Health Sentinel: a mobile crowdsourcing platform for self-reported surveys provides early detection of COVID-19 clusters in San Luis Potosí, Mexico

Overview of Revisions:

We thank the reviewers for their insightful comments with respect to our work. We would like to bring to the attention of the program committee the following major changes we've made in response to reviews:

- A thorough description of the VIRUS program in Section 2.1, including their core mission, resources, key tools and intended impact.
- An in-depth description of the informational resources provided by the CDS for participants. This has been added to Section 2.3.2
- A new Section (3.3) dedicated entirely to policy implications (past, present and future) of the CDS work. This includes ways in which the CDS directly affected policy taken by the SSSLP as well as future collaborations.
- Our discussion Section (4) has been split into 3 components. The first is dedicated to core results, with new material focusing on sampling bias. The second subsection (4.2) is dedicated entirely to our reflections on our experience with the research-to-practice pipeline in the Mexican context, and the final subsection contains material on future work and applications of the CDS technology and framework.

Health Sentinel: a mobile crowdsourcing platform for self-reported surveys provides early detection of COVID-19 clusters in San Luis Potosí, Mexico

Abstract

Background. The *Health Sentinel* (*Centinela de la Salud*, *CDS*), a mobile crowdsourcing platform that includes the *CDS* app, was deployed to assess its utility as a tool for COVID-19 surveillance in San Luis Potosí (SLP), Mexico.

Methods. The *CDS* app allowed anonymized individual surveys of demographic features and COVID-19 risk of transmission and exacerbation factors from users of the San Luis Potosí Metropolitan Area (SLPMA). The platform's data processing pipeline computed and geolocalized the risk index of each user, and enabled the analysis of the variables and their association. Point process analysis identified geographic clustering patterns of users at risk and compared these with patterns of COVID-19 cases confirmed by the State Health Services.
Results. A total of 1,554 COVID-19 surveys were administered through the *CDS* app. Among the respondents, 50.4% were men and 49.6% women, with an overall average age of 33.5 years. Overall risk index frequencies were, in descending order: no-risk 77.8%, low risk 10.6%, respiratory symptoms 6.7%, medium risk 1.4%, high risk 2.0%, very high risk 1.5%. Comorbidity was the most frequent vulnerability category (32.4%), followed by the inability to keep home lockdown (19.2%). Statistically significant risk clusters identified at a spatial scale between 5 and 730 meters coincided with those in neighborhoods containing substantial numbers of confirmed COVID-19 cases.

Conclusions. The *CDS* platform enables the analysis of the sociodemographic features and spatial distribution of individual risk indexes of COVID-19 transmission and exacerbation. It is a useful epidemiological surveillance and prediction tool because it detects statistically significant and consistent risk clusters in neighborhoods with a substantial number of confirmed COVID-19 cases. This tool was a key data point in decision making of public health for SLP during the initial stages of the pandemic. Through ongoing partnerships between our academics and policy makers at all levels, we hope that future iterations of the tool will continue to inform public health policy. Finally, we reflect upon our experience collaborating with public health officials in the Mexican context and share some observations and learning points for the future.

Keywords: COVID-19, SARS-CoV-2, personal risk index, vulnerability, mobile crowdsourcing platform, geolocalized survey, point pattern distribution, risk cluster, infection hotspot.

1. Introduction

The coronavirus disease 2019 (COVID-19) epidemic in Mexico (pop. 135 million) is part of the ongoing pandemic caused by the Severe Acute Respiratory Syndrome Coronavirus Virus 2, SARS-CoV-2 [1]. Mexican health authorities confirmed the first COVID-19 case in Mexico City on February 27, 2020 [2], and declared a health emergency on April 9, 2020, with interventions to avoid healthcare saturation and promote home lockdown to prevent viral spreading [3]. At the time of writing there have been over 642,860 confirmed COVID-19 cases, 451,159 recovered patients, 68,484 deaths, a recovery ratio of 70.2%, and a case fatality ratio (CFR) of 10.7% in Mexico, the country with the second-highest CFR in the world [4], that denote the pandemic's impact on the country. The disease quickly spread across Mexican cities, and on March 13, health authorities confirmed the first case in the San Luis Potosí Metropolitan Area (SLPMA, pop. 1.2 million). At the time of writing, the state of San Luis Potosí had 19,888 confirmed cases, 1,319 deaths, and 15,831 recovered patients.

Coronavirus infection spreads in clusters [5], as individuals pass the virus to many neighbors, resulting in geographically localized infections at the level of a city or a county [6–8]. Mapping was useful in China, Taiwan, Korea, and Israel to explore coronavirus transmission characterized by spatial features distinctive of early infection scattering, community spread, and full-scale outbreak [9,10]

In Mexico only 0.4 qRT-PCR COVID-19 tests per 1,000 inhabitants have been administered, a proportion 25 and 30 times lower than that of South Korea and the US, respectively [11]. The limited resources to conduct extensive diagnostic tests for early detection of SARS-CoV-2 infection hotspots in San Luis Potosí prompted us to develop the *Health Sentinel* (*CDS*), a mobile crowdsourcing platform with an app for real-time surveys of self-reported COVID-19 risk factors from the users. The *CDS* platform could predict actual COVID-19 clusters since the spatial distribution and evolution of the risk indexes from SLPMA users' coincided with those of the COVID-19 cases confirmed later by the local health authorities.

2. Methods

2.1 Study design

This study stemmed from the *VIRUS Program* (Surveillance of Respiratory Infections Creating Health Units, http://www.genomica.uaslp.mx/Research/COVID-19/Programa-Virus.pdf) of the School of Medicine of the Autonomous University of San Luis Potosí, started in March 2020, when the COVID-19 pandemic was declared in Mexico. It is an interventional model inspired by the United Nations Human Security methodology [12] to face the emergency caused by SARS-CoV-2 based on an action scheme of "Health Units" (construction of local health systems and community capabilities based on primary health care, health promotion, prevention, and protection) to reduce the risks of biological agents in vulnerable localities.

The *VIRUS Program* interventional model focuses on collective community actions to address the pandemic using an accumulated risk approach that considers the population's comorbidities and chemical/physical/biological threats to quantify the community's accumulated risk associated with COVID-19.
A multimodal vulnerability index enables the computation of accumulated risk to integrate four key elements: comorbidities, access to health services, economic status, and chemical, physical and biological co-threats. Comorbidity indicators include food security, absolute index of migratory intensity, municipal overcrowding, and prevalence of hypertension, obesity, diabetes, pneumonia/influenza, acute respiratory infections, and COVID-19. Based on the vulnerability index of each geographic area in the state of San Luis Potosí, a three-phase program was implemented with an initial focus on the Huasteca Potosina region and San Luis Potosí City.

The program's first phase consists of a detailed characterization of risk scenarios from vulnerable communities using data collected previously. The second phase consists of organizing a community-based health care system using two main strategies: effective risk communication with emphasis on human rights, and development of human resources through educational programs and technical training. The community-based health system integrates mobile technology (*CDS* app) for COVID-19 community triage, and a telehealth system to enable technical support to the community. The third phase consists of strengthening the creation of horizontal community networks, developing bottom-up solutions to problems, and facilitating the top-to-bottom flow of economic resources. All phases are led by a task force of four groups: *Epidemiological Intelligence, Total Health, Social Toxicology,* and *Health for Peace*. The *Epidemiological Intelligence* group focuses on envisioning epidemiological scenarios and developing tools to support community risk assessment programs for effective prevention and timely intervention (Diaz-Barriga, et al. 2020).

2.2 *CDS* architecture

The *CDS* platform has a Service-Oriented Architecture approach with three systems: a backend server enabling the system's logic and data flow control, the *CDS* mobile application, and the front-end server, which provided geographic information services and a web app. Two main system layers support a modular design of the backend. An API endpoint implements data acquisition rules that maintain data integrity and provide data transfer security [13]. The second layer enables the execution of business rules processes abstracted through Create-Read-Update-Delete operations [14]. The front-end server implementation follows software architecture patterns for data-driven applications facilitating the data structure's management and raw data storage. A Model-View-Presenter architecture implements the *CDS* app in a fashion suitable to efficiently handle background processes. The platform includes a web service that feeds a data processing pipeline, enabling the collected data's geospatial analysis. The *CDS* app was deployed on 1st April 2020 and became freely available through Google Play Store on 1st May 2020.

2.3 Setting and participants

2.3.1 Local context

The SLPMA includes the neighboring Soledad de Graciano Sánchez and other surrounding municipalities and has an estimated 1,221,526 inhabitants. SLPMA is subdivided into primary geostatistical areas (also known locally via their Spanish acronym, AGEBs), the basic units of the Mexican Geostatistical Framework which are the territorial extensions corresponding to subdivisions of urban or rural municipal geostatistical areas [15].

2.3.2 Participants

The *VIRUS Program* launched a social media campaign through Facebook and WhatsApp in mid-March 2020 to distribute the *CDS* app until it was available through the Google Store (early May). The campaign informed the public about the following key COVID-19 topics:

1. Signs and symptoms, personal and community risk factors.
2. Prevention of transmission at work, at home, within public transportation, at restaurants, and other means and locations.
3. Official information published by local and national health authorities.
4. National and regional public health campaigns, online conferences and workshops offered by health specialists.
5. Scientific advances published by national and international organizations (WHO, UNICEF, OIT, IHME, Johns Hopkins Coronavirus Resource Center, and others).
6. Risk scenarios developed by *VIRUS Program* specialists.
7. Feedback to *CDS* users.
8. Vaccine developments.
9. Guidelines to maintain physical and mental health.

This study includes the self-reported risk data collected by Android smartphone volunteer users of the SLPMA between 1st April 2020 and 31st July 2020. From the 9,971 anonymized COVID-19 cases confirmed through qRT-PCR tests by the state health authorities (known locally by the acronym SSSLP) between 12th March 2020 and 3rd August 2020, in the SLPMA, only 9,453 (94.8%) with known postal addresses were included in this study.

2.4 Personal COVID-19 risk index assessment

The *CDS* app enables the administration of 11 questions. Three correspond to the user's profile, and eight to risk items corresponding to the operational guidelines for COVID-19 epidemiological surveillance of the World Health Organization [16] and the Mexican System for Epidemiological Surveillance [17] that were available at the time of *CDS* development.

The personal profile includes the user's gender, age, and occupation. Risk index calculation is based on three categories of variables: contact, signs-and-symptoms, and

comorbidity. Contact variables consider whether the user recently traveled to a country with a COVID-19 epidemic, has had close contact with someone who recently returned from abroad, or is someone with suspected or confirmed infection. Signs and symptoms include cough and fever above 37.5 °C, six general symptoms, and two respiratory symptoms. Comorbidity variables include risk factors that can aggravate the disease and considers whether the user belongs to one or more risk groups or has one or more concomitant illnesses that can worsen COVID-19. Risk groups are those of pregnant women and people aged 65 years or more. Concomitant diseases include diabetes, hypertension, heart problems, liver disease, asthma, chronic pulmonary obstructive disease (COPD), HIV/AIDS, cancer, chronic kidney failure, obesity, and smoking. Two questions address social vulnerabilities: whether the user can keep lockdown and have a supportive network if affected by COVID-19.

An algorithm computes the personal risk index with the answers provided by the user. The values assigned to the risk factor categories are personal contact, signs and symptoms, and vulnerability. Personal risk indexes are, in order of increasing severity: risk-free, low-risk, moderate-risk, high- risk, severe-risk, and respiratory-risk. Based on the personal risk index, users were subdivided into risk-free cases, risk cases, and respiratory-risk cases (Table 1).
Depending on an individual's risk index, the *CDS* app issues health authorities' recommendations and provides a hotline phone and links to reliable COVID-19 information sources.

2.5 Georeferenced database

Data collected through the *CDS* app consists of the georeferenced contributions stored into a non-relational database within the *CDS* platform. A geographic information system enables the display of data layers, including population density, poverty index, overcrowding, medical center locations, business locations, and contributing users' location. A web application allows real- time data visualization through a dashboard (Fig. 1) and descriptive statistics of the survey variables.

2.6 Spatial data analysis

Point process analysis [17] was performed to determine *CDS* users' contributions' geographical features. Intensity functions were obtained for risk-free and risk cases using standard kernel density estimation techniques. Ripley's *K*-function and Monte Carlo simulations explored risk cases' geographic clustering. A K-function is a good indicator of spatial point structures [18] since its values for a given scale *h* are proportional to the average number of additional points within distance *h* of an arbitrary point in a point set. In practice, it is convenient to use the *L*- function for spatial analyses. This function, computed from the *K*-function, determines whether the point set shows clustering properties ($L > 0$) or dispersion properties ($L < 0$) at a given scale *h*. *K and L*-functions are useful to determine whether a point set fulfills the Complete Spatial Randomness (CSR) hypothesis. If data exhibit CSR, there is no underlying structure, and little to be gained from further analysis. The presence of point clusters ($L > 0$) disproves the CSR hypothesis.

Different CSR models are useful in practice, such as the CSR homogeneous model, which assumes a uniform probability measure describing a single random point [18]. A homogeneous set of points in the plane is such that approximately the same number of points occurs in any circular region of a given area; a set of points lacking homogeneity may show spatial clustering at a specific spatial scale. The CSR nonhomogeneous model assumes a probability measure distinct from the uniform distribution. A concrete example is the probability measure defined by the population distribution over a geographic area [19].

2.7 Ethical considerations

The SSSLP registered the *CDS* app in the Google Play Store after checking its compliance with institutional data privacy, security, and ethical requirements. To activate the app, users had to read and agree with the terms of use and consent form. The GPS location of users completing the survey was the only personally identifiable information collected. All *CDS* data remain in a Tier 3 data center of the National Supercomputer Center. Data storing followed strict security and privacy protocols compliant with Mexican law. Only researchers and health officials participating in the study had access to the pilot's data.

3. Results

Contributors to *CDS* surveys resided in Mexico (n = 2,125) and other countries (United States, Bolivia, Chile, Spain, Argentina, France, and Bangladesh; n = 64). Only SLPMA users (n = 1,554) were included in this study. The overall mean age ± SD of users was 33.5 ± 11.5 years, with a median of 30.0 years (Table 2).

3.1 Spatial analysis of the risk cases

3.1.1 Geographic location

The map in Fig. 2 shows the AGEBs with risk cases. Each AGEB is color-coded according to the number of contributions recorded within it. The color palette used in each map shows a specific risk level.

3.1.2 Intensity functions in geographic space

The point process analysis of the risk cases (represented in maps as black plus signs and shaded green areas) show darker areas indicating higher intensity. Fig. 3 depicts the spatial distribution of all users, no-risk cases, and risk cases. A red plus sign indicates the location of maximum empirical intensities. These functions consider kernel density estimation techniques with a bandwidth h = 1200 m. Lower bandwidth values are more likely to overfit data.

To check our results' consistency, intensity functions were computed in two periods along four months: Period I (first two months, April 1-May 31; 998 users), and Period II (all four months, April 1-July 31; 1,554 users).

The intensity functions computed for Period I are similar to those of Period II (Fig. 3). The Period I risk cases' maximum intensity point was located at 22° 8' 18.276" N, 100° 58' 15.708" W, quite close (300 m) to the maximum intensity location of the Period II data.

The maximum intensity area for risk cases is located across the street of the ISSSTE General Hospital (22° 8' 22.5456" N, 100° 58' 10.3476" W), a few meters from a highly-transited pedestrian street (Calzada de Guadalupe), a children's park, the Center for the Arts, the Red Cross, and the Military Zone garrison.

3.1.3 Spatial point patterns

Ripley's *K* and *L* functions were applied to test whether the point patterns of risk cases (Fig. 4) fulfill the nonhomogeneous CSR hypothesis. The reference probability measure was given by the AGEBs population density to avoid detecting cluster patterns reflecting population density features from Periods I and II.

3.1.3.1 Period I patterns

Fig. 4a shows the *L*-function (solid line) and upper and lower envelopes (dashed lines) for the risk cases. The *L* function is above the upper envelope at scales between 5 and 600 m, suggesting significant clustering features ($L > L_{up} > 0$, $p < 0.001$). Fig. 4b shows the *p*-values for each *h* value. The red dashed line represents a *p*-value of 0.01.

Using the K function over a grid combined with MC techniques enables computation of *p*-values over the geographic space to reveal point clusters at different scales. A reference probability measure given by the AGEB´s population densities was also considered. The top row of Fig. 5 shows color-labeled geographic areas with *p*-values calculated at different scales: (a) $h = 300$ m, (b) $h = 450$ m, and (c) $h = 600$ m. The central row of Fig. 5 shows the same geographic areas with $p < 0.05$ values to emphasize point clusters. Many clusters detected at $h = 300$ m grow across other scales (compare Figures 5d, 5e, 5f).

3.1.3.2 Period II patterns

Fig. 4c shows the *L*-function (solid line) and the upper and lower envelopes (dashed lines) for the risk cases. The *L* function is above the upper envelope at scales between 5 and 730 m, suggesting significant clustering features ($L > L_{up} > 0$, $p < 0.001$). Fig. 4d shows the *p*-values for each *h* value in the analysis range.

The bottom row of Fig. 5 shows point cluster analysis at different scales for Period II. the geographic areas for which $p < 0.05$ show clusters that expand across scales (compare Figures 5g, 5h, and 5i).

Many clusters detected using Period I data (central row of Fig. 5) intersect with those using more abundant Period II data (bottom row of Fig. 5) which, as expected, tend to be larger than those in the central figure.

3.2 Distribution of risk-clusters and confirmed COVID-19-clusters

The distribution of clusters identified from 9,971 anonymized COVID-19 cases confirmed by qRT-PCR between March 12 and August 31 was compared with the cluster distribution of risk cases surveyed in April-May (Period I). Only 9,453 confirmed COVID-19 patients (94.8%) residing in the SLPMA with known postal addresses were included in the comparison. Individual records included gender, age, risk factors, and the dates when symptoms started, and diagnoses were confirmed per individual. Several neighborhoods had COVID-19 cases confirmed between June and August (Table 3).

Risk case clusters A-E shown in Fig. 6 were compared with the quantitative cluster analysis of COVID-19 confirmed cases. Cluster A (274 confirmed COVID-19 cases) had several cases comparable to the top-five confirmed clusters (Tables 3 and 4). Cluster B (138 confirmed COVID-19 cases) intersects the Tequisquiapan area that ranks second among the confirmed SARS-CoV-2 hotspots. Cluster C (37 confirmed COVID-19 cases) intersects neighborhoods adjacent to the Tequisquiapan area. Cluster D (56 confirmed COVID-19 cases) includes the largest city park surrounded by a trail for strolling/jogging/biking. Cluster E (68 COVID-19 cases), on the northwest side of the park, intersect five neighborhoods. Cluster F (77 confirmed COVID-19 cases), intersects Graciano Sanchez's municipality, which includes seven neighborhoods.

3.3 Policy implications

CDS is the first mobile health (mhealth) platform used for real time epidemiological surveillance in the state of San Luis Potosí. Furthermore, our collaboration with the state health authorities has recently been reinforced by a formal agreement to improve epidemiological surveillance and analysis. On the other hand, we are currently analyzing the database of the more than 40,000 COVID-19 cases that had been confirmed in the SLPMA from March 2020 up to July 2021. In addition, we are also starting a project in collaboration with the state health and education authorities to implement a network of sensors to remotely monitor the carbon dioxide (CO_2) concentration in inner spaces, to assure their proper ventilation to prevent COVID-19 transmission.

The most important policy outcome of the CDS however is that it paved the way for transparent communication with key policy makers within the SSSLP. Though CDS results were not hard-coded into policy (by pre-emptively defining a specific relationship between CDS reports and concrete government actions), they certainly were a focal data point taken into consideration at every stage and level of the decision-making process of policies related to the pandemic. While we were not privy to discussions concerning top-level actions, such as extended lockdowns of entire neighbourhoods, the outcomes of these discussions often correlated with salient observations from the CDS, which were promptly shared with key decision makers.

In the future, we expect that the relevant data collected by subsequent iterations of the CDS, as well as the CO_2 remote sensing project will provide key inputs to policy decisions of a different nature, such as the allocation of various resources (medical, economic, etc.). Most notably, colleagues from the IPICYT and the University of Oxford have been developing algorithmic methods to allocate limited testing resources to a heterogeneous population in a way that balances viral spread and unnecessary self-isolation [22]. Key inputs to this algorithmic framework are parameters that represent rates of infection in distinct segments of a heterogeneous population, which can be estimated with methods similar to the CDS. In addition, the model incorporates co-transmission probabilities, for which CO_2 remote sensing (to be integrated in the CDS and SSSLP partnership framework) is a crucial proxy.

The CDS impact on policy extends far beyond the mere data it provides decision makers. The entire information pipeline which was fostered carefully over months of social and traditional media presence is a huge asset for maintaining open channels of communication with marginalized communities as the pandemic evolves.

4. Discussion

4.1 Core results

The coincidence of the spatial distribution of risk-case clusters generated with data from surveys of the first two-months with that of clusters containing substantial numbers of COVID-19 cases confirmed the following two-months indicates that the real-time analysis performed by the platform's processing pipeline of risk factors self-reported by *CDS* app users enables the early prediction of actual COVID-19 clusters.

Each survey recorded the geographic location of the *CDS* app and computed the risk index of the user based on its self-assessed signs, symptoms, and vulnerabilities. Analysis of 1,554 surveys of SLPMA users collected from April 1 to August 31 revealed statistically significant and consistent point-pattern and spatial clustering of risk cases independent of population density across geographical scales from 5 to 730 meters. Risk case clusters intersected neighborhoods with areas fostering social interactions/commercial activities and most of their locations coincided with those of SARS-CoV-2 infection hotspots with a substantial number of the SLPMA COVID-19 cases confirmed from March 12 to August 31 [17].

The exponential growth of COVID-19 cases in the SLPMA began in June, when public activities partially reopened, and SARS-CoV-2 hotspots increased significantly between May and August, while the number of confirmed COVID-19 cases in the top infection hotspots increased 81.1% during the first two months, and 82.0% during the last two months (Tables 3 and 4).

Some risk clusters inferred from data collected in April and May expanded and merged with clusters of COVID-19 cases confirmed the following three months. New risk clusters also appeared with time, and they merged and expanded to cover part of the largest infection hotspot containing 161 confirmed cases (Clusters A-D of Fig. 6).

Most CDS contributions corresponded to non-risk cases (Fig. 4b); risk cases amounted to 15.77%, and their maximum intensity was located in a densely populated residential area of high mobility and social interaction (Fig. 4c).

Risk cases from Period II (April-July) had statistically significant spatial clustering independent of population density across geographical scales from 5 to 730 m (Fig. 6 d-f) and risk clusters were computed at a scale of 730 m. Similar results were obtained for clusters of COVID-19 cases confirmed in Period I (April-May) at a scale of 600 m (Fig. 6 a-c). Many Period II clusters have significant overlap, are more prominent and expand several Period I clusters (Fig. 8).

Period II clusters also overlap city areas where the exceedance probabilities are close to one (Fig. 7). Risk case clusters in areas with a high incidence of risk of symptoms in Period I evolved to become the larger confirmed COVID-19 clusters of Period II (Fig. 8). Risk case clusters A-E (Fig. 8) were compared with the quantitative cluster analysis of COVID-19 confirmed cases. Cluster A (294 cases) had several cases comparable to the top-five confirmed neighborhoods. Cluster B (138 cases) and Cluster C (37 cases) intersect an area ranking second among confirmed SARS-CoV-2 hotspots. Cluster D (56 cases) includes the largest city park surrounded by a trail for strolling/jogging/biking. Cluster E (68 cases), on the northwest side of the park, intersects five neighborhoods. Cluster F (77 cases) intersects the Graciano Sanchez municipality, which includes seven neighborhoods.

Comparing the curve of cluster A showing the cumulative number of risk cases per week surveyed in Period II with that of confirmed cases in March 12 and August 3 (Fig. 9) we observed that risk cases started in Period I, before the transition to the exponential phase of confirmed COVID-19 cases.

4.1.1 Sampling bias

The sampling bias implicit in our mobile crowdsourcing approach restricts the assessment of epidemiological data, which are valid only for the surveyed population. The need to preserve the participants' privacy prevents individual trending analysis. On the other hand, it is difficult to quantify temporal disease trends from *CDS* data because surveys' timestamps encode temporal *CDS* usage patterns.

The bias sources include the age of the sampled population, the availability of technological resources and the availability of PCR COVID-19 tests.The highest proportion of *CDS* users were those of 20-29 years (44.02%) and 30-39 years (26.96%). The

predominant age groups were, in descending order, those of 10-19 years (18.65%), 20-29 years (17.54%), 0-9 years (16.65%), and 40-49-years (13.29%). Children under 10 years were not considered in this study. The predominant *CDS* groups are adults with a smartphone and internet connection. The elderly population is less likely to have access to and master the mobile technology deployed in the field; the analysis was directed to risk cases younger than 65 years. The *CDS* app was available for Android OS only; 68.3% of Mexico's population has access to a smartphone, with 82.5% Android OS users, and 17.1% iOS users; the rest use Samsung, Windows, and PlayStation devices. Asymptomatic individuals infected with SARS-CoV-2 were not eligible for COVID-19 testing due to the limited number of tests available and therefore are not considered in the SSSLP data.

4.2 Reflections on the Mexican context

The pandemic has been an unprecedented public health crisis. In the context of the CDS, we found it key to collaborate with existing teams—both in academia and on the ground—involved in combating viral spread and navigating ever-increasing public fears at the early stages of the pandemic. As mentioned previously, while we were developing the core functionality of the *CDS*, both the SSSLP and VIRUS program had epidemiological teams tasked with surveillance and information provision for marginalized communities (especially within the Huasteca Potosina). Our work tied in naturally with the pre-existing objectives of this team, and of note is the fact that the Potosinian Council of Science and Technology (COPOCYT) provided a crucial spark in connecting all our teams to work together. Indeed a strong recommendation we can provide for researchers seeking to bridge theory and practice in the Mexican context is to partner with or seek help from local research councils from the National Network of State Councils and Organizations of Science and Technology (REDNACECYT) of which COPOCYT forms a part. Collaborations in Mexico can often be accelerated when there are pre-existing personal relationships between stakeholders, and these very organizations can provide that initial step.

Indeed getting policy off the ground can be a lengthy process in the Mexican context, but a crucial component in the rapidity of the deployment of the CDS was the emphasis that we placed on civic engagement and participation in the platform. From the moment the app was created in early April 2020, through the first month of its availability on the Google play store (May 2020) and beyond, our team ensured public awareness through a series of social media campaigns. Most importantly, we focused on Facebook and Whatsapp (over Twitter, for example) as platforms since these are the most widely used communication channels amongst individuals in marginalized communities. These campaigns, coupled with the initial level of public fear during the pandemic provided with ample momentum to kick the project off, which ensured that policymakers from the SSSLP were willing to partner with us.

Finally, perhaps this was due to the unforeseen nature of the pandemic, but communication with the SSSLP, VIRUS program and the COPOCYT was very open throughout all stages of deployment of the CDS. This allowed results from the CDS to become key data points for government decisions and continues to prove a great asset as we further develop our technology and research to practice pipeline.

4.3 Future steps

To coordinate an adequate response to the COVID-19 pandemic it is imperative to have a fine- grained understanding of the heterogeneous nature of the population under study. Different intervention policies can have disparate effects depending on the underlying variation of multiple factors [20]. Variation in the outcomes of rate effects on target populations of pandemic interventions involves allocating extremely limited resources; as mentioned previously, since limited resources elicit different outcomes when allocated to different population sectors, the *CDS* platform's may be used as a valuable tool to optimize resource allocation.

Having a limited supply of COVID-19 tests is not a constraint unique to San Luis Potosí. Much of the Global South struggles with the same limitation. Whatever few tests available are often used to diagnose those already with symptoms, which is not as informative as to when part of the pandemic's scourge are the asymptomatic cases that can ravage a population. Whereas extensive contact tracing via testing has been touted as a success in South Korea and other countries, such a scale of testing is infeasible in Mexico. It is unclear whether approaching a given testing strategy is the optimal course of action due to the severe testing constraints faced. One possible solution involves using heterogeneous population factors to decide how to allocate testing kits differentially [21,22]. These approaches can profit from an underlying tool such as *CDS* to quantify risk and different citizen profiles in diverse geographies.

Information on the spatial distribution of cases during the spread of transmissible diseases is crucial to design and execute effective interventions. The *CDS* platform allows documentation and exploration of COVID-19 risk self-assessment data to enable epidemiological intelligence tasks and may be extended to address epidemiological events concurrent with COVID-19, such as the influenza A and B outbreaks expected in the coming months. Survey data could also provide useful information to allocate anti-COVID-19 vaccines across the state of San Luis Potosí and elsewhere.

References

[1] T.P. Velavan, C.G. Meyer, The COVID-19 epidemic, Trop. Med. Int. Heal. 25 (2020) 278– 280. https://doi.org/10.1111/tmi.13383.

[2] V. Suárez, M. Suarez Quezada, S. Oros Ruiz, E. Ronquillo De Jesús, Epidemiología de COVID-19 en México: del 27 de febrero al 30 de abril de 2020, Rev. Clínica Española.

(2020). https://doi.org/10.1016/J.RCE.2020.05.007.

[3] Gobierno de México, Inicia la fase 3 por COVID-19 - Coronavirus, April 21, 2020. (2020). https://coronavirus.gob.mx/2020/04/21/inicia-la-fase-3-por-covid-19-2/ (accessed June 24, 2020).
[4] Coronavirus Resource Center, Mortality Analyses - Johns Hopkins Coronavirus Resource Center, (2020). https://coronavirus.jhu.edu/data/mortality (accessed June 24, 2020).

[5] S.M. Kissler, Geographic and Demographic Transmission Patterns of the 2009 A/H1N1 Influenza Pandemic in the United States, University of Cambridge, 2017. https://www.repository.cam.ac.uk/handle/1810/273772.

[6] J. Cai, W. Sun, J. Huang, M. Gamber, J. Wu, G. He, Indirect Virus Transmission in Cluster of COVID-19 Cases, Wenzhou, China, 2020, Emerg. Infect. Dis. 26 (2020) 1343–1345. https://doi.org/10.3201/eid2606.200412.

[7] J.R. Koo, A.R. Cook, M. Park, Y. Sun, H. Sun, J.T. Lim, C. Tam, B.L. Dickens, Interventions to mitigate early spread of SARS-CoV-2 in Singapore: a modelling study, Lancet Infect. Dis. 20 (2020) 678–688. https://doi.org/10.1016/S1473-3099(20)30162-6.

[8] E.B. Hodcroft, Preliminary case report on the SARS-CoV-2 cluster in the UK, France, and Spain, Swiss Med. Wkly. 150 (2020). https://doi.org/10.4414/smw.2020.20212.

[9] H. Rossman, A. Keshet, S. Shilo, A. Gavrieli, T. Bauman, O. Cohen, E. Shelly, R. Balicer, B. Geiger, Y. Dor, E. Segal, A framework for identifying regional outbreak and spread of COVID-19 from one-minute population-wide surveys, Nat. Med. (2020) 1–4. https://doi.org/10.1038/s41591-020-0857-9.

[10] Z. Peng, R. Wang, L. Liu, H. Wu, Exploring Urban Spatial Features of COVID-19 Transmission in Wuhan Based on Social Media Data, ISPRS Int. J. Geo-Information. 9 (2020) 402. https://doi.org/10.3390/ijgi9060402.

[11] M. Roser, H. Ritchie, E. Ortiz-Ospina, Coronavirus Disease (COVID-19) – Statistics and Research, 2020.

[12] Human Security Unit of the United Nations, Human Security Handbook. An integrated approach for the realization of the SDG's, 2016. https://www.un.org/humansecurity/wp-content/uploads/2017/10/h2.pdf.

[13] M. Amundsen, S. Ruby, L. Richardson, Restful Web APIs: Services for a Changing World, O'Reilly M, 2013. https://www.oreilly.com/library/view/restful-web-apis/9781449359713/ (accessed October 13, 2020).

[14] A. Meier, M. Kaufmann, SQL & NoSQL Databases, Springer Fachmedien Wiesbaden, 2019. https://doi.org/10.1007/978-3-658-24549-8.

[15] INEGI, Catálogo Único de Claves de Áreas Geoestadísticas Estatales, Municipales y Localidades, 2020. https://www.inegi.org.mx/app/ageeml/#.

[16] World Health Organization, Digital technology for COVID-19 response, Newsletter. (2020) 1–1. https://www.who.int/news/item/03-04-2020-digital-technology-for-covid-19-response (accessed October 13, 2020).

[17] Sistema Nacional de Vigilancia Epidemiológica, Sitio oficial COVID-19 México Dirección General de Epidemiología - Gobierno Federal, (2020). https://covid19.sinave.gob.mx/ (accessed October 13, 2020).

[18] L.A. Waller, C.A. Gotway, Applied spatial statistics for public health data, John Wiley & Sons, 2004. https://www.wiley.com/en-mx/Applied+Spatial+Statistics+for+Public+Health+Data-p-9780471387718 (accessed June 4, 2020).

[19] M.A. Pett, Nonparametric Statistics for Health Care Research. Statistics for Small Samples and Unusual Distributions, 2nd ed., Sage Publications, Inc., 2015. https://us.sagepub.com/en-us/nam/nonparametric-statistics-for-health-care-research/book240591 (accessed October 13, 2020).

[20] M. Akbarpour, C. Cook, A. Marzuoli, S. Mongey, A. Nagaraj, M. Saccarola, P. Tebaldi, S. Vasserman, H. Yang, Socioeconomic Network Heterogeneity and Pandemic Policy Response, SSRN Electron. J. (2020). https://doi.org/10.2139/ssrn.3623593.

[21] A. Biswas, S. Bannur, P. Jain, S. Merugu, COVID-19: Strategies for Allocation of Test Kits, ArXiv Prepr. (2020). http://arxiv.org/abs/2004.01740.

[22] Jonnerby J, Lazos P, Lock E, Marmolejo-Cossío F, Ramsey CB, Sridhar D. Test and contain: A resource-optimal testing strategy for covid-19. In: AI for Social Good Workshop. (2020).

Table 1. Calculation of the users' personal risk index

Contact [a] (1-3 points)	Signs and symptoms [b] (4-6 points)			Vulnerability [c] (7-8 points)	Color coded risk index
No	No	No	No	No	Risk-free
No	No	No	No	Yes	
No	No	Yes	No	No	
No	No	Yes	No	Yes	
Yes	No	No	No	No	Low risk
Yes	No	No	No	Yes	
Yes	No	Yes	No	No	
Yes	No	Yes	No	Yes	
No	Yes	No	No	No	Moderate risk
No	Yes	Yes	No	No	
Yes	Yes	No	No	No	
Yes	Yes	Yes	No	No	
No	Yes	No	No	Yes	High risk
No	Yes	No	Yes	No	
No	Yes	Yes	No	Yes	
No	Yes	Yes	Yes	No	
Yes	No	No	Yes	No	
Yes	No	Yes	Yes	No	
Yes	Yes	No	Yes	No	
Yes	Yes	Yes	Yes	No	
No	Yes	No	Yes	Yes	Severe risk
No	Yes	Yes	Yes	Yes	
Yes	No	No	Yes	Yes	
Yes	No	Yes	Yes	Yes	
Yes	Yes	No	No	Yes	
Yes	Yes	No	Yes	Yes	
Yes	Yes	Yes	No	Yes	
Yes	Yes	Yes	Yes	Yes	
No	No	No	Yes	No	Respiratory risk
No	No	No	Yes	Yes	
No	No	Yes	Yes	No	
No	No	Yes	Yes	Yes	

[a] Contact factors: travel exposure, exposition to suspect or confirmed COVID-19 case.

[b] Signs and symptoms: cough and fever, general symptoms, respiratory symptoms.

[c] High risk and comorbidity factors.

Table 2. Age and gender distribution of risk cases

Age group (years)	Male (M)		Female (F)		All		M/F ratio
10-19	33	2.1%	22	1.4%	55	3.5%	1.50
20-29	346	22.3%	338	21.8%	684	44.0%	1.02
30-39	202	13.0%	217	14.0%	419	27.0%	0.93
40-49	115	7.4%	104	6.7%	219	14.1%	1.11
50-59	63	4.1%	61	3.9%	124	8.0%	1.03
60-69	21	1.4%	25	1.6%	46	3.0%	0.84
70-79	3	0.2%	4	0.3%	7	0.5%	0.75
Total	**783**	**50.4%**	**771**	**49.6%**	**1554**	**100.0%**	

Table 3. Top five SLPMA neighborhoods with confirmed COVID-19 case clusters (March-August, 2020)

Month	Accumulated neighborhoods	Accumulated cases	The top five COVID-19 case clusters per neighborhood				
			Neighborhood	Cases		Latitude	Longitude
				Number	%		
March (M1)	27	42	La Forestal	5	11.9	22.186728N	101.014444W
			La Loma	5	11.9	22.170625N	100.939759W
			Colonia 107	3	7.1	22.137671N	100.969575W
			Tangamanga	2	4.8	22.140680N	100.996548W
			San Pedro	2	4.8	22.157031N	100.998313W
April (M2)	85	147	Industrial Aviación	7	4.8	22.174750N	100.994764W
			Maya Mil	6	4.1	22.143625N	100.940586W
			21 de Marzo	6	4.1	22.147852N	100.926246W
			La Forestal	5	3.4	22.186887N	101.013832W
			La Loma	5	3.4	22.170625N	100.939759W
May (M3)	287	616	Tequisquiapan area	13	2.1	22.151746N	100.991206W
			Jardines del Sur	12	2.0	22.130786N	100.935747W
			21 de Marzo	12	2.0	22.147852N	100.926246W
			San Ángel	11	1.8	22.181965N	101.002597W
			San Ángel Inn	9	1.5	22.197854N	101.008545W
June (M4)	587	1827	Centro	29	1.6	22.152322N	100.976087W
			Tequisquiapan area	26	1.4	22.151746N	100.991206W
			San Ángel	26	1.4	22.181965N	101.002597W
			Jardines Sur	24	1.3	22.130786N	100.935747W
			Las Mercedes	22	1.2	22.119295N	100.887001W
July(M5)	1,076	6180	Centro	133	2.2	22.152322N	100.976087W
			Tequisquiapan area	106	1.7	22.151746N	100.991206W
			Simón Díaz	72	1.7	22.113709N	100.945591W
			Las Mercedes	69	1.1	22.119295N	100.887001W
			Satélite	66	1.1	22.106064N	100.950143W
August (M6)	1,315	8592	Centro	161	1.9	22.152322N	100.976087W
			Tequisquiapan area	138	1.6	22.151746N	100.991206W
			Progreso	107	1.2	22.126021N	100.949854W
			Simón Díaz	97	1.1	22.113709N	100.945591W
			Capricornio	97	1.1	22.144624N	100.953180W

Table 4. Persistence and growth of COVID-19 case clusters in the SLPMA

Neighborhoods with five top clusters	Cases accumulated	Cluster growth across time	Persistence (months)
Centro	161	82.0% (M4-M6)	3
Tequisquiapan area	138	81.1% (M3-M6)	4
Progreso	107	0.0% (M6)	1
Simón Díaz	97	25.8% (M5-M6)	2
Capricornio	97	0.0% (M6)	1
Las Mercedes	69	68.1% (M4-M5)	2
Satelite	66	0.0% (M5)	1
San Angel	26	57.7% (M3-M4)	2
Jardines Sur	24	50.0% (M3-M4)	2
21 de Marzo	12	50.0% (M2-M3)	2
San Angel Inn	9	0.0% (M3)	1
Industrial Aviación	7	0.0% (M2)	1
Maya Mil	6	0.0% (M2)	1
Forestal	5	0.0% (M1-M2)	2
La Loma	5	0.0% (M1-M2)	2
Tangamanga	2	0.0% (M1)	1
San Pedro	2	0.0% (M1)	1

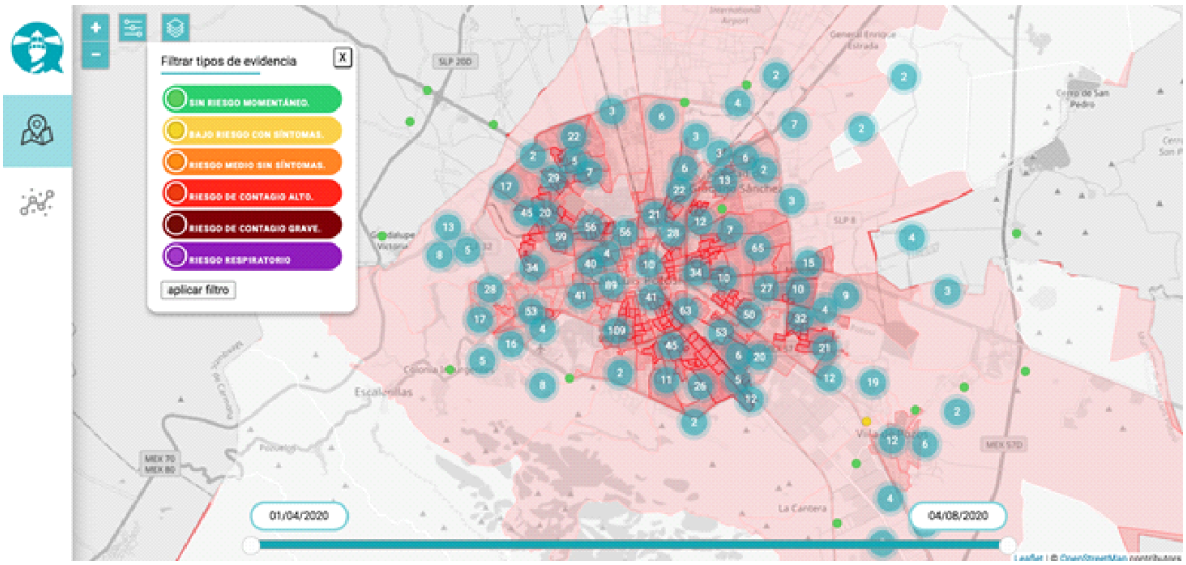

Figure 1. *CDS* platform front-end dashboard

The web application allows epidemiologists and public health specialists to visualize real-time data through a dashboard that facilitates decision-making. Besides displaying georeferenced data in a map, the dashboard shows descriptive statistics of all the variables associated with the *CDS* survey.

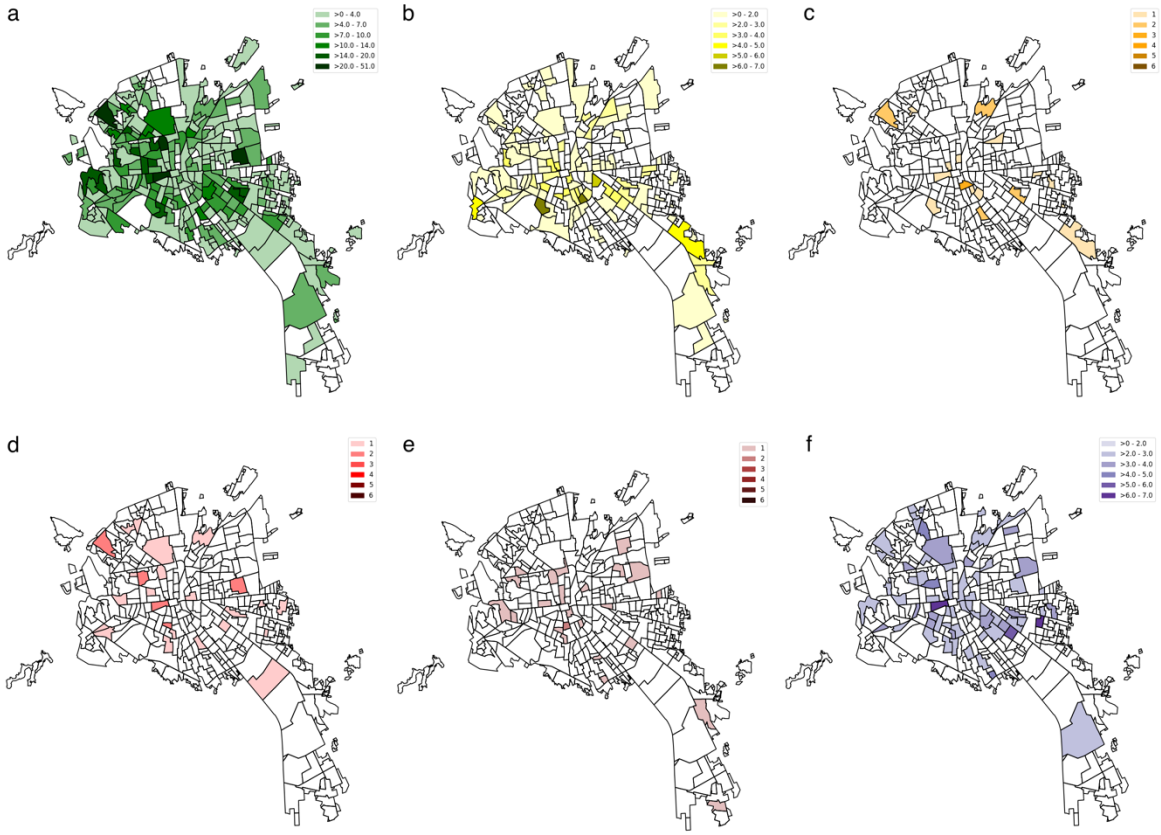

Figure 2. Geographic distribution of the risk index of the app users' contributions

AGEBs in the maps are color-coded based on the number of contributions recorded. Darker colors in a given AGEB indicate higher contributions count. The color palette used for each map indicates a specific risk level: (a) no-risk/no-symptoms, green; (b) low-risk, yellow; (c) medium-risk, orange; (d) high-risk, red; (e) very high-risk, brown; (f) respiratory symptoms, purple.

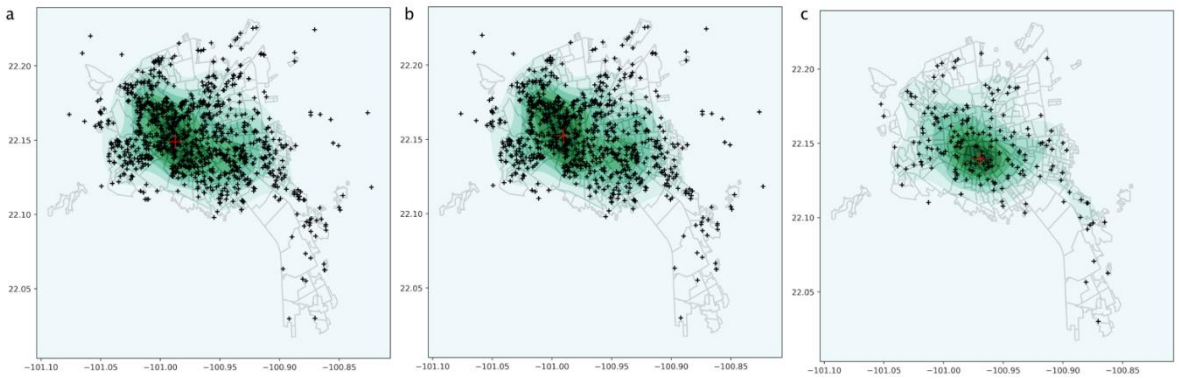

Figure 3. Intensity functions in the geographic space

Estimated sample intensity functions are shown as green-shaded areas in the maps for the point processes representing app users' contributions: (a) all users' contributions; (b) no-risk cases; (c) risk cases. Black plus signs represent app users' contributions. Darker green areas represent higher intensity. Locations with maximum empirical intensity values are shown as a red plus sign. Kernel density estimation techniques (with kernel bandwidth $h = 1200$ m) enabled calculation of intensity functions.

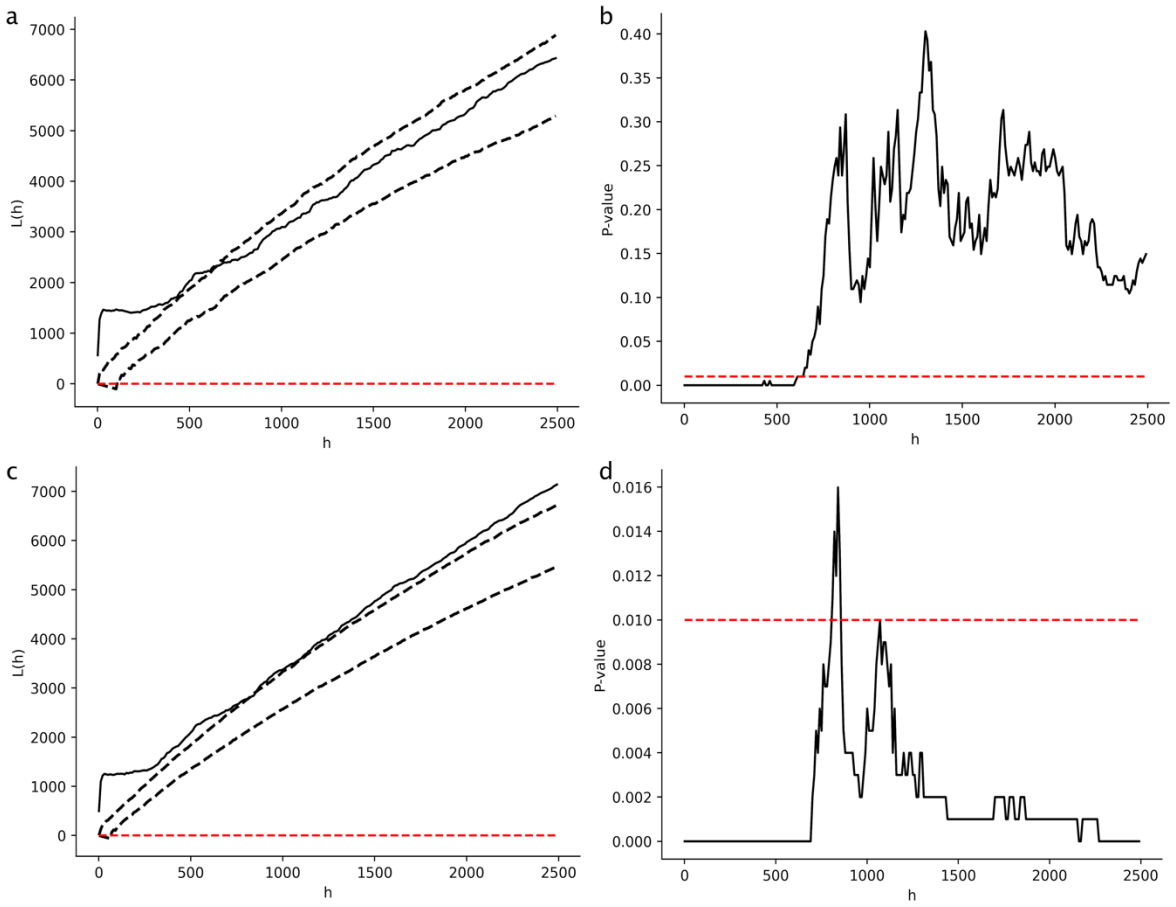

Figure 4. Spatial distribution patterns of risk cases

Period I: (a) *L*-function (solid line) and upper and lower envelopes (dashed lines) from Monte Carlo simulations for risk cases. The *L* function is above the upper envelope at scales between 5 and 600 m, suggesting significant clustering features ($L > Lup > 0$, $p < 0.001$) at these scales. (b) The corresponding *p*-values for each *h* in the analysis range. The red dashed line represents a *p*-value of 0.0. **Period II**: (c) *L*-function (solid line) and upper and lower envelopes (dashed lines) resulting from MC simulations for the risk cases. (d) The *L* function is above the upper envelope at scales between 5 and 730 m, suggesting significant clustering features ($L > Lup > 0$, $p < 0.001$) at these scales.

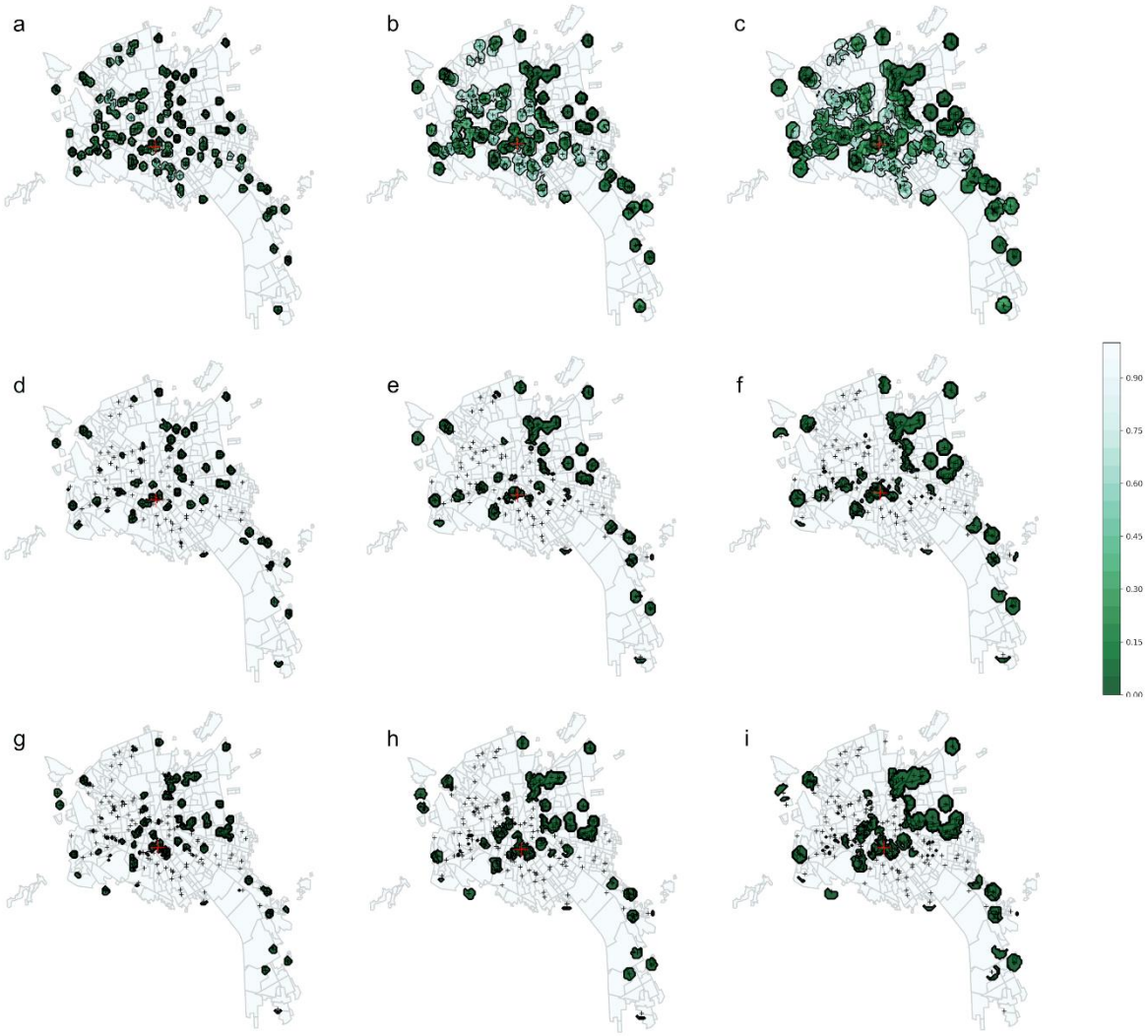

Figure 5. Point clusters at different spatial scales

Period I: Using Ripley's *K* function over a grid, combined with MC techniques, enables computation of *p*-values over the geographical space to reveal possible point clusters at different scales. The maps show the point clusters with *p* values < 0.05 at different scales, from left to right, *h* = 300, 450, and 600 m. (a-c). The top row shows the point clusters for the risk cases computed at the three scales. (d-f) The center row shows the geographical areas where *p*-values < 0.05. **Period II:** (g-i) The bottom row shows the point clusters at the three scales.

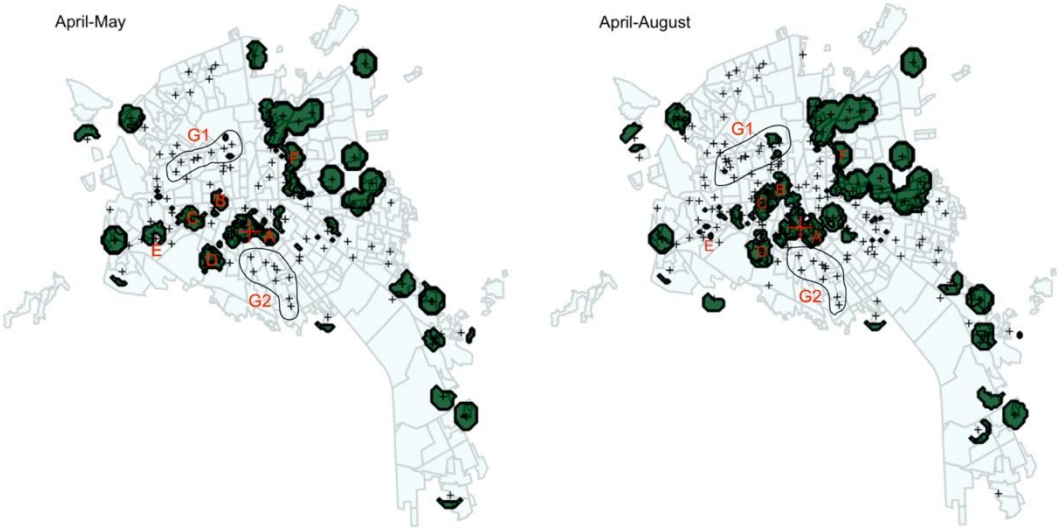

Figure 6. Socio-urban analysis of statistically significant point-pattern clusters

Supplementary tables

Table S 1. Personal risk index distribution by gender and age

Age group (years)	No risk				Low risk				Medium risk				High risk				Very high risk				Respiratory symptoms				All				M/F ratio
	Male		Female		Male		Female		Male		Female		Male		Female		Male		Female		Male		Female		Male		Female		
10-19	19	1.2%	15	1.0%	5	0.3%	4	0.3%	1	0.1%	0	0.0%	1	0.1%	0	0.0%	2	0.1%	0	0.0%	5	0.3%	3	0.2%	33	2.1%	22	1.4%	1.5
20-29	279	18.0%	249	16.0%	34	2.2%	33	2.1%	4	0.3%	6	0.4%	8	0.5%	7	0.5%	1	0.1%	4	0.3%	20	1.3%	39	2.5%	346	22.3%	338	21.8%	1.02
30-39	154	9.9%	171	11.0%	23	1.5%	28	1.8%	5	0.3%	1	0.1%	2	0.1%	4	0.3%	5	0.3%	4	0.3%	13	0.8%	9	0.6%	202	13.0%	217	14.0%	0.93
40-49	92	5.9%	91	5.9%	9	0.6%	6	0.4%	3	0.2%	1	0.1%	0	0.0%	3	0.2%	2	0.1%	1	0.1%	9	0.6%	2	0.1%	115	7.4%	104	6.7%	1.11
50-59	52	3.3%	44	2.8%	9	0.6%	10	0.6%	0	0.0%	1	0.1%	0	0.0%	2	0.1%	1	0.1%	2	0.1%	1	0.1%	2	0.1%	63	4.1%	61	3.9%	1.03
60-69	16	1.0%	20	1.3%	2	0.1%	2	0.1%	0	0.0%	0	0.0%	2	0.1%	2	0.1%	1	0.1%	0	0.0%	0	0.0%	1	0.1%	21	1.4%	25	1.6%	0.84
70-79	3	0.2%	4	0.3%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	3	0.2%	4	0.3%	0.75
Total	**615**	**39.6%**	**594**	**38.2%**	**82**	**5.3%**	**83**	**5.3%**	**13**	**0.8%**	**9**	**0.6%**	**13**	**0.8%**	**18**	**1.2%**	**12**	**0.8%**	**11**	**0.7%**	**48**	**3.1%**	**56**	**3.6%**	**783**	**50.4%**	**771**	**49.6%**	

Table S 2. Occupation distribution of risk cases by gender

Occupation	Male		Female		All		M/F ratio
Self-employed	163	10.5%	100	6.4%	263	16.9%	1.63
Business employee	260	16.7%	197	12.7%	457	29.4%	1.32
Student	195	12.5%	159	10.2%	354	22.8%	1.23
Government employee	141	9.1%	149	9.6%	290	18.7%	0.95
Unemployed	18	1.2%	65	4.2%	83	5.3%	0.28
Housekeeper	6	0.4%	101	6.5%	107	6.9%	0.06
Total	**783**	50.4%	**771**	49.6%	**1554**	100.0%	

Table S 3. Personal risk index distribution by gender

Risk index	Male		Female		All		M/F ratio
No risk	615	39.6%	594	38.2%	1209	77.8%	1.04
Low risk	82	5.3%	83	5.3%	165	10.6%	0.99
Medium risk	13	0.8%	9	0.6%	22	1.4%	1.44
High risk	13	0.8%	18	1.2%	31	2.0%	0.72
Very high risk	12	0.8%	11	0.7%	23	1.5%	1.09
Respiratory symptoms	48	3.1%	56	3.6%	104	6.7%	0.86
Total	**783**	**50.4%**	**771**	**49.6%**	**1554**	**100.0%**	

Table S 4. Overall distribution of risk factors

Risk category	Risk factor	Cases		Subtotal	
Contact	Contact with suspect COVID-9 case	152	9.8%	268	17.2%
	Contact with confirmed COVID-19 case	64	4.1%		
	Travel exposure	52	3.3%		
Signs and symptoms	General symptoms	526	33.8%	715	46.0%
	Respiratory symptoms	138	8.9%		
	Fever, dry cough	51	3.3%		
Vulnerability	Comorbidities	504	32.4%	1178	75.8%
	Risk conditions	267	17.2%		
	Cannot keep lockdown	298	19.2%		
	Lack of safety net	109	7.0%		

Table S 5. General features of confirmed COVID-19 cases in the state of San Luis Potosí (September 8, 2020)

Total cases		19,888 (100.0%)
Gender	Male	10,122 (50.9%)
	Female	9,766 (49.1%)
		218 pregnant (2.2%)
Age	Mean ± SD	43.5 ± 16.8 years
	Median	42 years
Contact with COVID-19 patient(s)		7,283 (36.6%)
Comorbidities	Hypertension	3,962 (19.9%)
	Obesity	3,848 (19.4%)
	Diabetes	3,026 (15.2%)
	Smoking	1,172 (5.9%)
	Asthma	612 (3.1%)
	Other	469 (2.4%)
	EPOC	349 (1.8%)
	Chronic renal insufficiency	375 (1.9%)
	Heart disease	312 (1.6%)
	Immunosuppression	205 (1.0%)
Pneumonia		2,394 (12%)
Medical attention	Ambulatory	16,969 (85.3%)
	Hospitalized	2,919 (14.7%)

Table S 6. Logit mortality by COVID-19 in SLPMA (n = 19,888)

Comorbidity/Vulnerability	Odds ratio	Confidence interval (95%)
Chronic renal insufficiency (Yes/No)	3.701**	2.84 - 4.81
Diabetes (Yes/No)	2.090**	1.83 - 2.39
Chronic obstructive pulmonary disease (Yes/No)	1.899**	1.45 - 2.49
Gender (male/female)	1.839**	1.62 - 2.09
Obesity (Yes/No)	1.487**	1.28 - 1.72
Hypertension (Yes/No)	1.244**	1.08 - 1.43
Immunosuppression (Yes/No)	1.182	0.73 - 1.92
Age (probability increase per additional year of age)	1.071**	1.07 - 1.08
Smoking (Yes/No)	1.071	0.84 - 1.36
Heart disease (Yes/No)	1.030	0.75 - 1.42
Asthma (Yes/No)	0.668	0.43 - 1.04

Exponentiated coefficients, * $p < 0.05$, ** $p < 0.01$.

Contributions of the study

Research Question	Contributions
1. Can we develop a mobile crowdsourcing platform as a community triage tool of COVID-19 through the assessment of self-reported contributions from the San Luis Potosí Metropolitan Area (SLPMA) residents?	Yes. We were able to develop the *CDS* platform with an app for data collection of COVID-19 signs, symptoms, and vulnerabilities self-reported by users, that enabled the computation of an anonymized COVID-19 risk index for each user through a data processing pipeline. These health mobile tools were deployed in the SLPMA to support health authorities and the *VIRUS Program* to fight the COVID-19 epidemic. The *CDS* app provides users with official information on COVID-19 to prevent transmission and ensure that those most in need get the right treatment. The platform enabled a community triage with geolocalization of the COVID-19 risk cases in a pilot survey of real-time self-assessment data from 1,554 users residing in the SLPMA. This triage was conducted under strict ethical guidelines to protect the privacy rights, health, and wellbeing of the participating volunteers. The data collected and the information distilled were available to health authorities through a web interface that enables a geographic information system's capability.
2. Data collected through the app can be used to characterize the sociodemographic features of the participants and to compute the COVID-19 risk index for each user?	Yes. Self-reported user data collected by the app such as gender, age, comorbidities, risk factors and vulnerabilities, was sufficient to meaningfully categorize these variables and computing the personal risk index for each user. Nearly one-third of the users reported one or more comorbidities, and 17.2% had at least one risk condition known to increase COVID-19 lethality. Data analysis also revealed statistically significant different bivariate associations.
3. Are the clustering and point-pattern distribution of risk indexes statistically significant and consistent?	Yes. Point pattern analysis revealed that risk cases exhibit statistically significant clustering features at spatial scales between 5 and 730 meters. The point clustering features analyzed over two intersecting time periods revealed consistent clusters across significant spatial scales. The socio-urban environment of the city areas covered by different risk case clusters suggests that their spatial distribution correlates with urban spaces fostering person-to-person contact such as commercial areas, shops, city spaces for community social interactions, recreation, religious events, and recreational sports.
4. Do risk index clusters and SARS-CoV-2 hotspots have similar spatial distributions?	Yes. Risk case clusters cover representative city areas. Those computed from *CDS* app data collected in April-May intersect city neighborhoods with substantial numbers of COVID-19 cases confirmed the next three months, including the Tequisquiapan area, with the second highest number of confirmed cases. Risk case clusters intersected neighborhoods containing 650 COVID-19 confirmed cases (7.6% of the total). Risk case clusters estimated with *CDS* data collected in April-May expanded, and some merged during the following three months.

	Risk case clusters appearing later intersected infection clusters such as that of the City center, the top infection hotspot. Growth of the risk case clusters determined by the CDS platform derives from the COVID-19 transmission dynamics.
5. Is the *CDS* platform useful in the fight against the COVID-19 pandemic?	Yes. Social media outlets reinforced the risk communication process to the *VIRUS Program* communities, *CDS* users, and general population. Population-wide surveys of risk factors administered through the *CDS* app provide a geospatial picture of COVID-19 risk and predict infection hotspots and potential outbreaks that help the *VIRUS Program* and health authorities intervene accordingly. Real-time information provided by the users complements COVID-19 field test results, which are often delayed up to 10 days. A risk case cluster analysis could help implement interventions and public health policies to fight COVID-19 and other epidemics in the SLPMA and elsewhere.