Estimation of COVID-19 cases in Mexico accounting for SARS-CoV-2 RT-PCR false negative results

Isaac J. Núñez, MD
Programa de especialidad en Medicina Interna, Departamento de Educación Médica, Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán
Vasco de Quiroga # 15, Belisario Domínguez Sección XVI, Mexico City, Mexico, PC 14080

Pablo F. Belaunzarán-Zamudio, MD
National Institute of Allergy & Infectious Diseases, National Institutes of Health
31 Center Dr # 7A03, Bethesda, Maryland, United States, PC 20892

Yanink Caro-Vega, PhD
Departamento de Infectología, Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán
Vasco de Quiroga # 15, Belisario Domínguez Sección XVI, Mexico City, Mexico, PC 14080

Corresponding author:
Isaac J. Núñez
isaac.nunezs@incmnzx.mx
Programa de especialidad en Medicina Interna, Departamento de Educación Médica, Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán
Vasco de Quiroga # 15, Belisario Domínguez Sección XVI, Mexico City, Mexico, PC 14080
(+52) 5554870999

ABSTRACT
Underestimation of the number of cases during the COVID-19 pandemic has been a constant concern worldwide. Case confirmation is based on identification of SARS-CoV-2 RNA using real time polymerase chain reaction (RT-PCR) in clinical samples. However, these tests have suboptimal sensitivity, especially during the early and late course of infection. Using open data, we estimated that among 1 343 730 people tested in Mexico since February 27th, there were 838 377 (95% CI 734 605 – 1 057 164) cases, compared with 604 376 considering only positive tests. ICU admissions and deaths were around 16% and 9% higher than reported. Thus, we show that accounting for the sensitivity of SARS-Cov-2 RT-PCR diagnostic tests is a simple way to improve estimations for the true number of COVID-19 cases in tested people, particularly in high-prevalence populations. This could aid to better inform public health measures and reopening policies.

INTRODUCTION
Around 25 485 000 confirmed cases of COVID-19 have been reported worldwide by country governments by August 31st, 2020.1 The Mexican government reported 599 560 confirmed cases to that date.2 Case confirmation is based on identification of SARS-CoV-2 RNA using real time polymerase chain reaction (RT-PCR) in clinical samples collected through nasopharyngeal or oropharyngeal swabs, saliva, or bronchoscopy. Dealing with the disease has proven extremely challenging for governments and health systems worldwide, partially due to the difficulties in case identification. Various measures have been proposed to reduce the viruses’ impact on the population, most of them rely on case identification for isolation and contact tracing.3-5 The World Health Organization (WHO) has highlighted the importance of generalized testing with the goal of early detection, quarantine, and contact tracing. Countries like South Korea and Iceland have been successful in implementing wide-spread testing, case-isolation and contact tracing, keeping the virus under control.6,7 In countries where tests are less available, focusing this resource to high risk people was deemed reasonable as a provisional strategy, with the urge to increase testing capacity. Mexico chose a different strategy, and a decision to use testing only for surveillance purposes...
was made early during the pandemic. Criteria for testing are applied as for the sentinel surveillance system for influenza, and the information provided allegedly used to estimate the total number of infections based on mathematical modelling. As anywhere else, underestimation of the number of cases has been a constant concern.4-7 Diagnostic tests rarely, if ever, are completely reliable, and RT-PCR for SARS-CoV-2 is no exception. Specificity almost always nears 100% in this kind of tests, but poor sensitivity has been an issue. Kucirka and collaborators estimated that test sensitivity is highest at the fourth day of symptoms onset (81% [95% confidence limits; 95%CL 71-88%]).8 Sensitivity is the lowest during the asymptomatic and late symptomatic periods (eg. 37%, 95%CL 26-49%) at day 21 after symptom onset. Other factors that could influence the accuracy of the test is the type of clinical specimen, severity of infection, and gene targets. A combination of these, and other factors, may account for the underestimation of the number of cases and attributable deaths worldwide.8-11 Until screening and diagnostic tests performance are optimized, applying mathematical modelling strategies can aid in estimating more accurately diseases occurrence for surveillance purposes. In this study, we aimed to provide corrected estimates of the number of cases among people that were tested for SARS-CoV-2 in Mexico between February 27th and August 31st, 2020 by taking into account the probability of RT-PCR false negative tests results.

METHODS

Study setting

In Mexico, the first COVID-19 confirmed case was tested on February 27th and reported on February 28th. Community transmission was declared on March 24th and mitigation country-wide measures were announced the same day. Social distancing was urged, and non-essential businesses and activities were suspended, initially until April 14th. The testing strategy was also published that day. A case definition was developed, and testing was recommended for one in ten patients seeking care due to a mild case of an influenza-like illness in a limited number of health services previously established to monitor seasonal influenza, and all of those requiring hospitalization. Also, heavy emphasis on voluntary quarantine if mild symptoms developed, urging people with co-morbidities and other high-risk conditions, such as older age, to search for health care. No accompanying contact tracing measure was spoken of, placing most of the responsibility at the individual level.12

Data sources and selection

We used the SARS-CoV-2 tests open datasets made public since April 12 by the Mexican government in their official coronavirus web page and updated daily.13 The datasets include every test done at public, but not private, laboratories. It contains state and municipality where the test was collected, sociodemographic information, dates of symptoms onset, date the patient was included in the database, ICU admission and death (if occurred) for everyone, with non-traceable, individual key identifiers. We assumed the date the patient was registered in the database was the date of testing, and we will refer to it as such from now on. We included in the analysis all tested individuals registered in the dataset between February 27th and August 31st. Patients with pending result, missing identification code, or more than 21 days with symptoms at the moment of the test were excluded.

We analyzed each information according to the date the tests results are reported in the database, regardless of the day it was collected to follow the format of the daily report by the Ministry of Health (See supplementary). This was not possible for people tested before April 12th and already had a result, so their result was included as the baseline count.

False negative estimation model

We used the method described by Kucirka and collaborators in their mathematical modelling study to calculate the false-negative rate of RT-PCR diagnostic tests. They calculated the sensitivity from day one of infection (assuming symptoms started at the 5th day of infection) until day 21 of infection. Since their estimates end at day 16 of symptoms, we replicated their analysis and estimated sensitivity up to day 21 of symptoms with 95% uncertainty bounds. Sensitivity varied depending on the day after infection, being higher during the symptomatic phase and reaching a maximum of 81% at the fourth day of symptoms (Supp Append Table S1). We used the mean estimate for the graphical representation but...
repeated the estimation with the upper and lower confidence bounds. Specificity for every test used by the Mexican government is reportedly 100%. \cite{8,14}

We used the following contingency table as the basis for our analysis:

<table>
<thead>
<tr>
<th>Test result</th>
<th>Yes</th>
<th>No</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>a</td>
<td>b</td>
<td>a+b</td>
</tr>
<tr>
<td>Negative</td>
<td>c</td>
<td>d</td>
<td>c+d</td>
</tr>
<tr>
<td>Total</td>
<td>a+c</td>
<td>b+d</td>
<td>a+b+c+d</td>
</tr>
</tbody>
</table>

Given a 100% of specificity, there are no false positive results then $b=0$, and all the positive results are true positives. From data set, each day we knew the number of true positives ‘a’ and the number of negative tests ‘c+d’. Our interest is to estimate the daily number of false negatives ‘c’. The probability of being false negative ‘p’ is defined in the equation (1):

$$p = \frac{c}{a+c}$$

$$c = \frac{pa}{1-p}$$

By reproducing Kucirca’s analysis, we know ‘p’ and ‘s’ (1-p) the test sensitivity for each time since symptom initiation reported. At each calendar day, we split individuals in groups, each corresponding to the number of days with symptoms when tested. For example, a day in which 100 people were tested, with 30 presenting on their 6th day and 70 in their 7th day with symptoms, two groups were created, each one with a test sensitivity (s) and the number of reported positive tests (a). Hence, we applied the equation to each group of every calendar day from February 27th to August 31st and added the false negatives calculated on every group. As the number of true COVID-19 cases is limited by the number of people tested, in case the estimation yielded a higher number of cases then the totality of people tested was used instead.

We also estimated the corrected number of ICU admission and the corrected number of deaths due to COVID-19 by calendar day. Assuming no difference in test precision among disease severity spectrum, we added the product of the proportion of negatives estimated to be false negatives and the number of ICU admissions or deaths among COVID negative patients. We applied the following equation to correct death and ICU admissions:

$$Estimated\ true\ COVID19\ deaths = official\ COVID19\ positive\ deaths + (proportion\ of\ false\ negatives * official\ COVID19\ negative\ deaths)$$

Thus, if in a given day there were 100 deaths and 50 ICU admissions among COVID-19 negative patients, and the estimated false negative proportion using the mean estimate of the test sensitivity was of 0.40, we would add 40 deaths and 20 ICU admissions to the COVID-19 positive group for that particular date.

**Statistical Analysis**

Applying results derived from the false negative estimation model on the government official dataset, we estimated the daily corrected number of cases. We performed the analysis at a National level and for each one of the 31 States and the country capital (Ciudad de México, formerly known as Distrito Federal).

To determine if positivity rates could be due to low testing per capita, we calculated Spearman’s rho of positivity rates and the number of tests done per 10,000 habitants by state. State population was obtained from the national statistics and geography institute’s (INEGI) most recent published data. \cite{15} We also calculated the 7-day moving average of time from
symptom onset to testing for the entire study period to determine if this could explain higher false negative rates during certain time periods.

All data analysis was done with R software version 4.0.0. The ethics committee of the Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán reviewed and approved the study. There was no sponsor involved in any step of the study.

RESULTS

There were 1,343,730 people tested between February 27th and August 31st according to the latest official database (August 31st). A detailed explanation of data selection is provided in supplementary materials. We included 1,280,910 patients that had an available result, were tested at less than or 21 days with symptoms and had no missing ID code. Of them, 604,376 (47.2%) were SARS-CoV-2 positive and 676,534 (52.8%) negative (Table 1).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Official count</th>
<th>Expected scenario</th>
<th>Best case scenario</th>
<th>Worst case scenario</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive tests</td>
<td>1,280,910</td>
<td>1,280,910</td>
<td>1,280,910</td>
<td>1,280,910</td>
</tr>
<tr>
<td>Negative tests</td>
<td>676,534 (52.8%)</td>
<td>442,533 (34.5%)</td>
<td>546,305 (42.6%)</td>
<td>223,746 (17.5%)</td>
</tr>
<tr>
<td>Estimated false negative tests</td>
<td>-</td>
<td>234,001 (18.3%)</td>
<td>130,229 (10.2%)</td>
<td>452,788 (35.3%)</td>
</tr>
<tr>
<td>Overall ICU count</td>
<td>13,038</td>
<td>15,085</td>
<td>14,167</td>
<td>17,008</td>
</tr>
<tr>
<td>Overall Death count</td>
<td>64,360</td>
<td>69,835</td>
<td>67,406</td>
<td>74,978</td>
</tr>
</tbody>
</table>

Note: Expected, best and worst-case scenario refer to estimations done with the mean test sensitivity and its upper and lower 95% confidence limits, respectively. Positive tests in the expected, best case and worst-case scenarios include estimated false negatives, while negative tests exclude them.

Estimated false negatives and corrected COVID-19 cases

We estimated a total of 838,377 (95% CI 734,605 – 1,057,164) positive cases (39% higher than official reported cases) (Figure 1). In our corrected estimates, 50,000 cases were reached by May 11th (May 7th-May 14th), while the official count reached that number at May 18th. Accumulated test positivity rate increased during the study period, being of 14.4% at March 24th, 36% at April 12th, and 50.1% by July 14th. Positivity rate also varied according to state and date (Table 2, Supp Material Figures SE). The states with the most cases were Ciudad de México, Estado de México, Guanajuato, Nuevo León, and Puebla, while those with the highest positivity rate were Veracruz, Oaxaca, Baja California, Quintana Roo and Hidalgo.

Spearman’s Rho for test positivity rate (taking false negatives into account) and tests performed per 10,000 people was of -0.41 (-0.40, -0.41). Time from symptom onset to testing increased over time (Supp Material Figure S1).

Corrected ICU hospitalization and deaths estimates

There were 13,038 ICU admissions and 64,360 deaths among COVID-19 positive patients during the study period (Table 1). The corrected estimate is of 15,085 ICU admissions (14,167 – 17,008) and 69,835 deaths (67,406 – 74,978). (Table 1). The magnitude of difference between official reports and corrected estimates varied between states (Table 2).

DISCUSSION

In this analysis, we estimated the corrected number of cases of COVID-19, ICU admissions and deaths due to COVID-19 in Mexico accounting for false negatives tests results using test sensitivities previously estimated by Kucirca and colleagues. These were estimated based on the day after the onset of symptoms when patients are tested. We identified that the number of cases of COVID-19 in Mexico based on RT-PCR testing might be almost 40% higher than currently registered, with 95% confidence limits of 21% and 75% depending on the test sensitivity. These differences vary widely by state and
period during the pandemic. Accordingly, the corrected number of ICU admission and deaths increased around 16% and 9%, but this increase might be as high as 25% in ICU and 17% in deaths. The magnitude of these differences may require important modifications in preparedness for response, which highlights the importance of accounting for the probability of false negative tests in public health estimations.

Figure 1. Estimated proportion of tested individuals with a false negative result during the study period

A) represents new daily test results, B) represents accumulated test results

The proportion of false negatives was similar to that found in other studies. \textsuperscript{16, 17} There were high heterogeneity in positivity tests proportions, corrected estimates, and confirmed and corrected estimations of ICU hospitalizations and deaths across States. A modelling study conducted in the United States showed that disease burden varied heavily among counties, both in an optimistic and in a pessimistic scenario. \textsuperscript{16} A recent seroepidemiological study conducted in Spain found considerable heterogeneity in seroprevalence among provinces, with >10% in the most heavily affected ones and less than one percent in the least. \textsuperscript{19} This data is consistent not only with the occurrence of “local epidemics” rather than a nation-wide epidemic; but also the fragmented response in Mexico with some regions faring better than others. On the last point, it may only reflect sound State-centered approaches, with resources being modified according to each states’ needs, though.

We also observed a weak negative correlation between test positivity and tests-per-capita. Considering that testing per-capita is very low in the whole country, this is not surprising. States that have particularly low testing rates are Chiapas, Chihuahua, Oaxaca, Veracruz, and Queretaro. This could be responsible for the high estimated positivity rate. Sonora and Baja California have more tests done per 10,000 habitants, but their estimated positivity rate is still of ~90%. Most likely the number of cases far outnumber the number of tests, with the relatively small differences in tests-per-capita being apparently inconsequential for the positivity rate. Interestingly, the time from symptomonset to testing did not appear to
change considerably over time. This suggests that increasing positivity rates most likely do not derive from changes in time of testing after symptoms onset along time but from high disease burden and insufficient testing worsening over time.

Considering that only a small proportion of symptomatic cases who search for healthcare are tested (not even considering asymptomatic individuals who have the virus) the true underestimation of COVID-19 cases can be huge. Given this, the daily number of cases will most likely grow on par with the number of tests.

### TABLE 2 Official and estimated confirmed state-wide COVID-19 cases

<table>
<thead>
<tr>
<th>State</th>
<th>Total tests</th>
<th>Tests per 10000 hab</th>
<th>Reported cases (positivity rate)</th>
<th>Reported cases (95% CI)</th>
<th>Estimated cases</th>
<th>Estimated cases (95% CI)</th>
<th>Estimated COVID-19 admissions</th>
<th>Estimated COVID-19 admissions (95% CI)</th>
<th>Estimated deaths</th>
<th>Estimated deaths (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>VERACRUZ</td>
<td>37,560</td>
<td></td>
<td>28,053 (67.3%)</td>
<td>28,053</td>
<td>27,956</td>
<td>27,956</td>
<td>27,063</td>
<td>27,063</td>
<td>191</td>
<td>191</td>
</tr>
<tr>
<td>OAXACA</td>
<td>13,781</td>
<td></td>
<td>11,204 (68.0%)</td>
<td>11,204</td>
<td>10,995</td>
<td>10,995</td>
<td>10,199</td>
<td>10,199</td>
<td>106</td>
<td>106</td>
</tr>
<tr>
<td>BAJA CALIFORNIA SUR</td>
<td>17,197</td>
<td></td>
<td>15,308 (60.3%)</td>
<td>15,308</td>
<td>14,904</td>
<td>14,904</td>
<td>14,100</td>
<td>14,100</td>
<td>106</td>
<td>106</td>
</tr>
<tr>
<td>COAHUILA</td>
<td>20,009</td>
<td></td>
<td>17,069 (85.8%)</td>
<td>17,069</td>
<td>16,579</td>
<td>16,579</td>
<td>15,785</td>
<td>15,785</td>
<td>126</td>
<td>126</td>
</tr>
<tr>
<td>QUERETARO</td>
<td>10,366</td>
<td></td>
<td>8,847 (65.1%)</td>
<td>8,847</td>
<td>8,459</td>
<td>8,459</td>
<td>7,881</td>
<td>7,881</td>
<td>104</td>
<td>104</td>
</tr>
<tr>
<td>Hidalgo</td>
<td>10,095</td>
<td></td>
<td>8,283 (82.0%)</td>
<td>8,283</td>
<td>7,953</td>
<td>7,953</td>
<td>7,421</td>
<td>7,421</td>
<td>122</td>
<td>122</td>
</tr>
<tr>
<td>Baja California Norte</td>
<td>16,191</td>
<td></td>
<td>13,680 (85.7%)</td>
<td>13,680</td>
<td>13,254</td>
<td>13,254</td>
<td>12,635</td>
<td>12,635</td>
<td>122</td>
<td>122</td>
</tr>
<tr>
<td>Sonora</td>
<td>21,210</td>
<td></td>
<td>17,818 (84.2%)</td>
<td>17,818</td>
<td>17,324</td>
<td>17,324</td>
<td>16,703</td>
<td>16,703</td>
<td>140</td>
<td>140</td>
</tr>
<tr>
<td>CHIHUAHUA</td>
<td>6,148</td>
<td></td>
<td>5,316 (86.1%)</td>
<td>5,316</td>
<td>4,944</td>
<td>4,944</td>
<td>4,535</td>
<td>4,535</td>
<td>134</td>
<td>134</td>
</tr>
<tr>
<td>GUERRERO</td>
<td>30,184</td>
<td></td>
<td>24,704 (82.1%)</td>
<td>24,704</td>
<td>23,670</td>
<td>23,670</td>
<td>22,528</td>
<td>22,528</td>
<td>167</td>
<td>167</td>
</tr>
<tr>
<td>TAMAULIPAS</td>
<td>28,331</td>
<td></td>
<td>22,749 (80.5%)</td>
<td>22,749</td>
<td>21,728</td>
<td>21,728</td>
<td>20,654</td>
<td>20,654</td>
<td>167</td>
<td>167</td>
</tr>
<tr>
<td>Coahuila</td>
<td>5,365</td>
<td></td>
<td>4,520 (84.8%)</td>
<td>4,520</td>
<td>4,207</td>
<td>4,207</td>
<td>4,012</td>
<td>4,012</td>
<td>66</td>
<td>66</td>
</tr>
<tr>
<td>ESTADO DE MEXICO</td>
<td>49,667</td>
<td></td>
<td>40,387 (81.3%)</td>
<td>40,387</td>
<td>38,665</td>
<td>38,665</td>
<td>37,319</td>
<td>37,319</td>
<td>224</td>
<td>224</td>
</tr>
<tr>
<td>BURGOS</td>
<td>24,146</td>
<td></td>
<td>19,697 (81.5%)</td>
<td>19,697</td>
<td>18,793</td>
<td>18,793</td>
<td>17,859</td>
<td>17,859</td>
<td>126</td>
<td>126</td>
</tr>
<tr>
<td>CHIHUAHUA</td>
<td>3,848</td>
<td></td>
<td>3,210 (83.1%)</td>
<td>3,210</td>
<td>3,104</td>
<td>3,104</td>
<td>2,926</td>
<td>2,926</td>
<td>28</td>
<td>28</td>
</tr>
<tr>
<td>ESTADO DE MEXICO</td>
<td>4,459</td>
<td></td>
<td>3,738 (83.9%)</td>
<td>3,738</td>
<td>3,665</td>
<td>3,665</td>
<td>3,516</td>
<td>3,516</td>
<td>34</td>
<td>34</td>
</tr>
<tr>
<td>CHIHUAHUA</td>
<td>2,818</td>
<td></td>
<td>2,380 (84.9%)</td>
<td>2,380</td>
<td>2,300</td>
<td>2,300</td>
<td>2,146</td>
<td>2,146</td>
<td>43</td>
<td>43</td>
</tr>
<tr>
<td>ESTADO DE MEXICO</td>
<td>2,818</td>
<td></td>
<td>2,380 (84.9%)</td>
<td>2,380</td>
<td>2,300</td>
<td>2,300</td>
<td>2,146</td>
<td>2,146</td>
<td>43</td>
<td>43</td>
</tr>
<tr>
<td>COAHUILA</td>
<td>2,818</td>
<td></td>
<td>2,380 (84.9%)</td>
<td>2,380</td>
<td>2,300</td>
<td>2,300</td>
<td>2,146</td>
<td>2,146</td>
<td>43</td>
<td>43</td>
</tr>
<tr>
<td>ESTADO DE MEXICO</td>
<td>2,818</td>
<td></td>
<td>2,380 (84.9%)</td>
<td>2,380</td>
<td>2,300</td>
<td>2,300</td>
<td>2,146</td>
<td>2,146</td>
<td>43</td>
<td>43</td>
</tr>
<tr>
<td>COAHUILA</td>
<td>2,818</td>
<td></td>
<td>2,380 (84.9%)</td>
<td>2,380</td>
<td>2,300</td>
<td>2,300</td>
<td>2,146</td>
<td>2,146</td>
<td>43</td>
<td>43</td>
</tr>
<tr>
<td>ESTADO DE MEXICO</td>
<td>2,818</td>
<td></td>
<td>2,380 (84.9%)</td>
<td>2,380</td>
<td>2,300</td>
<td>2,300</td>
<td>2,146</td>
<td>2,146</td>
<td>43</td>
<td>43</td>
</tr>
<tr>
<td>COAHUILA</td>
<td>2,818</td>
<td></td>
<td>2,380 (84.9%)</td>
<td>2,380</td>
<td>2,300</td>
<td>2,300</td>
<td>2,146</td>
<td>2,146</td>
<td>43</td>
<td>43</td>
</tr>
<tr>
<td>ESTADO DE MEXICO</td>
<td>2,818</td>
<td></td>
<td>2,380 (84.9%)</td>
<td>2,380</td>
<td>2,300</td>
<td>2,300</td>
<td>2,146</td>
<td>2,146</td>
<td>43</td>
<td>43</td>
</tr>
<tr>
<td>COAHUILA</td>
<td>2,818</td>
<td></td>
<td>2,380 (84.9%)</td>
<td>2,380</td>
<td>2,300</td>
<td>2,300</td>
<td>2,146</td>
<td>2,146</td>
<td>43</td>
<td>43</td>
</tr>
<tr>
<td>ESTADO DE MEXICO</td>
<td>2,818</td>
<td></td>
<td>2,380 (84.9%)</td>
<td>2,380</td>
<td>2,300</td>
<td>2,300</td>
<td>2,146</td>
<td>2,146</td>
<td>43</td>
<td>43</td>
</tr>
<tr>
<td>COAHUILA</td>
<td>2,818</td>
<td></td>
<td>2,380 (84.9%)</td>
<td>2,380</td>
<td>2,300</td>
<td>2,300</td>
<td>2,146</td>
<td>2,146</td>
<td>43</td>
<td>43</td>
</tr>
</tbody>
</table>

Note: State-wise per-occurrence date analysis, study period February 27th — August 31st. Ordered by estimated positivity rate (highest to lowest)
This means that with current testing capacity it is not possible to grasp the behavior of the pandemic, as the number of tests is so small and the positivity rate so high that it would be fully dependent on them. The World Health Organization recommends a positivity rate lower than 5%, among other criteria, to commence reopening, even if a sentinel system is being used. Mexico is currently far from a safe reopening, and further still if we consider false negatives.\textsuperscript{20, 21}

False negatives are accounted for in clinical medicine when a clinician suspects it in a patient that has a negative test but other disease indicators, such as suggestive lung images, that generate a convincing clinical scenario, and acts accordingly.\textsuperscript{22} As we show here, false negatives should also be accounted for in public health estimations, and it is also possible to act accordingly. Places with low prevalence, as in the states with low positivity rate (none in our case) or with massive testing strategies, will have a small amount of cases added to their official counts. This contrasts with the picture of Mexico as a whole, were the worrisome positivity rate increases even more when false negatives are considered.\textsuperscript{23, 24}

Currently, a reopening strategy based on a four-coloured traffic light (red, orange, yellow, and green) is being implemented, which assigns each state a colour based on several variables. Test positivity rate is among the criteria for changing the colour. Nonetheless, by not considering false negatives the positivity rate is being underestimated and it is likely this could influence the premature modification of the colour, and thus premature reopening, which could cause a new surge in cases.\textsuperscript{25-27} We observed a one-week delay in reaching a similar number of cases when mean false-negative test is not considered.

This approach as applied for Mexico has several limitations. As only a few symptomatic people are tested we cannot estimate the corrected number of cases among the total population. This would require a vastly higher number of tests. Thus, there should be caution when interpreting our results. Our estimation presents a limited view of how much more cases there are, and thus do not represent the actual number of cases in the country; in any case our corrected estimates are in the conservative side.\textsuperscript{14, 26, 26} We assumed that the date on which the information was captured in the system is the same on which the sample was procured, but there might be a small-time lag between one and another. We also identified limitations related to the diagnostic tests. For instance, there is a wide catalog of SARS-Cov-2 RT-PCR tests currently used in our country. Even if all of them have to be approved by the Institute of Diagnosis and Epidemiological Reference (INDRE) for surveillance purposes and all must comply with the Berlin protocol, we do not know how they compare to each other, and variation in diagnostic accuracy probably exists. We also do not have information on the anatomical test site, be it nasal or oropharyngeal swab, saliva or bronchoscopy sample, which could affect test sensitivity. We do not account for severity of the clinical picture on the test sensitivity. Sensitivity of the test in patients who develop pneumonia and/or critical illness could be higher than in those with less severe disease, and patients with severe disease is overrepresented in the testing strategy followed in Mexico. Even when considering all these limitations, the application of the test performance to correct the number of cases could certainly improve surveillance. The method we used can be easily adapted to other countries or areas. Our analysis can be updated if COVID-19 open data continues to be published, and thus be used to better inform decision making at the National and State level.

**Conclusion**

While it may very well be impossible to determine which patients had the false negative tests, taking the test precision into account is an effective way to improve estimations for the number of COVID-19 cases among a tested population, especially when the testing is done mainly in high-prevalence populations. We expect this could aid to better inform public health measures and reopening policies.

**Acknowledgements**

We would like to thank all the health-workers in our country for their tremendous work in caring for patients during this pandemic. Also, we thank the people behind the surveillance of the pandemic, as without them this work could not have come to be.
Declaration of interests

All authors state they have no conflict of interests.

Author contributions

Isaac J. Núñez: idea conception, literature search, study design, data curation, data analysis, data interpretation, writing

Pablo F. Belaunzarán-Zamudio: literature search, study design, data analysis and interpretation, writing

Yanink Caro-Vega: literature search, study design, data curation, data analysis, data interpretation, writing

Data sharing

All code utilized is available at https://github.com/isaac-nunez/covid_19_fn_estimates_mexico and all datasets are available at the official government COVID-19 webpage. 13

References


23. Versión estenográfica de la conferencia de prensa matutina, Gobierno de México. Tuesday, May 12th, 2020. Available at: https://www.gob.mx/presidencia/es/articulos/version-estenografica-de-la-conferencia-de-prensa-matutina-martes-12-de-mayo-de-2020?idiom=es


25. Datos Abiertos Ciudad de México, Capacidad hospitalaria ZMVM. Gobierno de la Ciudad de México. Available at: https://datos.cdmx.gob.mx/explore/dataset/capacidad-hospitalaria/table/ [Accessed 24th July 2020]


