CEDE
Centro de Estudios sobre Desarrollo Económico

Understanding how socioeconomic inequalities drive inequalities in SARS-CoV-2 infections

Documento CEDE

24

Rachid Laajaj, Duncan Webb,
Danilo Aristizabal, Eduardo Behrentz,
Raquel Bernal, Giancarlo Buitrago,
Zulma Cucunubá, Fernando de la Hoz,
Alejandro Gaviria, Luis Jorge Hernández,
Camilo De Los Rios, Andrea Ramírez Varela,
Silvia Restrepo, Norbert Schady,
Martha Vives

Abril de 2021

Universidad de los Andes Colombia
Facultad de Economía

Serie Documentos Cede, 2021-24 ISSN 1657-7191
Edición electrónica. Abril de 2021

infocede@uniandes.edu.co
http://economia.uniandes.edu.co

Impreso en Colombia – Printed in Colombia

Universidad de los Andes | Vigilada Mineducación Reconocimiento como Universidad: Decreto 1297 del 30 de mayo de 1964. Reconocimiento personería jurídica: Resolución 28 del 23 de febrero de 1949 Minjusticia.

CEDE
Centro de Estudios sobre Desarrollo Económico

Documento CEDE

Los documentos CEDE son producto de las investigaciones realizadas por al menos un profesor de planta de la Facultad de Economía o sus investigadores formalmente asociados.

Universidad de los Andes Colombia | Facultad de Economía

Understanding how socioeconomic inequalities drive inequalities in SARS-CoV-2 infections

Rachid Laajaj[†], Duncan Webb[†], Danilo Aristizabal, Eduardo Behrentz, Raquel Bernal, Giancarlo Buitrago, Zulma Cucunubá, Fernando de la Hoz, Alejandro Gaviria, Luis Jorge Hernández, Camilo De Los Rios, Andrea Ramírez Varela, Silvia Restrepo, Norbert Schady, Martha Vives[∓*]

[†]These authors contributed equally to this work.
[∓]On behalf of the CoVIDA working group.

*The authors acknowledge generous support from the Interamerican Development Bank, the Development Bank of Latin America (CAF), Universidad de Los Andes and 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, Paola Betancourt, Yessica Campaz, Marylin Hidalgo, Philip Keefer, Leonardo Leon, Diane Moyano, Juliana Quintero, Pablo Rodríguez, Diana Sofia Rios Oliveros, Andrés Felipe Patiño, Jose David Pinzon Ortiz, Elkin Osorio, Maribel Rincón, Rodrigo Rodriguez, Alejandro Segura, Leonardo Salas Zapata. Laajaj:Universidad de Los Andes, Bogotá, Colombia, r.laajaj@uniandes.edu.co; Webb, Paris School of Economics, Paris, France, duncan.webb@psemail.eu; Aristizabal: Universidad de Los Andes, Bogotá, Colombia, de.aristizabal411@uniandes.edu.co; Behrentz: Universidad de Los Andes, Bogotá, Colombia, ebehrent@uniandes.edu.co; Bernal: Universidad de Los Andes, Bogotá, Colombia, rbernal@uniandes.edu.co; Buitrago, Universidad Nacional de Colombia y Hospital Universitario Nacional de Colombia, Bogotá, Colombia, gbuitragog@unal.edu.co; Cucunubá, Imperial College London, London, UK, y Universidad Pontificia Javeriana, Bogota, Colombia zulma.cucunuba@imperial.ac.uk; de la Hoz, Universidad Nacional de Colombia, Bogotá, Colombia, fpdelahozr@unal.edu.co; Gaviria: Universidad de Los Andes, Bogotá, Colombia, agaviria@uniandes.edu.co; Hernández: Universidad de Los Andes, Bogotá, Colombia, luishern@uniandes.edu.co; ; De Los Rios: Inter-American Development Bank, Washington D.C, United States, cdelosriosru@gmail.com; Ramírez Varela: Universidad de Los Andes, Bogotá, Colombia, an-rami2@uniandes.edu.co; Restrepo: Universidad de Los Andes, Bogotá, Colombia, srestrep@uniandes.edu.co; Schady, Inter-American Development Bank, Washington D.C, United States, norberts@iadb.org; Vives: Universidad de Los Andes, Bogotá, Colombia, mvives@uniandes.edu.co. Author contributions: R.L, and D.W conceived the study; R.L., and D.W. analyzed the data; R.L., E.B., R.B., G.B., A.G., L.J.H., A.R.V., S.R. and M.V. implemented the CoVIDA project, including tests and primary data collection; D.A. and C.DLR. cured the data; R.L. and D.W. wrote the paper; all authors critically revised the various versions of the manuscript.

Abstract

Across the world, the SARS-CoV-2 (COVID-19) pandemic has disproportionately affected economically disadvantaged groups. This differential impact has numerous possible explanations, each with significantly different policy implications. We examine, for the first time in a low- or middle-income country, which mechanisms best explain the disproportionate impact of the virus on the poor. Combining an epidemiological model with rich data from Bogotá, Colombia, we show that total infections and inequalities in infections are largely driven by inequalities in the inability to work remotely and in within-home secondary attack rates. Inequalities in isolation behavior are less important but non-negligible, while access to testing and contract-tracing plays practically no role. Interventions that mitigate transmission are often more effective when targeted on socioeconomically disadvantaged groups.

Key words: COVID-19, inequality, infections, socioeconomic strata
JEL Classification: I14, I15, I18, O54

Entendiendo cómo inequidades socioeconómicas determinan desigualdad en infecciones por SARS-CoV-2

Rachid Laajaj[†], Duncan Webb[†], Danilo Aristizabal, Eduardo Behrentz, Raquel Bernal, Giancarlo Buitrago, Zulma Cucunubá, Fernando de la Hoz, Alejandro Gaviria, Luis Jorge Hernández, Camilo De Los Rios, Andrea Ramírez Varela, Silvia Restrepo, Norbert Schady, Martha Vives[∓*]

[†]Estos autores contribuyeron en igual proporción a este artículo.
[∓]En nombre del grupo de trabajo de CoVIDA.

*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, Paola Betancourt, Yessica Campaz, Marylin Hidalgo, Philip Keefer, Leonardo Leon, Diane Moyano, Juliana Quintero, Pablo Rodríguez, Diana Sofia Rios Oliveros, Andrés Felipe Patiño, Jose David Pinzon Ortiz, Elkin Osorio, Maribel Rincón, Rodrigo Rodriguez, Alejandro Segura, Leonardo Salas Zapata. Laajaj:Universidad de Los Andes, Bogotá, Colombia, r.laajaj@uniandes.edu.co; Webb, Paris School of Economics, Paris, France, dmbwebb@gmail.com; Aristizabal: Universidad de Los Andes, Bogotá, Colombia, de.aristizabal411@uniandes.edu.co; Behrentz: Universidad de Los Andes, Bogotá, Colombia, ebehrent@uniandes.edu.co; Bernal: Universidad de Los Andes, Bogotá, Colombia, rbernal@uniandes.edu.co; Buitrago, Universidad Nacional de Colombia y Hospital Universitario Nacional de Colombia, Bogotá, Colombia, gbuitragog@unal.edu.co; Cucunubá, Imperial College London, London, UK, y Universidad Pontificia Javeriana, Bogota, Colombia zulma.cucunuba@imperial.ac.uk; de la Hoz, Universidad Nacional de Colombia, Bogotá, Colombia, fpdelahozr@unal.edu.co; Gaviria: Universidad de Los Andes, Bogotá, Colombia, agaviria@uniandes.edu.co; Hernández: Universidad de Los Andes, Bogotá, Colombia, luishern@uniandes.edu.co; ; De Los Rios: Inter-American Development Bank, Washington D.C, United States, cdelosriosru@gmail.com; Ramírez Varela: Universidad de Los Andes, Bogotá, Colombia, an-rami2@uniandes.edu.co; Restrepo: Universidad de Los Andes, Bogotá, Colombia, srestrep@uniandes.edu.co; Schady, Inter-American Development Bank, Washington D.C, United States, norberts@iadb.org; Vives: Universidad de Los Andes, Bogotá, Colombia, mvives@uniandes.edu.co. Contribuciones de autores: R.L.,y D.W. concibieron el estudio; R.L., Y D.W. analizaron los datos; R.L, E.B., R.B., G.B., A.G., L.J.H., A.R.V., S.R. y M.V. implementaron el proyecto CoVIDA, incluyendo tests y recolección de datos primarios; D.A y C.DLR. limpiaron los datos; R.L. y D.W. escribieron el artículo; todos los autores hicieron una revisión crítica de las diferentes versiones del artículo.

Resumen

En todo el mundo, la pandemia de SARS-CoV-2 (COVID-19) ha afectado de forma desproporcionada a los grupos económicamente desfavorecidos. Este impacto diferencial tiene numerosas explicaciones posibles, cada una con implicaciones políticas significativamente diferentes. Examinamos, por primera vez en un país de ingresos bajos o medios, qué mecanismos explican mejor el impacto desproporcionado del virus en los pobres. Combinando un modelo epidemiológico con ricos datos de Bogotá, Colombia, mostramos que las infecciones totales y las desigualdades en las infecciones están impulsadas en gran medida por las desigualdades en la incapacidad de trabajar a distancia y en las tasas de ataques secundarios dentro del hogar. Las desigualdades en el comportamiento de aislamiento son menos importantes, pero no son insignificantes, mientras que el acceso a las pruebas y al seguimiento de los contratos no desempeña prácticamente ningún papel. Las intervenciones que mitigan la transmisión suelen ser más eficaces cuando se dirigen a grupos socioeconómicamente desfavorecidos.

Palabras clave: COVID-19, desigualdad, infecciones, estrato socioeconómico
Códigos JEL: I14, I15, I18, O54

With around 127 million confirmed cases around the world as of early April 2021, the COVID-19 pandemic has disproportionately affected disadvantaged groups. Evidence from the USA, Europe, and developing countries suggests that within each country, poor and minority groups are more likely to contract the disease Kim and Bostwick (2020), Figueroa et al. (2020), Millett et al. (2020). In Bogotá, Colombia, we estimate that individuals in the lowest socioeconomic strata (SES) are 3.7 times more likely to have been infected with COVID-19 than those in the highest strata as of March 3rd 2021. Addressing inequalities has been widely recommended to tackle the pandemic The Associated Press (2020), Stiglitz (2020). Some studies have documented factors that are likely to affect COVID-19 transmission patterns, including access to testing and contact tracing services Klinkenberg et al. (2006), Kucharski et al. (2020), Kretzschmar et al. (2020), Aleta et al. (2020), Grassly et al. (2020), Flaxman et al. (2020), Schmitt-Grohé et al. (2020), biological factors related to susceptibility and infectiousness Byrne et al. (2020), McAloon et al. (2020a), levels of exposure at work Brandily et al. (2020), Almagro et al. (2020), circumstances within the household Brandily et al. (2020), Almagro et al. (2020), Lee et al. (2020), Jing et al. (2020), lockdown and social distancing Kucharski et al. (2020), Flaxman et al. (2020), Hellewell et al. (2020), along with self-isolation behavior and compliance with regulations Flaxman et al. (2020), Smith et al. (2020), Williams et al. (2020). These factors are likely to differ by socioeconomic status, thereby driving inequality in COVID-19 infection rates across socioeconomic groups Brandily et al. (2020), Schmitt-Grohé et al. (2020), Rocha et al. (2021). Optimal policy design will vary significantly depending on which of these factors is key: targeted policies that focus on high-risk groups will reduce both inequalities and overall transmission more effectively if they concentrate on the most important dimensions of inequality. However, no studies have so far been able to examine which factors drive socioeconomic inequality in COVID-19 infections and compare how important each one is for explaining overall inequality. We thereby provide the first study of a low- or

middle-income country (LMIC) that (i) estimates the differences between socioeconomic groups in characteristics that can explain inequality in COVID-19 infections, and then (ii) incorporates these differences into an epidemiological model to tease out their impact on the spread of the pandemic.

1 Differences in characteristics between socioeconomic strata

We use primary data from the CoVIDA project led by the University of Los Andes, which includes the results of 59,770 RT-PCR tests in Bogotá, targeted on a mostly asymptomatic adult population from the beginning of June 2020 to March 3rd, 2021. We combine this with administrative data from the Health Secretary of Bogotá (HSB) that covers all reported cases in Bogotá (Supplementary Materials SI.1.1 includes a detailed data description). Both datasets include information on individuals' *socioeconomic stratum*, a classification commonly used as a proxy for economic welfare in Colombia. We use this six-level measure to create four SES groups for analysis, ranging from poorest to richest: 1&2, 3, 4, and 5&6.

Together, these data allow us to estimate a set of characteristics that are likely to determine infection rates, and to do so separately for each of the four socioeconomic groups. Table 1 displays the values of these characteristics for each group (and Table SI.6 includes standard errors, 95% CIs, source and estimation method). Throughout the paper, we classify the determinants of inequalities in infection into 4 dimensions: (i) contacts outside of the household, (ii) contacts within the household, (iii) isolation behavior, and (iv) testing and tracing.

The number of non-work-related contacts outside the home does not differ significantly across SES ($p = 0.06$). Secondary attack rate (SAR) for contacts outside the home also ex-

hibits no significant differences across SES ($p = 0.20$), with the overall average is estimated at 13%. This result is consistent with our finding that self-declared protection practices are not systematically better among higher SES; lower SES even appear to compensate for their inability to remain at home by wearing masks and using antibacterials more frequently (Table SI.1). By contrast, there is a large and significant difference in the number of days working outside of home during the 14 days prior to the survey, which varies from 2.5 days for SES 5&6 to 6.4 days for SES 1&2 ($p < 0.001$). This substantial difference is likely to reflect the well-documented variation in the ability to work remotely Dingel and Neiman (2020).

Characteristics related to infections inside of homes also reveal differences. First, mean household size shows modest variation, from approximately 2.5 in wealthier households to 3 in poorer households ($p < 0.001$ for the difference). Lower SES individuals therefore have more contacts within the household, which is known to be a particularly important setting for transmission Lee et al. (2020). There is a substantial difference in the SAR within household ($p = 0.02$), ranging from only 10% in SES 5&6 to around 27% in SES 1–3. Corroborating this result, the positive correlation between household size and infection probability is stronger for lower SES (Figure SI.9b), which may partly be explained by more crowded housing conditions, since the poor have fewer rooms per household (Figure SI.9c).

Self-reported isolation for individuals who have been tested positive is high (86%) and does not vary significantly by SES ($p = 0.61$). Other high-risk circumstances that require isolation, such as experiencing symptoms, lead to a substantial reduction in days worked outside of home for all groups. But richer groups are able to restrict their working activity in these circumstances significantly more than poorer groups.

Finally, access to testing and tracing could also affect infections if it leads to effective quarantine and isolation. We find differences in testing and tracing characteristics across

SES that may plausibly be explained by differing levels in health service quality that are correlated with income Cifuentes et al. (2021). The likelihood of being detected conditional on being infected varies substantially, from 15% in the SES 1&2 to almost double this in SES 4–6. There are moderate differences in the average delay in test consultations and results, which sum to 8.3 days for SES 5&6 and 9.5 days for SES 1&2. But average delays across *all* groups are very long. They clearly exceed recommendations for an effective Test, Trace and Isolate strategy, which suggest delays of no more than 5 days from onset of symptoms to the results of the test Kretzschmar et al. (2020).

2 A theoretical model that emphasizes differences between socioeconomic groups

In order to quantify how the differences shown in Table 1 translate into differences in COVID-19 infection patterns, we use the results as inputs for a novel branching-process model of the spread of SARS-CoV-2. (See Tables SI.4 and SI.5 for a complete list and description of the model parameters.)

The model is stochastic and individual-based, building on recent modelling work of the pandemic Kucharski et al. (2020), Kretzschmar et al. (2020), Hellewell et al. (2020). We structure the model by SES, allowing all parameter values found in Table 1 that are significantly different at the 5% level to be SES-specific.

An example of a transmission process is shown in Figure 1. It demonstrates some of the realistic features of the model, which include a distinction between contacts within and outside the home, assortative mixing, symptoms, testing, contact tracing, isolation, immunity, along with realistic distributions for all stochastically-generated timings.

Our baseline simulation of the epidemic uses the parameters as described in Tables SI.4 and SI.5. Figure 2 shows the infection patterns in each SES, both using data from

Bogotá (panels (a) and (d)) and comparing to the results of our model in the baseline scenario. We use two variations of the model. In the first (panels (b) and (e)), the average number of out-of-home contacts for each group stays constant over the course of the epidemic, leading to a one-wave pattern. In the second (panels (c) and (f)), we account for changes in mobility over time by scaling the number of out-of-home contacts by a time-varying constant, calibrated to match total confirmed incidence (see Supplementary Materials Section SI.1.3.4). This constant is the same for all groups, implying that all predictions of inequality result from the differences in characteristics described in Section 1, rather than the calibration process.

Panel (a) displays the per capita incidence over the preceding 2 weeks for each group based on data on *confirmed* cases from the Health Secretary of Bogotá. We see evidence of inequality between groups, particularly in the first wave, where SES 1&2 reaches a peak incidence rate of 0.72%, around double the level of SES 5&6. The model predictions of confirmed cases, seen in panels (b) and (c), match this observed pattern relatively well, with more detection of cases in the lower SES. The observed pattern of detected cases lies within the range of the confidence intervals of both models for all groups in first wave (Figure SI.10).

We also show that the model predictions match estimated inequalities in *true* infections, estimated with the CoVIDA data, which is a sample of mostly asymptomatic individuals and is thus less likely to be biased due to differential propensity to be tested. Panel (d) shows that the inequality in estimated true infections is even starker than that of confirmed cases: cumulative per capita incidence (during the entire period of the study) varies from 69% in strata 1&2 to only 19% in strata 5&6. Because the model with no mobility change (panel e) only captures the first wave of the epidemic, it underestimates the cumulative incidence rate in all groups, but it yields a prediction of the proportion of total cases that come from each SES (a proxy of inequality) that matches the CoVIDA

data well (Figure SI.11). When accounting for mobility change (panel f), the model gives very similar predictions to the CoVIDA data estimations, despite mobility being calibrated on only aggregated confirmed cases, with a difference of 41 percentage points in cumulative per capita incidence between the lowest and highest SES. Broadly, we are able to predict the macro-level differences in infections well by introducing micro-level inequality across SES into our epidemiological model.

3 Effects of an upward adjustment of inequalities in the various dimensions on incidence of the virus

In order to identify the key channels that drive overall inequality in infections between SES, we examine the effects of reducing inequality along 4 dimensions: (i) contacts outside of the household, (ii) contacts within the household, (iii) isolation behavior, and (iv) testing and tracing. For each of these dimensions, we first simulate a "100% upward adjustment" scenario, in which the characteristics of all SES are set at the level of the highest SES (5&6), and then a "50% upward adjustment" scenario (Figure 3), in which the differences with respect to SES 5&6 are reduced by half.

First, we find that in a simulation in which all SES have as few contacts outside the home as SES 5&6, the epidemic collapses, with a median cumulative incidence of less than 1% across all SES (column 2). This means that if every SES had the ability to work on average only 2.5 days every 2 weeks outside of home, then the $\mathcal{R}_t$ would fall below one, leading to a containment of the virus from early stages. When we adjust by only 50%, there is still a marked reduction both in infections across all groups and in the inequality between groups. This indicates that the differences in out-of-home contacts are a key driver of the inequality in COVID-19 infections between groups.

Second, attributing the within-home characteristics of strata 5&6 to all strata (column

3) leads to a reduction in infections that is as strong as the effect in the out-of-home scenario. Further analysis demonstrates that this effect is mostly driven by differences in the within-household SAR, while inequality in household size plays a significant but smaller role (Figure SI.9). There may thus be large benefits to policies that reduce within-home transmission for groups with crowded housing conditions, for example through recommending mask usage and social distancing within the home.

Third, we consider scenarios in which lower SES are just as able to isolate as SES 5&6 in high-risk circumstances (being in contact with or in the same household as a known case, or when presenting symptoms) (column 4). This leads to moderate reductions in infections and inequalities: for example, cumulative incidence among SES 1&2 would be reduced by 8 percentage points in the 100% scenario. Differences in isolation behavior are thus important, but contribute less to inequality in infections than the two previous channels. Reducing inequalities in isolation may nevertheless be more tractable for policy than changing characteristics like housing conditions or job-types. For example, using financial compensation to enable lower-income individuals to stay at home when symptomatic may be an effective strategy.

Finally, the effect of improving access to testing and contact tracing among low SES to the level of SES 5&6 has an effect on infections that is not significantly different from 0 (column 5). This is true despite the substantial inequality in access (see in Table 1). The absence of effect can be explained by the fact that, on average in Bogotá, delays in accessing testing, receiving results, and being contact traced are so severe *across all groups* that testing and tracing has little effect at all in mitigating the spread of the virus in this context.

The aim of these counterfactual scenarios is to build a quantitative understanding of which channels of inequality are key for explaining inequality in COVID-19 infections. However, the scenarios are not necessarily realistic. This leads to, for example, 100%

upward-adjustment scenarios that imply rather extreme results. This is partly because the model starts in the conditions of generalised lockdown in Bogotá, with an estimated R_q number of 1.22 in the baseline scenario. Thus, any scenario that is sufficient to reduce this R_q number to below 1 will lead to a total containment of the epidemic. In reality, mobility restrictions may have been loosened sooner if this was the case, leading to more infections than our model predicts. Such policy reactions, along with other relevant factors, such as continued imports of the virus, are not included in our model. Another factor making these scenarios more informative but less realistic is the fact that in our main results we use the baseline model with no mobility change over time and a "one-wave" epidemic. The model that allows for mobility change generates two epidemic waves, and while the upward-adjustment scenarios lead to similar reductions in inequalities and infections during the first wave, these reductions are offset by a larger second wave because immunity levels are lower when the second wave begins (see Figure SI.15). The mobility-change model is not more realistic, since it assumes, for example, that mobility would not have been restricted even in cases of extremely high incidence rates in the second wave. Nevertheless, the results provide a cautionary tale, common to many epidemiological models: measures that successfully reduce infections may only delay an epidemic, rather than eliminate it. In such a context, social inequalities may not predict who is more likely to get infected; rather they predict who gets infected first.

4 The role of inequalities, and alternative policy scenarios

To further examine the role of socioeconomic inequality, we isolate the effect of a reduction in inequalities, by reducing the dispersion of all the characteristics that were found to be significantly different across SES while preserving the mean of each variable (Figure 4).

We find that if these inequalities are fully collapsed, total infections would be reduced from 38.2% to 35.9% of the population. The effect is moderate but statistically significant ($p < 0.001$). Inequality *in and of itself* can lead to more widespread infection, even when holding constant the average characteristics of the population.

We next examine simulations of policy-style scenarios that indicate (i) what types of policies may be effective in combating the epidemic, and (ii) whether targeting policies on low socioeconomic groups can reduce the spread of the virus. Figure 5 shows the main results, while Figure SI.16 shows the results when varying the intensity of each policy. The first set of simulations describes the results of (i) 10% population immunity (e.g. due to vaccinations at an early stage), (ii) a reduction of 1 in outside-home contacts (e.g. due to restricting economic activity, or a policy facilitating or enforcing remote work), and (iii) an increase in ability to isolate (e.g. due to financial support for those required to isolate). Increasing immunity and reducing out-of-home contacts lead to large reductions in infections; increasing isolation leads to more modest effects. Even when holding constant the number of beneficiaries, reductions are between 28% and 49% larger when targeted on the lowest SES ($p < 0.001$ for all differences). This highlights that policies are likely to be substantially more effective when targeted on socioeconomically disadvantaged populations.

By contrast, increased access to testing for symptomatic individuals has only a small impact on total infections, and this impact does not differ significantly if targeted or not. Access to tests does not reduce infections in Bogotá, due to both severe delays across all SES and limited initial coverage.

In keeping with this claim, even a complete removal of testing and tracing (the "No testing" scenario) only increases the overall incidence rate by a mean of 1.0 percentage points ($p < 0.001$), while increasing the speed of testing to an average of 2 days from symptoms to detection results in a modest 4 percentage point reduction in incidence rate

from the baseline scenario. Combining this "Fast testing" scenario with an increase in the probability of being tested leads to more substantial reductions in transmission, with the overall incidence rate reducing to below 30%. Targeting the improvements in self-testing on SES 1&2 leads to an additional mean reduction of 2.7 percentage points compared to the non-targeted scenario (p-value of the difference= 0.04).

5 Discussion

This paper documents differences across SES along multiple dimensions that are relevant to the spread of COVID-19. We provide one of the first studies that estimates the relative importance of each form of inequality for explaining disparities in COVID-19 infections. Such estimates are important to improve policy design. Disparities in types of jobs and the ability to work from home are shown to be a key factor behind inequalities. Additional channels that have seen less emphasis in existing research include the major role of differences in within-home SAR and inequalities in the ability to isolate when required (in particular when one has symptoms, a detected individual in the household, or a recent contact with a person tested positive). Finally, while poor individuals do have substantially less access to tests, and a lower chance of being detected and traced, this does not drive inequalities in incidence rates because testing and tracing in Bogotá is too slow to contribute to the mitigation of the virus.

We find that improving the conditions of lower SES translates to a reduction in incidence rate that is almost proportional across all groups. Despite some degree of assortative mixing, in which each group is somewhat more likely to contact their own group, even the highest SES benefit from improvements of the conditions faced by the lowest SES. This supports the message that improving conditions for disadvantaged groups is necessary to tackle the pandemic and would benefit the entire population.

Even when maintaining the same mean characteristics across the whole population, a reduction in inequality can reduce the overall spread of the virus. As a consequence, policy measures are likely to be more effective if they target disadvantaged socioeconomic groups that are typically at higher risk. This would include targeting vaccine roll-outs on these groups, but also targeting other non-pharmaceutical interventions. Our results suggest placing a particular emphasis on (i) maximising the ability to work remotely for lower socioeconomic groups where possible, or temporary and targeted economic shutdown measures in the absence of other alternatives, and (ii) raising awareness that within-home infections are a major source of transmission, but that this transmission may be avoidable, and may be mitigated using within-house preventative measures such as mask use. Immediate financial compensation for individuals required to isolate, including close contacts, housemates of infected persons, and anyone with COVID-19-related symptoms, may also be an important and tractable policy lever. Our results call for a rethinking of testing, tracing, and isolation strategies in developing country contexts: although testing systems provide valuable information about the spread of the virus, they may be so slow that they have little mitigating effect on the transmission. In these cases, policymakers must be transparent about testing delays, and consider a dramatic speed-up of the testing system where feasible.

Our findings provide new evidence on the importance of different channels that drive inequality in COVID-19 infections, and show that improving the circumstances of the most disadvantaged groups, including by targeting interventions on the poor, can have benefits for all. Inequalities must be addressed in order to better handle both the COVID-19 crisis and potential future epidemics.

6 Figures

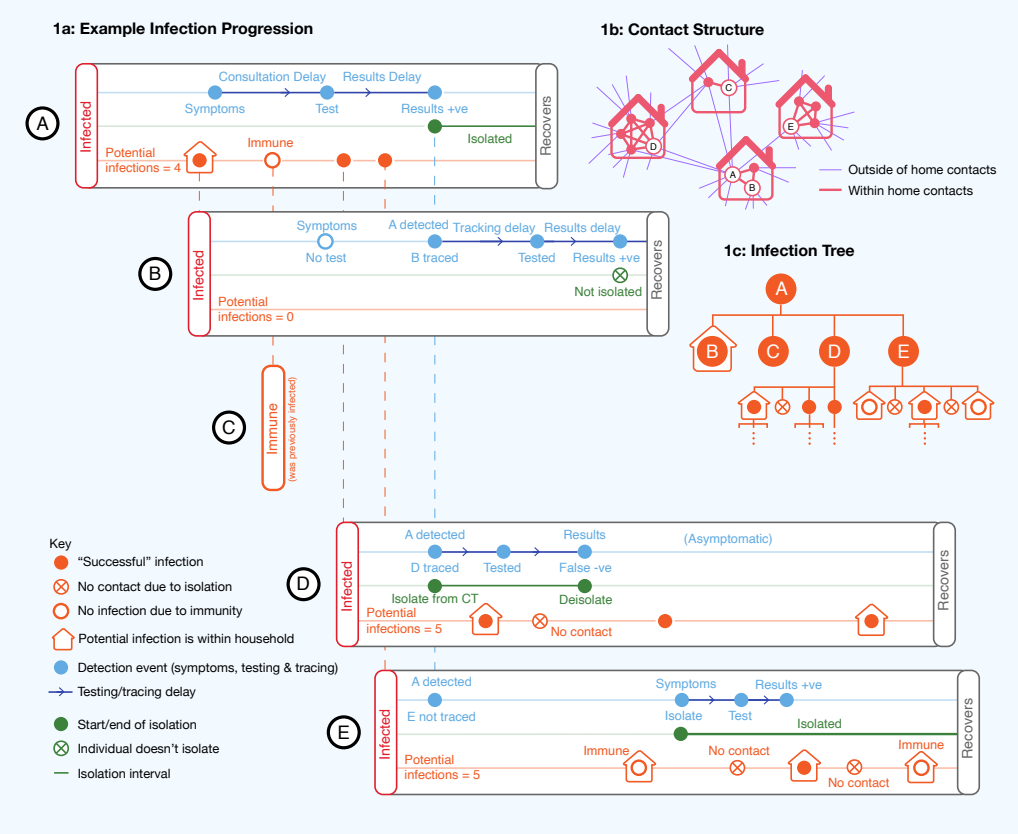


Figure 1. **Visual representation of the theoretical model**. An initial infection A potentially infects four other individuals, called B, C, D, and E. **1a:** A successfully infects B, D, and E. C does not get infected by A because she has already been infected previously, and is thus immune to further infections. A gets tested upon experiencing symptoms, and isolates upon receiving a positive test result. This begins a process of contact tracing, through which B and D (but not E) are tested. Individuals in the model may or may not be symptomatic, get tested, be contact traced, and they may isolate for a variety of reasons. **1b:** Each individual has contacts both within the home and outside the home. Every individual within the home is a contact, and the number of outside-of-home contacts is drawn from a negative binomial distribution. A contact becomes a potential infection with a probability equal to the secondary attack rate, which differs for within-home and outside-of-home contacts, and by SES. **1c:** The infection tree summarises the "branching process" in the model, i.e. the first and second generation potential infections caused by A.

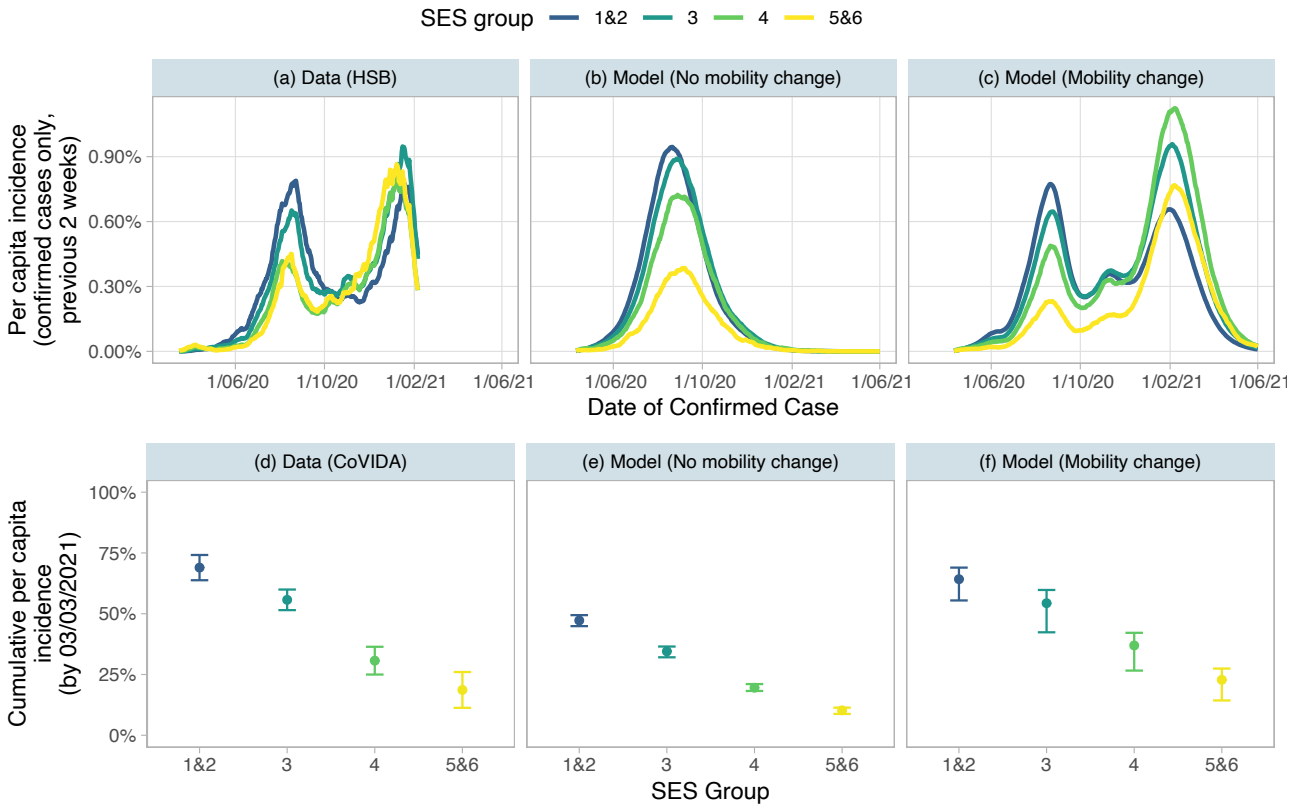


Figure 2. **Estimations of incidence rate using data and baseline simulations**. Panels (a), (b), and (c) show the per capita incidence over the previous 2 weeks based on *confirmed cases* (those who test positive) for each SES at each date. Panel (a) is based on the administrative data from the HSB on the number of confirmed cases at each date. Panel (b) is calculated using the number of infected individuals that test positive in the model simulation with no mobility change, while panel (c) uses the same calculation for the model simulation that allows for mobility to change over the course of the epidemic (in a way that best predicts total detected cases). Panels (d), (e) and (f) show the cumulative per capita incidence (including *both* confirmed and unconfirmed cases) by the 3rd March 2021 (the most recent date for which the CoVIDA data is available). Panel (d) uses positivity in CoVIDA data to calculate incidence, see Section SI.1.1. Panel (e) and (f) includes all infections in the versions of the model without and with mobility change respectively. All model results are calculated by taking the median value over 50 simulations. Model and actual dates are aligned by taking the model time period for which the model-predicted 2 week total per capita incidence is the same as the actual value on June 1st 2020, and setting this time period to be June 1st 2020.

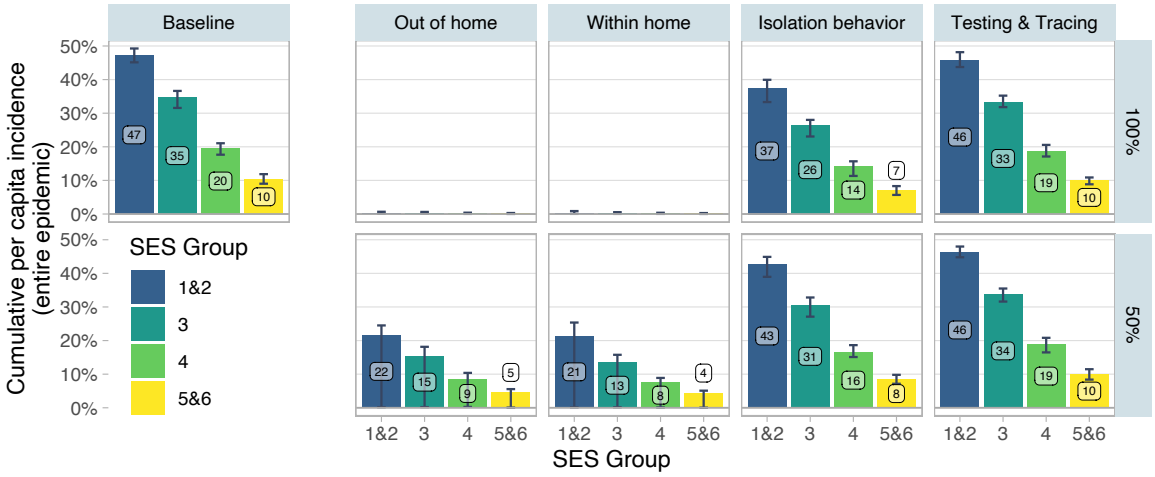


Figure 3. **Upward Adjustment Scenarios**. Baseline indicates the model with the parameters of Table 1 and no adjustment. The panels in columns 2 to 6 are the results of upward adjustment scenarios. In the top row of columns 2 to 6 (100% adjustment), the set of parameters indicated in the column heading is adjusted so that all SES have the same value as SES 5&6. In the bottom row (50% adjustment), all SES other than 5&6 have their parameters adjusted to move halfway to the value of 5&6. Parameters adjusted in each set are as follows: *out of home* (number of contacts outside the home), *within home* (within-household SAR, household size), *isolation behavior* (probability of isolating conditional on observing symptoms, testing positive, being contact traced, and probability of quarantining as a household), *testing & tracing* (probability of self testing, delay in test consultation, delay in test results, and probability of being contact traced). Point estimates denote the median of 50 simulations. Error bars indicate the 0.025 and 0.975 quantiles of the 50 simulations.

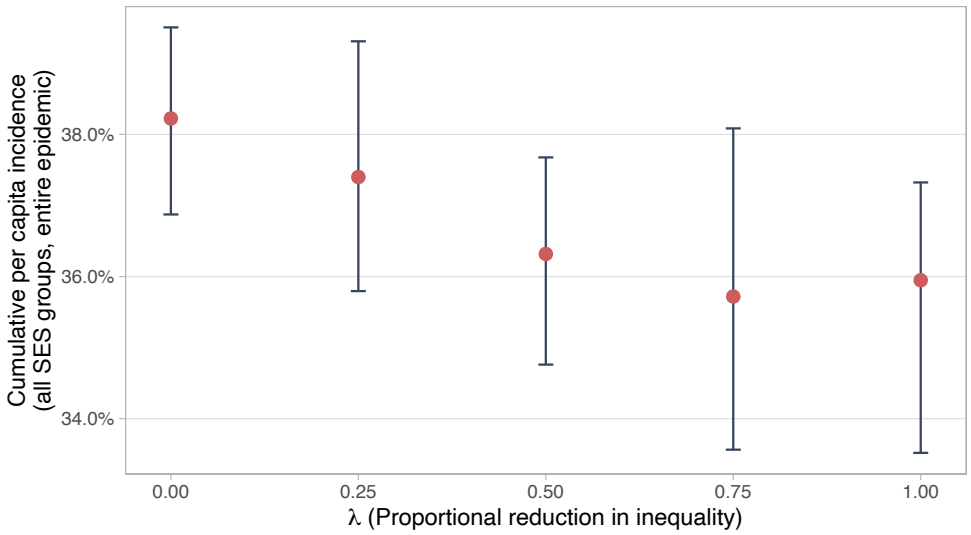


Figure 4. **Mean-preserving reduction in inequalities.** Describes the effect of reducing inequalities in all parameters simultaneously while preserving the mean of all parameters. The value of parameter k for an SES j in the baseline simulation can be written as $v_{jk} = \overline{v}_k + \varepsilon_{jk}$, where $\overline{v}_k$ is the (population weighted) mean value for the parameter across all groups, and ε_{jk} is some deviation. The graph plots the results of adjusting all parameters to the value $v^*_{jk}(\lambda) = \overline{v}_k + (1 - \lambda)\varepsilon_{jk}$. The outcome variable is the median cumulative per capita incidence across all SES over the course of the entire simulated epidemic in 50 models with no mobility change. Error bars indicate the 0.025 and 0.975 quantiles of the 50 simulations.

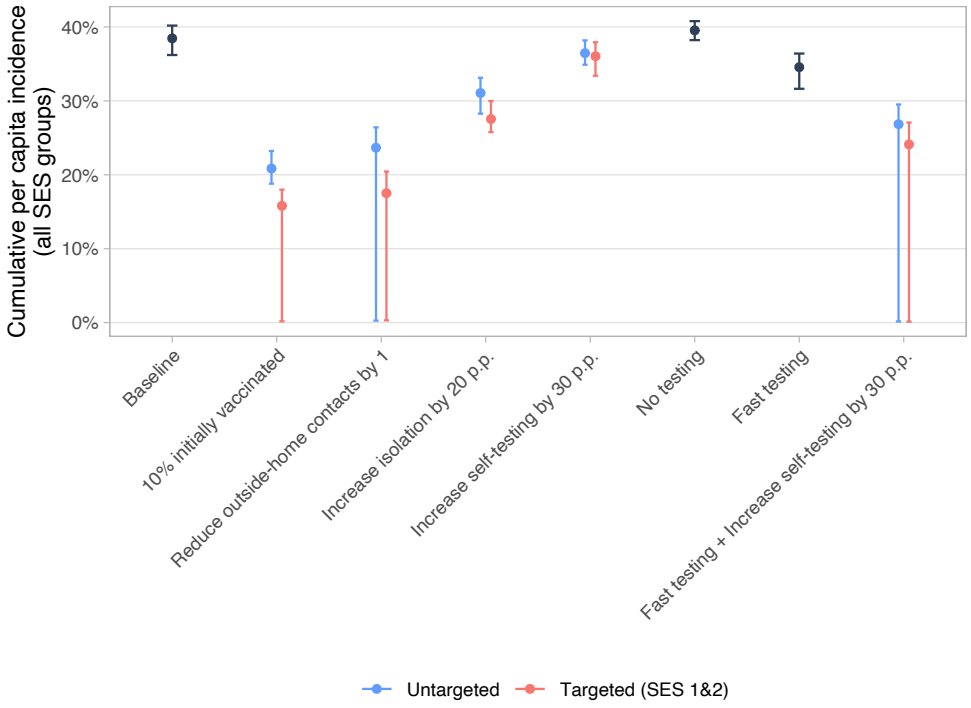


Figure 5. **Policy-style scenarios.** In "Untargeted" scenarios, policy adjustments affect all groups equally. In "Targeted" scenarios, only the parameters of SES 1&2 are adjusted, but adjustments in this group are greater, such that the mean adjustment across the whole population is the same as in the untargeted scenario. "10% initially vaccinated": 10% of the population are immune to the virus from the start of the epidemic. "Reduce outside-home contacts by 1": mean reduction of 1 in contacts outside the home. "Increase isolation by 20 p.p.": mean increase of 20 percentage points in probability of isolating conditional on being symptomatic and being contact traced. "Increase self-testing by 30 p.p.": mean increase of 30 percentage points in the probability of being tested after observing symptoms. "No testing": probability of self-testing and of being contact traced are set to 0. "Fast testing": all tests have a consultation delay and a results delay of 1 day. Outcome variable is the median cumulative per capita incidence across all SES for 50 simulated epidemics with no mobility change. Error bars indicate the 0.025 and 0.975 quantiles of the 50 simulations.

Table 1. Potential Determinants of Infection Estimated by SES

(a) All measures

Channel	Measure	SES Group 1&2	3	4	5&6	Full population	p-val, diff. between SES
Infections outside home	Days working outside home (in last 14 days)	6.4	4.8	3.2	2.5	4.6	<0.001
	Number of non-work contacts outside home (in last 14 days)	1.108	1.392	1.506	1.423	1.314	0.063
	Secondary attack rate (outside home)	15%	13%	8%	12%	13%	0.2
	Contact matrix structure	[see Panel (b)]					
Infections inside home	Household size	2.99	2.81	2.50	2.48	2.84	<0.001
	Secondary attack rate (inside home)	26%	27%	24%	11%	26%	0.02
Isolation behaviour	Isolation rate after positive test result	0.87	0.85	0.86	0.87	0.86	0.61
	# days worked when has symptoms	3.03	2.29	2.4	1.5	2.6	<0.001
	# days worked when knowing about positive contact	4.5	3.4	3.5	2.2	3.9	0.016
	# days worked when someone is tested positive in same household	2.8	2.4	2.4	1.9	2.5	0.0040
Testing & tracing	Share detected among positive	11.7%	15.2%	22.2%	21.3%	16.1%	<0.001
	Test consultation delay in days	5.56	5.59	5.41	5.26	5.55	<0.001
	Test results delay in days	3.94	3.57	3.28	3.05	3.72	<0.001
	Average number of contacts traced	1.73	1.74	1.75	1.75	1.74	
	Proportion of infections that are contact traced	81%	84%	88%	89%	83%	
	Population size in Bogota	4,063,470	2,857,861	757,923	365,459	8,044,713	
	Sample size in CoVIDA Survey Data	22,171	31,636	14,608	7,539	75,954	
	Sample size wih PCR test in CoVIDA Data	15,818	24,450	11,759	6,158	58,185	

(b) Contact matrix

	Contact stratum				
Index case stratum	1&2	3	4	5&6	Total
1&2	206	98	14	2	320
3	126	418	69	17	630
4	9	52	58	25	144
5&6	5	8	18	16	47
Total	346	576	159	60	1141

Panel (a): The table displays variables that capture various determinants of infection, sorted in four categories, followed by population and sample sizes. It provides the average value for each SES and for the population all-together. The last column presents the p-value of the F-test of difference between the 4 SES. A p-value below 0.05 means that one can reject at the 95% confidence level that the variable has population average that is equal for all SES (two-sided test). Standard deviations, Confidence Intervals, data sources and explanations of the calculation methods are presented in Table SI.6. **Panel (b)**: The contact matrix enumerates the number of cases for each possible of combination of stratum of the index case and its contacts. Positive cases in the CoVIDA study were traced, from this data, We use the self-declared stratum of the index cases and all their non-household contacts to count the number of contacts within each cell.

References

Aleta, A., Martin-Corral, D., y Piontti, A. P., Ajelli, M., Litvinova, M., Chinazzi, M., Dean, N. E., Halloran, M. E., Longini Jr, I. M., Merler, S., et al. (2020). Modelling the impact of testing, contact tracing and household quarantine on second waves of covid-19. *Nature Human Behaviour*, 4(9):964–971.

Almagro, M., Coven, J., Gupta, A., and Orane-Hutchinson, A. (2020). Racial disparities in frontline workers and housing crowding during covid-19: Evidence from geolocation data. *Available at SSRN 3695249*.

Azzalini, A. and work(s):, A. D. V. R. (1996). The Multivariate Skew-Normal Distribution. *Biometrika*, 83(4):715–726.

Bi, Q., Wu, Y., Mei, S., Ye, C., Zou, X., Zhang, Z., Liu, X., Wei, L., Truelove, S. A., Zhang, T., Gao, W., Cheng, C., Tang, X., Wu, X., Wu, Y., Sun, B., Huang, S., Sun, Y., Zhang, J., Ma, T., Lessler, J., and Feng, T. (2020). Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: A retrospective cohort study. *The Lancet Infectious Diseases*, 20(8):911–919.

Brandily, P., Brébion, C., Briole, S., and Khoury, L. (2020). A poorly understood disease? the unequal distribution of excess mortality due to covid-19 across french municipalities. *NHH Dept. of Economics Discussion Paper*, (15).

Buitrago-Garcia, D., Egli-Gany, D., Counotte, M. J., Hossmann, S., Imeri, H., Ipekci, A. M., Salanti, G., and Low, N. (2020). Occurrence and transmission potential of asymptomatic and presymptomatic SARS-CoV-2 infections: A living systematic review and meta-analysis. *PLOS Medicine*, 17(9):e1003346.

Byrne, A. W., McEvoy, D., Collins, A. B., Hunt, K., Casey, M., Barber, A., Butler, F., Griffin, J., Lane, E. A., McAloon, C., et al. (2020). Inferred duration of infectious period of sars-cov-2: rapid scoping review and analysis of available evidence for asymptomatic and symptomatic covid-19 cases. *BMJ open*, 10(8):e039856.

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*.

Cifuentes, M. P., Rodriguez-Villamizar, L. A., Rojas-Botero, M. L., Alvarez-Moreno, C. A., and Fernández-Niño, J. A. (2021). Socioeconomic inequalities associated with mortality for covid-19 in colombia: a cohort nationwide study. *J Epidemiol Community Health*.

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

Figueroa, J. F., Wadhera, R. K., Lee, D., Yeh, R. W., and Sommers, B. D. (2020). Community-Level Factors Associated With Racial And Ethnic Disparities In COVID-19 Rates In Massachusetts: Study examines community-level factors associated with racial and ethnic disparities in COVID-19 rates in Massachusetts. *Health Affairs*, page 10.1377/hlthaff.

Flaxman, S., Mishra, S., Gandy, A., Unwin, H. J. T., Mellan, T. A., Coupland, H., Whittaker, C., Zhu, H., Berah, T., Eaton, J. W., et al. (2020). Estimating the effects of non-pharmaceutical interventions on covid-19 in europe. *Nature*, 584(7820):257–261.

Grassly, N. C., Pons-Salort, M., Parker, E. P. K., White, P. J., Ferguson, N. M., Ainslie, K., Baguelin, M., Bhatt, S., Boonyasiri, A., Brazeau, N., Cattarino, L., Coupland, H., Cucunuba, Z., Cuomo-Dannenburg, G., Dighe, A., Donnelly, C., van Elsland, S. L., FitzJohn, R., Flaxman, S., Fraser, K., Gaythorpe, K., Green, W., Hamlet, A., Hinsley,

W., Imai, N., Knock, E., Laydon, D., Mellan, T., Mishra, S., Nedjati-Gilani, G., Nouvellet, P., Okell, L., Ragonnet-Cronin, M., Thompson, H. A., Unwin, H. J. T., Vollmer, M., Volz, E., Walters, C., Wang, Y., Watson, O. J., Whittaker, C., Whittles, L., and Xi, X. (2020). Comparison of molecular testing strategies for COVID-19 control: A mathematical modelling study. *The Lancet Infectious Diseases*, page S1473309920306307.

He, X., Lau, E. H. Y., Wu, P., Deng, X., Wang, J., Hao, X., Lau, Y. C., Wong, J. Y., Guan, Y., Tan, X., Mo, X., Chen, Y., Liao, B., Chen, W., Hu, F., Zhang, Q., Zhong, M., Wu, Y., Zhao, L., Zhang, F., Cowling, B. J., Li, F., and Leung, G. M. (2020). Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nature Medicine*, 26(5):672–675.

Hellewell, J., Abbott, S., Gimma, A., Bosse, N. I., Jarvis, C. I., Russell, T. W., Munday, J. D., Kucharski, A. J., Edmunds, W. J., Sun, F., et al. (2020). Feasibility of controlling covid-19 outbreaks by isolation of cases and contacts. *The Lancet Global Health*, 8(4):e488–e496.

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.

Jing, Q.-L., Liu, M.-J., Zhang, Z.-B., Fang, L.-Q., Yuan, J., Zhang, A.-R., Dean, N. E., Luo, L., Ma, M.-M., Longini, I., et al. (2020). Household secondary attack rate of covid-19 and associated determinants in guangzhou, china: a retrospective cohort study. *The Lancet Infectious Diseases*, 20(10):1141–1150.

Johnson, S. G. (2020). The nlopt nonlinear-optimization package, v1.2.2.2.

Kim, S. J. and Bostwick, W. (2020). Social Vulnerability and Racial Inequality in COVID-19 Deaths in Chicago. *Health Education & Behavior*, 47(4):509–513.

Klinkenberg, D., Fraser, C., and Heesterbeek, H. (2006). The effectiveness of contact tracing in emerging epidemics. *PloS ONE*, 1(1):e12.

Kretzschmar, M. E., Rozhnova, G., Bootsma, M. C., van Boven, M., van de Wijgert, J. H., and Bonten, M. J. (2020). Impact of delays on effectiveness of contact tracing strategies for covid-19: a modelling study. *The Lancet Public Health*, 5(8):e452–e459.

Kucharski, A. J., Klepac, P., Conlan, A. J., Kissler, S. M., Tang, M. L., Fry, H., Gog, J. R., Edmunds, W. J., Emery, J. C., Medley, G., et al. (2020). Effectiveness of isolation, testing, contact tracing, and physical distancing on reducing transmission of sars-cov-2 in different settings: a mathematical modelling study. *The Lancet Infectious Diseases*, 20(10):1151–1160.

Laajaj, R., De Los Rios, C., Sarmiento-Barbieri, I., Aristizabal, D., Behrentz, E., Bernal, R., Buitrago, G., Cucunubá, Z., de la Hoz, F., Gaviria, A., Hernández, L. J., León, L., Moyano, D., Osorio, E., Ramírez Varela, A., Restrepo, S., Rodríguez, R., Schady, N., Vives, M., and Webb, D. (2021). SARS-CoV-2 spread, detection, and dynamics in a megacity in Latin America. *CEDE Working Paper*.

Lee, E. C., Wada, N. I., Grabowski, M. K., Gurley, E. S., and Lessler, J. (2020). The engines of sars-cov-2 spread. *Science*, 370(6515):406–407.

Lee, S. X. and McLachlan, G. J. (2013). EMMIXuskew: An R Package for Fitting Mixtures of Multivariate Skew t Distributions via the EM Algorithm. *Journal of Statistical Software*, 55(12).

McAloon, C., Collins, Á., Hunt, K., Barber, A., Byrne, A. W., Butler, F., Casey, M., Griffin, J., Lane, E., McEvoy, D., et al. (2020a). Incubation period of covid-19: a rapid systematic review and meta-analysis of observational research. *BMJ open*, 10(8):e039652.

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. (2020b). Incubation period of COVID-19: A rapid systematic review and meta-analysis of observational research. *BMJ Open*, 10(8):e039652.

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.

Millett, G. A., Jones, A. T., Benkeser, D., Baral, S., Mercer, L., Beyrer, C., Honermann, B., Lankiewicz, E., Mena, L., Crowley, J. S., Sherwood, J., and Sullivan, P. S. (2020). Assessing differential impacts of COVID-19 on black communities. *Annals of Epidemiology*, 47:37–44.

Rocha, R., Atun, R., Massuda, A., Rache, B., Spinola, P., Nunes, L., Lago, M., and Castro, M. C. (2021). Effect of socioeconomic inequalities and vulnerabilities on health-system preparedness and response to covid-19 in brazil: a comprehensive analysis. *The Lancet Global Health*.

Schmitt-Grohé, S., Teoh, K., and Uribe, M. (2020). Covid-19: Testing Inequality in New York City. Technical Report w27019, National Bureau of Economic Research, Cambridge, MA.

Smith, L. E., Amlt, R., Lambert, H., Oliver, I., Robin, C., Yardley, L., and Rubin, G. J. (2020). Factors associated with adherence to self-isolation and lockdown measures in the uk: a cross-sectional survey. *Public Health*, 187:41–52.

Stiglitz, J. (2020). Conquering the great divide. *Finance & Development*, 57(3):17–19.

The Associated Press (2020). Biden adviser marcella nunez-smith says race central to virus fight.

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.

Williams, S. N., Armitage, C. J., Tampe, T., and Dienes, K. (2020). Public perceptions and experiences of social distancing and social isolation during the covid-19 pandemic: A uk-based focus group study. *BMJ open*, 10(7):e039334.

Supplementary Information

SI.1 Materials and Methods

SI.1.1 Data Description

SI.1.1.1 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 with 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 stratum, and address.

Two main strategies were employed to recruit participants, and approximately one half of the total sample comes from each strategy. First, through 74 agreements with institutions and companies, we obtained long-lists that we used to contact and invite participants. Most lists were specific to a given occupation, based on the employees of a large company or on a list of individuals who were signed up to a specific mobile app. We also used some lists of residents based on beneficiaries of social programs. We 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. This means that it is relatively close to a population-based sampling, but with an over-representation of some occupations that were prioritized in the CoVIDA project, in particular because they were expected to be more exposed (which is why we re-weight by occupation to maintain representativity of actual occupations in Bogotá).

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.

For estimations of cumulative incidence with the CoVIDA data, 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). Hence in order to obtain the number of cases per day and per inhabitant, one needs to divide the positivity rate by 17. Intuitively, if, any person that gets infected can be tested positive during 17 days on average, then the positivity rate should be 17 times higher than the numbers of daily new infections. See our companion paper Laajaj et al. (2021) for more details on this calculation.

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.

Both databases include information on an individual's *socioeconomic stratum*, a classification that is based on the neighbourhood of residence and is commonly used in Colombia as a proxy for the household's economic living 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, leaving us with the four following groups: SES 1&2, 3, 4 and 5&6.

SI.1.2 Model details

In the model, each infected individual comes into contact with other individuals and potentially infects them. An example of this process is seen in Figure 1. A potential secondary infection is defined as a contact that would become infected unless prevented by immunity or isolation. The stochastic number of potential secondary infections each infected person generates depends on (i) whether they are symptomatic or asymptomatic, (ii) the number of contacts they have during their infectious period, and (iii) the secondary attack rate. We distinguish between contacts within the household and outside the household (Figure 1 panel (b)) and each type of contact has a different secondary attack rate. All members within the household are assumed to be contacts and to be from the same SES. Contacts from outside the household are sampled randomly from the entire population with sampling weights that reflect our estimated contact matrix between SES (see Table 1b, Panel b). The model therefore permits assortative mixing; contacts are typically more likely to occur within a SES, but are not restricted to the same SES.

Potential infections may not translate into actual infections for two reasons. First, because of immunity, a potential secondary case that has already been infected in the past will not develop into a new infection (e.g. person C). Second, any out-of-home contact is prevented if the individual is isolating at the time of the potential infection (while within-home transmissions are unaffected by isolation). An individual may isolate for four reasons: (i) after experiencing symptoms, (ii) after receiving a positive test result, (iii) after being contact-traced, or (iv) if the entire household quarantines because one member received a positive test result.

An individual may be tested either because they decide to take a test themselves after experiencing symptoms (person A and person E in Figure 1), or because they are contact traced. An individual who is "detected" (tests positive) is subjected to a process of contact-tracing with a probability lower than 1, leading to the possible testing and detection of

each of their secondary cases. For example, in Figure 1, A is detected, leading to B and D (but not E) being traced and tested.

The model is based on a branching process structure. We initiate the model with 1/5000 of the population infected, and iterate over one-period cycles (that are equivalent to one day). In each period, infected individuals can transmit the virus to others, and other events such as isolation, testing, and tracing can also occur. These events, and how they are simulated, are described in detail below. In all results, the model is simulated with a total population of 100,000, and the proportion of initial cases in each SES is set to be equal to the SES's population as a proportion of total population. The parameters used in the model are summarised in Table SI.4, which describes the parameters common across all SES, and in Table SI.5, which describes the parameters that were potentially allowed to differ by SES.

SI.1.2.1 Secondary Infections

Infected individuals come into contact with other simulated individuals. Contacts are divided into two types, "household" and "external" (outside the household). The number of contacts within the household for individual i, δ_i^{hh}, is assumed to be equal to $(householdsize_i - 1)$. The number of contacts outside the household for an individual i, δ_i^{ext}, is drawn from a negative binomial distribution with a dispersion parameter $k = 0.58$ (taken from Bi et al. (2020)) and a mean of μ_j, where j is i's SES (see Figure SI.6). The number of "potential" infections[1] that i may generate in each category is then given by a binomial distribution with size equal to the number of contacts and probability of infection from the group-specific secondary attack rate (SAR):

$$p_{ij}{}^{hh} \sim Binom(\delta_i^{hh},\ SAR_j^{hh})$$
$$p_{ij}{}^{ext} \sim Binom(\delta_i^{ext},\ SAR_j^{ext})$$

People in the same household are assumed to be from the same SES. External contacts are selected randomly from the wider population with sampling weights that depend on the estimated contact matrix seen in Table SI.5, panel (b). This allows for phenomena such as assortative mixing (i.e. individuals are more likely to encounter someone from their own group).

SI.1.2.2 Symptoms

Based on data from a recent review Buitrago-Garcia et al. (2020), we assume that 20% of infected individuals remain completely asymptomatic throughout the course of infection. We also follow this review in assuming that these individuals are less infectious than symptomatics, with a relative risk of 0.35 (i.e. the secondary attack rate is $0.35 \times SAR_j$).

[1]Potential infections are the number of infections that would happen, absent considerations of any isolation behavior or immunity. Some potential transmissions will not take place because the infector is isolating or because the potential infectee is immune.

The remaining 80% of infected individuals present symptoms according to the timing described in the next subsection. Symptomatic individuals may get tested upon observing their symptoms (see the Section SI.1.2.5), whereas asymptomatic individuals will never get tested through this channel.

SI.1.2.3 Timings

The incubation periods of both infector and infectee are both assumed to be distributed lognormally with parameters drawn from a meta-analysis of the literature on the incubation period for COVID-19 ($\mu = 1.63$ and $\sigma = 0.5$) McAloon et al. (2020b).

The serial interval is assumed to be gamma-distributed with parameters $\alpha = 8.12$, $\beta = 0.64$ (and with the distribution function translated by Δx = -7.5 to allow for negative values), all taken from He et al. (2020). This distribution implies that 10.1% of serial intervals will be negative. The mean serial interval is 5.2 days.

If the incubation periods and the serial interval are assumed to be independent, then the implied generation interval is often significantly below 0, which is epidemiologically impossible. Therefore, to make the distributions of the incubation periods and serial interval consistent with a realistic distribution of the generation interval, we allow for correlation between the values of the two incubation periods and the serial interval. To do this, we first draw values from a trivariate skewed-normal distribution Azzalini and work(s): (1996), Lee and McLachlan (2013), and then transform the resulting variables to follow the lognormal and shifted-gamma distribution described above. The variance-covariance matrix and the skew parameters of the trivariate skewed-normal are chosen to match the resulting generation interval distribution to a gamma distribution with a mean of 5.2 days (the same as the mean serial interval) and with a freely varying shape parameter. This ensures that the generation interval is realistic and that negative generation intervals are rarely drawn.[2] The matching process is carried out by numerically minimizing the Kolmogorov-Smirnoff test statistic. The shape parameter of the gamma distribution is estimated to be 4.79. The mean of the generation interval distribution, 5.2 days, is consistent with 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. The estimation process results in a shape parameter of 4.79. The probability distributions of all timing variables can be found in Figures SI.2, SI.3, SI.4, and SI.5. All timing variables are assumed follow the same distribution for both household infections and external infections.

SI.1.2.4 Immunity and isolation

A "potential" infection in the model may not become an actual new infection for three reasons:

[2]In approximately 0.36% of cases, generation intervals less than 1 day are drawn using this procedure. Secondary infections are assumed to only be possible 12 hours after infection, so in such cases we redraw the incubation period and serial interval values using the same procedure until the generation interval is greater than or equal to 1 day.

1. The potential infectee has already been infected. All individuals who have been infected once are assumed to be immune indefinitely.

2. The potential infector is isolating at the time of potential infection.

3. The potential infectee is isolating at the time of potential infection.

If the potential infectee or infector is isolating at the time of potential infection, then this reduces the probability of infection to 0 for out-of-home infections, while leaving the probability of within-home infections unchanged.

Potential infectors may isolate for one (or more) of 4 reasons:

1. Symptoms. Upon symptom onset, individuals will isolate with some probability that depends on their SES.

2. Positive test result. Upon receiving a positive test result (see below), individuals will isolate with some probability conditional on their SES.

3. Contact tracing call. Upon being told by a contact tracing team that they have been in contact with an infected individual, individuals will isolate with some probability conditional on their SES.

4. Household quarantine. A proportion ω_j of households in each group j are quarantining-types. This means that if at least one person in the household isolates because they receive a positive test result, then all the members of the household also isolate for the same period as the detected individual.

All the relevant probabilities can be found in Table SI.5.

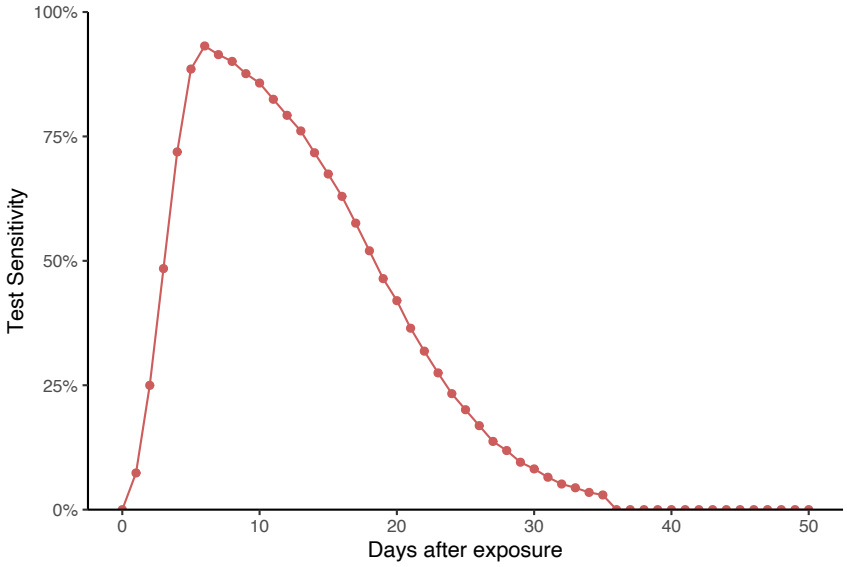
Potential infectees also reduce transmission by isolating through the household quarantining channel, although this plays a minor empirical role in reducing overall transmissions in the model simulations. In most cases, isolating individuals stop isolating when they recover, which is assumed to be 10 days after experiencing symptoms. However, there are some cases in which individuals will "deisolate" before recovery. Deisolation occurs only in the case when these conditions hold: (i) individual was isolating due to a contact tracing call, (ii) the individual then receives a negative test result (either because they were tested before they were infected, or because the test was a false negative), and (iii) the individual is not isolating for any other reason. This feature of the model avoids overestimating the efficacy of contact tracing in cases where tests are carried out very quickly.

SI.1.2.5 Testing and Contact Tracing

There are two reasons for which an individual can be tested in the model:

1. "Self" testing. Upon symptom onset (if symptomatic), individuals will be tested with some probability conditional on their SES.

Figure SI.1. Test Sensitivity (Source: Grassly et al 2020)



2. Contact tracing. If an infected individual from group j is "detected" (is tested positive), then everyone they infected has some probability π_j to be called by a contact tracing team and then tested themselves.

The relevant probabilities can be found in Table SI.5.

Tests are imperfectly sensitive, and test sensitivity is assumed to depend on time since infection, following the data from Grassly et al. (2020) (see Figure SI.1, in which test sensitivity increases rapidly over the first few days of infection and then decreases after around day 5.

All testing processes include relevant delay times. The "test consultation delay" denotes the time it takes for individuals to access a test after symptom onset (if they self-test). The "test results delay" denotes the time it takes for results to be given to an individual after they are tested (for all three types of test). And the "contact tracing delay" denotes the time it takes to contact and test an individual after the original infector is detected (these are assumed to take place simultaneously).

SI.1.3 Calibration exercises for out-of-home contacts

In order to calculate the full matrix of out-of-home contacts to be input into the model, we first estimate the average number of non-work contacts outside the home for each SES j (from individuals traced in the CoVIDA project), and add it to contacts due to working outside of home, estimated from average number of days of work in a 14-day period for group (from the full data) j. Together, it provides the average number of contacts of an individual from group j, called μ_j. We then use the COVIDA contact tracing data to infer the proportion of the contacts of an individual from group j that come from group k, called q_{jk}. This basically assumes that people are "missing at random" i.e. probability of

Figure SI.2. Incubation Period

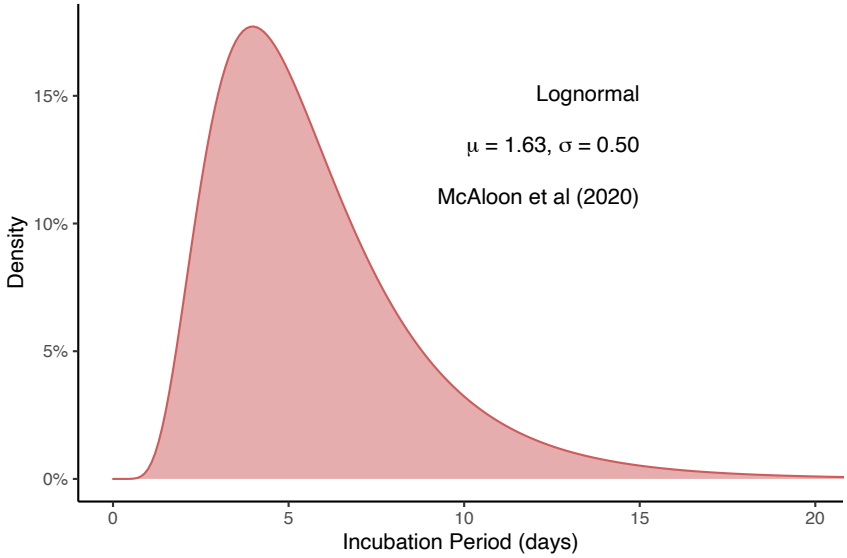


Figure SI.3. Serial Interval

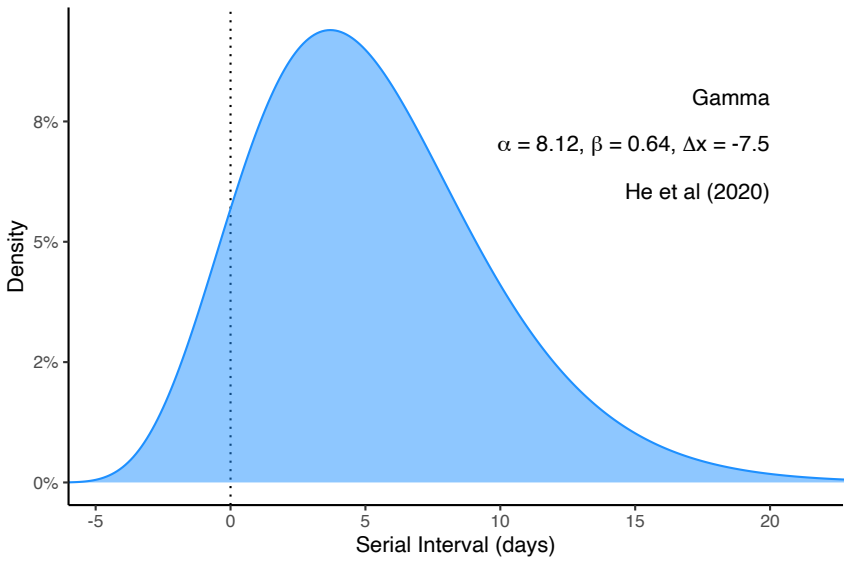


Figure SI.4. Implied Generation Interval

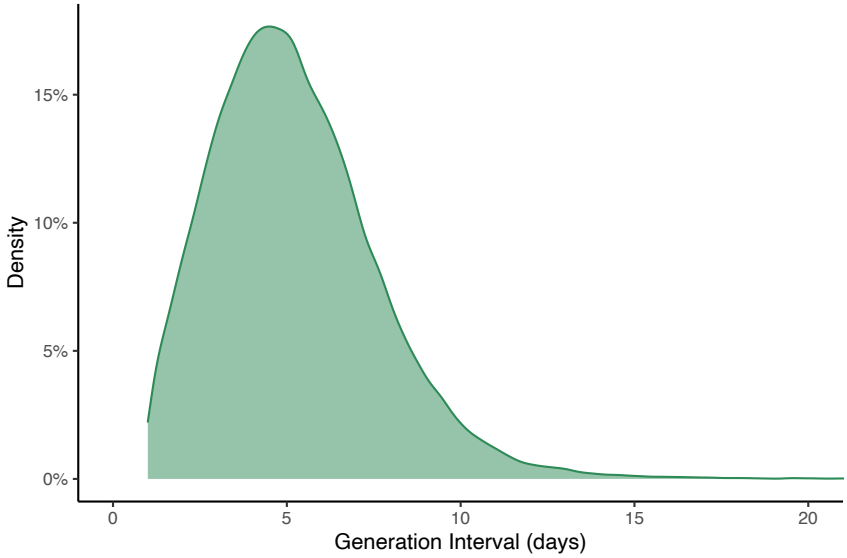


Figure SI.5. Implied Infectiousness Profile

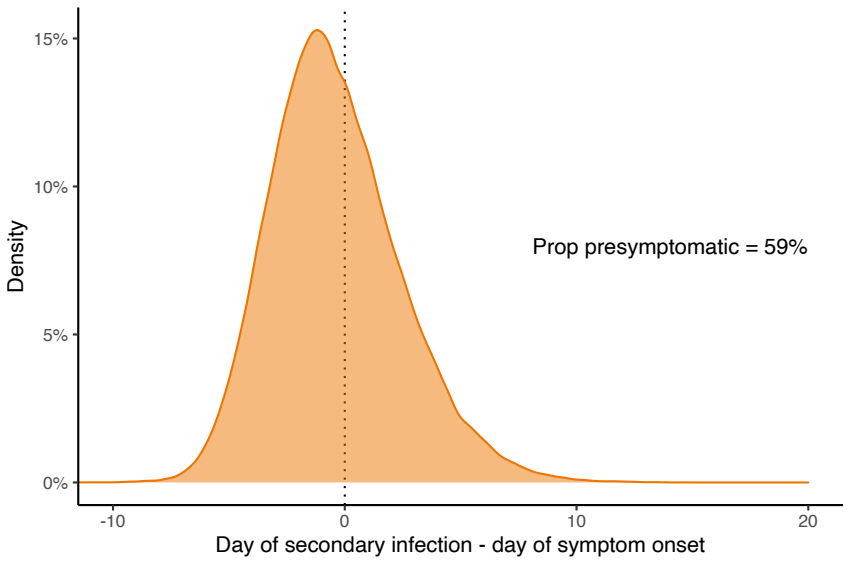
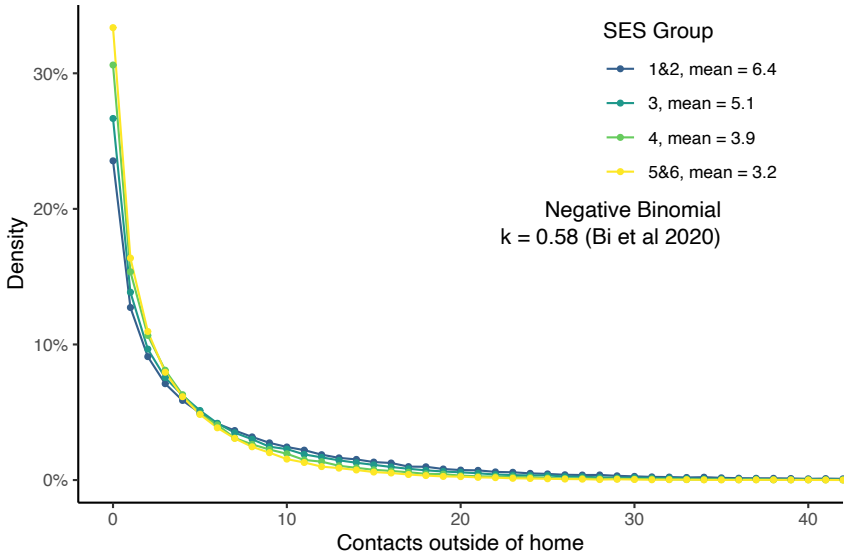


Figure SI.6. Contacts outside of home (distribution)



being missing is independent of the group. Together these pieces of information allow us to calculate the full symmetric contact matrix $\mathbf{\Delta}$ that describes the number of contacts in an average infectious period from group j to group k (absent any isolation behavior). The steps carried out are described in more detail below.

SI.1.3.1 Contacts outside the home

We assume that the mean number of contacts an individual from group j has outside of home during an average infectious period is given by μ_j, defined as:

$$\mu_j(\lambda) = v_j + \lambda w_j$$

Where v_j is the mean number of *non-work* contacts from outside of home for group j, and w_j is the mean number of days of work for group j. Both of these are estimated directly from the CoVIDA data (see Table 1) in the main text. λ is an unknown parameter (the "work factor") that describes the relationship between the number of days at spent at work and the number of out-of-home contacts during an infectious period (we assume this relationship to be linear). We choose a value of λ by calibrating the model to the true value of $\mathcal{R}_q$ estimated from the data. More specifically, we carry out the following steps:

1. For each candidate value λ_0 in the set of candidate values $\{0.05, 0.1, 0.15, 0.2, \ldots, 2.0\}$:
   (a) Calculate the implied value of $\mu_j(\lambda_0)$.
   (b) Use the values of $\mu_j(\lambda_0)$ to calculate the implied full contact matrix $\mathbf{\Delta}$ using the maximum likelihood process outlined in Section SI.1.3.2.
   (c) Run 50 simulations of the first 100 periods of the model.

(d) Use these simulations to calculate the implied value of $\mathcal{R}_q$ for this value of λ_0 following the steps in Section SI.1.3.3. Call this $\mathcal{R}_q(\lambda_0)$.

2. Set λ equal to λ^* , where λ^* is the value of λ_0 that minimizes $|\mathcal{R}_q(\lambda_0) - \mathcal{R}_q^*|$, with $\mathcal{R}_q^*$ being the true value of the basic reproduction number calculated based on the real data using the methodology in Section SI.1.3.3.

3. Use $\hat{\mathbf{\Delta}}_{ML}(\lambda^*)$ as calculated using the method in Section SI.1.3.2 as the value for the contact matrix in the model simulations.

When we carry out the above steps in this way, we find that $\lambda^* = 0.8$, implying that the value for the work factor that best matches the observed growth in early cases in Bogotá's epidemic is 0.8. In other words, the model assumes that the mean number of out-of-home contacts during the average infectious period increases by 0.8 with each additional day of work outside of home.

Using this work factor, the estimated contact matrix based on the maximum likelihood procedure in Section SI.1.3.2 for a population of 100,000 is:

$$\hat{\mathbf{\Delta}}_{ML}(\lambda^*) = \begin{bmatrix} 272,155 & 49,440 & 5,162 & 2,432 \\ 49,440 & 102,967 & 16,864 & 2,729 \\ 5,162 & 16,864 & 8,431 & 6,022 \\ 2,432 & 2,729 & 6,022 & 5,247 \end{bmatrix}$$

The estimated $\mathbf{\Delta}$ matrix tells us the number of contacts that occur between each pair of groups in an average infectious period (which is the same across all groups). For example, the value in the first column and second row indicates that 49,440 contacts occur between individuals of groups 1 and 2 (corresponding to SES 1&2 and 3 respectively) during an average infectious period. Contacts are assumed to be mutual, so that the matrix is symmetric. We can see from the estimation result that there is strong assortative mixing: each group is more likely than random to contact another individual from the same group. (Note that smaller values for the 3$^{\text{rd}}$ and 4$^{\text{th}}$ column are due to smaller population sizes in these groups).

When using this estimated contact matrix $\hat{\mathbf{\Delta}}_{ML}(\lambda^*)$, the estimated value of $\mathcal{R}_q(0.8)$ is 1.222, close to the estimated value in the data $\mathcal{R}_q^* = 1.216\,[95\%CI : 1.170, 1.263]$. We use the matrix $\hat{\mathbf{\Delta}}_{ML}(\lambda^*)$ as an input to the baseline simulations of the model seen in the main results.

Another way of viewing this matrix is by looking at the proportion of contacts for an individual from group j that are from group k:

$$\hat{\mathbf{Q}}(\lambda^*) := [q_{jk}] = \begin{bmatrix} 0.827 & 0.150 & 0.016 & 0.007 \\ 0.287 & 0.599 & 0.098 & 0.016 \\ 0.142 & 0.462 & 0.231 & 0.165 \\ 0.148 & 0.166 & 0.367 & 0.319 \end{bmatrix}$$

This tells us, for example, that 14.8% of the contacts of an individual from 4 come from group 1. These values are then directly used in the model to set the probability that a potential infection generated by someone from group j is from group k.

In the following section, we describe in detail the maximum likelihood process used to estimate $\hat{\mathbf{\Delta}}_{ML}(\lambda_0)$ for each of the candidate values λ_0 of the work factor. Then, in Section SI.1.3.3 we describe the how we calculate the value of $\mathcal{R}_q$ both in the data (to calculate $\mathcal{R}_q^*$) and in the models (to calculate $\mathcal{R}_q(\lambda_0)$. Finally, in Section SI.1.3.4 we describe the "mobility matching" process, in which we allow the $\mathbf{\Delta}$ matrix to be scaled by a time-varying constant over the course of the epidemic in order to account for changes in mobility in Bogotá.

SI.1.3.2 Maximum likelihood process for contact matrix

The aim of the following maximum likelihood process is to estimate the symmetric contact matrix $\mathbf{\Delta}$, where the element c_{jk} describes the *total* number of out-of-home contacts (across the whole population) between individuals from group $j \in \{1, 2, 3, 4\}$ and individuals from group $k \in \{1, 2, 3, 4\}$ over the course of an average infectious period:

$$\mathbf{\Delta} := \begin{bmatrix} c_{11} & c_{12} & c_{13} & c_{14} \\ c_{21} & c_{22} & c_{23} & c_{24} \\ c_{31} & c_{32} & c_{33} & c_{34} \\ c_{41} & c_{42} & c_{43} & c_{44} \end{bmatrix}$$

Let μ_j be the mean number of out-of-home contacts for an individual from group j across all groups during an average infectious period. For the purposes of this estimation process, we assume that $\boldsymbol{\mu}(\lambda_0) := [\mu_1(\lambda_0), \mu_2(\lambda_0), \mu_3(\lambda_0), \mu_4(\lambda_0)]'$ is given, because it has been calculated in step (1a) in the previous section using a candidate value for the work factor of λ_0. What follows is thus the estimation process for $\widehat{\mathbf{\Delta}}_{ML}(\lambda_0)$. We also treat as given $\mathbf{n} := [n_1, \ldots, n_4]'$ where n_j is the number of individuals in group j in the population, because this is known from the data. Given $\boldsymbol{\mu}(\lambda_0)$ and $\mathbf{n}$, we use maximum likelihood to estimate the 6 remaining parameters of $\mathbf{\Delta}$, which we denote using the vector $\mathbf{d} = [d_{11}, d_{12}, d_{13}, d_{22}, d_{23}, d_{33}]'$.

The values c_{jk} are filled out by the parameter vectors $\boldsymbol{\mu}(\lambda_0), \mathbf{n}$, and $\mathbf{d}$. In particular, we define $\mathbf{\Delta}(\mathbf{d}, \boldsymbol{\mu}(\lambda_0), \mathbf{n})$ as the matrix with elements c_{jk} defined in the following way:

1. First, the $\mathbf{d}$ vector directly specifies some elements of the $\mathbf{\Delta}$ matrix, so that:

$$c_{jk} = d_{jk} \quad \text{for all } d_{jk} \in \mathbf{d}$$

2. Then, we enforce symmetry because contacts between each group are symmetrical so happen at the same rate, so set :

$$c_{kj} = d_{jk} \quad \text{for all } d_{jk} \in \mathbf{d}$$

3. Finally, the last row and column are pinned down by the μ vector and the population size since the sum of each row j in the $\boldsymbol{\Delta}$ matrix needs to sum to $\mu_j n_j$:

$$\sum_{k=1}^{4} c_{jk} = \mu_j(\lambda_0) n_j$$

$$\Rightarrow c_{j4} = \mu_j(\lambda_0) n_j - \sum_{k=1}^{3} c_{jk} \quad \text{for all } j \in \{1,2,3,4\}$$

To understand how we estimate the vector $\mathbf{d}$, note first that the set of vectors $\{\mathbf{d}, \boldsymbol{\mu}(\lambda_0), \mathbf{n}\}$ also pin down the probability q_{ik} that a given contact of a person from group i is from group j, defined in the following way: $q_{jk}(\mathbf{d}, \boldsymbol{\mu}(\lambda_0), \boldsymbol{n}) :=$ Pr(Secondary contact is from group $k \mid$ Individual is from group j)

$$= \frac{[\boldsymbol{\Delta}(\mathbf{d}, \boldsymbol{\mu}(\lambda_0), \mathbf{n})]_{jk}}{\mu_j(\lambda_0) n_j}$$

We can use this probability to define the conditional likelihood function for d given the contact tracing data we observe, and the known parameters $\boldsymbol{\mu}(\lambda_0)$ and $\mathbf{n}$. The contact tracing data is a set of observations of contacts from one group to another $Z = \{(x_1, y_1), (x_2, y_2), \ldots, (x_M, y_M)\}$, where $x_m \in \{1,2,3,4\}$ denotes the group of an surveyed infected individual and $y_m \in \{1,2,3,4\}$ denotes the group of the reported contact (both defined for $m \in \{1, \ldots, M\}$). (Each surveyed individual can have multiple reported contacts, although this does not affect the estimation procedure.) The conditional probability of observing a given observation pair (x_m, y_m) is given by:

$$\Pr\left(Y = y_m \mid X = x_m; \mathbf{d}, \boldsymbol{\mu}(\lambda_0), \mathbf{n}\right) = q_{x_m, y_m}(\mathbf{d}, \boldsymbol{\mu}(\lambda_0), \mathbf{n})$$

So, assuming that the observations are independently and identically distributed, the conditional likelihood of observing the entire dataset is:

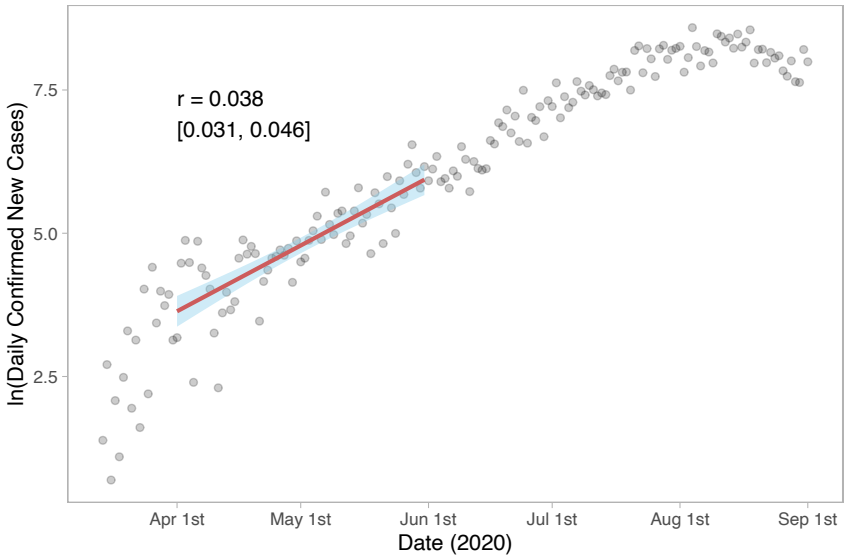
$$\mathcal{L}(\mathbf{d}; \boldsymbol{Z}, \boldsymbol{\mu}(\lambda_0), \mathbf{n}) = \prod_{m=1}^{M} q_{x_m, y_m}(\mathbf{d}, \boldsymbol{\mu}(\lambda_0), \mathbf{n})$$

Therefore, to run the maximum likelihood estimation procedure, we input the values of $\mu_j(\lambda_0)$ from the procedure described in Section SI.1.3.1, along with the known values of n. We then follow the maximisation programme below, maximising the log likelihood function while imposing the constraint that all elements of $\boldsymbol{\Delta}$ are positive:

$$\max_{\boldsymbol{k}} \quad \ln \mathcal{L}(\mathbf{d}; \boldsymbol{Z}, \boldsymbol{\mu}(\lambda_0), \mathbf{n}) \quad \text{s.t.} \quad [\boldsymbol{\Delta}(\mathbf{d}, \boldsymbol{\mu}(\lambda_0), \mathbf{n})]_{ij} > 0 \quad \forall i, j$$

We use the *nloptr* package in R Johnson (2020) to run this optimization numerically, using a local COBYLA (Constrained Optimization BY Linear Approximations) algorithm. This

Figure SI.7. Calculation of exponential growth rate of cases in early stage



yields a maximum likelihood estimate of the full symmetric contact matrix, which we denote $\widehat{\mathbf{\Delta}}_{ML}(\mathbf{d}, \boldsymbol{\mu}(\lambda_0), \mathbf{n})$, sometimes denoted $\widehat{\mathbf{\Delta}}_{ML}(\lambda_0)$ for brevity.

SI.1.3.3 Calculating $\mathcal{R}_q$

Here we calculate the value of $\mathcal{R}_q$, 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 (or very close to 1). We treat this value as a constant in the early phase of the quarantine. To calculate the value of $\mathcal{R}_q$, 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 $\mathcal{R}_q$.

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.7 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.

Then, we take the estimated generation interval from Figure SI.4, denoted $g(a)$ where a is the number of days since infection. This and use this to calculate the initial value of $\mathcal{R}_q$ using the Lotka-Euler equation Wallinga and Lipsitch (2007):

$$\frac{1}{\mathcal{R}_q} = \int_{a=0}^{\infty} e^{-ra} g(a) da$$

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 SI.4, this yields an estimate of $\mathcal{R}_q^* = 1.216\,[95\%CI : 1.170, 1.263]$. Note that this estimate comes during a period of strict lockdown in Bogotá, which explains why our estimate of $\mathcal{R}_q$ 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)).

When calculating the $\mathcal{R}_q(\lambda_0)$ values based on the model simulations, we carry out exactly the same steps apart from the fact that we add simulation fixed effects to the OLS regression for calculating r. This ensures that between-simulation variation in new cases does not affect the estimates of the rate of the exponential growth in new cases.

SI.1.3.4 Mobility calibration

In some model specifications, we allow for changes in the number of out of home contacts over time in the model, in order to account for changes in the level of mobility over time in Bogotá (in particular due to changes in policies such as stay-at-home orders and other mobility restrictions).

To do this, we allow for a generalized mobility factor $m(t)$ to change over time throughout the epidemic, so that at each time t, the contact matrix input into the model is equal to:

$$\hat{\mathbf{\Delta}}_{ML}(\lambda^*) \ \times m(t)$$

$m(t)$ scales the contacts of all groups similarly, and does not lead to differential mobility changes across groups through the course of the epidemic.

We estimate $m(t)$ by using an iterative process to match the model predictions to the observed pattern of new cases seen in Bogotá. This matching process calibrates the total confirmed[3] new cases in the model (summing all groups) to the total confirmed new cases in the data (summing all groups). This means that any inequality between groups is a result of the other parameters we input into the model described elsewhere. The precise quantity used to match is the total (all group) per capita new confirmed cases per day, smoothed by taking a 2-week rolling average. Call this quantity it $X_k(t; m(t))$ when it is calculated from an individual simulation k. And call this same quantity $Y(d)$ when it is calculated from the data with dates d.

The iterative matching process follows the steps below:

1. Let $m_1(t)$ be 1 for all periods t.

[3]A case is deemed confirmed when an individual receives a positive test result (both in the model and the data).

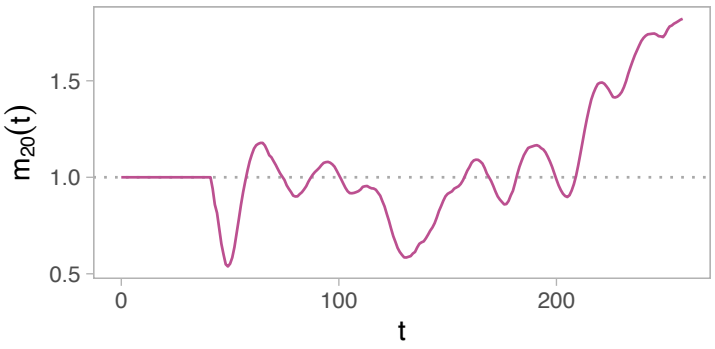


Figure SI.8. **Results of mobility calibration process**

2. For iteration l in 1:20:

   (a) Run a set of *100* simulations with the postulated values for $m_l(t)$.

   (b) Calculate the mean across all simulations of the smoothed number of per capita new confirmed cases per day for each time t:

$$\overline{X}_k(t; m_l(t)) := \frac{1}{50} \sum_{k=1}^{50} X_k(t; m_l(t))$$

   (c) Match the model periods $t \in \{ 0, 1, \dots \}$ to real dates d by choosing a lower bound date early in the epidemic (01/06/2020) and finding the time when new cases per day is equal in the model and the data, i.e. when Y (d = 01/06/2020) $= \overline{X}_k(t; m_l(t))$. This defines the quantities $Y(t)$.

   (d) For all t after t corresponding to the lower bound date, calculate the deviation quantity $\Delta(t)$ that indicates the extent to which the model predictions deviate from the data:

$$\Delta(t) = \ln \left[\overline{X}_k(t; m_l(t))\right] - \ln \left[Y(t)\right]$$

   To avoid sharp changes in mobility, we calculated a smoothed version of this function $\widetilde{\Delta}(t)$ by taking the 1 week rolling average of the deviations.

   (e) Making use of the fact that the average time from infection to receiving positive test results is approximately 3 weeks, we set the value of $m_{l+1}(t)$ in order to account for the deviations from the 3 weeks later:

$$m_{l+1}(t) = m_l(t) - \gamma \cdot \widetilde{\Delta}(t \ + 21)$$

   where the adjustment factor γ was chosen to be 1/3, so that a 1 log deviation at $t + 21$ would lead to an adjustment of 1/3 on the value of $m_{l+1}(t)$. We then use this value of $m_{l+1}(t)$ as the input to step (a) for loop $l + 1$.

3. Use $m_{20}(t)$ as the final value for $m(t)$ in the "mobility change" models.

The results of this mobility-matching process, i.e. the values of $m_{20}(t)$, are plotted in Figure SI.8.

SI.2 Figures

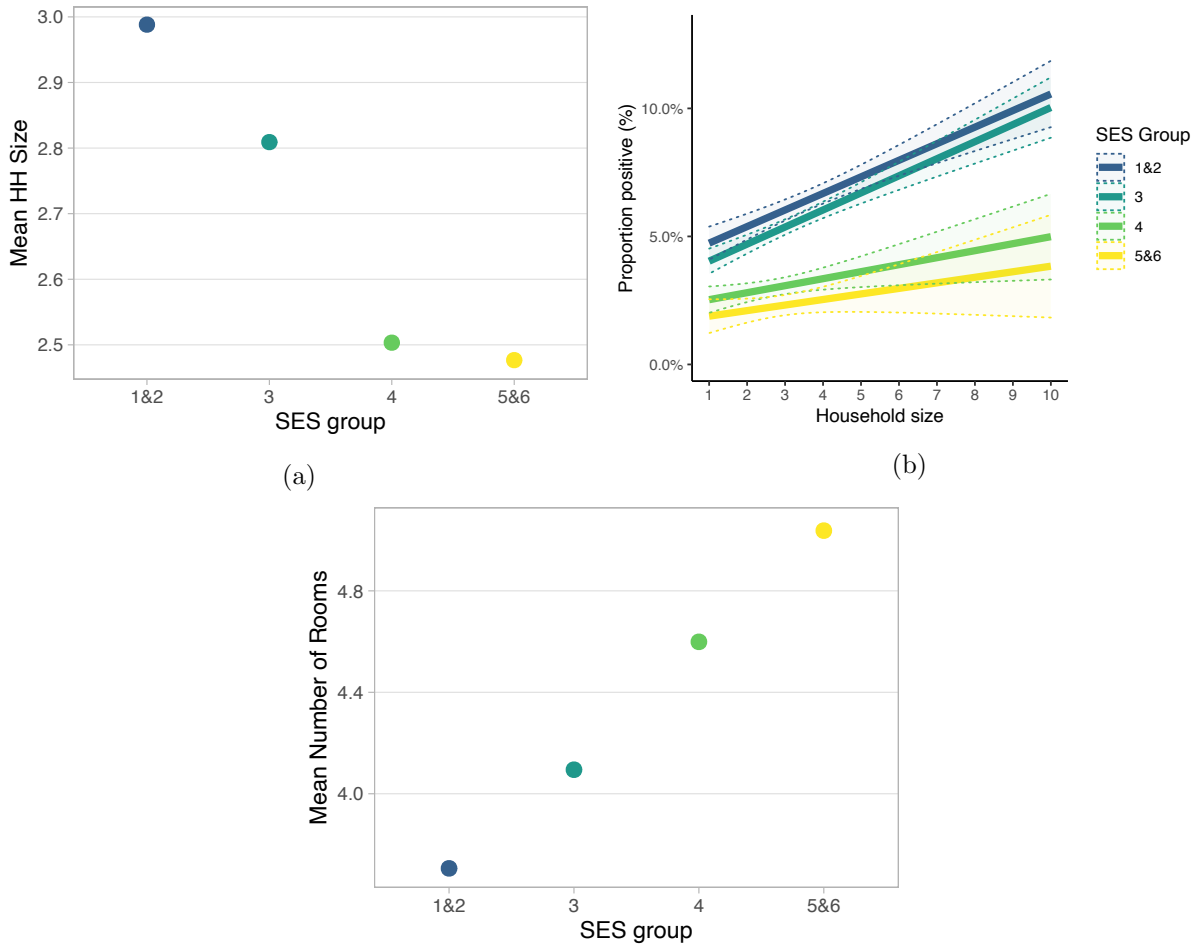


Figure SI.9. **Variation in within-household conditions across SES.** Panel (a) shows the mean household size for an individual in each SES. Panel (b) shows the mean number of rooms in the household of an individual in each SES. Both (a) and (b) are derived from census data. 95% confidence intervals are too small to be seen at this scale; all differences in means are significant at the 1% level. Panel (c) shows the linear fit of the relationship between household size and the probability of being infected conditional on being tested in the CoVIDA data by SES, with 95% shaded confidence intervals. The slopes of the effect of household size on positivity for strata 1&2, 3, 4 and 5&6 are 0.56, 0.58, 0.13 and 0.13 respectively, The p-value of the F-test of difference between these slopes is p=0.0033

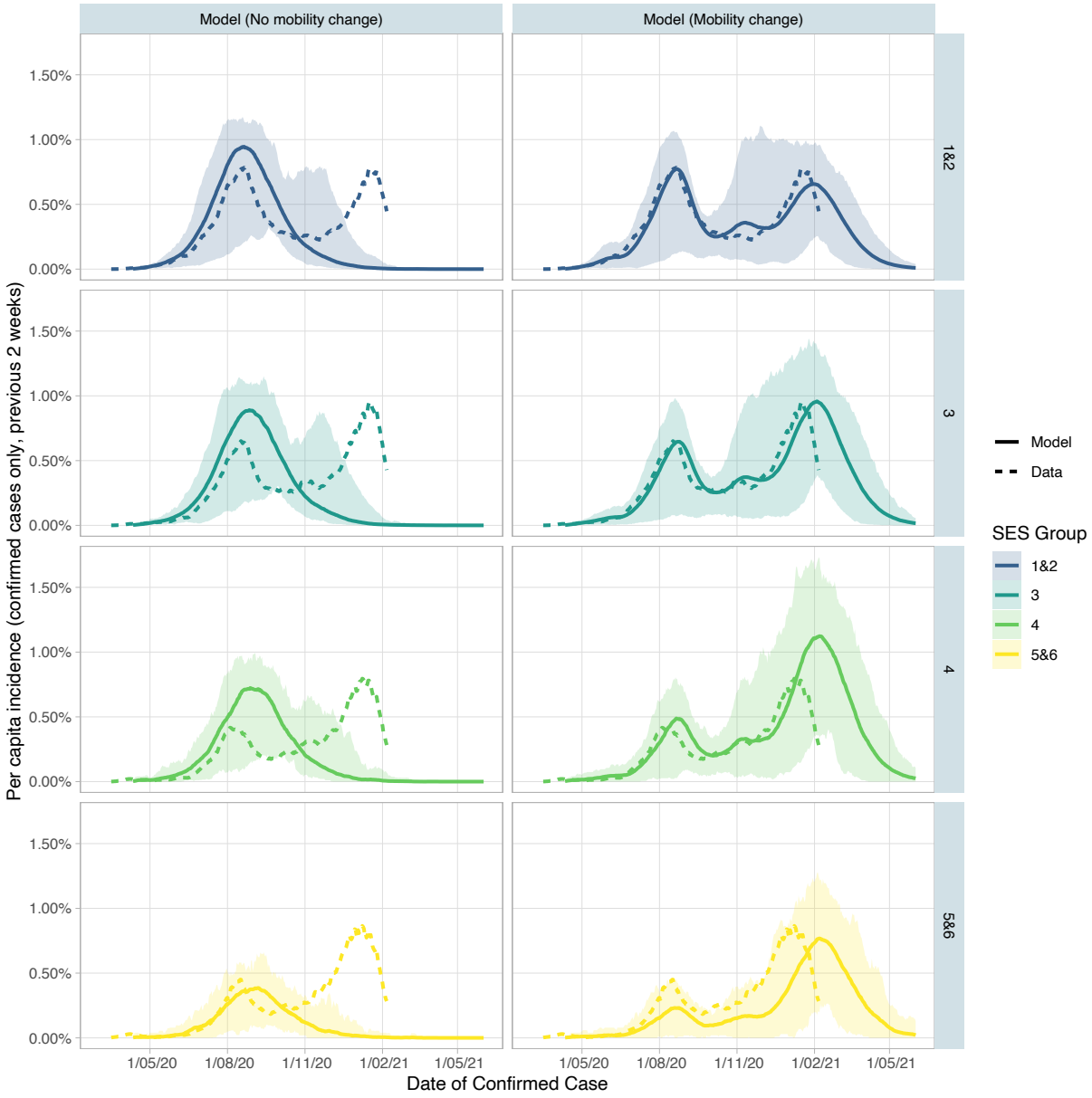


Figure SI.10. **Model match based on confirmed cases with confidence intervals.** This shows the same results as panels (a), (b) and (c) in Figure 2. The dashed lines represent the observed data on confirmed cases from the Health Secretary of Bogotá (HSB). The solid lines represent the median predictions over 50 model simulations of per capita confirmed incidence. The shaded intervals represent the 0.025 and 0.975 quantiles of the per capita confirmed incidence over these same 50 model simulations.

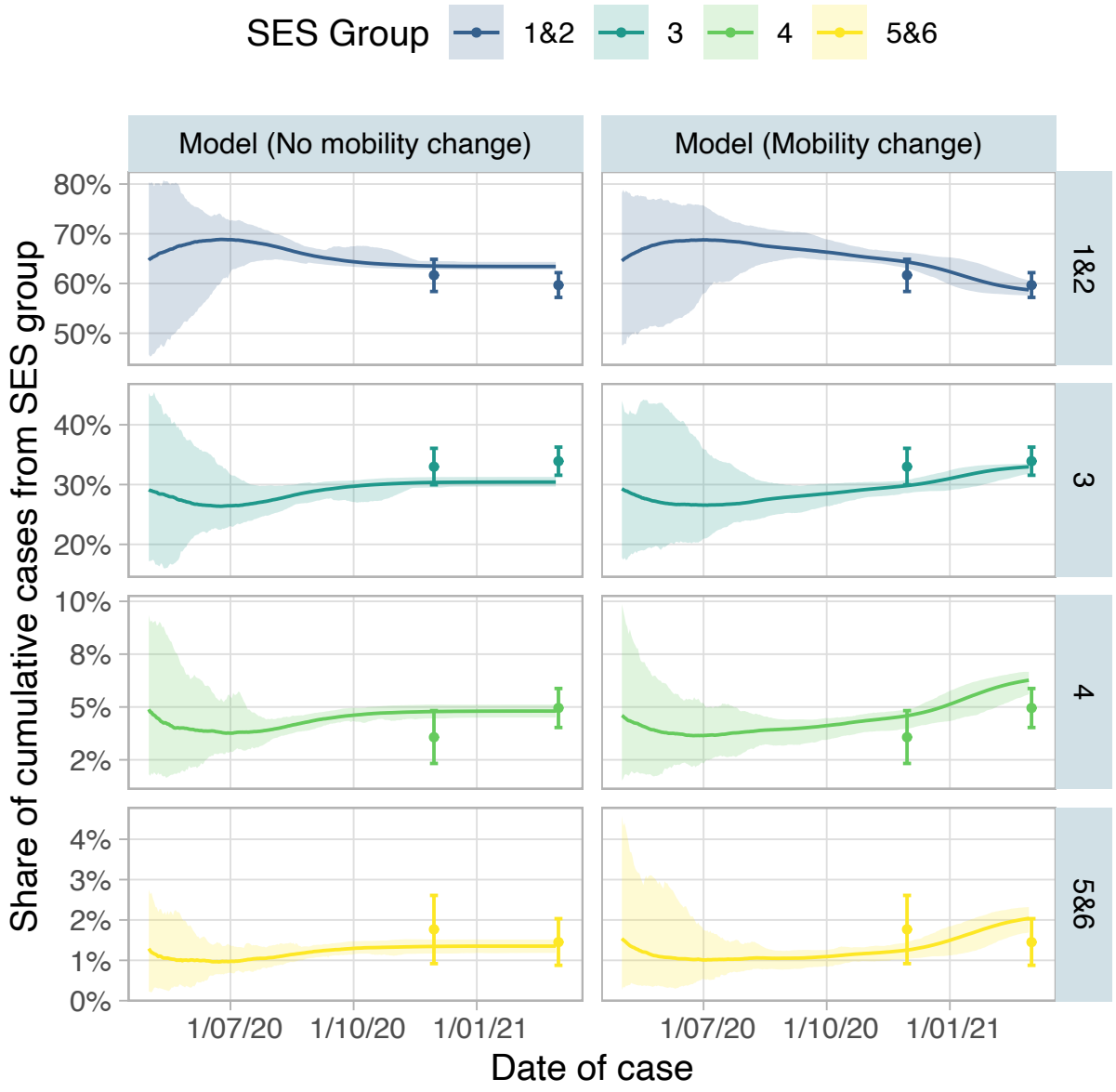
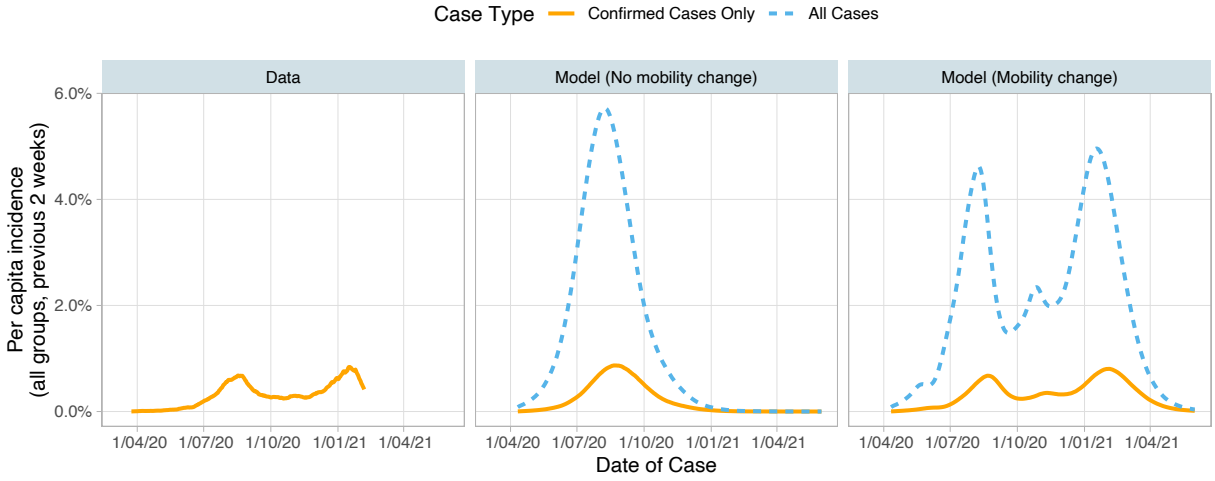
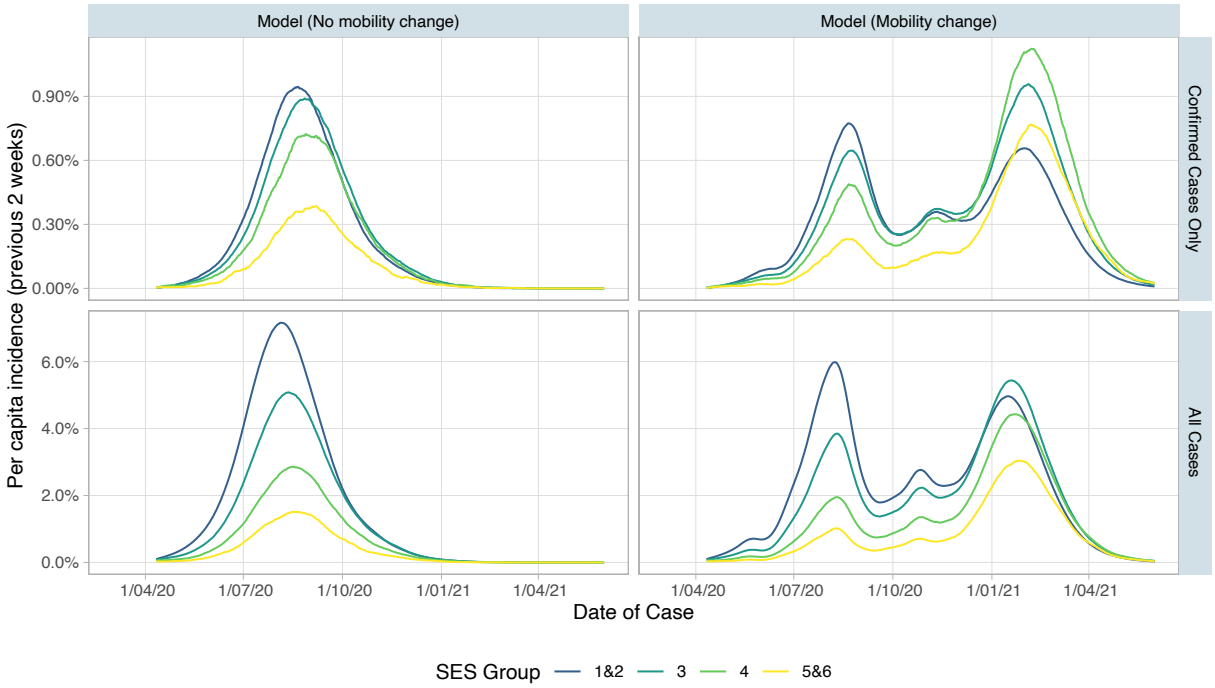


Figure SI.11. **Share of cases from each group**. The y-axis denotes the number of cumulative cases from the specific SES shown divided by the total number of cumulative cases across all groups. The solid line is the median, and the shaded areas denote the 0.025 and 0.975 quantiles of the model results. The points and surrounding error bars are estimations from the CoVIDA data, calculated using the methodology described in subsection SI.1.1.



(a)



(b)

Figure SI.12. **Detected cases compared to confirmed cases**. Panel (a) shows the per capita incidence over the last 2 weeks for confirmed cases (orange solid line) compared to all cases (blue dashed lane). Data comes from the Health Secretary of Bogotá. Panel (b) shows the per capita incidence by SES in the model simulations, using confirmed cases in the top row and all cases in the bottom row. Note the differing scales on the Y-axes. All model results are the median for 50 simulations.

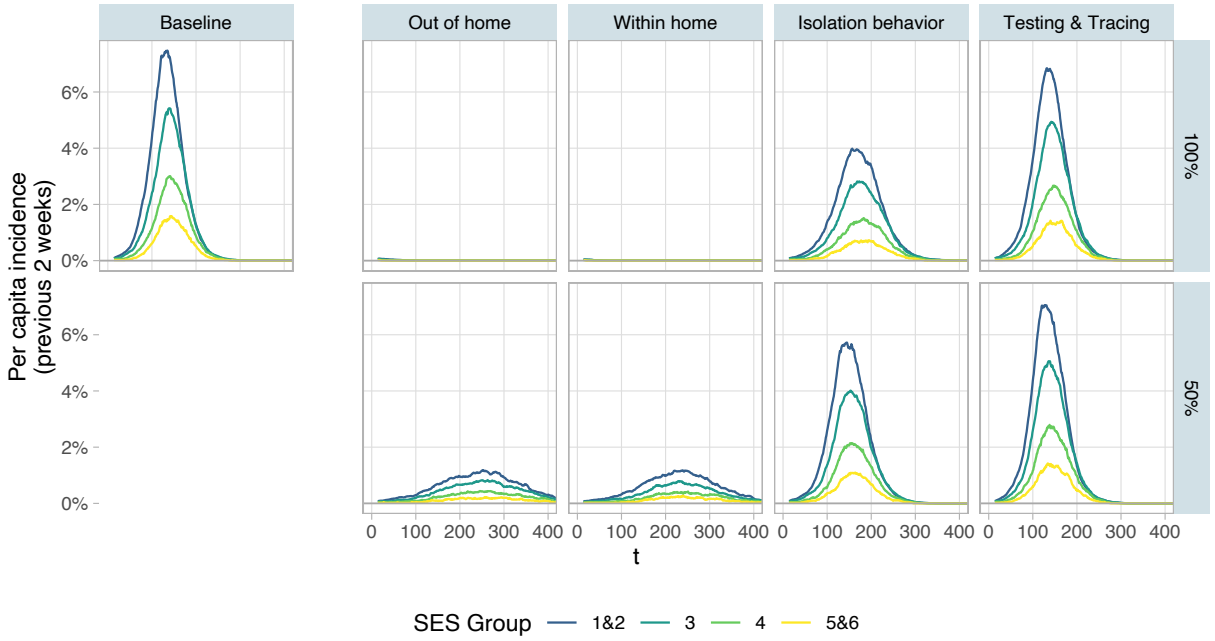


Figure SI.13. **Upward adjustment scenarios epidemic curves**. Additional results of the upward-adjustment scenarios, in which the parameters of all SES are adjusted to match the values of SES 5&6. The variable on the y-axis is the per capita incidence over the last 2 weeks in each SES. Parameters adjusted in each set are as follows: *out of home* (number of contacts outside the home), *within home* (within-household SAR, household size), *isolation behavior* (probability of isolating conditional on observing symptoms, testing positive, being contact traced, and probability of quarantining as a household), *testing & tracing* (probability of self testing, delay in test consultation, delay in test results, and probability of being contact traced).

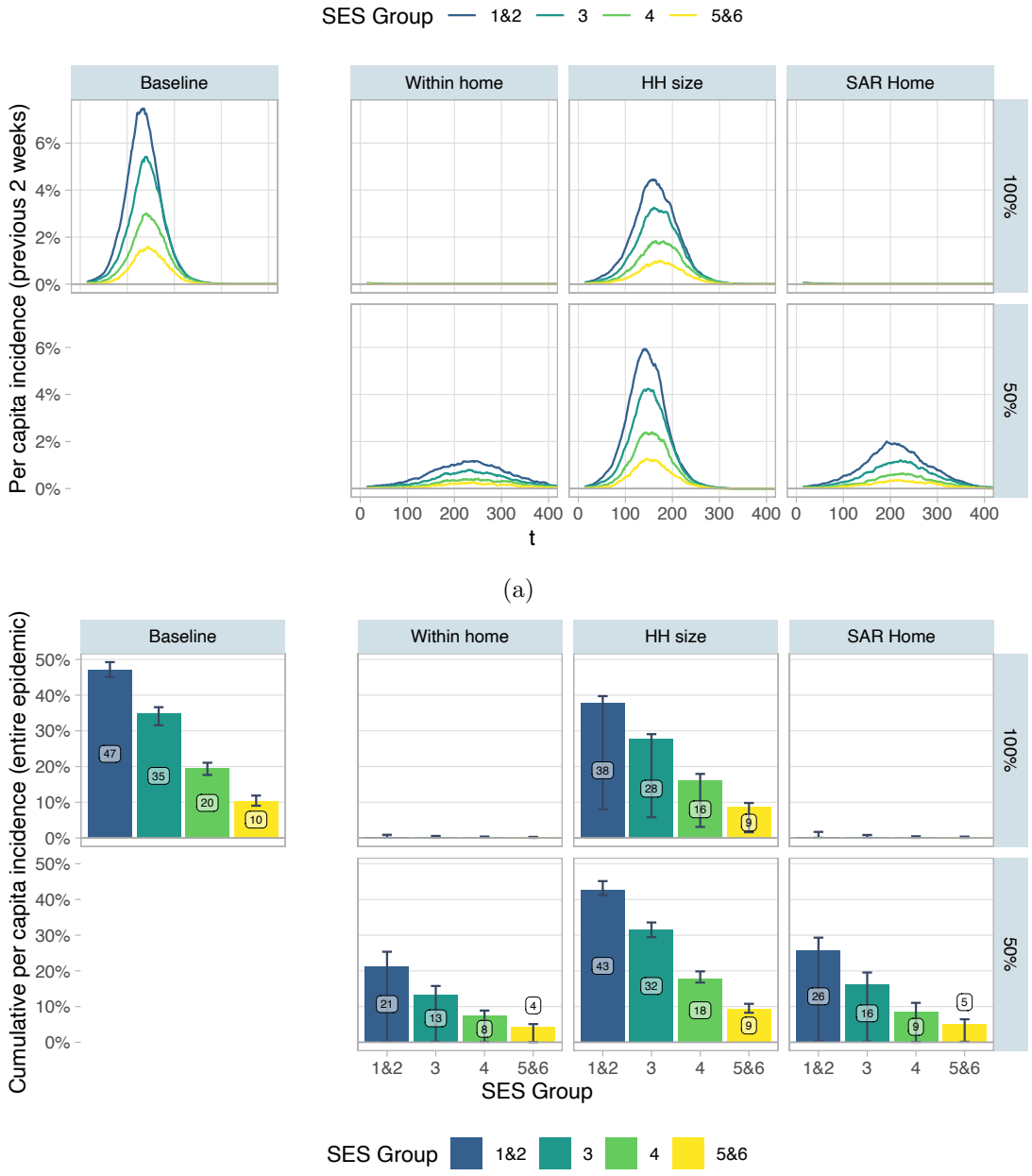
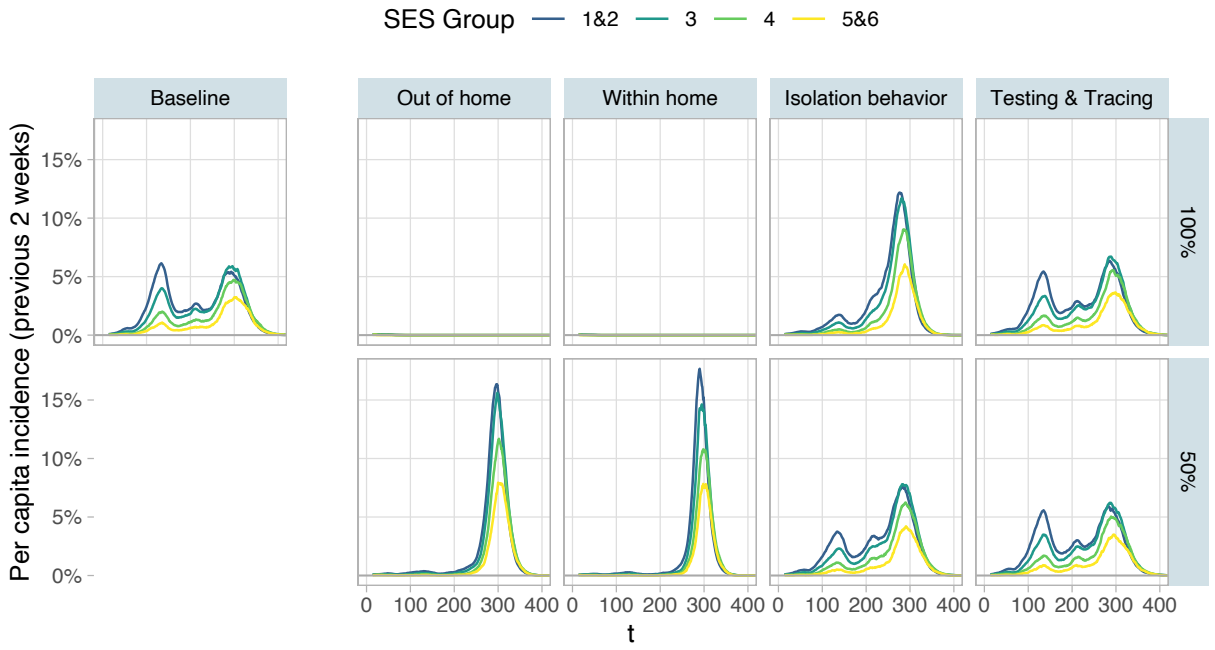
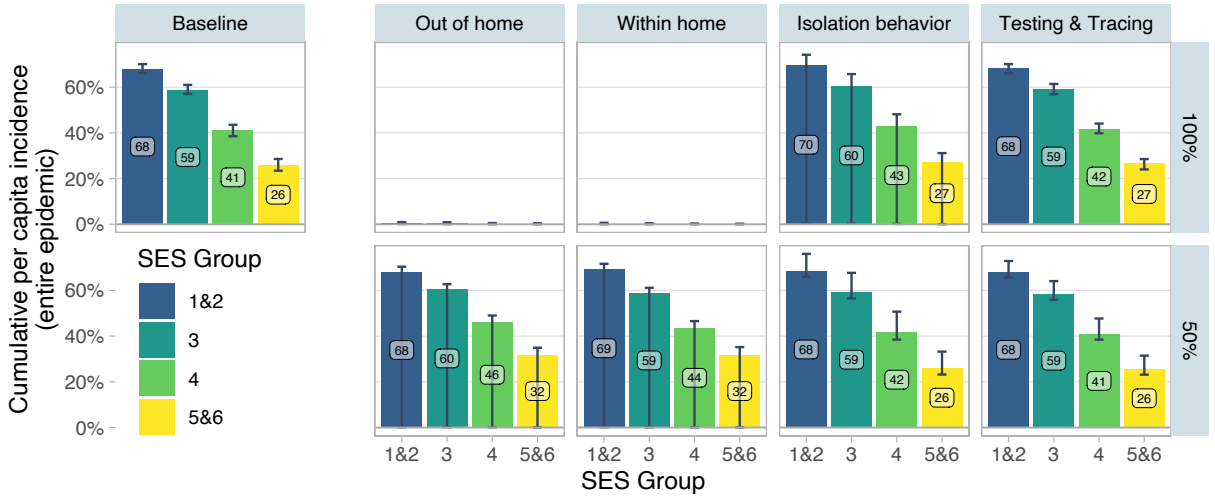


Figure SI.14. **Upward adjustment scenarios when changing within-home SAR and household size**. Results of additional upward-adjustment scenarios, in which the parameters of all SES are adjusted to match the values of SES 5&6. "Within-home" adjusts both household size and within-home SAR. "HH size" and "SAR Home" adjust only the single named parameter at a time. Panel (a) shows the two-weekly incidence for each group in each of the upward adjustment scenarios. Panel (b) shows the overall cumulative incidence for each group in these scenarios.



(a)



(b)

Figure SI.15. **Upward adjustment scenarios with mobility change**. Panel (a) shows the two-weekly incidence for each group in each of the upward adjustment scenarios. Panel (b) shows the overall cumulative incidence for each group in these scenarios. Apart from in simulations where the epidemic is completely contained (e.g. Out of home (100%) and Within home (100%)), reductions in transmission early on in the epidemic lead to much more severe second waves, leading to little overall change in the overall cumulative incidence.

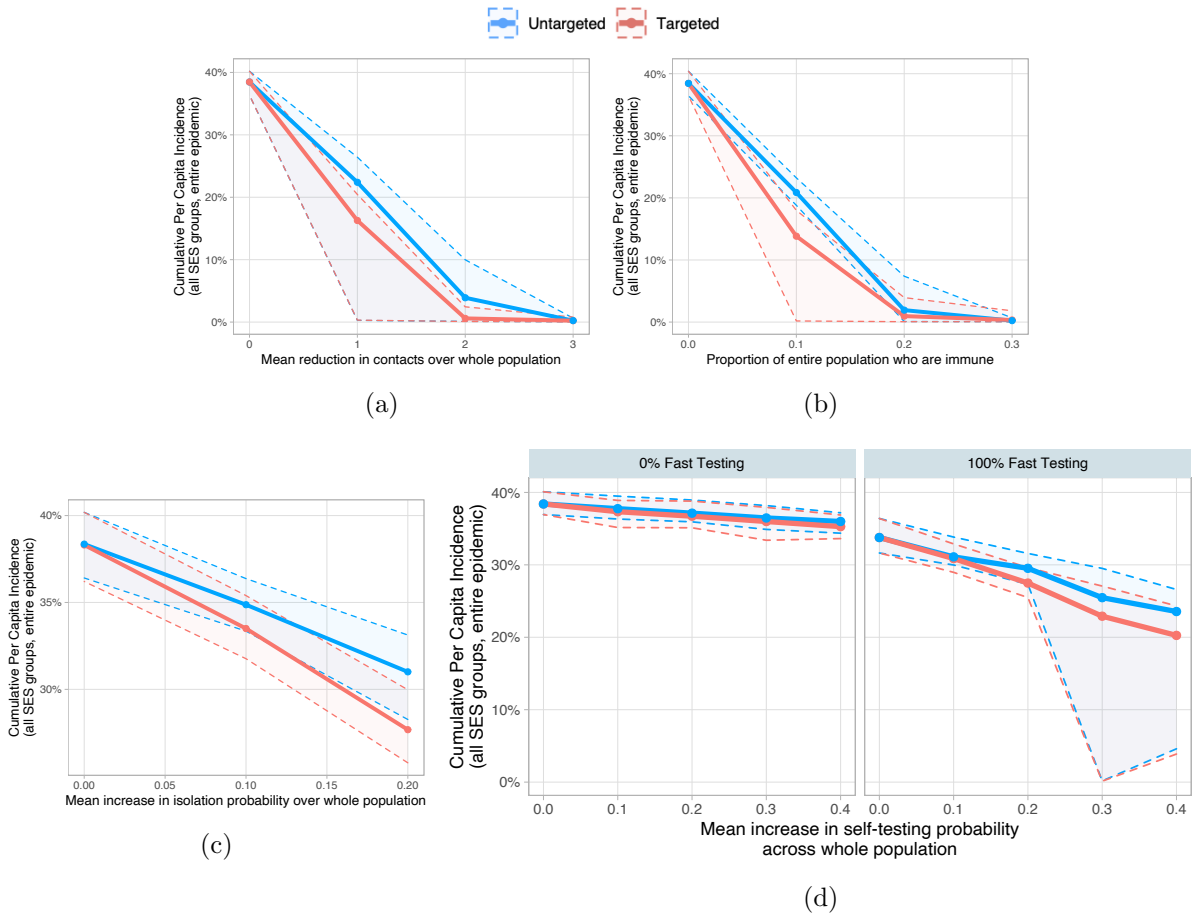


Figure SI.16. **Targeted and untargeted policies with varying intensity**. The outcome variable in all cases is the cumulative per capita incidence over the course of the whole epidemic for models with no mobility change. In "Targeted" scenarios, only the parameters of SES 1&2 are adjusted, but adjustments in this group are greater, such that the mean adjustment across the whole population is the same as in the untargeted scenario. Panel (a): outside-of-home contacts are reduced by 1, 2, and 3 relative to baeline scenario. Panel (b): 10%, 20% and 30% of the population are immune to the virus from the start of the epidemic. Panel (c): mean increase of 10 and 20 percentage points in select isolation parameters (probability of isolating conditional on being symptomatic and being contact traced). Panel (d): mean increase of 10, 20, 30, and 40 percentage points in tehthe probability of being tested after observing symptoms. 0% fast testing indicates the same testing delays as in the baseline scenario. 100% fast testing indicates that all tests have a consultation delay of 1 day and a results delay of 1 day. The estimates represent the median of 50 simulations, while the confidence intervals represent the 0.025 and 0.975 quantiles of these simulations.

SI.3 Tables

Table SI.1. Self-declared prevention practices

Dependent variable:	No Trips (last 14 days)	Frequency of hand washing	Duration of hand washing	Frequency of mask usage	Frequency of antibacterial usage
Stratum 3	-0.041***	0.028***	0.0093	-0.034***	-0.00078
	[0.0029]	[0.0093]	[0.0088]	[0.0050]	[0.0085]
Stratum 4	-0.093***	0.060***	0.0065	-0.095***	-0.066***
	[0.0040]	[0.011]	[0.011]	[0.0068]	[0.010]
Strata 5&6	-0.13***	0.051***	-0.045***	-0.13***	-0.067***
	[0.0054]	[0.014]	[0.013]	[0.0090]	[0.013]
Constant	0.89***	2.83***	3.00***	3.77***	2.78***
(Mean SES 1&2)	[0.0021]	[0.0072]	[0.0069]	[0.0037]	[0.0066]
Observations	74,345	74,613	74,612	74,544	74,467
p-value of F-test of joint significance	0	0.0000005	0.00016	0	0

Robust standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1, all p-values and F-test are double-sided tests. Linear regressions. The outcome variable "No trips" is a dummy indicating that the individual didn't take any trips in the 14 days preceding the survey. Outcome variables in columns 2 to 5 are transformed into an index from 1 to 4 where higher values of the outcome variable always means higher level of protection. The last line presents the p-value of the joint significance of the three SES dummies in the regression, hence it tests the nul hypothesis that the average value of the outcome variable is the same for all SES.

Table SI.2. Number of days worked outside home as a function of sickness and SES

	Days worked outside home (in last 14 days)
Stratum 3	-0.65*
	[0.33]
Stratum 4	-1.12***
	[0.39]
Strata 5&6	-2.13***
	[0.46]
Strata 1&2 × fraction sick	-1.92***
	[0.52]
Stratum 3 × fraction sick	-2.01***
	[0.37]
Stratum 4 × fraction sick	-1.43**
	[0.57]
Stratum 5&6 × fraction sick	-1.32
	[0.83]
Constant	4.95***
(Mean for SES 1&2 if 0 days sick)	[0.27]
Observations	6,798

Robust standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1, all p-values and F-test are double-sided tests. The sample is restricted to individuals that have been sick for at least one day and observations are weighted by occupation. Linear regression where the explained variable is the respondent's answer to the question "In the last 14 days, how many days have you worked outside of home?". "fraction sick" is equal to the number of days with symptoms in the past 14 days / 14; it goes from 0 to someone who had no symptoms to 1 for someone who had symptoms during the past 14 days.

Table SI.3. Statistical predictions of the number of days worked using results from Table SI.2

Stratum	Prediction if 0 days sick	Prediction with 14/14 days sick
1 & 2	4.95	3.03
3	4.30	2.29
4	3.83	2.40
5 & 6	2.82	1.50

Example of calculations of statistical linear predictions, for stratum 3: 4.95 -0.65 = 4.30 provides the average number of days worked for a person in stratum 3 if the person was sick during 0 days out of 14. 4.95 -0.65 -2.01 =4.30 provides the average number of days worked for a person in stratum 3 if the person was sick during 14 days out of 14. The difference (-2.01) is the effect of being 14 days sick on the number of days worked for stratum 3.

Table SI.4. Model parameters common across all SES

Model Parameter	**Value**	**Source**
Proportion of asymptomatic infections	20%	Buitrago-Garcia et al. (2020)
Asymptomatic Relative Risk	35%	Buitrago-Garcia et al. (2020)
Out-of-home contacts distribution	Negative binomial with $k = 0.58$	Bi et al. (2020)
Incubation period distribution	Lognormal with $\mu = 1.63$ and $\sigma = 0.5$	McAloon et al. (2020)
Serial interval distribution	Gamma with $\alpha = 8.12$, $\beta = 0.64$, and translated by Δx = -7.5	He et al. (2020)
Generation interval distribution	Empirical distribution matching gamma with mean of 5.2 days, shape = 4.72	Inferred from incubation and serial interval distribution
Test sensitivity distribution	Empirical distribution	Grassly et al. (2020)

Table SI.5. Model parameters for each SES

(a) Parameter Summary

Model Parameter	SES 1&2	SES 3	SES 4	SES 5&6	Source / Construction
P(Self Test|Symptoms)	0.188	0.244	0.356	0.341	$d/((1-A)(1-F))$ where d = share detected among positive from Table SI.6, A = share of asymptomatics in the model (20%), F is the estimated share of false negatives in a typical model (22%)
P(Isolation|Symptoms)	0.526	0.520	0.261	0.395	$1-(w_s/w)$, where w_s = the avg. number of days worked after having been sick for 14 days, inferred from Table SI.3, and w is the avg. number of days worked
P(Isolation|Contact traced)	0.300	0.290	0.000	0.106	$1-(w_{ct}/w)$, where w_{ct} = the avg. number of days worked when individual knows about contact with an infected person We assume this is weakly greater than 0
P(Isolation|Test)	0.858	0.858	0.858	0.858	$1-(w_t/w)$, where w_t = the avg. number of days worked when an individual has received a positive test
P(Household quarantine)	0.561	0.503	0.270	0.254	$1-(w_{hh}/w)$, where w_{hh} = the avg. number of days worked when someone in the same household is tested positive (does not vary by SES)
P(Contact traced)	0.189	0.227	0.291	0.325	$c/(n+\lambda^* w)$, where c = average number of contacts traced from Table SI.6, n = Number of non-work contacts outside of home from Table 1, $\lambda^* = 0.8$ outside work factor as described in Section 3, and w is days worked outside of home in Table 1
SAR (Home) for symptomatics	0.146	0.146	0.146	0.146	$SAR_{hh}/(\rho A+(1-A))$ where SAR_{hh} = estimated SAR (Home) from Table SI.6, ρ = the relative risk of asymptomatics (0.35), A is the share of asymptomatics (20%)
SAR (Outside home) for symptomatics	0.303	0.310	0.270	0.121	$SAR_{out}/(\rho A+(1-A))$ where SAR_{out} = estimated SAR (out) from Table SI.6, ρ = the relative risk of asymptomatics (0.35), A is the share of asymptomatics (20%)
K matrix	[See Panel (b)]				Estimated using process seen in Section SI.1.3.1
Household size	[full distribution used]				Observed distribution from census data
Test consultation delay	[full distribution used]				Observed distribution from HSB data
Test results delay	[full distribution used]				Observed distribution from HSB data
Contact tracing delay	[full distribution used]				Assumed to be same for all SES groups and drawn from the observed distribution for test consultation delays

(b) K matrix

Index Case SES	Contact Case SES 1&2	3	4	5&6	Total
1&2	272,155	49,440	5,162	2,432	329,190
3	49,440	102,967	16,864	2,729	172,000
4	5,162	16,864	8,431	6,022	36,480
5&6	2,432	2,729	6,022	5,247	16,430
Total	329,190	172,000	36,480	16,430	554,100