panels (a) and (b)) from CoVIDA data and smoothed SARS-CoV-2 daily cases per 100,000 inhabitants (panels (c) and (d)) from the Health Secretary of Bogotá (HSB). Positivity rates were calculated using worker's occupation weights. Daily rates and cases were smoothed using a local polynomial regression (loess) with a smoothing parameter of 0.7. Population by workers category was obtained from a detail review of official records in several sources.

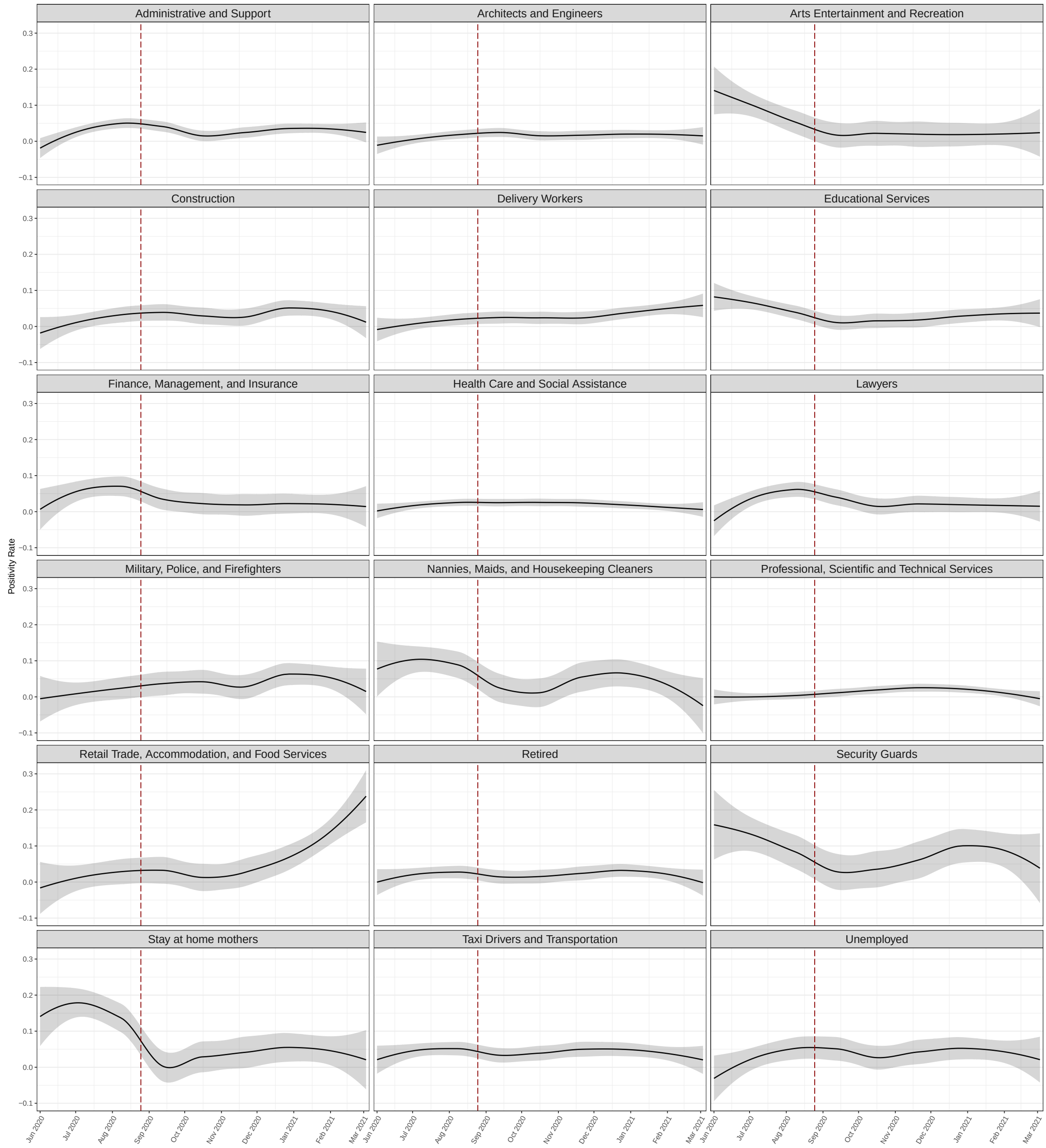
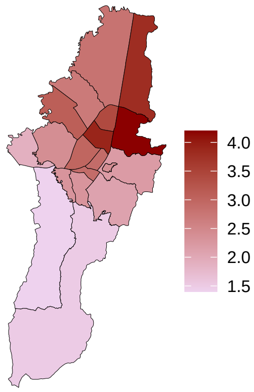
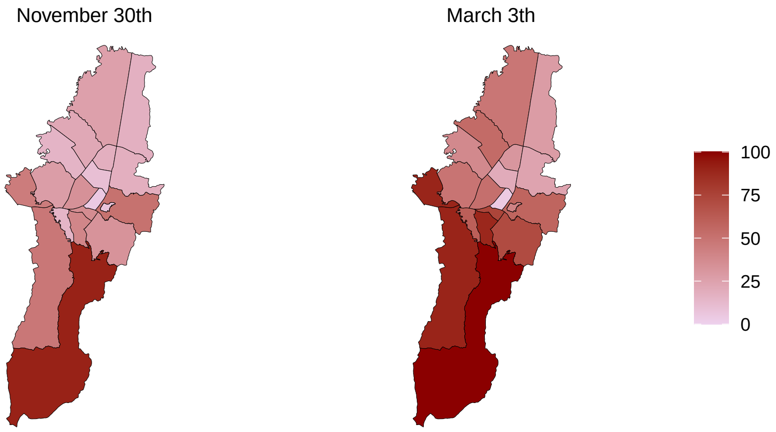


Figure SI.8. Daily Dynamics by Occupation. The Figure desegregates the occupational groups shown in 3a. It shows for each occupations described in SI.3 smoothed SARS-CoV-2 daily positivy rates. Daily rates were smoothed using a local polynomial regression (loess) with a smoothing parameter of 0.7. The vertical dashed line marks the end of quarantine on August 25, 2020. Shaded regions denote 95% confidence intervals in gray



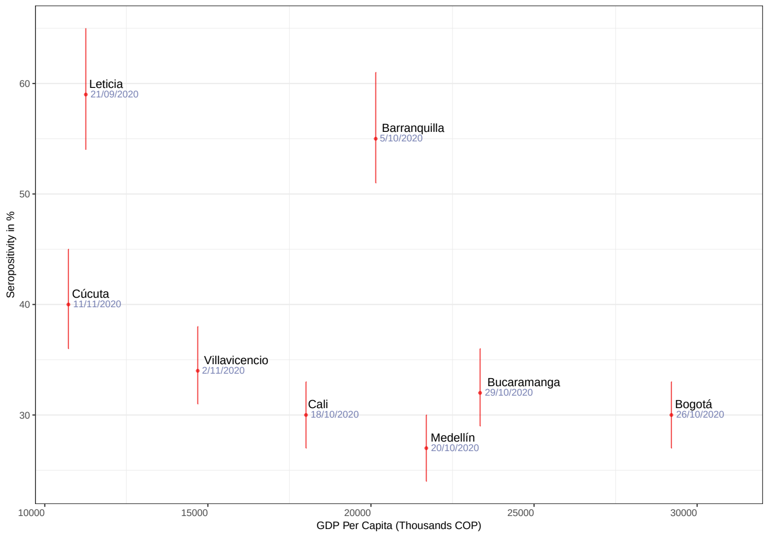
(a) Mean Socio-economic Strata



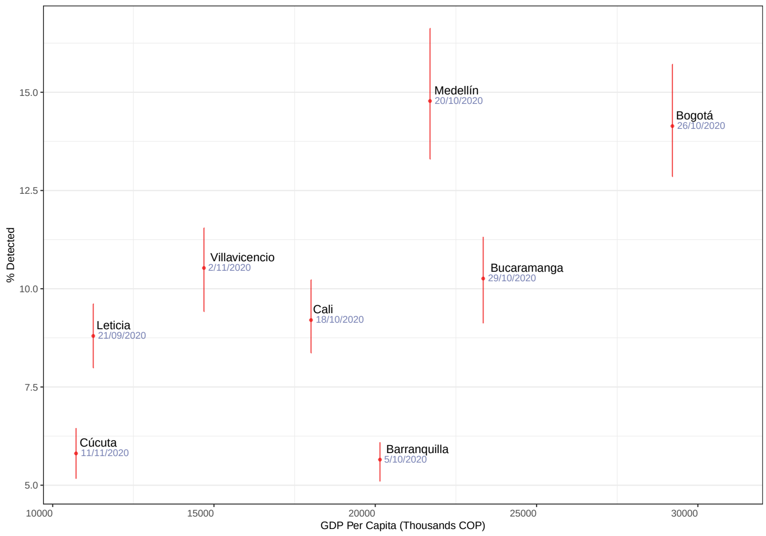
(b) CoVIDA data Accumulated Cases

Figure SI.9. Dynamics by Locality Strata. This figure shows the accumulated SARS-COV-2 cases in every locality of Bogotá using CoVIDA data. It was calculated using a weighted average positivity for the whole period. Weights were based on worker's occupation. Population by workers category was obtained from a detail review of official records in several sources. Mean stratum was calculated based on official data from Bogotá's mayor office, it shows the average socio economic stratum by locality using the population of each stratum living in the locality.

SI.3.2.2 SARS-CoV-2 across Colombia



(a) Seropositivity



(b) Detection Rate

Figure SI.10. SARS-CoV-2 Seropositivity and Detection Rate across Colombia. This figure shows the detection rate in other cities of Colombia using official data from the National Health Institute of Colombia (NHI). Calculations ere performed using the adjusted seropositivity estimates from NHI. For the case of Bogotá we adjusted the total number of cases to those reported to us by the Health Secretary of Bogotá (HSB).