STUDY PROTOCOL

Maré cohort-profile: a prospective cohort study based in a socially vulnerable community during the COVID-19 pandemic in Rio de Janeiro, Brazil [version 1; peer review: 1 approved with reservations]

Amanda A Batista-da-Silva, Olivia T Ranzani, Daniela M de Paulo, Mariana L Braunstein, Helena R Bozza, Ronald Fischer, Grazielle V Ramos, Carolina M C Dias, Everton P da Silva, Luna E Arouca, Leonardo L S Bastos, Otavio T Ranzani, Silvio Hamacher, Fernando A Bozza

1Department of Industrial Engineering, Pontifical Catholic University of Rio de Janeiro, Rio de Janeiro, Brazil
2Tecgraf Institute, Pontifical Catholic University of Rio de Janeiro, Rio de Janeiro, Brazil
3Institute of Economics, Federal University of Rio de Janeiro, Rio de Janeiro, RJ, Brazil
4D’Or Institute for Research and Education, Rio de Janeiro, Brazil
5School of Psychology, Victoria University of Wellington, Wellington, New Zealand
6Associação Redes de Desenvolvimento da Maré, Rio de Janeiro, Brazil
7Graduate Program on Sociology and Anthropology, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil
8Barcelona Institute for Global Health, ISGlobal, Universitat Pompeu Fabra, CIBER Epidemiología y Salud Pública, Barcelona, Spain
9Pulmonary Division, Heart Institute (InCor), Hospital das Clínicas HCFMUSP, Faculdade de Medicina, Universidade de Sao Paulo, São Paulo, Brazil
10National Institute of Infectious Disease Evandro Chagas, Oswaldo Cruz Foundation, Rio de Janeiro, Brazil

Abstract

Background: Socially vulnerable populations were vastly affected by the COVID-19 pandemic. The pandemic significantly impacted Brazil, pressuring its healthcare system for several months, with high mortality rates, even among the youngest population. Cohort studies combining disease surveillance are essential for understanding virus circulation in the community, surrogates of protection, vaccine effectiveness, and demand for health resources.

Methods: Here, we present the protocol for a community-based prospective cohort study in the largest complex of favelas (slums) in Rio de Janeiro, Brazil (Complexo da Maré). The study participants are residents initially recruited during a massive vaccination campaign in the community. Five waves of data collection at approximately six-month intervals were planned. The first two waves have been completed at the time of writing this study protocol, and the third is...
underway. The protocol comprises interviews, blood sampling, and records linkage with secondary data to enrich the profiles of cohort participants and community information. We will describe COVID-19 seroprevalence, socio-demographic characteristics, and the burden of COVID-19, followed by estimating the association of socioeconomic factors and the burden of disease with seroprevalence.

Discussion: The primary aims of the study are to assess COVID-19 clinical, epidemiological and genomic profiles and outcomes in residents from Maré, including vaccine effectiveness, surrogates of immune protection, virus transmission in households, and the overall burden of the pandemic.

Keywords
cohort, COVID-19, serosurvey, LMIC, community, vaccine, household

This article is included in the Coronavirus (COVID-19) collection.
Introduction
Since December 2019, the world population, health workers, and governments have been dealing with the challenges imposed by the COVID-19 pandemic. The burden across the globe was not only perceived through high morbidity and mortality rates but also by its effects on economic and social dimensions. Vulnerable populations, defined as groups of individuals susceptible to physical, biological, psychological, and socioeconomic stressors while lacking adequate resources to cope with these stressors, were vastly affected by the COVID-19 pandemic, both in high-income and low-and-middle-income countries (LMICs).

One of the countries most affected by the pandemic was Brazil, where the health system was under increased demand and pressure for several months, and the mortality rate was high, even among the young population. During the pandemic, initiatives articulated between social organisations and academia were essential to provide health and socioeconomic support to vulnerable populations, such as those living in favelas or “slums” (a Brazilian urban community, sharing some features with slums in other countries, comprising highly populated areas, with different levels of urbanisation and informal settlements, historically neglected from govern) - usually crowded areas with several basic needs.

Cohort studies associated with disease surveillance are essential for understanding pathogens’ circulation, surrogates of protection, disease burden, and demand for health services, especially in poor communities in LMIC. By following individuals over time, cohorts provide granular data for monitoring disease trends and the long-term impacts of the pandemic. Prospective studies may fill knowledge gaps regarding the burden of disease and risk factors in a vulnerable population.

Here, we describe the protocol and the baseline demographic profile of a cohort design, which started recruiting during the COVID-19 pandemic in Complexo da Maré, a neighborhood of Rio de Janeiro and one of the largest groups of favelas in Brazil. The Maré Cohort is embedded in several initiatives recently implemented in Maré, such as socioeconomic support, health care, and surveillance, managed by an innovative organisational arrangement involving non-governmental organisations (NGOs), research institutions, private funders, and local government.

Objectives
We aim to assess clinical, epidemiological and genomic profiles, outcomes and the impact of the COVID-19 pandemic in the residents from Maré. The specific objectives of the Maré Cohort are:
1. To evaluate the effectiveness of the COVID-19 vaccines and the temporal course of seroprevalence of SARS-CoV-2 antibodies
2. To perform genomic surveillance and identify the transmission chains of SARS-CoV-2 in the community
3. To estimate long-term health impacts of COVID-19 (long COVID-19 and mental health)
4. To evaluate the association of socioeconomic aspects, access to primary care services and the burden COVID-19 pandemic

Protocol
Study Design
The study is a community-based prospective cohort study in Maré, Rio de Janeiro, Brazil. Ethical approval was obtained from the National Research Ethics Committee (IRB/CONEP) (CAEE: 49726921.6.0000.5248 and 44180821.1.0000.5249).

Study Status
We planned five waves at approximately six-month intervals. The study is planned to finish by the beginning of 2024. Data collection is ongoing; the first and second waves have been completed, and the third was initiated in December 2022.

Setting
This prospective cohort study was designed to be conducted among the residents of Maré in Rio de Janeiro municipality, an area composed of 16 favelas, one of the largest set of favelas in Brazil. Maré is the city’s ninth most populous neighbourhood with more than 140,000 inhabitants spread over 4.3 km². Its population is mainly young (51.9% under 30 years old) and almost twice the size of Rio’s second largest favela.

Maré is located near the city centre and is surrounded by major highways. Although the region presents informal settlements with substandard housing, most residences have running water, access to a sewage network and public lighting. The chaotic urbanisation of favelas has resulted in a heterogeneous population spread within the limited area, generating a high population density and vulnerable regions. With respective educational and healthcare facilities, 44 public schools are attending the territory in 2013, and the public health system provides seven primary healthcare units, three psychosocial centres, a mobile outreach clinic, an emergency unit, and a reference public hospital to the area.

The community is among the most vulnerable neighbourhoods in the municipality (Table 1), evidenced by its low Human Development Index (HDI: 0.686), ranked 123rd out of 126 in Rio de Janeiro. The GINI index of 0.39 may suggest a more homogenous and less unequal community. Maré has nearly three times as many illiterate residents and a one-third monthly per capita income of USD 294 compared with the municipality estimates. In terms of health indicators, Maré has higher incidences of tuberculosis (155.7 per 100,000 population), higher infant mortality rates (16.7 per 100,000 live births), and higher mortality rates due to external causes (105.8 standardised deaths per 100,000 population) than the municipality.

Governance and Management
This cohort study is part of a platform composed of several initiatives developed in the territory to assist residents during
the COVID-19 pandemic\textsuperscript{8-10}, with diverse actions and partnerships (Figure 2). A steering committee (SC) composed of the principal investigator and one member of four institutions was assembled to manage the cohort. The SC has held weekly meetings to plan and decide on support action in the territory. Furthermore, two coordination units report directly to the SC, one for field-related actions and the other for data-related activities. A communication team was constituted to perform direct engagement and targeted communication activities, composed of different professionals: coordinator, designer, illustrator, social media manager and content producer, web designer, photographer, video maker, and image editor. The field team consists of 17 community mobilisers, seven of them with technical training in nursing, and 14 community health workers (CHWs) responsible for direct contact with the cohort participants, interviewing, laboratory testing and clarifying questions about the project, and reporting directly to the field coordination. The CHWs were recruited specifically for the project on a part-time basis. Firstly, five candidates were selected by the manager of each primary healthcare unit. Then, after an interviewing process with the field coordinator, two were recruited/selected. The recruitment process was based on nursing technical training and the largest work experience in the unit.

In addition, the data working group is divided into management and analytics teams, who reports to the data coordination team. The former is responsible for curating and integrating the data and the latter consists of epidemiologists, data scientists and vaccinologists performing research and data analysis.

The corresponding author is the study principal investigator (PI) and part of the steering committee, along with three co-authors. Two authors are part of the field coordination team, and eight compose the data working group.

The laboratory testing structure has two major components to support the surveillance actions in Maré. “Dados do Bem”, the Brazilian symptom tracking app\textsuperscript{14}, was designed by the PI of this project with a team of developers (Zoox Smart Data) to detect symptomatic cases, track contacts, and identify areas at increased risk of SARS-CoV-2 infection. The app was free of charge and available in iOS and Android systems. The app manages the user registration and SARS-CoV-2 testing process,
Table 1. Territorial, socioeconomic and health indicators of Maré neighbourhood and Rio de Janeiro municipality based on the 2010 National Census.

<table>
<thead>
<tr>
<th>Territorial dimension</th>
<th>Complexo da Maré</th>
<th>Rio de Janeiro Municipality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Area - 2020 (a)</td>
<td>4.3 km²</td>
<td>1,204 km²</td>
</tr>
<tr>
<td>Population - 2010 (b)</td>
<td>129,770</td>
<td>6,320,446</td>
</tr>
<tr>
<td>Population density - 2010 (b)</td>
<td>30,179/km²</td>
<td>5,266/km²</td>
</tr>
<tr>
<td>Number of households - 2010 (b)</td>
<td>41,731</td>
<td>2,144,445</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Socio-economic indicators</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Human Development Index (HDI) - 2010 (c)</td>
<td>0.674</td>
<td>0.799</td>
</tr>
<tr>
<td>GINI index - 2010 (c)</td>
<td>0.39</td>
<td>0.62</td>
</tr>
<tr>
<td>Ageing index - 2010 (b)</td>
<td>27</td>
<td>77</td>
</tr>
<tr>
<td>Monthly Per capita income* - 2010 (c)</td>
<td>USD 294.0</td>
<td>USD 895.8</td>
</tr>
<tr>
<td>Workers with an income of up to 2 minimum wages - 2010 (c)</td>
<td>79%</td>
<td>54%</td>
</tr>
<tr>
<td>Illiteracy (15 years of age or older) - 2010 (b)</td>
<td>7.7%</td>
<td>2.9%</td>
</tr>
<tr>
<td>Household sewage disposal - 2010 (c)</td>
<td>99%</td>
<td>99%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Health indicators - 2019</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Tuberculosis incidence per 100,000 (d)</td>
<td>155.7</td>
<td>140.1</td>
</tr>
<tr>
<td>Infant mortality per 1,000 live births (d)</td>
<td>16.7</td>
<td>12.1</td>
</tr>
<tr>
<td>Mortality rate for infectious disease (Chapter I ICD-10) (d)</td>
<td>59.4**</td>
<td>59.8</td>
</tr>
<tr>
<td>Mortality rate for neoplasms (tumours) (Chapter II ICD-10) (d)</td>
<td>26.4**</td>
<td>158.0</td>
</tr>
<tr>
<td>Mortality rate for circulatory system (Chapter IX ICD-10) (d)</td>
<td>277.6**</td>
<td>269.3</td>
</tr>
<tr>
<td>Mortality rate for external causes (Chapter XX ICD-10) (d)</td>
<td>105.8**</td>
<td>79.3</td>
</tr>
</tbody>
</table>

(a) 2020 carried out by the Pereira Passos Municipal Urbanism Institute (IPP)
(b) 2010 census carried out by the Brazilian Institute of Geography and Statistics (IBGE), published on the Pereira Passos Municipal Urbanism Institute (IPP)
(c) 2010 census carried out by the Brazilian Institute of Geography and Statistics (IBGE), published on Atlas do Desenvolvimento Humano no Brasil
(d) Health systems’ databases provided by the Municipal Health Department
(*) Average nominal monthly household income per capita. The minimum wage in 2010 was $306.1
(**) The rate in Maré standardised for Rio de Janeiro municipality - per 100,000

working as a platform for hosting the data of serological and RT-qPCR test results for cohort participants. The COVID-19 diagnosis unit (Unidade de Apoio ao Diagnóstico da COVID-19, UNADIG-FIOCRUZ) provides laboratory support with the distribution and analysis of serological and RT-qPCR results from participants and Maré residents. Tests were free of charge to participants.

Population selection and participants
The cohort comprises participants that reside in Maré. The inclusion criteria considered all applicants must be permanent residents of Maré and complete an Informed Consent (IC) form before enrolment. We excluded residents with incomplete identification data, such as invalid unique identifier numbers or incomplete addresses and telephone numbers.

We calculated the required number of responders using the formula for prevalence studies\(^1\). The appropriate sample size required for different prevalence and precision levels is shown in Figure 3. We estimated that an overall minimum sample size of approximately 2,400 participants, with an alpha of 0.05 and precision of 2%, is sufficient to ensure the outcome is significant and provides meaningful results.
Participant recruitment and consent
The initial round of recruitment occurred during the large-scale COVID-19 vaccination campaign from July 29th, 2021 to August 1st, 2021, when approximately 36,000 first doses of ChAdOx1 were administered. Since then, study participants have been monitored via household follow-up.

The enrolment strategy has three stages. In the first stage, a trained interviewer invites participants, explains the research purpose and testing procedures, clarifies questions, and obtains informed consent - in the case of participants who are illiterate, the researcher reads the consent inform supervised by a family member of the participant, and a thumbprint is taken. In the second stage, the participant answers a questionnaire filled out by a trained mobiliser. In the third stage, they register at the app - in the case of illiterate participants, the mobiliser completes the registration - and takes a serological test - a specialised technician collects the material (dry blood spot). For children, parental consent is obtained, and the guardian answers a simplified and adapted questionnaire. Participants can withdraw from the study anytime, even after providing written, informed consent. This enrolment strategy will be used in all waves of data collection.

The Maré cohort is based on household units. Following the enrolment of an index adult, the mobilisers and CHWs conducted home visits to map and recruit new family members, following the same steps executed during the recruitment of the first adult (i.e., applying an informed consent form to the new family member and collecting information as a new entry).

Representativeness
Participants were pragmatically included in the cohort since recruitment occurred during the pandemic. Our efforts at mitigating selection bias and achieving representativeness combined social, spatial, and demographic characteristics, recruiting participants at different times and days of the week and spatial locations within Maré. We plan to adjust cohort estimates using sampling weights derived from characteristics reported in the 2013 Maré survey, where we have a complete demographic cross-sectional evaluation of the entire community.

Follow-up
This cohort study will be divided into waves, each corresponding to different stages of analyses and data collection (Figure 4). The study was initially planned to hold five waves at approximately six-month intervals and is expected to end at the beginning of the first half of 2024. The first wave occurred from August 2021 to January 2022, and we evaluated infection by COVID-19 and the socio-demographic characteristics of the target population. The second wave ran from February to July 2022, and the analyses focused on socioeconomic factors and COVID-19 response to the vaccine.

The first two waves have been completed at the time of writing this study protocol, and the third is underway. Following a similar enrolment and data collection strategy from the first and second waves, the third wave focuses on understanding the residents’ perspective about vaccination, primary care services, and socioeconomic aspects during the pandemic. Future (fourth and fifth) waves are planned to address the conduct of COVID-19 surveillance and other health problems prevalent in vulnerable communities in low-and-middle-income countries. These new waves will follow the same strategy of recruitment, data collection, communication, and engagement used in waves one-three. The questionnaire will be adapted to incorporate the dimensions of other diseases and the follow-up questions from previous waves, using the same teams and structure already in place in the territory.
Data Collection
The protocol includes interviews, blood sampling and record linkage to secondary data associated with cohort participants.

Data access
Once the data has been collected, access to individual data will require a formal request to the authors and the cohort’s Steering Committee. This route of access follows the guidelines and practices according to the Brazilian General Data Protection Regulation (Lei Geral de Proteção de Dados 13.709/2018) for data integrity and security.

Questionnaires Development
A group of 11 specialists, consisting of clinical researchers, epidemiologists, psychologists, social workers, and vaccinologists, developed the standardised questionnaires to collect data from the cohort participants. They were responsible for creating the protocol, and eight form this paper authorship. The questionnaires were validated with local community members, followed by interviewers’ training. Questionnaire data was collected directly on a standardised form at each contact via an electronic data capture system (REDCap – Research Electronic Data Capture)\(^1\). The data was collected using a mobile device (tablet) during the interview, uploaded into REDCap at the end of the day, and then deleted from the device to preserve the integrity of the participants.

Questionnaires applied to adults included variables of socio-demographic characteristics, socioeconomic levels, general health, vaccines (including adverse events), well-being, search for medical assistance, the self-reported status of COVID-19 infection, and risk factors for exposure to COVID-19. Well-being questions addressed generalised anxiety and depression using the Patient Health Questionnaire-4 (PHQ-4)\(^1\), a standardised screening instrument. We also included an item measuring happiness\(^1\) and two items from the Life Orientation Test\(^2\), indicating overall optimism levels. A specific questionnaire for those under 18 years contains variables of socio-demographic characteristics, education level status, general health and medical assistance.

Data from different sources were linked directly using a unique identifier or probabilistic linkage when applicable. After data preparation and integration, participants were de-identified and anonymised for analysis. Data governance and curation activities were described in a Data Protection Assessment document that followed the Brazilian General Data Protection Regulation (Lei Geral da Proteção de Dados)\(^3\).

Testing Strategy
We collected five blood spots on filter paper with a lancet for finger-prick in each wave of data collection. The primary data analysed were the serological tests for anti-SARS-CoV-2 IgG antibodies. We used the Perkin-Elmer (Rio de Janeiro, RJ, Brazil) kit to identify IgG immunoglobulin class antibodies to SARS-CoV-2, a fully automated system dried blood spot based able to test a large population efficiently\(^2\),\(^3\).

Vaccination campaign
On January 17\(^{th}\) 2021, the Brazilian government launched a nationwide vaccination campaign against COVID–19 following an age-based priority\(^4\). The Brazilian National Immunisation Programme (PNI) is responsible for the vaccination schedule and the administration of free doses of the following vaccines: Sinovac-CoronaVac, ChAdOx1-S-nCoV-19 (Oxford-AstraZeneca), BNT162b2 (Pfizer/BioNTech) and Ad26.COV2.S (Janssen). Six months after the vaccination roll-out started in the country (July 2021), a mass vaccination campaign (Vacina Maré) was carried out in Complexo da Maré, which increased the first dose coverage from 38% of to 93% in four days. Along with this campaign, the cohort study started the recruitment phase in the territory.

The COVID-19 immunisation data from cohort participants were made available for research and analysis purposes through the PNI Information System (Sistema de Informação do Programa Nacional de Imunização, [SI-PNI]), a secondary data made available by the city health department. SI-PNI is Brazil’s central data archive for vaccine doses administered. It contains information on demographics (age, sex and place of residence), vaccine platform, the number of doses received, date of administration and location.

Secondary data
We plan to perform a deterministic followed by probabilistic record linkage between cohort participants and administrative data collected from the Brazilian health systems, such as COVID-19-related surveillance systems for mild and severe cases.
Preliminary data
Overall, 6,429 residents in Maré answered the questionnaire during the first wave, 5,923 (92%) were adults, and 506 (8%) were under 18 (Table 2). The sample is adherent to the size predicted, considering an alpha of 0.05 and precision of 2%.

A brief description of the baseline characteristics of the cohort, Maré neighbourhood Census 2019 and Rio de Janeiro municipality Census 2010, is shown in Table 2. The majority of participants are up to 39 years old. The cohort has a lower proportion of children (newborn-nine years) and a higher proportion of the elderly than Maré. Females are the majority of the three samples; almost half of the cohort and Maré are self-declared Pardo.

Evaluating anti-SARS-CoV-2 antibody prevalence is part of the next step of this cohort analysis, and at the time of writing this study protocol, we have results from 6,272 (97.6% of 6,429) tests for IgG. A total of 157 (2.4%) individuals not tested or with inconclusive results were excluded at this stage (Figure 5).

Immunisation information was obtained for 5,780 participants. We excluded 92 adults based on system registry inconsistencies data (5,688 with complete immunisation status).

Full baseline data will be included in the final dataset.

Outcomes
The primary outcome of this study will be the seroprevalence of SARS-CoV-2 antibodies in the population. Secondary outcomes will include symptomatic and asymptomatic COVID-19 infections and severe cases with hospitalisations and deaths related to COVID-19, and long-term impacts, such as the prevalence of long COVID-19 symptoms and mental health conditions.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Cohort¹ N = 6,429</th>
<th>Maré² N = 129,770</th>
<th>Rio de Janeiro³ N = 6,320,446</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age group, years</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0–9</td>
<td>211 (3%)</td>
<td>21,260 (16%)</td>
<td>759,791 (12%)</td>
</tr>
<tr>
<td>10–19</td>
<td>646 (11%)</td>
<td>23,473 (18%)</td>
<td>930,717 (14%)</td>
</tr>
<tr>
<td>20–29</td>
<td>1,697 (26%)</td>
<td>25,746 (20%)</td>
<td>1059,810 (17%)</td>
</tr>
<tr>
<td>30–39</td>
<td>1,175 (18%)</td>
<td>22,951 (18%)</td>
<td>992,986 (16%)</td>
</tr>
<tr>
<td>40–49</td>
<td>862 (13%)</td>
<td>16,514 (13%)</td>
<td>876,487 (14%)</td>
</tr>
<tr>
<td>50–59</td>
<td>822 (13%)</td>
<td>10,800 (8%)</td>
<td>759,804 (12%)</td>
</tr>
<tr>
<td>60+</td>
<td>1,016 (16%)</td>
<td>9,026 (7%)</td>
<td>940,851 (15%)</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>4,456 (69%)</td>
<td>66,027 (51%)</td>
<td>3360,629 (53%)</td>
</tr>
<tr>
<td>Male</td>
<td>1,973 (31%)</td>
<td>63,743 (49%)</td>
<td>2959,817 (47%)</td>
</tr>
<tr>
<td><strong>Self-reported skin colour³</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brown</td>
<td>3,108 (48%)</td>
<td>62,156 (48%)</td>
<td>2318,675 (37%)</td>
</tr>
<tr>
<td>White</td>
<td>1,777 (28%)</td>
<td>51,580 (40%)</td>
<td>3239,888 (51%)</td>
</tr>
<tr>
<td>Black</td>
<td>1,215 (19%)</td>
<td>14,341 (11%)</td>
<td>708,148 (11%)</td>
</tr>
<tr>
<td>Asian</td>
<td>154 (2%)</td>
<td>1,445 (1%)</td>
<td>45,913 (1%)</td>
</tr>
<tr>
<td>Indigenous</td>
<td>30 (1%)</td>
<td>178 (0%)</td>
<td>5,981 (0%)</td>
</tr>
<tr>
<td>Missing</td>
<td>145 (2%)</td>
<td>69 (0%)</td>
<td>1,842 (0%)</td>
</tr>
</tbody>
</table>

¹Data collected from August 2021 to January 2022.
²2010 census carried out by the Brazilian Institute of Geography and Statistics (IBGE)
³Skin colour was recorded as self-reported race or skin colour, classified as Branco (White), Preto (Black), Pardo (Brown), Amarelo (Asian), or Indígena (Indigenous).
Data analysis
We will describe seroprevalence, socio-demographic characteristics, and the burden of the disease for the baseline and each wave using descriptive statistics. Categorical variables will be displayed in absolute and relative frequencies, and continuous variables will be described as means and standard deviation (SD) or medians and interquartile range (p25-p75). When applicable, missing values will be imputed using multiple imputations with chained equations (MICE).

Direct comparisons between baseline and waves characteristics will be performed using statistical hypothesis testing, such as Pearson’s chi-square test, Fisher’s exact test, Student’s t-test or Wilcoxon-Mann-Whitney test. Correlation between variables will be calculated using Spearman’s rho correlation.

We will use regression modelling approaches to estimate the association of socioeconomic characteristics and seroprevalence of SARS-CoV-2 infection. The response variable will be the seroprevalence outcome; therefore, we will apply generalised linear models accounting for repeated measures when necessary. We have considered study and work status, number of healthcare workers, the density of residents in households, use of public transport, and number of vaccinated residents as exposure variables. Numerical variables will be modelled using cubic splines. All statistical analyses are to be performed in R. A significance level of 0.05 will be considered. We will follow STROBE recommendations. Dictionaries data from the first and the second waves are available on the extended data.

Dissemination of study results
Monthly meetings with local leaders, agents from the NGO Redes Maré and the primary healthcare units’ workers are to be held throughout the study to notify and discuss, for example, the main self-reported comorbidities, vaccine status, number of tests and data collection at each phase. The meeting’s objectives are to show the scientific results from the cohort to stakeholders and create a space for scientific discussions about how this study can improve healthcare in a vulnerable population. Newsletters addressing the pandemic context and initiatives which happened in the community were regularly disclosed on print and digital platforms, targeting the cohort participants, healthcare workers and the Maré population.

Conclusions/discussion
Cohort studies conducted in urban poor communities are essential to understand the spread of pathogens, protective strategies, disease burden and health resource needs in areas where epidemiological information is usually limited.

This study aims to contribute a new source of data on infectious and poverty-related diseases for researchers, policymakers and programme implementers who wish to improve health services for a vulnerable population in an urban area of a LMIC. The results will be used to support the establishment of a public health data repository and communication and mobilisation activities that can promote data-driven decision-making and public health action in other vulnerable communities, especially in LMICs.

Data and results will be limited to the follow-up of participants during wave collections. To mitigate potential dropout, engagement and communication activities will be performed with Maré residents. The communication team will be in charge of the dissemination and information in different platforms, with weekly actions in social media, newsletters posters and cards. In parallel, the field team will be in direct contact with the participants during waves. We also plan to conduct events such as seminars to disseminate the insights from the study.
Moreover, records linkage with secondary data to enrich the cohort participants and community information may present inconsistencies since external databases offer different input patterns and data quality. It should be emphasised that participants enrolled in the cohort will be mostly vaccinated, as enrolment started after the vaccination campaign.

The engagement process is challenging in a vulnerable population due to difficulties in maintaining contact with participants, as they frequently change their mobile phone numbers and addresses. This is combined with low adherence to other COVID-19 vaccine doses and the potential refusal to participate in the subsequent waves of data collection.

Data availability
Underlying data
No underlying data are associated with this article.

Extended data

This project contains the following extended data:

- Interview questions:
  - W1_adults_questionnaire_EN.pdf (Questionnaire to adult residents used in the first wave - English);
  - W1_minors_questionnaire_EN.pdf (Questionnaire to residents aged under 18 years used in the first wave - English);
  - W2_adults_questionnaire_EN.pdf (Questionnaire to adult residents used in the second wave - English);
  - W2_minors_questionnaire_EN.pdf (Questionnaire to residents aged under 18 years used in the second wave - English)

- Data dictionaries:
  - Wave1_dictionary_data.xlsx (Data dictionary from the first wave);
  - Wave2_dictionary_data.xlsx (Data dictionary from the second wave)

- Informed Consent forms:
  - Adult_Consent_Form_PTBR.pdf (Informed consent form for adult residents - Portuguese);
  - Legal_Responsible_Consent_Form_PTBR.pdf (Legal responsibility consent form for adults - Portuguese);
  - Minor_Assent_Form_7_11_years_PTBR (Informed agreement form for residents aged 7 to 11 years - Portuguese);
  - Minor_Assent_Form_12_17_years_PTBR.pdf (Informed agreement form for residents aged 12 to 17 years - Portuguese)

- Newsletter on vaccine studies carried out in the Maré:
  - Newsletter_Mare_Covid_studies_EN.pdf (Newsletter – COVID-19 cohort studies in Maré - English)

Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0).

Acknowledgements
We acknowledge support from the National Council for Scientific and Technological Development (CNPq), the Coordination for the Improvement of Higher Education Personnel (CAPES) - Finance Code 001, Carlos Chagas Filho Foundation for Research Support of the State of Rio de Janeiro (FAPERJ), and the Pontifical Catholic University of Rio de Janeiro.

Otavio T Ranzani acknowledges indirect support from the Instituto de Salud Carlos III (CD19/00110) by a Sara Borrell grant, from the Spanish Ministry of Science and Innovation and State Research Agency through the “Centro de Excelencia Severo Ochoa 2019–2023” Program (CEX2018-000806-S), and the Generalitat de Catalunya through the CERCA Program received a research grant from the Health Effects Institute for research unrelated to this manuscript.

We would like to thank the Unidade de Apoio ao Diagnóstico da Covid-19 (UNADIG-FIOCRUZ) for their support in the test diagnosis.

The authors also would like to thank Maré residents and all the professionals who engaged in the project. Without them, this work would not have been possible. Thank you for all the hard work!

References
3. de Souza CDF, Machado MF, do Carmo RF: Human development, social


21. Lei Geral de Proteção de Dados Pessoais (LGPD). Reference Source


Open Peer Review

Current Peer Review Status: Version 1

Reviewer Report 24 July 2023

https://doi.org/10.21956/gatesopenres.15333.r33959

© 2023 Umakanthan S. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Srikanth Umakanthan
Paraclinical Sciences, The University of the West Indies, St. Augustine, Trinidad and Tobago

Abstract: The methodology explanation should be curtailed, and more emphasis should be given to the results. Results and conclusion should be added.

Keywords: should be different from the title.

Introduction and discussion can be strengthened by incorporating the following points:

1. Include a sentence on origin, transmission of COVID-19 (refer and cite: Umakanthan et al., 2019).
2. In brief about Post acute(long) COVID-19 (refer and cite: Umakanthan et al., 2023).
3. COVID-19 awareness among medical professionals (refer and cite: Umakanthan et al., 2022).
4. The role of social environmental predictors of COVID-19 in general (refer and cite: Umakanthan et al., 2022).
   - List the inclusion and exclusion criteria.
   - Were the factors classified based on variables?
   - How efficient was the survey tool in eliminating/minimizing the bias?
   - Was ethical clearance obtained before data extraction?
   - Was the online survey tested using a pilot study before the inception to the survey population?
   - Indicate the strengths and limitations of your study.

References
PubMed Abstract | Publisher Full Text


**Is the rationale for, and objectives of, the study clearly described?**
Yes

**Is the study design appropriate for the research question?**
Yes

**Are sufficient details of the methods provided to allow replication by others?**
Yes

**Are the datasets clearly presented in a useable and accessible format?**
Partly

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Covid-19, Lung cancer, molecular genetics, breast cancer,

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

---

**Comments on this article**

**Version 1**

Reader Comment 21 Mar 2023

Vesselin Blagoev, Management and Marketing, Varna University of Management, Sofia, Bulgaria
Excellent methodology and protocol. All important elements of the research are stated correctly. The sample is representative. The statistical methods, which have been used so far, and will continue to be used, are relevant and guarantee correct results of the research. Problematic, to some extent, is that a part of the respondents change and it will be tricky to get as close as possible to the results which would represent the numbers if the sample of respondents would be about the same. However, such a problem is always expected when the survey population and sampling frame are like in this case. Altogether, I find the research going very well and with promising correct expected results.

Vesselin Blagoev

**Competing Interests:** I am not a participant, or co-author, or linked in whatever way to the sponsors of this research. Vesselin Blagoev