ABSTRACT
Introduction Since 2020, the world has been going through a viral pandemic with a high morbidity and mortality rate along with the potential to evolve from an acute infection to post-acute and long-COVID, which is still in the process of elucidation. Diagnostic and prognostic research is essential to understand the complexity of factors and contexts involving the illness's process. This protocol introduces a study strategy to analyse predictors, sequelae, and repercussions of COVID-19 in adults and older adults with different disease severities in the State of Paraná, Brazil.

Methods and analysis A mixed-methods sequential explanatory design. The quantitative data will be conducted by an ambispective cohort study, which will explore the manifestations of COVID-19 for 18 months, with nearly 3000 participants with confirmed diagnoses of COVID-19 (reverse transcription-PCR test) between March and December of 2020, retrieved from national disease reporting databases, over 18 years old, living in a Brazilian State (Paraná) and who survived the viral infection after being discharged from a health service. Data collection will be conducted through telephone interviews, at two different occasions: the first will be a recall 12 months after the acute phase as a retrospective follow-up, and the second will be another prospective interview, with data of the following 6 months. For the qualitative step, Grounded Theory will be used; participants will be selected from the cohort population. The first sample group will be composed of people who were discharged from the intensive care unit, and other sample groups will be composed according to sociodemographic factors and the health risk of the cohort members and the severity of COVID-19 symptoms.

STRENGTHS AND LIMITATIONS OF THIS STUDY
⇒ The study stands out for using an ambispective cohort design to follow COVID-19 survivors in the State of Paraná, Brazil for up to 18 months, and for qualitatively assessing different social, epidemiological and clinical aspects of the long-term consequences of COVID-19.
⇒ The design allows comparing the course of the post-COVID-19 syndrome in the state’s macro-regions, between adults and older adults after treatment in an outpatient, ward or intensive care unit, associated with the severity of the disease, in line with the WHO Ordinal Disease Outcome Scale.
⇒ The stratification and use of multivariate regression models—Poisson, Cox and linear logistics—allow estimating the long-term prognosis of the disease, according to sociodemographic factors and the health risk of the cohort members and the severity of COVID-19 symptoms.
⇒ Possible limitations of the study include its observational type and the quality of data available in the disease-reporting system database, as well as the use of severity-based groups for dose–response effect comparison.

State of Paraná, under opinion number: 4 214 589 and CAAE: 34787020.0.3001.5225 on 15 August 2020. The participants will verbally consent to the research, their consent will be recorded, and the informed consent form will be sent by mail or email. Outcomes will be widely disseminated through peer-reviewed manuscripts, conference presentations, media and reports to related authorities.

Correspondence to
Dr Maria Aparecida Salci
masalci@uem.br
INTRODUCTION
The COVID-19 pandemic, caused by the new coronavirus (SARS-CoV-2), continues to threaten global public health. On 1 October 2021, more than 233.160 million COVID-19 cases and more than 4.7 million deaths due to the disease were confirmed worldwide. In Brazil, confirmed cases up to the same date exceeded 21.3 million people, about 10% of the country’s population, and more than 595,000 deaths were reported. As of 1 October 2021, Brazil is the country with the third highest number of confirmed COVID-19 cases, and the second highest number of COVID-19-related deaths.

SARS-CoV-2 infection may cause several clinical signs, ranging from asymptomatic conditions to cases that progress to multiple organ failure, leading to the individual’s death. Recent evidence highlights a multisystemic effect of the virus on the human body, with respiratory, cardiovascular, musculoskeletal and neuropsychological damage simultaneously. However, much discussion has increased around the long-term effects of COVID-19, as many months after the acute course of the infection, patients continue to showcase manifestations of the disease. Some authors are proposing to distinguish between the short-term and long-term consequences of COVID-19 to better understand and research the effects of the disease over time. It is proposed to describe post-acute COVID as symptoms extending beyond 3 weeks, and long-COVID as symptoms extending beyond 12 weeks, after the onset of initial symptoms. Despite the advances of research on clinical and epidemiological characteristics of the disease and its complications, there are still many questions about the physical and psychological consequences of the so-called post-acute COVID, post-COVID-19 syndrome or long-COVID.

Throughout 2020, the priority of health services was to respond to the acute demands of COVID-19, with a focus on hospital and intensive care. However, as the pandemic persisted, providers and institutions other than hospitals, especially including outpatient services and primary healthcare providers, had growing demand from the increasing number of COVID-19 survivors. Functional and radiological abnormalities of different degrees have been reported in people who have recovered from the severe form of COVID-19. Overall satisfaction and quality of life in patients with severe and long-term complications from the disease may decrease, whether due to physical or mental health limitations. However, the impairment of their physical or mental capacity for social and economic interaction should also be considered as a preponderant factor for this study.

The several health problems associated with long-term COVID-19 require complex and joint responses by all countries. In Brazil, it is especially challenging to respond efficiently to the country’s health system needs due to its wide geographical span, the historical inequalities in the provision of health services and the chaotic national management of the pandemic. To address this issue, the study designed two cohorts in order to investigate the clinical, epidemiological, and social manifestations of post-acute and long-COVID between adults and older adults discharged from outpatient, ward, or intensive care unit (ICU) services in the State of Paraná, Brazil.

OBJECTIVES
The study’s main objective is to analyse the long-term consequences on the health and well-being conditions of adults and older adults with post-acute and long-COVID, in the State of Paraná, Brazil, according to the severity of the disease, defined by the type of treatment in an outpatient, ward or ICU facility.

The following are the primary objectives of the study:
1. Identify the signs and symptoms of COVID-19 that occurred during the acute onset of the disease.
2. Monitor the path of physical and emotional sequelae in the short, medium and long term, according to the severity of the disease.
3. Examine predictive risk factors for the development of the severe form of COVID-19 and the occurrence and worsening of post-acute and long-COVID.
4. Evaluate the effects of different degrees of disease severity and post-acute and long-COVID on productivity and economic conditions in adults and older adults in the cohort and their families.

METHODS AND ANALYSIS
Design
The mixed-methods (MM) sequential explanatory design was employed, in order to improve the combination of quantitative and qualitative approaches. When investigating an unknown problem—post-acute and long-COVID—through the combination of both methods, there will be a chance to offer more complex and better answers than those achieved with a single research strategy. In the sequential explanatory MM (QUAN → qual) design, quantitative data have greater emphasis and are collected before qualitative data. The two types of collection are independent, but connected, as they involve individuals from the same cohort under follow-up.

The long-term effects on the health and well-being will be followed by an ambispective prognostic cohort of adults and older adults in the State of Paraná, Brazil, with a confirmed diagnosis of COVID-19 and discharged from an outpatient health service, hospital ward or ICU from March to December 2020.

A control group would be valuable, but the source of study information is the notifications of COVID-19 in the State of Paraná, which do not provide information for people without the disease. However, as the study intends to observe the variability of problems, depending on the severity of the disease, the three groups of comparison are able to show it. Thus, despite the limitation, it will be possible to understand the differences in the prevalence, duration and intensity of persistent symptoms according
to the severity of the disease. The study groups will make possible to establish a dose–response effect on the relationship between the severity of COVID-19 in the acute phase and its manifestations in the post-acute and long phases.

The clinical, epidemiological, and social manifestations of patients with post-acute and long-COVID will be followed for 18 months, through two consecutive phases. The initial contact 12 months after the acute infection will involve an interview looking retrospectively at symptoms reported over the period, emphasising the conditions at 1st, 3rd, 6th, 12th and 18th months after discharge from an outpatient health service, a hospital ward or an ICU. After that, a second interview will ask the participants about the symptoms they present at 18 months after the acute infection, the prospective component, and a further retrospective description on the symptoms between 12 and 18 months.

The qualitative step will be guided by the Constructivist Grounded Theory (GT) by Charmaz. A subsample of cohort participants will be selected, including members from the groups who were under outpatient, ward or ICU treatment. As the research progresses, other groups will be composed to create analytical categories, respecting the theoretical principle.

The study is under development with the collaboration of national and international research, teaching and extension centres, within the scope of the Brazil-US COVID-19 Network, articulated internationally at Duke University and nationally at the State University of Maringá, Federal University of Pelotas and Health Department of the State of Paraná.

Setting
Brazil has an estimated population of 214.5 million inhabitants, with a territory of 8,514,876 km², and is the fifth largest country in the world, by area. Its territory is divided into five major regions (North, Northeast, Southeast, South and Midwest), 26 states and a Federal District.

The State of Paraná is in the southern region of the country, with an estimated population of over 11.5 million inhabitants. It has 399 municipalities, grouped into 22 regional health departments, divided into 4 macro-regional health departments (figure 1). Paraná ranks the 18th state in Brazil in the absolute number of confirmed cases of COVID-19 and 5th in the number of deaths due to the disease according to data from August 2021. There are more than 1.4 million cases and over 36,800 deaths from the disease in the state. Also, there is a high risk of dissemination of SARS-CoV-2 in this state, due to the migration of poor individuals from neighbouring countries (Paraguay and Argentina).

Outcomes
The main outcomes will be signs and symptoms of COVID-19 in the acute phase of the disease and its persistence in the 1st, 3rd, 6th, 12th and 18th months after discharge from an outpatient, ward or ICU. Besides monitoring the path of physical and emotional sequelae in the short, medium and long term, depending on the severity of the disease, the epidemiological and social consequences of the post-COVID-19 syndrome will also be outcomes under investigation (online supplemental table 1).

The protocol adopts the term ‘post-acute COVID’ to characterise symptoms that extend beyond 3 weeks, and ‘long-term COVID’ to include symptoms that extend beyond 12 weeks, after the initial symptoms of the acute infection.

Outcomes from a qualitative analytical perspective will also be included to answer the study objectives and other projects under development aiming to understand the process of contracting and living with the consequences of COVID-19 based on the participants’ experiences.

Procedures for the cohort sample selection
The cohort sample selection was developed considering the sources of the potential population for the study and the sample size calculation, as illustrated in figure 2 and described below.

Sources of the potential population for the study
Considering the treatment as a severity indicator of the acute disease, the initial selection of potential cases to be included was conducted in two official databases of reported cases of COVID-19 in Brazil, intentionally chosen for meeting this requirement.

The outpatient (mild) cases were obtained from the official database of the State of Paraná, entitled ‘NOTIFICA COVID-19 PARANÁ’. This database contains each of the municipalities’ reports of COVID-19 cases if they meet the requirement (reverse transcription (RT)-PCR test), including individuals who tested positive for COVID-19 and did not need hospitalisation. Regarding the cases that required a hospital ward (moderate) and intensive care (severe) treatments, the data source was the Influenza Epidemiological Surveillance Information System (SIVEP-Influenza), the country’s official database from the Ministry of Health. The SIVEP-Influenza is responsible for assembling records of severe acute respiratory infection cases in the national territory. Since the beginning of the pandemic, the database also started to report COVID-19 cases, upon laboratory confirmation (RT-PCR test).

As selection criteria, the participants should be 18 years of age or older at the time the data were reported, and should be residing and being treated in the State of Paraná. Pregnant or postpartum women were excluded from the sample.

Considering the overlapping of information from the State of Paraná database with the national database and chance of double records, we conducted a comparative analysis to exclude duplicates. Cases between databases were linked using variables previously included: (1) individual’s name; (2) date of birth; (3) sex and (4) mother’s full name. In the case of a double report in one of the databases, the first report remains. In the case of a double
report across different databases (outpatients and hospitalisation), the hospitalisation report remains.

Sample size calculation
In a context of multiple outcomes and great variability in the prevalence of persistent symptoms, the sample size for association was defined according to Fleiss’s continuity correction method. A confidence level of 95% and a statistical power of 80% were defined to detect a relative risk ratio of at least 2.0 in individuals with the severe form of COVID-19 for a prevalence of symptoms of at least 4.8% to those who were seen in an outpatient clinic (table 1).

The statistical power of the study will be even greater to detect a lower risk ratio in individuals hospitalised in the ICU, in the case of symptoms with a higher prevalence to those with the mild form of COVID-19.

For the whole cohort, we will select 1500 adults (500 outpatient, 500 ward and 500 intensive care) and 1500 older adults (500 outpatient, 500 ward and 500 intensive care).

Sample selection
For the sample selection process, we used a proportional stratified approach to avoid compromising its homogeneity, considering: (1) health macro-region (North, Northwest, East and West) of the individual’s city of residence; and (2) month in which the individual’s contraction of the disease was reported (March–December 2020) or the month in which the patient was discharged from the hospital. Our approach considered the subdivision of the management and organisation mechanisms of health services in the State of Paraná, including differences in infrastructure, human resources, support and normative specificities for coping with COVID-19 throughout 2020.

Data collection
The data collection procedures will be conducted by a team of interviewers, using adequate tools and resources to the expected processes. The data collection team will be composed of professionals and students within the health field—undergraduate and graduate (n=40)—trained by
a ‘training and professional qualification of the research project’ lasting 20 hours.

Regarding the tools, each interviewer will receive a smartphone with a working phone line and with the CubeACR and WhatsApp applications installed to record calls, and the website and password to access the electronic registration form collection in the JotForm website, available at: https://sites.google.com/uem.br/coortecovid19pr/.

The list of potential participants will be randomly assigned to each interviewer, and within a given time, interviewers must return to the cohort supervisors all the materials collected and the interview reports (spreadsheet created in Microsoft Office Excel V.2016 software with the information collected from each interview).

The technique adopted for data collection will be the interview, characterised by Tele-collection, which will take place preferably by telephone by a voice or video call, at the participant’s discretion, for the application of the electronic interview form. The main operational procedures for conducting the cohort interview are described in table 2.

The standard electronic form is organised into four thematic sections, to verify the objectives proposed by the study. The four thematic sections cover the aspects of the cohort regarding its identification, history (signs and symptoms of COVID-19 and post-COVID-19 syndromes; history of COVID-19 and post-COVID-19 and

Table 1  

<table>
<thead>
<tr>
<th>Prevalence</th>
<th>Mild COVID-19 outpatient clinic (n=500)</th>
<th>Severe COVID-19 intensive care (n=500)</th>
<th>Risk ratio</th>
<th>Statistical power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Symptom 1</td>
<td>4.8%</td>
<td>9.6%</td>
<td>2</td>
<td>80.4%</td>
</tr>
<tr>
<td>Symptom 2</td>
<td>6.5%</td>
<td>12%</td>
<td>1.8</td>
<td>82.5%</td>
</tr>
<tr>
<td>Symptom 3</td>
<td>9.5%</td>
<td>16%</td>
<td>1.7</td>
<td>84.9%</td>
</tr>
<tr>
<td>Symptom 4</td>
<td>9.7%</td>
<td>16%</td>
<td>1.6</td>
<td>82.3%</td>
</tr>
<tr>
<td>Symptom 5</td>
<td>14%</td>
<td>22%</td>
<td>1.6</td>
<td>89.6%</td>
</tr>
</tbody>
</table>

Figure 2  Procedures for sample calculation and selection; Paraná, Brazil, 2021.
Qualitative data collection will also be conducted by Tele-collection, using the intensive interview technique, which allows for detailed clarification and interpretation of the participants regarding their experience. It will be directed by a guide of questions about persistent signs and symptoms and people’s health needs in a longitudinal follow-up with the cohort participants. The initial interview guide is composed of open-ended questions, analysed by researchers with expertise in qualitative research and/or in the study theme. After collection and analysis of each interview and each sample group formed, new questions may be incorporated into the guide, as allowed by the GT method. All interviews will be audio-recorded and later transcribed in full.

Data processing and analysis
The data resulting from the application of the electronic form and the intentional interviews will be extracted from the platform in a virtual environment through reports and compilation spreadsheets generated by the website and analysed following the methodological approaches of the study.

Quantitative data
Quantitative data collected by the electronic form will be exported to Microsoft Office Excel V.2016 and R V.4.0.4 (R Core Team, 2021) software for data processing, such as variable coding, variable labels, missing variables and others.

Obtaining adequate descriptive statistics for the variables under study (mean and SD, or absolute and relative frequencies), especially for the main outcomes of acute and persistent manifestations of COVID-19, will be part of the first stage of data analysis.

Regarding analytical statistics, the association between variables will be tested using a \( \chi^2 \) test with a significance level of 5% to verify potential interactions between prognostic factors and the long-term consequences of COVID-19. Multivariate logistic regression models will be tested to check the association of independent variables with increased risks of signs and symptoms of COVID-19, according to patients’ treatment in outpatient services, hospital ward and ICU. Multivariate models to verify the association of independent variables with the outcomes of severity and time of signs and symptoms of the post-COVID-19 syndrome will use Poisson, Cox and linear regression methods, according to the adequacy and explanatory capacity.

All analyses will be adjusted for macro-regional context, age, sex, education, socioeconomic status, health conditions before COVID-19 and other potential confounding factors.

Factors associated with disease severity will be assessed using the \( \chi^2 \) test, relative risk, OR and prevalence ratio, according to the regression models used. The stratification will allow defining the path of the disease and its long-term consequences, according to its severity, in
adults (aged 18–59 years) and older adults (60 years and older).

Qualitative data
After conducting the qualitative interview and audio transcription, the final narrative data will be analysed according to the coding steps in the GT theoretical framework, including:

- **Initial coding**: the researcher makes a rigorous study of the data, conceptualising the participants’ ideas by creating codes that can be described word by word, line by line or incident by incident.
- **Focused coding**: at this step, the codes produced at the beginning of coding are directed, selected and conceptualised, generating a synthesis that explains the amount of data produced. This step requires decision-making and allows the researcher to choose the most relevant codes for better analytical understanding and categorisation of data concisely and completely.
- **Axial coding**: allows the researcher to relate categories to subcategories, specifying the properties and dimensions of a category. It is considered the time to regroup the data that were fragmented in the initial step, giving coherence to the emerging analysis.
- **Theoretical coding**: as the most sophisticated level of coding, allows the researcher to conceptualise how codes relate to each other and form hypotheses to integrate the fundamental theory of the data.

To organise and analyse the data, the MAXQDA software will be used. This system allows the researcher to categorise relevant information using codes, colours, symbols and emoticons, and make notes that can be attached to documents, codes or projects.

**Patient and public involvement**
Patients and the public were not directly involved in the development of the research question or study design. The results of this project will be publicly available on the research website and through press releases elaborated by the group, which the patients are encouraged to follow.

**ETHICS AND DISSEMINATION**
Research is developed following the guidelines of Resolutions 466/12 and 510/16 of the Brazilian National Health Council for research with human beings. Authorisation was obtained from the Health Department of the State of Paraná and approval was granted by two Permanent Human Research Ethics Committees, one from the State University of Maringá under opinion number: 4.165,272 and CAAE: 34787020.0.0000.0104 on 21 July 2020, and the other from the Hospital do Trabalhador (Worker’s Hospital), which is accountable for the Health Department of the State of Paraná, under opinion number: 4,214,589 and CAAE: 34787020.0.3001.5225 on 15 August 2020.

At the beginning of the telephone contact and informed consent procedures, the participants will be introduced to the study and asked if they agree to participate. Permission or denial to conduct the interview will be recorded by the interviewer. If the answer is ‘No’, the form will be closed immediately, without any information from the participant. If the participant’s answer is ‘Yes’, the interview will continue and the questions will be presented through paused reading, allowing time for the answer. If the participant refuses to answer any question, their choice will be respected and their authorisation will be requested again to continue the interview, presenting the subsequent questions. A copy of the informed consent form will be sent to the participant by mail or email after the conclusion of the interview.

Scientific articles will be submitted to high-quality national and international journals. Theses, dissertations and academic monographs of the study’s researchers will be presented to graduate programmes and undergraduate courses, especially in the field of nursing and the State University of Maringá. The study and its outcomes will also be disseminated to the press and social networks for the lay public, along with presentations to the academic community, managers, and health professionals at meetings, seminars, and events.

**DISCUSSION**
By carefully classifying the monitored population, the disease severity profile and therapeutic strategies, the study under development supports the advancement of coping with post-acute and long-COVID not only in the Brazilian Unified Health System but also in the international setting. The study will support the development of health services that strengthen equity and promote the quality of care for all. With robust scientific evidence, it also supports decision-making in health policies in the short, medium and long term.

A possible limitation is the use of 7-point, self-referred Likert scales to approach the disease’s severity rather than validated questionnaires and scales. The use of Likert scales is to determine the patient’s perceived symptom severity, looking forward to the qualitative step, where the perceptions of the disease and treatment will be assessed in depth. We highlight the lack of validated scales in Brazil that could address all the symptoms at the time of the research’s design.

A control group would also be valuable, but the source of study information is the notifications of COVID-19 in the State of Paraná, which do not provide information for people without the disease. However, as the study intends to observe the variability of problems, depending on the severity of the disease, the three groups of comparison are able to show it, in a dose–response effect.

Thus, despite the limitations, it will be possible to understand the differences in the prevalence, duration and intensity of persistent symptoms according to the severity of COVID-19.


