ERASMUS UNIVERSITY ROTTERDAM, Erasmus School of Economics

Bachelor Thesis International Bachelor's in Economics and Business Economics

Inequality in Brazil's Universal Healthcare System: How Distance to a Healthcare Facility increases Patients' COVID-19 Mortality Risk through Delay in Care.

Name student: Joana Geisler
Student ID number: 508145

Supervisor: Dr. Eddy Van Doorslaer
Second assessor: Dr. Zhiling Wang

Date final version: 07.07.2022

Abstract

With the COVID-19 pandemic putting unprecedented strain on Brazil's healthcare services, inequalities in healthcare access have been made more apparent than ever. In this paper, the impact of delay in treatment for COVID-19 due to distance to a healthcare facility on the mortality of a patient is assessed using an Instrumental Variable (IV) approach, with data made available by Brazil's public healthcare system (*SUS*). The results show that distance-attributed delay is associated with a large increase in mortality, though the small impact of distance on delay suggests that few patients delay care for longer periods of time due to distance alone. No significant difference in delay is found between Brazil's Northern and Central-Southern regions, despite longer distances in the former. The results emphasize the importance of how barriers to healthcare can create inequalities in health outcomes, even in healthcare systems where everyone has the right to treatment.

List of Abbreviations:

SUS: Sistema Único de Saúde, Brazil's universal healthcare system

ICU: Intensive care unit

IV: Instrumental Variable

FE: Fixed Effects

OVB: Omitted Variable Bias

LATE: Local Average Treatment Effect

Contents

1. Introduction

In 1988 Brazil introduced a universal healthcare system, the *Sistema Único de Saúde* (SUS), which has enabled access to healthcare for those least able to afford it. However, the COVID-19 pandemic has highlighted that even though everyone has the right to healthcare, access remains unequally distributed within Brazil (Baqui et al., 2020, Rocha et al, 2021).

Brazil is a diverse and divided country. The Northern regions, namely the North and Northeast macroregions, are significantly poorer and less educated, with a larger share of the population identifying as Black or *Pardo* (mixed race), than the Central-South regions, consisting of the Central-West, South and Southeast (Bacqui et al., 2020; Rocha et al., 2021). While the population in the Northern regions of the country generally have fewer comorbidities, they have a shorter (healthy) life expectancy (Szwarcwald et al., 2016). During the pandemic, this manifested itself in the fact that of the 619,334 COVID-19 deaths that occurred until the end of 2021 (Ritchie et al., 2020), the Northern regions suffered a disproportionate burden; COVID-19 mortality rates were double those of the Central-South (Rocha et al., 2021).

Various studies suggest that the explanatory factor is the differential access to healthcare services between Brazil's states and cities (Rocha et al., 2021; Bacqui et al., 2020; Pereira et al, 2021). While these disparities in healthcare certainly predate the pandemic (Amaral et al., 2017; Albuquerque et al., 2017), the increased demand for intensive care units (ICUs), and other healthcare services during this crisis, accentuated some of the regional inequalities that Brazil faces. This is highlighted by the fact that, even though the first cases of COVID-19 were high-income Brazilians who returned to São Paulo from intercontinental travels, it was lower socioeconomic groups that eventually suffered most, both in terms of mortality (Rocha et al., 2021) and income (Masri et al., 2021).

An interrelated but nonetheless distinct aspect of regional inequality in patients' survival probability is distance to a healthcare facility (Bacqui et al., 2021). In Brazil, distance to a healthcare facility varies starkly within the population. This has implications for accessibility to healthcare and hence survival probabilities of the population, particularly in remote areas. Given that the distribution of healthcare services has played a significant role in discrepancies in mortality between the Northern and Central-Southern regions, this paper will investigate how distance to a healthcare facility, through its impact on delay in care, has affected COVID-19 mortality. Hence, the research question will be:

How did the delay in treatment for COVID-19 due to distance to the healthcare facility affect in-hospital mortality risk in Brazil in 2020 and 2021?

The impact of distance to a healthcare facility on delay in care and its association with mortality risk is well known from pre-pandemic studies on other illnesses. Distance has been found to increase delay in hospital admission, for example in the case of myocardial infarctions in Sweden (Nilsson, 2016) and tuberculosis in Ethiopia (Demissie, 2002). This is because distance acts as a disincentive as well as a barrier to accessing healthcare (Thaddeus & Maine, 1994). Through delay, distance to a healthcare facility has been associated with a higher mortality risk for neonatal mortality in Vietnam (Målqvist et al., 2010) and early childhood mortality in Burkina Faso (Schoeps, 2011). However, with the COVID-19 pandemic putting unprecedented strain on healthcare systems, it is likely that other factors are important for explaining how distance interacts with mortality through a delay in care. Thaddeus and Maine (1994) suggest that the impact of distance on delay in accessing healthcare also depends on the perception of the illness and of healthcare providers. As the population in the Northern regions of Brazil are generally further away from a healthcare facility, some of the analysis will be dedicated to whether part of the increased mortality burden in the Northern states can be attributed to greater distance to healthcare facilities.

While there exists an abundance of literature documenting the delay in other healthcare services such as cancer treatment in Brazil (Santos Costa et al., 2021; Mahl et al., 2020), in the U.S. (Czeisler, 2020; Birkmeyer et al., 2020), and Europe (Vintura, 2021), there have been few studies considering distance as a relevant risk factor for COVID-19 mortality through delay in care.

In an analysis of 14 European countries, the importance of travel time to healthcare facility for mortality was found to be negligible, with greater importance given to local demand for ICUs (Bauer et al., 2020). However, greater distances and worse transport infrastructure likely imply a different result for Brazil. Studies that cover spatial inequalities in healthcare in Brazil focus on the distribution of healthcare services at a state level, not based on distances (Rocha et al, 2021; Baqui et al., 2020; Silva et al., 2021; Souza et al., 2021). Bacqui et al. (2020) and Bacqui et al. (2021) show that distance to a healthcare facility is a relevant risk factor for in-hospital mortality in Brazil, though they do not analyze delay as a mechanism. This is relevant as distance to a healthcare facility can affect mortality through other factors, as distance to a healthcare facility is generally greater in areas with relatively lower socioeconomic status, and thus may affect groups that have a greater risk irrespective of distance (Noronha et al., 2020). Another study comparing 20 Brazilian cities showed that healthcare services were unequally distributed between and within cities. However, they did not investigate how that translated into mortality risk (Pereira et al., 2020).

Improving the understanding of how distance to a healthcare facility affects patients' decisions of when to access care during public health crises can, for one, inform policymakers of where greater availability of advanced healthcare equipment is needed. However, a more contextually relevant impact of this research might be its emphasis on the importance of how to communicate with the population – especially vulnerable groups – about the health risks associated with COVID-19. This is because part of the delay in seeking care is likely attributable to the disincentive that distance presents, and hence the result of a decision to delay care, for example, if the illness is not perceived as severe. It is therefore essential that elected officials stress the risks and urgency that a severe COVID-19 infection presents to reduce delay in care.

To analyze the impact of distance-induced admission delay on mortality, an Ordinary Least Squares (OLS) regression of mortality on delay and distance will be estimated. However, as most of such delay appears to be determined by disease severity, potentially causing downward bias in the OLS estimator, an Instrumental Variable (IV) approach is warranted. To estimate the impact of distance on in-hospital mortality through admission delay in Brazil, we will use distance as an instrument. This has been commonly applied in healthcare research since its early implementation by McClellan et al. (1994), in which the authors estimate the effect of intensive treatment of myocardial infarction on survival using the instrument of differential distance between a hospital with and without intensive treatment. By using an IV approach, one can estimate how the part of delay caused by distance affects a patient's mortality risk. Although exogeneity of the instrument of distance cannot be ensured, including various control variables and hospital and municipality Fixed Effects (FE) can help reduce the bias from confounding variables such as socioeconomic characteristics of the patient and hospital quality.

What will follow is a theoretical framework laying out the rationale behind the hypothesis that distance, through delay in care, increases in-hospital mortality. This will be complemented with an analysis of whether this can partly explain why the Northern regions suffered from higher in-hospital mortality rates. How these hypotheses will be tested, and the data used, will be laid out in the data and methods section. Finally, the results will be presented and discussed, followed by a conclusion.

2. Theoretical Framework

2.1. Distance, Delay, and Death

To elucidate the mechanisms through which distance affects mortality through delay, this section will dive into the framework developed by Thaddeus and Maine (1994) to formulate hypotheses. Understanding how the impact of distance on delay in care affects the risk of mortality can be divided into two sequences of events: a) the distance to the healthcare facility acts as a disincentive and barrier to accessing healthcare, leading to a delay (Thaddeus & Maine, 1994) and b) the delay in care leads to an increased risk of mortality by allowing the illness to advance to a later stage.

2.1.1. Distance to a Healthcare Facility and its Impact on Delay in Treatment

The "three delays" described by Thaddeus and Maine (1994) outline three phases that contribute to a delay in maternal care in low-income countries. First, there is the delay of decision, followed by a delay in arrival at the facility and delay in receiving treatment. Since its publication, the framework has been widely applied in healthcare research in low- and middle-income countries on maternal health (Mgawadere, 2017) and other contexts, including respiratory infections (Pajuelo et al., 2018). The analysis of this paper only observes the first two phases jointly – decision and arrival – with a focus on how distance impacts delay in care. It should be noted that the data used in this paper only observes the delay in hospital care and its impact on mortality, and therefore does not consider patients that do not reach healthcare because of distance-attributed delay.

Phase I Delay: Decision to seek care

In their study, the authors describe the various factors that contribute to a delay in the decision to seek care. This is based on the observation that increasing the availability of healthcare facilities is not always associated with an increase in their utilization (Thaddeus & Maine, 1994). The authors cite various explanatory factors from the literature, namely: distance, cost, quality of care and sociocultural factors. Importantly, it is the perception of these factors that affects the delay in decision-making of when the patient goes to the hospital. Regarding distance, the authors point to a phenomenon described as "distance decay", whereby rates of healthcare usage decrease as distances to the healthcare facility increase (Kelly et al., 2016). This suggests that distance as a disincentive to accessing healthcare services might not only cause underutilization, but also delay.

The impact of distance on the decision to delay care is also cited to depend on other factors such as perception of the disease. Patients who perceive themselves to have more severe symptoms usually delay care for shorter periods of time when it comes to stroke (Teuschl & Brainin, 2010), mental health (Wang et al., 2004) and cancer (Caplan, 2014). Brazilian President Jair Bolsonaro is likely to have impacted how patients and their families perceived the disease by downplaying its severity (Londoño, 2021). Estimating the impact of delay in treatment for COVID-19 must disentangle the effect that the perception of the severity of disease has on delay in care from other factors. To estimate the impact of delay on mortality, the impact of disease severity will be circumvented by focusing on distance-attributed delay. Moreover, perceived quality of care, and specifically reports and social media posts showing overwhelmed hospitals (da Silva, 2021; Coutinho, 2020; Globo, 2020) might have discouraged patients to seek care until the disease had progressed to a more severe stage. Hence, the perception of distance and other factors have implications for a patient's decision whether and when to seek treatment for COVID-19.

Phase II Delay: Delay in Arrival

These four factors – distance, cost, quality of care and socioeconomic status – only work as disincentives because they can present legitimate barriers to accessing healthcare. Delays in arrival are attributable to the distribution of facilities, travel distances, and the means of transportation (Thaddeus & Maine, 1994), thus relating to the concept of transport disadvantage. Transport disadvantage is often described as having few and/or poor transport options and is frequently measured using distance or travel time to goods and services, car ownership, and public transport access (Lucas, 2012; Currie & Delbosc, 2010).

As healthcare facilities in Brazil tend to be concentrated in major urban centers (Pereira et al., 2020; Bacqui et al., 2020), many areas in Brazil are underserved by the healthcare system. While there have been attempts at bridging the healthcare gaps between Brazil's most and least advantaged areas, such policies have focused on primary care (Szwarcwald et al., 2016; Benevenuto et al., 2019). However, with the pandemic overwhelming healthcare systems, more advanced healthcare services, such as ventilators, have proven more important than ever. Thus, this underserved population segment of remote Brazilians is significantly transport disadvantaged regarding healthcare access compared to their urban counterparts, considering the differential in distances that patients travel to access a healthcare facility.

Transportation means as a potential source of transport disadvantage is difficult to assess. While some socioeconomic groups such as women and people of color have been found to be transport disadvantaged in urban areas (Vilela, 2020; Redação Portal, 2022), it is not clear how populations outside major urban areas compare to the population living close to a healthcare facility.

Phase III Delay: Receiving care

Though not the interest of this analysis, delay in receiving care is a reality that many Brazilian patients faced during the COVID-19 pandemic, with reports of patients dying in queues for ICU admission (da Silva, 2021; Coutinho, 2020; Globo, 2020). Delay in receiving care is attributed to factors related to the healthcare facility that the patient attends, that is, the staff and equipment (Thaddeus & Maine, 1994). However, this delay is more often described as a function of the service areas of a hospital as well as the infection rates observed in those areas and is therefore not patient-specific.

The poorer quality of healthcare and more dispersed population in the Northern compared to the Central-Southern regions of Brazil (Rocha et al., 2021; Pereira et al., 2020; Bacqui et al., 2020; IBGE, 2010), imply that the distance that patients travel might partly predict the delay in receiving care. However, this association can be controlled for using hospital FE, thus isolating Phase I and II delays.

2.1.2. Delay in Treatment and Mortality

How the delay in care – whether due to distance or other reasons – impacts patients' survival probability has been widely discussed in the literature. Delay in care is evidently detrimental to emergency healthcare (De Luca et al., 2004; Bugiardini et al., 2017; Hirose et al., 2015), but has also been reported to increase mortality risk in non-urgent conditions such as cancer (Hanna et al., 2020; Kutikov et al., 2020), tuberculosis (Demissie, 2002) and, importantly, COVID-19 (Le Terrier et al., 2022). This is because delaying care allows the disease to progress in severe cases, thus making it harder to treat once at the facility and increasing the mortality risk (Thaddeus & Maine, 1994). Focusing on the impact of distance on delay, its effect on mortality gives way to the following hypothesis:

H1: Delay in care due to distance to a healthcare facility increases the mortality risk for COVID-19 in Brazil.

This hypothesis suggests that the delay caused by distance, due to Phase I and/or Phase II delays, increases the mortality risk for COVID-19. Consequently, even in public healthcare, patients are unequal; where you live can determine the quality and accessibility of healthcare and consequently your survival probability (Zaman et al., 2014).

While patients travelling further to a healthcare facility might experience greater delays between symptom onset and hospital admission, it is likely that there are other factors related to distance that translate into a higher mortality risk, a priori. Within cities, residents further away from hospitals have been found to be poorer and older (Pereira et al., 2021). Similarly, remote populations have higher rates of poverty, and a greater share of the population is Black, which has been associated with a higher mortality risk (Benevenuto et al., 2019; Albuquerque et al., 2017; Bacqui et al., 2020). In other words, the increased mortality risk associated with distance can be due to other factors such as education, which independently affects how far a patient lives from a hospital, and their mortality risk. This is illustrated by the fact that, before the pandemic, remote areas had a life expectancy 3.4 years lower than urban areas (Benevenuto et al., 2019). An analysis of the impact of distance on mortality through delay must consider factors associated with distance to a healthcare facility and patients' mortality risk to reduce bias of the estimator of delay due to distance. This will be done using various control variables as well as FE.

2.2.Regional Differences in Population Density and Mortality

Given stark socioeconomic differences between the Northern and Central-Southern regions of Brazil it is unsurprising that the former experience mortality rates twice those of the Central-Southern regions (Bacqui et al., 2020; Bacqui et al., 2021; Rocha et al., 2020). Moreover, lower population density in the Northern regions (IBGE, n.d.) begs the question whether part of the divide in Brazil's mortality rates might be due to the distance to the healthcare facility, and by extension, delay in care. This is likely accentuated by the fact that the Northern regions are known to have poorer transport networks, which should cause distance to weigh more heavily due to higher travel time per km (Amann et al., 2016), implying a steeper distance decay. Hence, it can be hypothesized that:

H2: In Brazil, part of the higher COVID-19 mortality burden of the Northern, compared to the Central-Southern regions, is due to longer delays attributable to distance to healthcare facilities.

Part of the analysis will distinguish between the North and Central-South regions of Brazil, as it can be expected that accessibility presents a greater obstacle to healthcare in the North, which is more rural, poorer, and has worse transport infrastructure. (Bacqui et al., 2020; Rocha et al., 2020). As a result, it is likely that for the same distance, the population in the Northern regions is, on average, more likely to be transport disadvantaged. Distinguishing between these areas of Brazil can generate insight on whether distance has a different impact on delay, and by extension, mortality. In other words, it can provide insight into whether the distance decay function differs between regions of Brazil.

3. Data

Distance from the healthcare facility and its impact on delay in care and in-hospital mortality rates are analyzed using the SIVEP-Gripe dataset (Sistema de Informação de Vigilância Epidemiológica da Gripe), which is provided by the *SUS* and made publicly available on the Datasus portal (Datasus, n.d.). The dataset exists since 2009 and includes all SARS (severe acute respiratory syndrome) cases in public and private hospitals in Brazil. According to the World Health Organization (WHO), SARS refers to "all viral respiratory disease caused by a SARS-associated coronavirus" (WHO, n.d.a). Since 2020, SARS-Cov-2, the virus causing COVID-19, has been added as a cause of illness. For this analysis, the data from 2020 and 2021 are merged to form a cross-sectional dataset. COVID-19 and non-COVID-19 cases together account for approximately 3 million observations. The sample is restricted to patients where COVID-19 is classified as the final cause of illness by a medical professional, which accounts to 1,907,617 observations.

To reduce measurement error, people older than 100 or with more than 20 days of delay are excluded from the sample, leading to the deletion of 1,535 and 158,621 observations, respectively. This is supported by findings suggesting that patients wait on average three to ten days between symptom onset and hospitalization (Faes et al., 2020). Approximately 91.7% of patients access a hospital within 20 days, and 72.6% within 10 days after symptom onset.

It is worth noting that the dataset only includes COVID-19 cases and deaths within hospitals, making sample selection likely. Hence, underreporting is a concern as not all COVID-19 patients make it to a hospital, especially in more remote areas. Moreover, missing data on specific controls is likely to create some bias, for example if patients prefer not to report their level of education due to social perception of their socioeconomic status. The number of observations considered is 1,747,461, but missing values on some variables imply that only 574,237 observations are used in the analysis.

3.1. Main Variables of Interest

The variables relevant for estimating whether distance to hospital is associated with an increase in mortality risk through delay in care are the outcome of the case, delay in care, and distance to hospital in km. As an IV approach has two stages, there are two outcome variables. The first stage predicts delay, and the second stage predicts mortality of the patient (McClellan et al., 1994). To understand regional heterogeneity in Brazil, patients are differentiated based on whether they are from the Northern or Central-Southern regions.

Outcome of the case is a dummy variable describing whether the patient dies from COVID-19. This is determined by a medical professional based on testing or other clinical methods such as imagery. Delay in care in days is calculated as the difference between onset of symptoms and admission to the healthcare facility. Distance is defined as the km distance the patient travels to the hospital to receive care, calculated on a municipal level. The region where the patient lives is divided into two regions; the Northern regions, which consist of the macroregions North and Northeast and the Central-South, which includes the macroregions the Central-West, Southeast and South (for a detailed overview of the macroregions and states in Brazil, see Table A1 in the Appendix).

For patients that travel to another municipality to access hospital services, approximately 27% of patients, distance is approximated as the distance between the municipality centroid of residence and municipality centroid of the hospital (Bacqui et al., 2021).

Data from the IBGE on Brazil's municipalities can be retrieved from the institute's website in shapefile format. These data are imported to QGIS, a geospatial information system (GIS), to calculate the centroids of each municipality. For patients for whom the hospital's municipality are missing, the reporting entity and its municipality is assumed to indicate the relevant treatment location, if available. Accurate locational data on Brazil's over 5000 hospitals is not available.

The coordinates of the municipality centroids and their corresponding six-figure municipality code are exported to Stata and merged with the SIVEP-Gripe dataset based on the code of the municipality of residence and treatment of each patient. All municipality codes in the federal district (DF) in the SIVEP-Gripe dataset are recoded into one municipality, as the IBGE classifies the DF as one municipality, thus reducing missing values in distance. Using the community-contributed command *geodist* (Picard, 2010), km distance between the municipality of residence and treatment – more specifically their centroids – can be calculated for each patient (Bacqui et al., 2021; Noronha et al., 2020). The command uses Vincenty's (1975) formula, which calculates distance with reference to an ellipsoid.

While calculating distance this way leads to an approximation instead of a precise estimate, it is nonetheless useful as there is expected to be a significant amount of variation in how far people travel to access a hospital (Bacqui et al., 2021). Although data of Brazil's road network is available, the size of Brazil implies that calculating a distance matrix on the network using the QNEAT3 plug-in on QGIS is not feasible due to the computing power required. Moreover, there is no Stata command that allows distance on a network to be calculated between two specified centroids, only between specified nodes on a network, with similar computing power issues. Hence, for patients travelling to a different municipality to access treatment for COVID-19, distance will be defined as the linear km distance between the centroid of the patient's municipality of residence to the centroid of the hospital of treatment.

The remaining 73% of patients visit a healthcare facility in their own municipality. To account for large variations in municipality size, particularly between the Northern and Central-Southern regions (see, for example, Figure 3), distance to hospital is approximated by calculating the minimum distance from the centroid of their municipality to the border of their municipality. This is done on QGIS using the NNJoin plugin by joining the lines from the municipality polygon border with the centroids of the municipality. The plugin calculates the distance to the "nearest neighbor" from the centroid to the polygon border, that is, the shortest distance to the municipal border. While this approximation is far from perfect, it is likely to capture a part of the variation in size between municipalities.

These calculations produce 37,806 unique values of distance. The municipality with the greatest number of patients attributed to same distance is in São Paulo and accounts for 141,369 patients, out of 1,747,461 patients. However, approximately 1.25 million patients are grouped by distance with fewer than 10,000 other patients. This suggests that the calculations poorly account for heterogeneity within municipalities and are therefore likely to produce imprecise estimators at small units of measurement. However, they are useful when comparing large differences in distance, for example when comparing between municipalities. To reduce the influence of outliers due to large variation in the km that patients travel, distance will undergo logarithmic transformation (Benoit, 2011).

3.2.Control Variables

Relevant and available control variables are age, gender, race, whether the person lives in an urban or rural area, education, whether the patient has comorbidities, and severity of the COVID-19 infection.

Age & Gender

Age, ranging from 1 to 100, will be divided into five age groups: 1 to 59, 60 to 69, 70 to 79 and 80 to 89 and 90 to 100. COVID-19 mortality risk increases substantially with age, with only a fraction of deaths being observed in people under 60 (Demombynes, 2020; Islam et al., 2021). Moreover, gender will be coded as a dummy variable for male, given reports of higher COVID-19 mortality amongst males (Islam et al., 2021). These variables are relevant regarding delay and distance-attributed delay, as males have lower rates of healthcare utilization (Schünemann et al., 2017) and consequently longer delays. Age might be attributed to less sensitivity to delay given poorer health and greater healthcare utilization at older ages.

Ethnicity

Another relevant control variable is self-reported race (Ranzani et al., 2020). This is relevant as studies have demonstrated wide-ranging mortality disparities between racial groups in Brazil (Bacqui et al., 2020; Sousa et al., 2020). A dummy is included for each group: White, *Pardo* (mixed race), Black, East Asian, and Indigenous. It is likely that the population distribution partly affects the mortality risk related to distance, as rural populations are more likely to be Black or *Pardo*, who generally have a higher mortality risk irrespective of distance (Bacqui et al, 2020; Benevenuto et al., 2019).

Education

Education is included as a dummy for higher education, which indicates whether the patient has at least a high school diploma (approximately 48.8 % of the sample). As education is generally lower in more remote areas, it is likely that some of the increased mortality associated with distance is due to education (Benevenuto et al., 2019). This is because a patient's level of education might affect where they live. Simultaneously, lower education is often associated with poorer health, leading to a heightened mortality risk (Conti et al., 2010). Moreover, low levels of education might increase Phase I delay of decision making if awareness of the risks is low (Thaddeus & Maine, 1994).

Urban/Rural

Whether the patient lives in an urban or rural area within the municipality is included in the SIVEP-Gripe dataset. While a distinction is made between rural and peri-urban areas, these will be merged as accessibility is likely to present similar challenges to these areas. Henceforth, these areas will be referred to as non-urban or rural areas. It can be expected that rural populations suffer from higher accessibility issues due to greater distance to a healthcare facility, while also being in poorer health and thus at higher risk of dying from COVID-19 (Amann et al., 2016; Benevenuto et al., 2019).

Comorbidities

A control variable capturing whether the patient has pre-existing comorbidities that increase the risk of COVID-19 is included in the model. The medical professional registers whether patients have a condition which constitutes a risk factor for COVID-19. People with comorbidities are at higher risk of severe illness and death from COVID-19 (Zhou et al. 2020; Sanyaolu et al. 2020). Comorbidities are likely associated with distance to a healthcare facility in one way or another; for example, populations in Central-Southern Brazil have higher rates of comorbidities but are on average closer to a hospital (see Figure 3) (Rocha et al., 2021).

Severe Respiratory COVID-19 Symptoms

Patients with a more severe COVID-19 infection are likely to seek care earlier as they might be more concerned. Severity of disease is measured as a dummy of whether the patient presents one of the following symptoms, recorded by a medical professional: painful breathing, difficulty breathing and low oxygen content. These symptoms are relevant given that most COVID-19 patients die from respiratory complications (Zhang et al., 2020). Disease severity is included as it affects both the time of delay between symptoms and treatment, and the mortality risk of a patient. Moreover, including COVID-19 severity might account for a part of unobserved variables, such as income.

3.3.Municipality & Hospital Fixed Effects

Fixed Effects (FE) are a central part of the analysis, as they can account for factors that are the same for all individuals in a hospital and municipality. This can help address bias, as patients' residence region and treatment hospital have proven crucial to their survival probability (Rocha et al., 2021; Bacqui et al., 2020; Bacqui et al., 2021). Given that average Phase III delay and hospital quality would be the same for all patients in a hospital, including hospital FE can isolate Phase I and II delays. FE require information on a patient's residence and the hospital, which is provided through the municipality code and hospital code.

4. Methodology

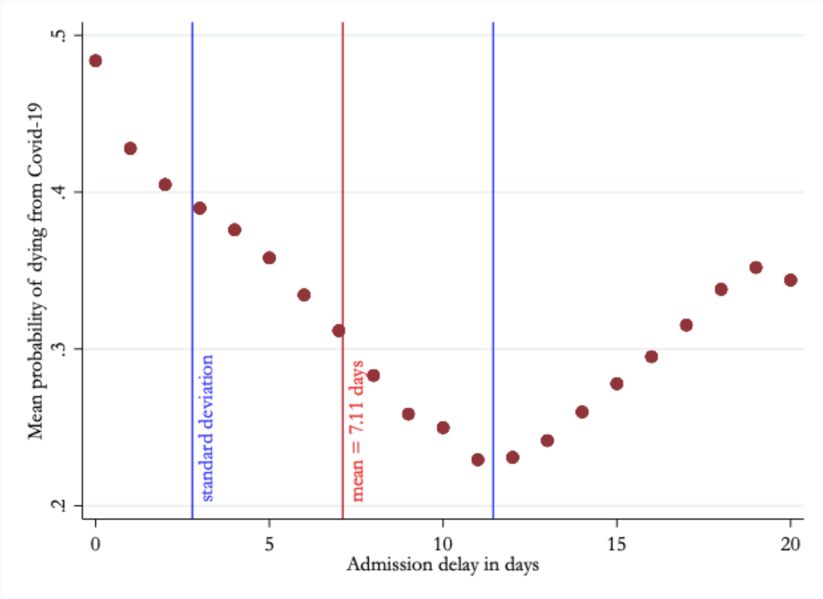
To investigate how distance affects in-hospital COVID-19 mortality through delay in care, we will consider individual mortality risk. In particular, we will explore whether the impact of distance on delay, and hence mortality, differs between Brazil's regions and between 2020 and 2021. First, we will focus on providing context using descriptive statistics as presented by Bacqui et al. (2020) by describing patient characteristics based on survival and mortality.

An OLS model using delay to predict the mortality risk of a patient is specified in Equation 1, where mortality is the dependent variable and delay is the variable of interest. Controls are included to reduce Omitted Variable Bias (OVB), as not accounting for factors that affect both delay and the outcome, mortality, can induce bias by assigning the effect of the omitted variable onto delay. For example, if a patient has completed higher education, they are less likely to delay seeking out a healthcare facility than people with lower education (Thaddeus & Maine, 1994). Since higher educated people are generally also in better health (WHO, n.d.b), controlling for educational attainment means that you compare patients within an educational category, not between educational categories, which in this case would reduce upward bias.

It is worth noting that delay follows a non-linear relationship with mortality, as average mortality decreases monotonically up to 11 days of delay and increases monotonically up to 19 days of delay thereafter (see Figure 2). To model mortality and delay accurately, a squared term of delay must be included in the OLS model. With delay being skewed to the right, one expects a negative relationship between mortality and delay, as most patients wait fewer than 10 days.

Figure 1

Mean Mortality Rate by Day of Delay



Note. This figure shows the non-linear relationship between average mortality from COVID-19 by each day of delay. Delay is measured as the difference in days between first symptoms and hospital admission. Data can be retrieved from the Datasus portal.

The OLS model of the relationship between mortality and delay with a squared term of delay is outlined below in Equation 1. It predicts the mortality probability of a patient due to delay in care, controlling for a range of factors and including hospital and municipality Fixed Effects (FE).

$$Mortality_{i,h,m} = \varphi \ + \rho_1 * delay_{\ i} + \rho_2 * (delay_{\ i})^2 + \sum_{c=k}^{K} * \chi_c X'_{c,i} + FE_h + FE_m + \varepsilon_i \qquad (1)$$

Mortality is predicted by individual patient, indicated by the subscript i. Delay, the explanatory variable, is patient-specific and is included once with and once without a squared term. K levels of controls are indicated by X', and include: distance to healthcare facility (Thaddeus & Maine, 1994), age (Demombynes, 2020), gender (Islam et al., 2021), education (Conti et al., 2010), ethnicity (Bacqui et al., 2020), comorbidities (Zhou et al. 2020; Sanyaolu et al. 2020) and severity of disease (Zhang et al, 2020). These are relevant control variables as they are likely to cause OVB, that is, they are likely to affect both mortality and delay.

Unobserved hospital-specific and municipality-specific factors are represented by subscripts h and m respectively. They are included as Fixed Effects (FE) to control for factors that are the same for everyone in a municipality of residence and in a hospital. Since about a quarter of patients receive treatment at a healthcare facility in a municipality that is not their own, municipality FE can control for things such as quality of schooling, which is associated with delay in treatment for COVID-19 and mortality (Conti et al., 2010). The large discrepancy of hospital quality in Brazil also implies significant variation in mortality risk (Rocha et al., 2021). By essentially including a control variable for each hospital, we can isolate hospital-specific factors that are hospital-invariant; everyone in the hospital is subject to the same staff and equipment, which affects patients' mortality probability. This model is implemented on Stata using the user-contributed command *reghdfe* (Correia, 2017).

While this OLS model might be able to predict mortality risk, it cannot inform how the delay caused by distance increases patients' mortality risk. To do this, an Instrumental Variable (IV) approach estimated using Two-Stage-Least-Squares (2SLS) is more appropriate. Instrumental variables are a widely used econometric model to estimate a causal effect in two stages: first, the variable of interest (X) is regressed on the instrument (Z). In the second stage, the outcome variable (Y), is regressed on the instrument Z (Angrist & Krueger, 2001). The causal estimator of the variable of interest is calculated by dividing the instrument estimator from the second stage by the instrument estimator from the first stage. Alternatively, the estimator of the variable of interest can be obtained by regressing the outcome variable Y on the predicted values of X, shown in Equation 3.

In the context of this paper, the instrument Z is the log of distance to the healthcare facility, X is the variable of interest, namely delay, and Y is the outcome variable mortality. An IV approach is better suited than an OLS model to estimate the effect of delay from distance on mortality, as Figure 1 suggests that patients with more severe COVID-19 seek treatment earlier after first developing symptoms. However, it is not because they have a shorter delay, but rather because they have a more severe case of COVID-19, that they have a higher mortality risk. While the model in Equation 1 accounts for whether the patient suffers from severe respiratory symptoms, it does not account for the degree of severity. This will be done using distance as an instrument in the IV model, estimated in two stages, as it can be expected that severity of COVID-19 does not influence how far the patient lives from the healthcare facility. However, there might be other factors associated with distance that make a severe COVID-19 infection more likely. These are captured by control variables.

The first stage estimator can be defined as the prediction of delay based on the log of distance and controlling for various patient characteristics, hospital FE, and municipality FE.

$$\widehat{delay}_{i,h,m} = \alpha \quad + \beta_1 \log(distance) + \sum_{c=k}^{K} * \gamma_c X'_{c,i} + FE_h + FE_m + \varepsilon_i \tag{2}$$

Delay is predicted by individual patient, indicated by the subscript i. The same control variables and FE are included as in Equation 1, with the exception of severity of disease. This is because including it can cause collider bias, as both delay and the log of distance affect severity of disease, causing a spurious relationship between distance and delay (Deuchert & Huber, 2017). After estimation, delay is predicted for each individual patient based on the log of distance, the control variables, and the FE.

The second stage uses the prediction of delay from Equation 2 to estimate the impact of delay due to distance on the mortality of a patient.

$$Mortality_{i,h,m} = \theta + \delta_1 \widehat{delay}_{i,h,m} + \sum_{c=k}^{K} * \lambda_c X'_{c,i} + FE_h + FE_m + \varepsilon_i \quad (3)$$

The second stage follows the same subscripts as in Equations 1 and 2. However, the controls and FE are re-estimated. Since the same control variables need to be included in the second stage, severity of disease is again excluded. Hypothesis 2 will be tested by adding an additional interaction term in the two stages for the Northern regions to see whether the effect of distance and delay differs by region. This model will be implemented on Stata with the user-contributed command *ivreghdfe* (Correia, 2018).

However, to obtain a causal effect using the IV approach, the instrument needs to be valid, that is, it needs to fulfill the assumptions of Relevance, Independence, Exclusion, and Monotonicity (Angrist & Krueger, 2001; Angrist et al., 1996).

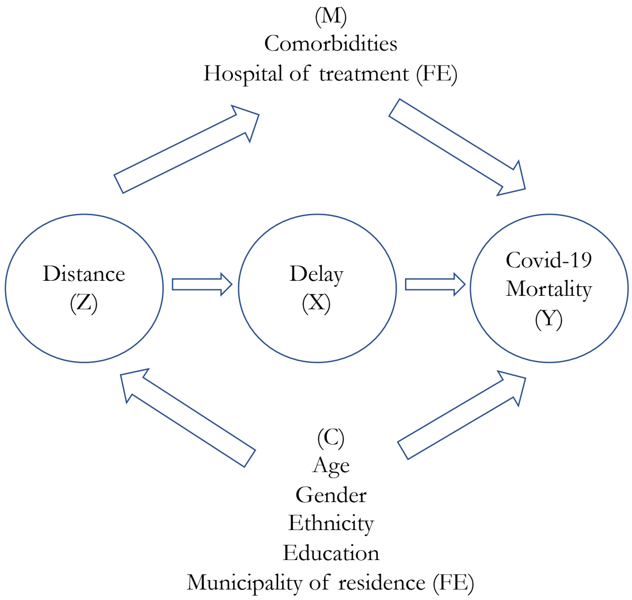
First, an instrument is relevant if it is sufficiently related to the variable of interest. That is, the log of distance needs to have sufficient explanatory power for delay in the first stage to produce consistent results (Angrist et al., 1996). This can be checked based on the rule of thumb of an F-Statistic above 10 (Stock & Yogo, 2005). With an F-Statistic of 39.47 the relevance of the instrument affirms conclusions from the literature discussed in the theoretical framework.

Next, an instrument should satisfy Independence, that is, it should be random (Angrist et al., 1996). This assumption is often referred to as instrument exogeneity, as it claims that the instrument should not be correlated with the error term. Therefore, the log of distance should not be associated with distance in any way other than delay. The biggest obstacle in the OLS model is that patients can influence delay based on the severity of their case of COVID-19, leading to OVB. An advantage of the IV model is that patients do not choose how far to live from a hospital based on the severity of their COVID-19 infection. However, the fulfilment of the Independence assumption is not guaranteed, given potential factors that are associated with distance that can affect a patient's severity of disease.

Figure 2 shows that there are two ways in which distance might be associated with the outcome, mortality, through ways other than delay.

Figure 2

IV model estimating the impact of delay from distance on in-hospital COVID-19 mortality.



Note. This figure visualizes the implementation of the Instrumental Variable (IV) approach of the analysis. To estimate the impact of delay from distance, distance is used as an instrument (Z), delay is the variable of interest (X) and the outcome variable is mortality (whether the patient dies) (Y). The instrument is not exogenous: it is correlated with the outcome through mechanisms (M) and confounders (C). Using control variables can help reduce bias, but they alone cannot rule out Omitted Variable Bias (OVB).

For one, there might be mechanisms (denoted by M), referring to variables affected by distance that also affect the outcome (Deuchert & Huber, 2017). For example, distance to a healthcare facility might mean that the patient is in poorer health, for example if they cannot regularly get check-ups. Moreover, distance to hospital is likely to determine which hospital you attend, especially in remote areas where choice of hospital might be limited. Unlike in a linear regression, it is useful to control for M, as it ensures that the second stage estimator of distance on mortality only captures the part of distance that operates through delay, and not through other factors.

Similarly, distance can be associated with the outcome through confounders, denoted by C. Confounders affect both Z and Y, and hence need to be controlled for to reduce bias in the estimator of the log of distance. When confounder C affects distance and mortality in the same direction, leaving them out of the model will cause upward bias in the estimator of the log of distance (Deuchert & Huber, 2017). The overestimation of the impact of distance on mortality, and hence predicted delay on mortality, is because people living further away from healthcare centers are likely to be in worse health to begin with and therefore have a higher COVID-19 mortality risk, for reasons such as lower levels of education (Conti et al., 2010), age (Braga et al., 2019) and being *Pardo* or Black (Benevenuto et al., 2019), as these populations have been shown to have a higher COVID-19 mortality risk (Bacqui et al., 2020). Lastly, a patient's municipality largely defines the distance a patient must travel to access healthcare, though some patients might choose a facility that is further to seek better care. Municipality is also likely to affect a patient's mortality risk considering various social determinants of health (WHO, n.d.b). Controlling for these variables can help reduce endogeneity, as the characteristics are in the model instead of in the error term. However, whether the Independence assumption is satisfied cannot be tested when there is only one instrumental variable (Baum et al., 2002).

Another related but nonetheless distinct assumption is the Exclusion restriction (Angrist et al., 1996). This assumes that the instrument Z has no direct effect on Y. This is unlikely to be a risk given that it is not distance itself, but rather factors associated with distance, that can influence the mortality risk of the patient.

Lastly, the IV model should satisfy Monotonicity (Angrist & Krueger, 2001); there should be no 'defiers', that is, patients that delay care less if they live further away, and more if they live closer to the healthcare facility. This is difficult to affirm; patients that live closer to the healthcare facility might delay healthcare for longer because they know they can easily access the healthcare facility once it's necessary, though the case for distance as a disincentive is also sensible (Thaddeus & Maine, 1994).

Angrist & Krueger (2001) conclude that only when these assumptions hold, a causal effect of delay from distance can be affirmed. While controlling for various factors can help reduce bias in estimating delay and mortality in the first and second stage respectively, whether the Independence assumption is satisfied cannot be ensured. Moreover, Monotonicity is likely but not guaranteed to hold, given that distance might have either a disincentivizing or procrastinating effect.

Irrespective of whether a causal effect of delay from distance can be assured, an IV approach to this issue is nonetheless useful. Paradoxically, this is due to a commonly cited disadvantage of the IV approach, namely that it estimates a Local Average Treatment Effect (LATE) (Angrist & Krueger, 2009). When investigating the causal effect of X on Y, using instrument Z means that one only estimates the causal effect of X explained by Z. Hence, one cannot make general conclusions about the causal effect of X, but only the part of X explained by Z. In this analysis, it is exactly the LATE that is of interest, as it is not the delay in care generally, but the delay in care caused by distance that is considered. The LATE allows for an estimation of how an increased delay due to distance increases COVID-19 mortality risk in Brazil. Even if causality cannot be assured, one can 'narrow down' the analysis to investigate only how distance-predicted delay is associated with mortality.

While the binary outcome variable makes fit for a probit model, including a probit model eliminates the possibility of using Fixed Effects due to non-linearity of the estimators. Fixed Effects are possible to implement in a logit model, which however does not offer the possibility of implementing an IV approach.

5. Results & Discussion

5.1.Descriptive Statistics

The characteristic means and frequencies presented in Table 1 allow for a comparison between COVID-19 patients (Bacqui et al., 2020). Approximately two-thirds of patients that seek treatment for COVID-19 do not leave the healthcare facility alive. How this differs by characteristic is outlined in this section.

Non-surviving patients live on average one km further away from the hospital than survivors. Based on Table A2 in the Appendix, the Northern population travel further to reach a healthcare facility compared to their Central-Southern counterparts.

Moreover, non-survivors tend to delay treatment by 6.28 days, compared to 7.52 days for survivors. This confirms that people with more severe cases of COVID-19 are more likely to seek treatment earlier, as shown in Figure 3. This supports the IV approach, as part of the bias stemming from COVID-19 severity can be circumvented; distance might indirectly cause severity of disease, but this bias is likely smaller than in an OLS model. There is no statistically significant difference in delay between Brazil's regions (see Table A2 in the Appendix).

In line with previous findings, the Northern regions suffered a disproportionate mortality burden compared to the Central-Southern regions (Rocha et al., 2021; Bacqui et al., 2020; Bacqui et al., 2021).

Non-survivors are about 12 years older, in line with existing literature (Islam et al., 2020). Moreover, while the in-hospital mortality rate between men and women does not differ, there are 26% more males than females being treated for COVID-19. The greater susceptibility of males has been widely reported in the literature. This result suggests that while males might more often develop a severe case of COVID-19 (Islam et al., 2020), there is no gender differential in mortality in severe cases of COVID-19.

There is a less than one percentage point difference between urban and rural patients in terms of their in-hospital mortality risk. Contrary to expectations based on previous literature documenting rural populations to be of lower socioeconomic status (Benevenuto et al., 2019), there is an insignificant difference in mortality between rural and urban patients (see also regression results in Table 2).

As reported by Bacqui et al. (2020), Black, *Pardo* and indigenous Brazilians bear a disproportionate mortality burden as seen by the higher in-hospital mortality rate compared to White Brazilians and Brazilians with East Asian heritage. A significant proportion of this is likely to be attributable to the fact that non-White ethnic groups make up a greater share of the North Brazilian population, where poorer healthcare equipment causes higher mortality rates than in the Central-South of Brazil (Bacqui et al., 2020; Rocha et al., 2021).

However, the most significant risk factor based on Table 1 is educational attainment. Of patients without a high school diploma (approximately 45.6% of observations for whom information on education is available), 41.8% died from COVID-19 in the healthcare facility, compared to 26.6% of patients with at least a high school diploma. This confirms the importance of educational attainment for mortality, before the pandemic (Conti et al., 2010; Thaddeus & Maine, 1994; Ribeiro et al., 2021) and during it (Ribeiro et al., 2021a; Feldman et al., 2021; Dalsiana et al., 2022). This might be due to poorer prior health in lower educated groups (Rocha et al., 2021; Conti et al., 2010), but could also be due to behavioral differences, for example, due to different interpretations of illness (Thaddeus & Maine, 1994).

Confirming previous studies, 39.2% of patients with pre-existing conditions died from COVID-19 compared to 22% without comorbidities (Zhou et al. 2020; Sanyaolu et al. 2020). While the relative discrepancy between patients with and without comorbidities is the same as between lower and higher educated patients, being a lower educated patient is nonetheless associated with the highest mortality risk out of all characteristics.

Lastly, severe respiratory symptoms are associated with a higher mortality risk of about ten percentage points. Thirty-three percent of patients with severe respiratory symptoms die from COVID-19.

Table 1

Baseline Characteristics: Differences in Characteristics between Survivors and Non-Survivors

	Hospital admission survival of Patient				
	Survivors		Died		Mean difference (died minus survived)
Mean					
Distance to Hospital	23.33		24.43		1.1
SD	105.31		89.89		-15.42
Admission delay in days	7.52		6.28		-1.24
SD	4.22		4.43		0.21
Age	53.91		65.94		12.03
SD	17.07		15.71		-1.36
Factor variable frequency	Count	Frequency	Count	Frequency	
North	71011	63.58%	40668	36.42%	-27.16%
Northeast	167273	63.64%	95582	36.36%	-27.28%
Central-west	116008	70.03%	49644	29.97%	-40.06%
Southeast	564786	67.91%	266908	32.09%	-35.82%
South	133116	71.48%	53119	28.52%	-42.96%
Female	498653	67.83%	236472	32.17%	-35.66%

Male	625984	67.34%	303657	32.66%	-34.68%
Non-Urban	172953	66.83%	85830	33.17%	-33.66%
Urban	951684	67.69%	454299	32.31%	-35.38%
White	482152	67.37%	233480	32.63%	-34.74%
Black	43924	61.82%	27129	38.18%	-23.64%
East Asian	11138	68.77%	5058	31.23%	-37.54%
Parda	364998	65.77%	189941	34.23%	-31.54%
Indigenous	2122	65.94%	1096	34.06%	-31.88%
Lower Educated	179382	58.16%	129033	41.84%	-16.32%
Higher Educated	213721	73.35%	77669	26.65%	-46.70%
No Pre-existing Conditions	509662	77.99%	143850	22.01%	-55.98%
Pre-existing Conditions	614975	60.81%	396279	39.19%	-21.62%
Severe Respiratory Symptoms	959550	66.20%	489881	33.80%	-32.40%
No Severe Symptoms	165087	76.67%	50248	23.33%	-53.34%

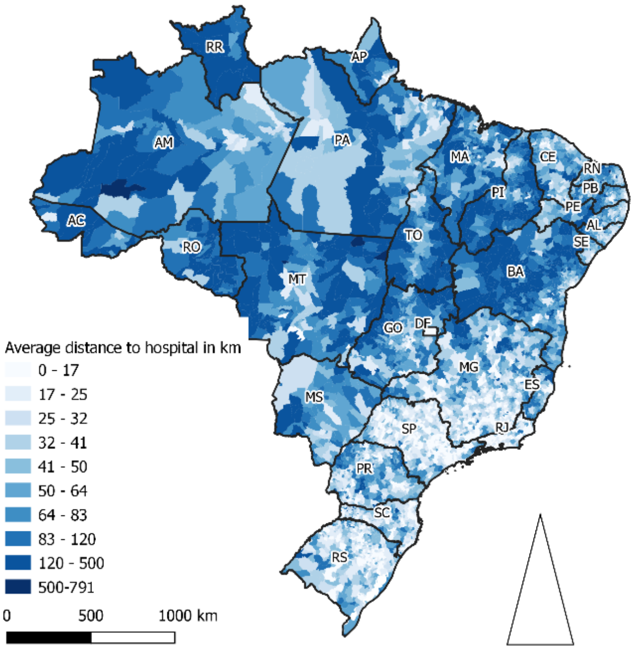
Note: This table includes the frequencies of baseline characteristics of COVID-19 survivors and non-survivors in Brazil's hospitals. The sample excludes people with a delay between first symptoms and admission of over 20 days and excludes people aged over 100.

While Table 1 considers the entirety of Brazil, more complex patterns emerge when analyzing regional variation in distance, delay, and mortality, visualized in Figures 3 to 5. Despite the pandemic originating in the Central-Southern parts of Brazil, the Northern parts suffered higher in-hospital mortality rates in the first year of the pandemic (see Figure A1 in the Appendix) (Rocha et al., 2021). This has been largely attributed to a more poorly equipped healthcare system in the Northern compared to Central-Southern regions of Brazil (Rocha et al., 2021; Bacqui et al., 2020). Whether different patterns of population dispersal can help explain this discrepancy is discussed in the remainder of this section.

Figure 3 shows the average km distance to the hospital by municipality. The darker shades of blue in the North clearly indicate that patients from the Northern regions, on average, travelled further to receive treatment for COVID-19. The exception is Mato Grosso (MT) and Goiás (GO) (Center-west) and Espírito Santo (ES) (Southeast), as they show a visible proportion of their municipalities with large distances to a healthcare facility.

Figure 3

Average km Distance to Hospital by Municipality

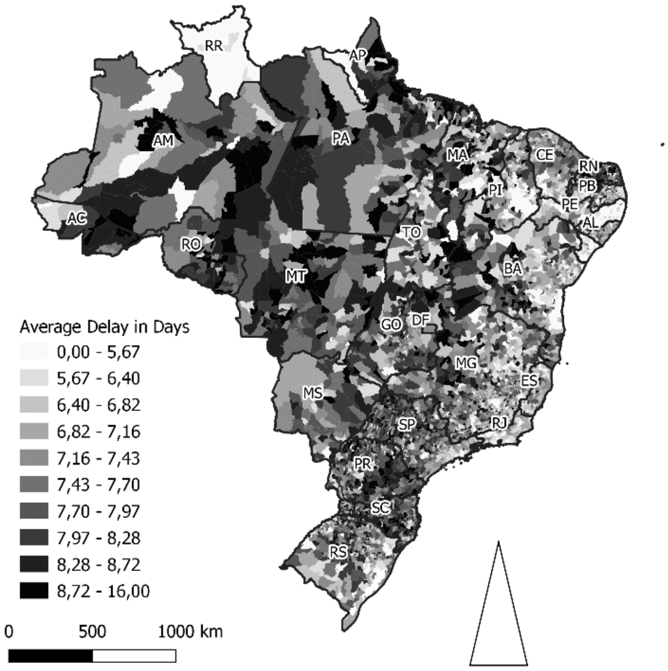


Note: This figure shows the average distance that a patient has to travel to access the hospital in which they are seeking treatment for COVID-19. Darker shades of blue represent further distances to the hospital. Distance is calculated as the km distance between the centroid of the patient's municipality of residence and treatment. For patients who visit a hospital in their municipality, distance is calculated as the minimum distance from the municipality centroid to the municipal border. The Northern regions consist of the North (Acre (AC), Amazonas (AM), Rondonia (RO), Roraima (RR), Pará, Amapá (AP), Tocatins (TO)) and Northeast (Maranhao (MA), Piauí (PI), Ceará (CE), Rio Grande do Norte (RN), Paraíba (PB), Pernambuco (PE), Alagoas (AL), Segipe (SE), Bahia (BA)). The Central-South includes the Central-West (Mato Grosso (MT), Goiás (GO), Distrito Federal (DF), Mato Grosso do Sul (MS)), Southeast (Minas Gerais (MG), Espírito Santo (ES), Rio de Janeiro (RJ), Sao Paulo (SP)) and South (Paraná (PR), Santa Catarina (SC), Rio Grande do Sul (RS)).

However, while patients from the Northern regions of Brazil are generally further away from the healthcare facility in which they are treated, Figure 4, showing the average delay between symptom onset and hospital admission by municipality, does not suggest regional inequality in delay in care. The darker shades of grey, indicating greater delay, seem somewhat uniformly distributed within the country. Therefore, the relationship between distance to a healthcare facility and delay in care is not immediately clear.

Figure 4

Average Delay between Symptom Onset and Hospital Admission by Municipality

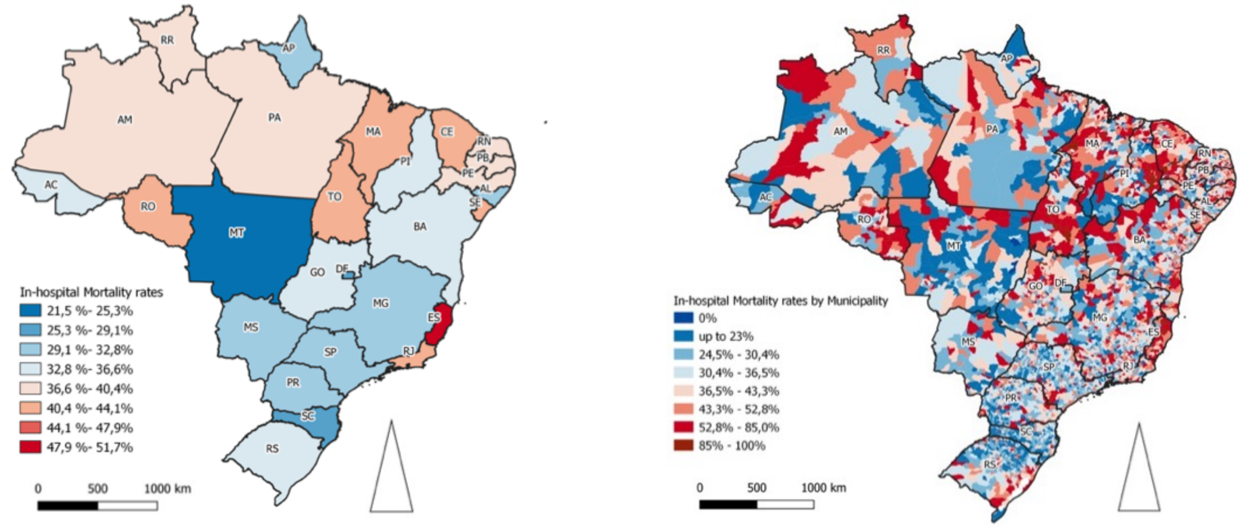


Note: Delay is measured as days between first symptoms and admission. The average delay is taken by municipality of residence of patients. Darker shades of grey represent longer average days of delay. The Northern regions consist of the North (Acre (AC), Amazonas (AM), Rondonia (RO), Roraima (RR), Pará, Amapá (AP), Tocatins (TO)) and Northeast (Maranhao (MA), Piauí (PI), Ceará (CE), Rio Grande do Norte (RN), Paraíba (PB), Pernambuco (PE), Alagoas (AL), Segipe (SE), Bahia (BA)). The Central-South includes the Central-West (Mato Grosso (MT), Goiás (GO), Distrito Federal (DF), Mato Grosso do Sul (MS)), Southeast (Minas Gerais (MG), Espírito Santo (ES), Rio de Janeiro (RJ), Sao Paulo (SP)) and South (Paraná (PR), Santa Catarina (SC), Rio Grande do Sul (RS)).

However, Figure 5a clearly shows that in-hospital mortality risk is higher in the Northern states. This is supported by a wide range of literature on the COVID-19 pandemic in Brazil (Bacqui et al., 2020; Bacqui et al., 2021; Pereira et al., 2020; Rocha et al., 2021). The exceptions are the Southeastern states of Espírito Santo (ES) and Rio de Janeiro (RJ), the former having the highest in-hospital mortality rates in the country. This is also consistent with previous literature (Rocha et al., 2021). To a lesser extent, the pattern is visible on a municipal level too, as most of the municipalities with high mortality rates are concentrated in the Northern regions (see Figure 5b).

Figure 5

In-hospital COVID-19 Mortality Rates by State and Municipality



(a) State-level in-hospital mortality rates

(b) Municipal-level in-hospital mortality rates

Note. This figure shows the in-hospital COVID-19 mortality rates by state and municipality in Brazil, in percentages. Mortality rates are calculated based on state and municipality of residence. Shades of blue represent lower than average mortality, whereas red shades represent higher than average rates of mortality. The Northern regions consist of the North (Acre (AC), Amazonas (AM), Rondonia (RO), Roraima (RR), Pará, Amapá (AP), Tocatins (TO)) and Northeast (Maranhao (MA), Piauí (PI), Ceará (CE), Rio Grande do Norte (RN), Paraíba (PB), Pernambuco (PE), Alagoas (AL), Segipe (SE), Bahia (BA)). The Central-South includes the Central-West (Mato Grosso (MT), Goiás (GO), Distrito Federal (DF), Mato Grosso do Sul (MS)), Southeast (Minas Gerais (MG), Espírito Santo (ES), Rio de Janeiro (RJ), Sao Paulo (SP)) and South (Paraná (PR), Santa Catarina (SC), Rio Grande do Sul (RS)).

While mapping the relevant variables can provide some insight on regional variation in Brazil, these should be interpreted with caution. For one, these averages do not account for local sociodemographic characteristics such as age. Moreover, the map does not account for the fact that some municipalities have a smaller population and/or larger area size, which can distort which patterns get visual emphasis. Therefore, the hypotheses from Section 2 will be empirically tested in the remainder of this paper.

5.2. Hypothesis 1: Distance-attributed Delay and Mortality

Hypothesis 1, that delay due to distance leads to an increased mortality risk, is evaluated using the models proposed in Section 4. The relationship between delay and mortality is laid out in two OLS models (Models 1 & 2) and one IV model (Model 3).

The results from Models 1 and 2 in Table 2 show that regressing mortality on delay can cause OVB. Despite delay being a factor known to increase mortality risk (Thaddeus & Maine, 1994; Hanna et al., 2020; Kutikov et al., 2020; Demissie, 2002; Le Terrier et al., 2022) the results suggest a negative relationship between mortality and delay, even after controlling for other factors. While both models show a negative estimated coefficient for delay, Model 2, which accounts for the non-linearity in delay using the squared term of delay, shows a steeper slope of delay until the 20th day of delay, which is when the sum of the coefficient of delay and squared delay exceed the coefficient for delay of Model 1.

Given the expectation of a positive estimator of delay based on the literature, the estimator of delay is likely to face downward bias. One possible source of OVB is income; income has exhibited negative association with health status and hence positive association with COVID-19 mortality risk in Brazil (Rocha et al., 2021). In turn, lower income might act as a disincentive for seeking care, for example if people cannot afford the journey. Another factor might be disease severity; while variables accounting for whether the patient has severe symptoms is found to increase mortality risk in Models 1 and 2, it does not capture the heterogeneity in symptom severity, which is likely to be a dominant factor for the timing with which patients seek out treatment for COVID-19.

Given that Models 1 and 2 are likely to suffer from significant OVB, the question remains whether this can be addressed, in part, by the IV approach in Model 3. The IV model is estimated in two stages: Stage 1 predicts delay using the instrument distance, while Stage 2 estimates individual mortality risk using predicted delay. Stage 1, column 3 of Table 2, confirms what has been reported by the literature in low- and middle-income countries; distance acts as a disincentive and barrier to seeking healthcare for various conditions (Thaddeus & Maine, 1994; Mgawadere, 2017; Pajuelo et al., 2018; Målqvist et al., 2010). However, the impact is small; each 1% increase in km distance is associated with an increase of approximately 0.0005 days of delay. This is calculated by dividing the coefficient for log of distance by 100 (Benoit, 2011). While distance is a relevant predictor of delay given the first stage F-Statistic of 39.47, its effect on delay nonetheless remains small.

This can also in part explain the large estimate of the coefficient of delay in the second stage, column 4 of Table 1. According to the IV results of Model 3, each day of delay from distance is associated with an increase in a patient's mortality risk of over 32.7 percentage points. This might seem large, but it is important to keep in mind that distance only explains a very small part of why patients delay care. It is unsurprising then, that the delay predicted in the first stage ranges between 4.4 and 8.7, a much narrower range than in the observed data. Given reports of patients dying in hospital queues (da Silva, 2021; Coutinho, 2020; Globo, 2020), it is not unreasonable to think that delaying care for 24 hours because of distance, all else equal, can have serious implications for patients' survival probability.

As Model 3 is our preferred model given that it is likely to reduce the bias that causes the spurious relationship between mortality and delay, our discussion of the effects of covariates is limited to the second stage of the IV model (column 3 & 4 in Table 2).

There is no statistically significant difference between urban and rural patients, neither in delay, nor in mortality. For one, the impact of being rural on delay should already be captured by the distance to the healthcare facility (column 3, Table 2). Moreover, since in-hospital mortality rates and not population mortality rates are considered, population density is likely to be of lesser importance (column 4, Table 1).

Males are significantly more prone to delay care by about 0.2 days, which is unsurprising given lower levels of healthcare utilization than females (Schünemann et al., 2017). Interestingly, the second stage predicts males to have a lower mortality risk than females. This might be because males and females have similar in-hospital mortality rates (see Table 1) and including other control variables such as education might expose inequalities in healthcare access favoring males (Thaddeus & Maine, 1994).

Regarding ethnic differences in healthcare-seeking behaviour, all non-White Brazilians delay care for less time than White Brazilians, as seen by a negative coefficient in the first stage (column 3, Table 2). Except for East Asian Brazilians, all non-White Brazilians have a higher in-hospital mortality risk than their White counterparts, in line with Bacqui et al. (2020). Patients with an indigenous background are least likely to delay care, which might be due to more severe cases of COVID-19 (Silva et al., 2021). This is made evident by a mortality risk 21 percentage points higher than White Brazilians, the highest relative mortality risk of all ethnic groups.

As age has also been cited as a risk factor for COVID-19, it is unsurprising that, as age increases, both delay and mortality increase (Bacqui et al., 2020; Demombynes, 2020; Islam et al., 2021). This suggests that COVID-19 patients that have, or at least have the potential to get a more severe infection, take this into account when making decisions regarding whether to delay care.

Despite education being associated with greater awareness of health (Conti et al., 2010), delay in care is shorter for the lower educated group by 0.3 days at the 1% significance level. This suggests that the greater severity of COVID-19 in people with lower educational attainment drove the decision to shorten delay in treatment. This is supported by a mortality probability that is 12.6 percentage points greater in lower compared to higher educated patients. A similar interpretation can be applied to the people with comorbidities; they delay care by 0.34 days less but suffer a mortality risk that is 21.3 percentage points greater than patients without comorbidities.

Lastly, delay was higher by about 0.77 days in 2021 compared to 2020, while mortality was lower by about 19.1 percentage points. This might point to the arrival of vaccines in 2021, which reduce the chances of a severe COVID-19 infection and hence leads to greater delays.

Table 2

OLS and IV Regressions with Hospital and Municipality FE predicting in-Hospital Mortality Probability

	(1)	(2)	(7)	(8)
	Model 1	Model 2	Model 3: First Stage	Model 3: Second Stage
	Mortality	Mortality	Delay	Mortality
VARIABLES	OLS Regression	OLS Regression with squared estimators	2SLS Regression: First Stage	2SLS regression
Admission delay in days	-0.00865***	-0.0166***		0.327***
	(0.000135)	(0.000431)		(0.0578)
delay squared		0.000434***		
		(2.39e-05)		
logdistance	0.0165***	0.00641***	0.0496***	
	(0.000826)	(0.00236)	(0.00818)	
c.logdistance#c.delay		0.000489***		
		(8.87e-05)		
lungsymp	0.0933***	0.0951***		
	(0.00180)	(0.00180)		
Urban	-0.00575***	-0.00564***	-0.00493	-0.00324
	(0.00205)	(0.00205)	(0.0203)	(0.00712)
Male	0.0374***	0.0374***	0.192***	-0.0262**
	(0.00109)	(0.00109)	(0.0108)	(0.0118)
Black	0.00551**	0.00528**	-0.291***	0.103***
	(0.00243)	(0.00242)	(0.0240)	(0.0188)
East Asian	-0.0206***	-0.0207***	-0.163***	0.0335
	(0.00553)	(0.00553)	(0.0548)	(0.0214)
Parda	-0.0118***	-0.0119***	-0.110***	0.0249***
	(0.00149)	(0.00148)	(0.0147)	(0.00819)
Indigenous	0.0148	0.0141	-0.606***	0.217***
	(0.0119)	(0.0119)	(0.118)	(0.0540)
60-69	0.133***	0.132***	-0.102***	0.167***
	(0.00148)	(0.00148)	(0.0146)	(0.00783)
70-79	0.213***	0.211***	-0.667***	0.437***
	(0.00167)	(0.00167)	(0.0165)	(0.0391)
80-89	0.291***	0.289***	-1.285***	0.723***
	(0.00210)	(0.00211)	(0.0207)	(0.0747)
90-100	0.362***	0.359***	-1.738***	0.947***
	(0.00384)	(0.00385)	(0.0380)	(0.101)
Higher Educated	-0.0257***	-0.0249***	0.300***	-0.126***
	(0.00125)	(0.00125)	(0.0124)	(0.0179)
Comorbidities	0.0951***	0.0948***	-0.342***	0.213***
	(0.00121)	(0.00121)	(0.0120)	(0.0202)
2021	0.0682***	0.0691***	0.776***	-0.191***
	(0.00125)	(0.00125)	(0.0124)	(0.0451)
Constant	0.104***	0.135***	7.009***	-2.044[1]
	(0.00343)	(0.00413)	(0.0293)	

First Stage F-Statistic			39.470	
Observations	574,237	574,237	574,237	574,237
R-squared	0.297	0.298	0.143	-9.690
Hospital FE	YES	YES	YES	YES
Municipality FE	YES	YES	YES	YES

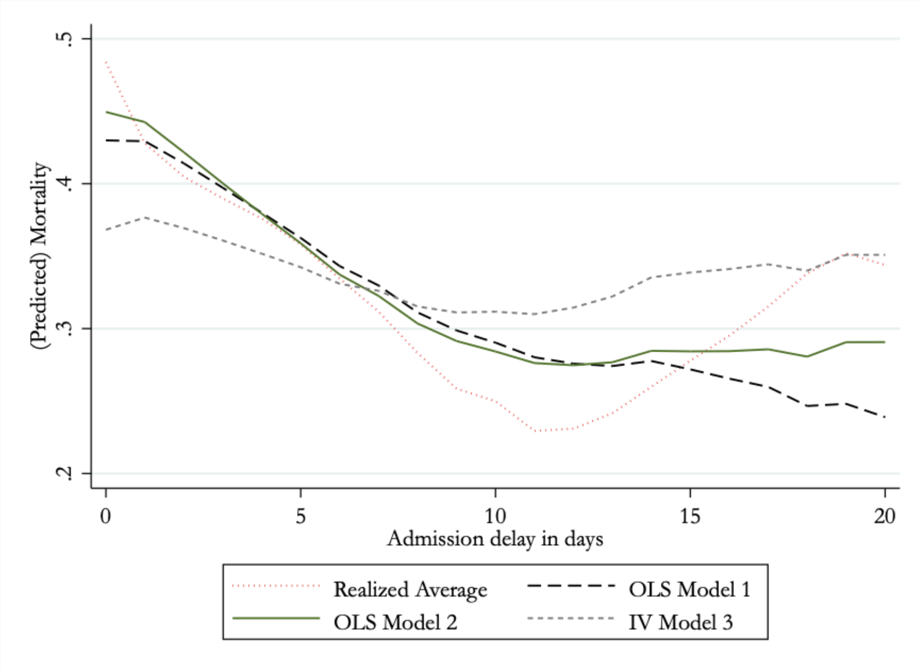
Note. This table shows the regression results from OLS and IV estimates of mortality on delay. Standard errors in parentheses. Significance levels indicated by *** p<0.01, ** p<0.05, *p<0.1. Constant in column 4 ([1]) is from own Two-Stage calculations, therefore there are no significance stars. The OLS model predicts mortality using delay and various controls, while the IV model instruments distance to predict mortality based on distance-attributed delay.

While Model 3 is the preferred model based on its ability to remove much if not most of the bias stemming from COVID-19 severity, some OVB is likely to remain. For example, in the first stage, distance to a healthcare facility might be an important determinant of income, if one considers it as distance to opportunities. Simultaneously, as previously discussed, income might in part affect people's decision to delay care. OVB might in part also cause a large estimator of delay in the second stage if there is downward bias in the first stage or upward bias in the second stage. As the Independence of the instrument distance cannot be ensured, causal inference of the estimators cannot be ensured. The control variables included in the models will have reduced OVB, but its elimination is not guaranteed. Nevertheless, it is reasonable to say that delay from distance is a better estimator of how delay affects mortality, given that delay in Models 1 and 2 is likely to be mostly determined by disease severity, something which cannot affect distance.

While Models 1 and 2 might be less informative for estimating the impact of delay on mortality, Figure 6 below suggests ways in which they are nonetheless useful. The OLS models, and particularly Model 2, seem more adept at predicting individual mortality risk of patients for each day of delay than Model 3. These models are useful at capturing associations between variables but are less equipped to isolate the impact of delay on patients' mortality risk. For this, Model 3 is more useful for the reasons explained above. This also helps explain why Model 3 has a flatter curve, given that it only accounts for a part of delay due to distance, even though delay appears to contain an important predictive component of mortality (Figure 3).

Figure 6

Illustration of Models 1 to 3.



Note. This figure shows the prediction of mortality of Models 1 to 3 for each day of delay. Admission delay is counted as the days between first symptoms and hospital admission. Mortality is predicted for each individual patient based on various explanatory factors, which is then plotted against the actual admission delay of the patient.

Based on the results from Model 3, one cannot reject the first hypothesis that as delay in COVID-19 treatment increases, mortality risk increases. This is because of the positive statistically significant coefficient for delay from distance in the second stage of the IV model. While the literature strongly suggests that delay in care is associated with an increased mortality risk (Thaddeus & Maine, 1994; Hanna et al., 2020; Kutikov et al., 2020; Demissie, 2002; Le Terrier et al., 2022), it is possible that the positive effect of delay due to distance on mortality is over- or under-estimated due to OVB. As delay due to distance is relatively small, few patients are expected to delay care by an entire day due to distance alone. In fact, the negative relationship between mortality and delay in Models 1 and 2, as well as shorter treatment delays by more vulnerable groups in the first stage of Model 3 suggest that COVID-19 severity is the dominant factor determining the duration of delay.

Nevertheless, the estimator of delay in Model 3 reaffirms the importance of reducing barriers to accessing healthcare, both due to distance and other factors (Thaddeus & Maine, 1994). Reducing barriers leads to shorter Phase I and II delays, as there are fewer disincentives as well as physical barriers that prevent timely treatment for COVID-19. This must include increasing the reach of healthcare services from Brazil's major urban centers to less densely populated areas (Pereira et al., 2020; Rocha et al., 2021). Moreover, reducing transport disadvantage might include investing more in crucial transport links that enable timely and reliable access to healthcare facilities. Reducing severity of COVID-19 infection at time of admission may also reduce pressure on healthcare systems, as less severe cases of COVID-19 tend to recover more quickly, thus freeing up healthcare services (Le Terrier et al., 2022).

5.3.Hypothesis 2: Regional Differences in Mortality

The second hypothesis claims that part of the greater in-hospital mortality in the Northern, compared to the Central-Southern regions of Brazil is due to greater distances to the healthcare facility in the former. Results from the previous section already suggest that the impact of distance on delay is small. This also explains why, despite the Northern population of Brazil having to travel significantly further than their Central-Southern counterparts to access a healthcare facility (see Figure 4 and Table A2 in the Appendix), there is no statistically significant difference in the delay in treatment. This section will discuss the regression results for Hypothesis 2. Control variables are included in the model but excluded from Table 3. For the full table, see Table A3 in the Appendix.

Results from Table 3 confirm that there is no statistically significant difference between Brazil's regions in terms of delay-attributed mortality that stems from distance to the healthcare facility. While this might stand in contrast with insights drawn from Hypothesis 1, there are a few possible explanations. While there are differences in distance between Northern and Central-Southern regions, there is more variation within than between the regions. This is relevant considering that results from Table 2 suggest that delays due to distance are small. Additionally, the higher mortality burden on the Northern regions has been accompanied by greater adherence to regulations in those regions (Rocha et al., 2021), suggesting greater vigilance regarding the virus and shorter Phase I delays. Related to this is lower support for Bolsonaro, Brazil's President, in the Northern regions of Brazil (BBC News, 2018). Less concern about COVID-19 in the Central-Southern regions of Brazil might have increased phase I delay of decision-making, if patients believed, in Bolsonaro's words, that COVID-19 was a *gripezinha* ('little flu') (Walsh, 2020), thus equalizing delay between Brazil's regions.

Table 3

Main Regression Results from Model 3 with Interaction Effect of Northern Regions

	(1)	(2)
VARIABLES	Delay 2SLS regression: First Stage	Mortality 2SLS regression: Second Stage
Admission delay in days		0.316***
		(0.0579)
Northern Region # Delay		0.0599
		(0.112)
Log of km distance	0.0544***	

	(0.00911)	
Northern Region # Distance	-0.0268	
	(0.0224)	
First Stage F-Statistic	16.533	
Observations	574,237	574,237
R-squared		-9.838
Hospital FE	YES	YES
Municipality FE	YES	YES

Note. Standard errors in parentheses. The significance levels are indicated by *** p<0.01, ** p<0.05, * p<0.1. This is the summarized form of the regression model. The full version, with controls, is included in the Appendix in Table A3.

These findings suggests that most of the in-hospital mortality discrepancy between the Northern and Central-Southern regions of Brazil was due to different levels in hospital quality (Bacqui et al., 2020; Rocha et al., 2021). Hence, Hypothesis 2 is rejected due to an insignificant difference in mortality from distance-attributed delay between the Northern and Central-Southern regions. Despite significant differences in distance to healthcare facilities, these are unlikely to have played a part in the mortality discrepancy between Brazil's regions.

6. Conclusion

This paper has investigated how distance-attributed delay has affected in-hospital mortality risk of patients in Brazil in 2020 and 2021. Given the dominance of disease severity in patients' decisions of when to seek treatment for COVID-19, an IV approach using distance as an instrument was implemented. Various control variables as well as hospital and municipality FE help reduce the bias in the estimator of distance-attributed delay, as distance can still affect disease severity indirectly. However, the bias in the IV model is likely to be significantly smaller than in the OLS model, as the control variables can help account for the association between disease severity and distance.

There are two major findings. First, distance was associated with greater delay in seeking treatment for COVID-19, which was accompanied by an increased mortality risk, thus providing evidence for the first hypothesis. However, a causal effect cannot be ensured due to potential violation of the Independence assumption. Second, the small association between distance and delay suggests that the estimated impact of delay on mortality is exaggerated, as few people would delay care by one day due to distance alone. It is this, amongst other things, that can help explain why greater average distance to a hospital in the Northern regions of Brazil was not associated with an increased mortality risk, leading to the rejection of the second hypothesis. The mortality inequality between regions is much more likely to be due to the poorer healthcare infrastructure in the North (Bacqui et al., 2020; Rocha et al., 2021).

There are some important limitations of the validity and robustness of the results. Firstly, the measurement of distance to the healthcare facility is imprecise, as it approximates distance on a municipal level. This leads to little variation on the municipal level, which might also partly explain why the prediction of delay due to distance is small. Despite imprecision in the measurement of distance, it is the best approximation given the dataset, which covers over 1.7 million patients. This is because it accounts for between-municipality variation by scaling distance travelled to the size of each municipality. Hence, the distance approximation creates concerns of validity and robustness, given that a more precise measurement might produce different results. As inequality is generally greater when measured at the individual level than the municipal level (Bär et al., 2020), it is possible that there would be a greater impact of distance with more accurate travel data.

Next, the validity of the instrument implies that no causal effect can be inferred from the results. As the instrument distance is likely associated with other factors which affect delay and mortality, it is likely that there is bias in the estimator of distance in the first stage, and bias in the estimator of delay in the second stage. Hence, the estimators should be interpreted with caution and not be seen as a causal effect, but rather an association. Even though controlling for various factors can help reduce OVB, an endogenous instrument means the elimination of bias cannot be ensured.

To address these limitations, future research could collaborate with the *SUS* to gain access to more accurate geospatial data on patients, as the address is included in the patient questionnaire, but not in the public dataset. Having more accurate geospatial data can help to estimate more precisely the interaction between distance and healthcare access. Moreover, while finding a valid instrument is difficult, researchers more familiar with the context could improve on this research by implementing the methods using an exogenous instrument such as hospital closures (Walker et al., 2011). This would give the results causal interpretability, which is of essence for policy makers implementing scarce resources. Lastly, future research could test for heterogenous effects, for example by educational attainment, as sensitivity to distance is likely to differ between certain groups.

To conclude, this research has shown that while offering healthcare services to all is essential, there are important barriers to healthcare access which can be detrimental to patients' survival. It suggests that, even in public healthcare, there is inequality in access. Moreover, these barriers can put an even greater strain on healthcare systems during crises, as patients arrive to the healthcare facility with the illness having advanced to a more severe stage. It is crucial that healthcare services and governments take a holistic approach to providing healthcare, as the existence of a healthcare facility alone does not ensure equitable access.

7. References

Albuquerque, M. V. D., Viana, A. L. D. Á., Lima, L. D. D., Ferreira, M. P., Fusaro, E. R., & Iozzi, F. L. (2017). Regional health inequalities: changes observed in Brazil from 2000-2016. *Ciência & Saúde Coletiva*, *22*, 1055-1064.

Amann, E., Baer, W., Trebat, T., & Lora, J. V. (2016). Infrastructure and its role in Brazil's development process. *The Quarterly Review of Economics and Finance*, *62*, 66-73.

Amaral P.V., Rocha T.A.H., Barbosa A.C.Q., Vissoci J.R.N. Spatially balanced provision of health equipment: a cross-sectional study oriented to the identification of challenges to access promotion. *Int. J. Equity Health.* 2017;16(1):209. doi: 10.1186/s12939-017-0704-x

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American statistical Association*, *91*(434), 444-455.

Angrist, J. D., & Krueger, A. B. (2001). Instrumental variables and the search for identification: From supply and demand to natural experiments. *Journal of Economic perspectives*, *15*(4), 69-85.

Baqui, P., Bica, I., Marra, V., Ercole, A., & van Der Schaar, M. (2020). Ethnic and regional variations in hospital mortality from COVID-19 in Brazil: a cross-sectional observational study. *The Lancet Global Health*, *8*(8), e1018-e1026.

Baqui, P., Marra, V., Alaa, A. M., Bica, I., Ercole, A., & van Der Schaar, M. (2021). Comparing COVID-19 risk factors in Brazil using machine learning: the importance of socioeconomic, demographic and structural factors. *Scientific reports*, *11*(1), 1-10.

Bär, M., Wouterse, B., Riumallo Herl, C., Van Ourti, T., & Van Doorslaer, E. (2021). Diverging Mortality Inequality Trends among Young and Old in the Netherlands. *Fiscal Studies*, *42*(1), 79-101.

Bauer, J., Brüggmann, D., Klingelhöfer, D., Maier, W., Schwettmann, L., Weiss, D. J., & Groneberg, D. A. (2020). Access to intensive care in 14 European countries: a spatial analysis of intensive care need and capacity in the light of COVID-19. *Intensive care medicine*, *46*(11), 2026-2034.

Baum, C., Schaffer, M., & Stillman, S. (2015). IVREG210: Stata module for extended instrumental variables/2SLS and GMM estimation (v10).

BBC News. (2018, 8 October). Brazil election: Far-right Bolsonaro "will not soften rhetoric". *BBC News.* https://www.bbc.com/news/world-latin-america-45791811

Benevenuto, R. G., Azevedo, I. C. C., & Caulfield, B. (2019). Assessing the spatial burden in health care accessibility of low-income families in rural Northeast Brazil. *Journal of Transport & Health*, *14*, 100595.

Benoit, K. (2011). Linear regression models with logarithmic transformations. *London School of Economics, London*, *22*(1), 23-36.

Birkmeyer, J. D., Barnato, A., Birkmeyer, N., Bessler, R., & Skinner, J. (2020). The impact of the COVID-19 pandemic on hospital admissions in the United States: study examines trends in US hospital admissions during the COVID-19 pandemic. *Health Affairs*, *39*(11), 2010-2017.

Braga, L. D. S., Caiaffa, W. T., Ceolin, A. P. R., de Andrade, F. B., & Lima-Costa, M. F. (2019). Perceived discrimination among older adults living in urban and rural areas in Brazil: a national study (ELSI-Brazil). *BMC geriatrics*, *19*(1), 1-11.

Bugiardini, R., Ricci, B., Cenko, E., Vasiljevic, Z., Kedev, S., Davidovic, G., ... & Badimon, L. (2017). Delayed care and mortality among women and men with myocardial infarction. *Journal of the American Heart Association*, *6*(8), e005968.

Caplan, L. (2014). Delay in breast cancer: implications for stage at diagnosis and survival. *Frontiers in public health*, *2*, 87.

Conti, G., Heckman, J., & Urzua, S. (2010). The education-health gradient. *American economic review*, *100*(2), 234-38.

Correia, S. (2017). reghdfe: Stata module for linear and instrumental-variable/gmm regression absorbing multiple levels of fixed effects. Statistical Software Components s457874, Boston College Department of Economics. https://ideas.repec.org/c/boc/bocode/s457874.html

Correia, S. 2018. "**IVREGHDFE: Stata module for extended instrumental variable regressions with multiple levels of fixed effects**," Statistical Software Components S458530, Boston College Department of Economics.

Coutinho, R. (2020, December 8). Hospitais municipais do Rio estão lotados e não têm mais leitos de UTI para COVID-19. *G1*. https://g1.globo.com/rj/rio-de-janeiro/noticia/2020/12/08/hospitais-municipais-do-rio-estao-lotados-e-nao-tem-mais-leitos-de-uti-para-COVID-19.ghtml

Currie, G., & Delbosc, A. (2010). Modelling the social and psychological impacts of transport disadvantage. *Transportation*, *37*(6), 953-966.

Czeisler, M, É,, Marynak, K,, Clarke, K, E,, Salah, Z,, Shakya, I,, Thierry, J, M,, ,,, & Howard, M, E, (2020), Delay or avoidance of medical care because of COVID-19–related concerns—United States, June 2020, *Morbidity and mortality weekly report*, *69*(36), 1250,

Dalsania, A. K., Fastiggi, M. J., Kahlam, A., Shah, R., Patel, K., Shiau, S., ... & DallaPiazza, M. (2022). The relationship between social determinants of health and racial disparities in COVID-19 mortality. *Journal of racial and ethnic health disparities*, *9*(1), 288-295.

Datasus. (n.d.) Brasilia: Ministério da Saúde—Sistema de Informações Ambulatoriais do SUS (SIA/SUS). http://tabnet.datasus.gov.br/cgi/tabcgi.exe?sia/cnv/qauf.def

De Luca, G., Suryapranata, H., Ottervanger, J. P., & Antman, E. M. (2004). Time delay to treatment and mortality in primary angioplasty for acute myocardial infarction: every minute of delay counts. *Circulation*, *109*(10), 1223-1225.

Demissie, M., Lindtjorn, B., & Berhane, Y. (2002). Patient and health service delay in the diagnosis of pulmonary tuberculosis in Ethiopia. *BMC public health*, *2*(1), 1-7.

Demombynes, G. (2020). COVID-19 Age-Mortality Curves Are Flatter in Developing Countries. *Policy Research Working Paper;No. 9313. World Bank, Washington, DC. © World Bank.* https://openknowledge.worldbank.org/handle/10986/34028 License: CC BY 3.0 IGO

Deuchert, E., & Huber, M. (2017). A cautionary tale about control variables in iv estimation. *Oxford Bulletin of Economics and Statistics*, *79*(3), 411–425. https://doi.org/10.1111/obes.12177

Faes, C., Abrams, S., Van Beckhoven, D., Meyfroidt, G., Vlieghe, E., Hens, N., & Belgian Collaborative Group on COVID-19 Hospital Surveillance (2020). Time between Symptom Onset, Hospitalisation and Recovery or Death: Statistical Analysis of Belgian COVID-19 Patients. *International journal of environmental research and public health*, *17*(20), 7560. https://doi.org/10.3390/ijerph17207560

Feldman, J. M., & Bassett, M. T. (2021). Variation in COVID-19 mortality in the US by race and ethnicity and educational attainment. *JAMA network open*, *4*(11), e2135967-e2135967.

Globo. (2020, April 20). Superlotação das UTIs: Fantástico mostra a situação crítica em capitais por causa da COVID-19. *G1*. https://g1.globo.com/fantastico/noticia/2020/04/19/superlotacao-das-utis-fantastico-mostra-a-situacao-critica-em-capitais-por-causa-da-COVID-19.ghtml

Hanna, T. P., King, W. D., Thibodeau, S., Jalink, M., Paulin, G. A., Harvey-Jones, E., ... & Aggarwal, A. (2020). Mortality due to cancer treatment delay: systematic review and meta-analysis. *bmj*, *371*.

Hirose, A., Borchert, M., Cox, J., Alkozai, A. S., & Filippi, V. (2015). Determinants of delays in travelling to an emergency obstetric care facility in Herat, Afghanistan: an analysis of cross-sectional survey data and spatial modelling. *BMC pregnancy and childbirth*, *15*(1), 1-13.

Instituto Brasileiro de Geografia e Estastatisticas (IBGE). (n.d.). Population Census. https://www.ibge.gov.br/en/statistics/social/population/22836-2020-census-censo4.html?=&t=resultados

Instituto Brasileiro de Geografía e Estatísticas. (2010). Densidade demografica. https://geoftp.ibge.gov.br/cartas_e_mapas/mapas_do_brasil/sociedade_e_economia/mapas_murais/densidade_populacional_2010.pdf]

Islam N, Shkolnikov V M, Acosta R J, Klimkin I, Kawachi I, Irizarry R A et al, (2021) Excess deaths associated with COVID-19 pandemic in 2020: age and sex disaggregated time series analysis in 29 high income countries *BMJ*; 373 :n1137 doi:10,1136/bmj,n1137

Kelly, C., Hulme, C., Farragher, T., & Clarke, G. (2016). Are differences in travel time or distance to healthcare for adults in global north countries associated with an impact on health outcomes? A systematic review. *BMJ open*, *6*(11), e013059.

Kim, K., Ghorbanzadeh, M., Horner, M. W., & Ozguven, E. E. (2021). Identifying areas of potential critical healthcare shortages: A case study of spatial accessibility to ICU beds during the COVID-19 pandemic in Florida. *Transport Policy*, *110*, 478-486.

Kutikov, A., Weinberg, D. S., Edelman, M. J., Horwitz, E. M., Uzzo, R. G., & Fisher, R. I. (2020). A war on two fronts: cancer care in the time of COVID-19. *Annals of internal medicine*, *172*(11), 756-758.

Le Terrier, C., Suh, N., Wozniak, H., Boroli, F., Giudicelli-Bailly, A., Sangla, F., ... & Pugin, J. (2022). Delayed intubation is associated with mortality in patients with severe COVID-19: A single-centre observational study in Switzerland. *Anaesthesia Critical Care & Pain Medicine*, *41*(4), 101092.

Lobo, S. M., Creutzfeldt, C. J., Maia, I. S., Town, J. A., Amorim, E., Kross, E. K., ... & Wahlster, S. (2022). Perceptions of critical care shortages, resource use, and provider well-being during the COVID-19 pandemic: a survey of 1,985 health care providers in Brazil. *Chest*.

Londoño, E. (2021, September 22). Bolsonaro's Rejection of COVID Vaccine Creates Awkwardness in N.Y. *The New York Times*. https://www.nytimes.com/2021/09/21/world/americas/bolsonaro-COVID-vaccine-un.html

Lucas, K. (2012). Transport and social exclusion: Where are we now?. *Transport policy*, *20*, 105-113.

Mgawadere, F., Unkels, R., Kazembe, A., & van den Broek, N. (2017). Factors associated with maternal mortality in Malawi: application of the three delays model. *BMC pregnancy and childbirth*, *17*(1), 1-9.

Mahl, C., Melo, L. R. S. D., Almeida, M. H. A., Carvalho, C. S., Santos, L. L. S., Nunes, P. S., ... & Martins-Filho, P. R. (2020). Delay in head and neck cancer care during the COVID-19 pandemic and its impact on health outcomes. *Brazilian oral research*, *34*.

Målqvist, M., Sohel, N., Do, T. T., Eriksson, L., & Persson, L. Å. (2010). Distance decay in delivery care utilisation associated with neonatal mortality. A case referent study in northern Vietnam. *BMC Public Health*, *10*(1), 1-9.

Masri, D. A., Flamini, V., & Toscani, F. (2021). The short-term impact of COVID-19 on labor markets, poverty and inequality in Brazil. *IMF Working Papers.*

Mayumi Sakamoto, L. (n.d.) The impacts of COVID-19 on access to private healthcare in Brazil. *International Bar Association*.

McClellan, M., McNeil, B. J., & Newhouse, J. P. (1994). Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality?: analysis using instrumental variables. *Jama*, *272*(11), 859-866.

Miller, I. F., Becker, A. D., Grenfell, B. T., & Metcalf, C. J. E. (2020). Disease and healthcare burden of COVID-19 in the United States. *Nature Medicine*, *26*(8), 1212-1217.

Mishra, V., Seyedzenouzi, G., Almohtadi, A., Chowdhury, T., Khashkhusha, A., Axiaq, A., ... & Harky, A. (2021). Health inequalities during COVID-19 and their effects on morbidity and mortality. *Journal of healthcare leadership*, *13*, 19.

Nascimento, O. A., Camelier, A., Rosa, F. W., Menezes, A. M. B., Pérez-Padilla, R., & Jardim, J. R. (2007). Chronic obstructive pulmonary disease is underdiagnosed and undertreated in São Paulo (Brazil): results of the PLATINO study. *Brazilian Journal of Medical and Biological Research*, *40*, 887-895.

Nilsson, G., Mooe, T., Söderström, L., & Samuelsson, E. (2016). Pre-hospital delay in patients with first time myocardial infarction: an observational study in a northern Swedish population. *BMC cardiovascular disorders*, *16*(1), 1-10.

Noronha, K. V. M. D. S., Guedes, G. R., Turra, C. M., Andrade, M. V., Botega, L., Nogueira, D., ... & Ferreira, M. F. (2020). The COVID-19 pandemic in Brazil: analysis of supply and demand of hospital and ICU beds and mechanical ventilators under different scenarios. *Cadernos de Saúde Pública*, *36*.

Patanavanich, R., & Glantz, S. A. (2020). Smoking is associated with COVID-19 progression: a meta-analysis. *Nicotine and tobacco research*, *22*(9), 1653-1656.

Pajuelo, M. J., Anticona Huaynate, C., Correa, M., Mayta Malpartida, H., Ramal Asayag, C., Seminario, J. R., ... & Paz-Soldan, V. A. (2018). Delays in seeking and receiving health care services for pneumonia in children under five in the Peruvian Amazon: a mixed-methods study on caregivers' perceptions. *BMC health services research*, *18*(1), 1-11.

Pereira, R. H., Braga, C. K. V., Servo, L. M., Serra, B., Amaral, P., Gouveia, N., & Paez, A. (2021). Geographic access to COVID-19 healthcare in Brazil using a balanced float catchment area approach. *Social Science & Medicine*, *273*, 113773.

Phillips, T. (2021, January 18). Bolsonaro rival hails COVID vaccinations as "triumph of science against denialists." *The Guardian*. https://www.theguardian.com/world/2021/jan/18/bolsonaro-rival-hails-COVID-vaccinations-as-triumph-of-science-against-denialists

Picard, R. (2010). GEODIST: Stata module to compute geodetic distances. *Boston College Working Papers in Economics*.

Queiroz, M. C. D. C. A. M. D., Moreira, M. A. C., & Rabahi, M. F. (2012). Underdiagnosis of COPD at primary health care clinics in the city of Aparecida de Goiânia, Brazil. *Jornal Brasileiro de Pneumologia*, *38*, 692-699.

Ranzani, O. T., Bastos, L. S., Gelli, J. G. M., Marchesi, J. F., Baião, F., Hamacher, S., & Bozza, F. A. (2021). Characterisation of the first 250 000 hospital admissions for COVID-19 in Brazil: a retrospective analysis of nationwide data. *The Lancet Respiratory Medicine*, *9*(4), 407-418.

Redação Portal. (2022, March 10). A mobilidade urbana como instrument de desigualidade social. *Transito web*. https://transitoweb.com.br/a-mobilidade-urbana-como-instrumento-de-desigualdade-social/

Ribeiro, K. B., Ribeiro, A. F., Veras, M. A. D. S. M., & de Castro, M. C. (2021a). Social inequalities and COVID-19 mortality in the city of São Paulo, Brazil. *International Journal of Epidemiology*, *50*(3), 732-742.

Ribeiro, M. M., Turra, C. M., & Pinto, C. C. D. X. (2021b). Adult mortality by education level in São Paulo: comparative analysis based on different methodological strategies. *Revista Brasileira de Estudos de População*, *38*.

Ritchie, H. Mathieu, E., Rodés-Guirao, L. Appel, C. Giattino, C., Ortiz-Ospina, E., Hasell, J., Macdonald, B., Beltekian, D. & Roser, M (2020) - "Coronavirus Pandemic (COVID-19)". *Published online at OurWorldInData.org*. Retrieved from: 'https://ourworldindata.org/coronavirus' [Online Resource]

Rocha, R., Atun, R., Massuda, A., Rache, B., Spinola, P., Nunes, L., ... & Castro, M. C. (2021). Effect of socioeconomic inequalities and vulnerabilities on health-system preparedness and response to COVID-19 in Brazil: a comprehensive analysis. *The Lancet Global Health*, *9*(6), e782-e792.

dos Santos Costa, A. C., Hasan, M. M., Xenophontos, E., Mohanan, P., Bassey, E. E., Hashim, H. T., ... & Essar, M. Y. (2021). COVID-19 and Zika: an emerging dilemma for Brazil. *Journal of Medical Virology*, *93*(7), 4124.

Sanyaolu, A., Okorie, C., Marinkovic, A., Patidar, R., Younis, K., Desai, P., ... & Altaf, M. (2020). Comorbidity and its impact on patients with COVID-19. *SN comprehensive clinical medicine*, *2*(8), 1069-1076.

Schoeps, A., Gabrysch, S., Niamba, L., Sié, A., & Becher, H. (2011). The effect of distance to health-care facilities on childhood mortality in rural Burkina Faso. *American journal of epidemiology*, *173*(5), 492-498.

Schünemann, J., Strulik, H., & Trimborn, T. (2017). The gender gap in mortality: How much is explained by behavior?. *Journal of health economics*, *54*, 79-90.

Silva, L. L., de Carvalho Dutra, A., de Andrade, L., Iora, P. H., Ramajo, G. L. R., Gualda, I. A. P., ... & Cardoso, R. F. (2021). Emergency Care Gap in Brazil: Geographical Accessibility as a Proxy of Response Capacity to Tackle COVID-19. *Frontiers in public health*, *9*.

Silva, L. L. D., Nascimento, P. E., Araújo, O. C. G., & Pereira, T. M. G. (2021). The articulation of the Indigenous Peoples of Brazil in facing the COVID-19 pandemic. *Frontiers in Sociology*, *6*, 611336.

da Silva, R. C. M. F. E. (2021, March 20). Mortes na fila por um leito de UTI, falta de insumos e funerárias sem férias: os sinais do colapso na saúde brasileira. *G1*. https://g1.globo.com/bemestar/coronavirus/noticia/2021/03/20/mortes-na-fila-por-um-leito-de-uti-falta-de-insumos-e-funerarias-sem-ferias-os-sinais-do-colapso-na-saude-brasileira.ghtml

Sousa BLA, Brentani A, Costa Ribeiro CC, Dolhnikoff M, Grisi SJFE, Ferrer APS, Ferraro AA. (2021). Non-communicable diseases, sociodemographic vulnerability and the risk of mortality in hospitalised children and adolescents with COVID-19 in Brazil: a cross-sectional observational study. BMJ Open. 2021 Sep 6;11(9):e050724. doi: 10.1136/bmjopen-2021-050724. PMID: 34489291; PMCID: PMC8423513.

Souza, C. D. F. D., Santana, G. B. D. A., Leal, T. C., Paiva, J. P. S. D., Silva, L. F. D., Santos, L. G., ... & Carmo, R. F. D. (2021). Relationship between the intensive care unit beds and mortality by COVID-19 in Brazil. *Revista da Associação Médica Brasileira*, *67*, 645-649.

Statista. (2018). Income distribution inequality based on Gini coefficient in Latin America as of 2018, by country. https://www.statista.com/statistics/980285/income-distribution-gini-coefficient-latin-america-caribbean-country/

Stock, J. & Yogo, M. (2005). Testing for Weak Instruments in Linear IV Regression. In: *Andrews DWK Identification and Inference for Econometric Models*. New York: Cambridge University Press ; 2005. pp. 80-108.

Szwarcwald, C. L., Souza Júnior, P. R. B. D., Marques, A. P., Almeida, W. D. S. D., & Montilla, D. E. R. (2016). Inequalities in healthy life expectancy by Brazilian geographic regions: findings from the National Health Su rvey, 2013. *International journal for equity in health*, *15*(1), 1-9.

TeuschI, Y., & Brainin, M. (2010). Stroke education: discrepancies among factors influencing prehospital delay and stroke knowledge. *International Journal of Stroke*, *5*(3), 187-208.

Thaddeus S, Maine D. (1994). Too far to walk: maternal mortality in context. *Soc Sci Med.* Apr;38(8):1091-110. doi: 10.1016/0277-9536(94)90226-7. PMID: 8042057.

Vilela, P. R. (2020, September 27). Desigualdades de gênero, raça e renda marcam uso do transporte no DF. *Agência Brasil.* https://agenciabrasil.ebc.com.br/geral/noticia/2020-09/desigualdades-de-genero-raca-e-renda-marcam-uso-do-transporte-no-df

Vincenty, T. (1975). Direct and inverse solutions of geodesics on the ellipsoid with application of nested equations. *Survey Review, 23(176), 88–93.* doi:10.1179/sre.1975.23.176.88

Vintura (2021), Every day counts – The impact of COVID-19 on patient access to cancer care in Europe.

Walker, K. O., Leng, M., Liang, L. J., Forge, N., Morales, L., Jones, L., & Brown, A. (2011). Increased patient delays in care after the closure of Martin Luther King Hospital: implications for monitoring health system changes. *Ethnicity & disease*, *21*(3), 356–360.

Walsh, N. J. P. S. (2020, 25 mei). Bolsonaro calls coronavirus a "little flu." Inside Brazil's hospitals, doctors know the horrifying reality. *CNN.* https://edition.cnn.com/2020/05/23/americas/brazil-coronavirus-hospitals-intl/index.html

Wang, P. S., Berglund, P. A., Olfson, M., & Kessler, R. C. (2004). Delays in initial treatment contact after first onset of a mental disorder. *Health services research*, *39*(2), 393-416.

WHO. (n.d.a). Severe Acute Respiratory Syndrome (SARS). https://www.who.int/health-topics/severe-acute-respiratory-syndrome#tab=tab_1

WHO. (n.d.b) Social determinants of health. https://www.who.int/health-topics/social-determinants-of-health#tab=tab_1

Zaman S.M.A., Cox J., Enwere G.C., Cutts F.T. The effect of distance on observed mortality, childhood pneumonia and vaccine efficacy in rural Gambia. *Epidemiol. Infect.* 2014;142(12):2491–2500. doi: 10.1017/S0950268814000314.

Zhang, B., Zhou, X., Qiu, Y., Song, Y., Feng, F., Feng, J., ... & Wang, J. (2020). Clinical characteristics of 82 cases of death from COVID-19. *PloS one*, *15*(7), e0235458.

Zhou, F., Yu, T., Du, R., Fan, G., Liu, Y., Liu, Z., ... & Cao, B. (2020). Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *The lancet*, *395*(10229), 1054-1062.

8. Appendix

Table A1

Overview of Brazil's Federal States and Macroregions, classified into Northern and Central-South Regions

	Macroregion	Federal States
Northern Regions	North	Acre (AC), Amazonas (AM), Rondonia (RO), Roraima (RR), Pará, Amapá (AP), Tocatins (TO)
	Northeast	Maranhão (MA), Piauí (PI), Ceará (CE), Rio Grande do Norte (RN), Paraíba (PB), Pernambuco (PE), Alagoas (AL), Segipe (SE), Bahia (BA)
Central-Southern Regions	Center-West	Mato Grosso (MT), Goiás (GO), Distrito Federal (DF), Mato Grosso do Sul (MS)
	Southeast	Minas Gerais (MG), Espírito Santo (ES), Rio de Janeiro (RJ), Sao Paulo (SP)
	South	Paraná (PR), Santa Catarina (SC), Rio Grande do Sul (RS)

Note. This table summarizes the composition of Brazil's regions (as suggested by Bacqui et al., 2020), macroregions and states. There are five macroregions and 26 states.

Table A2

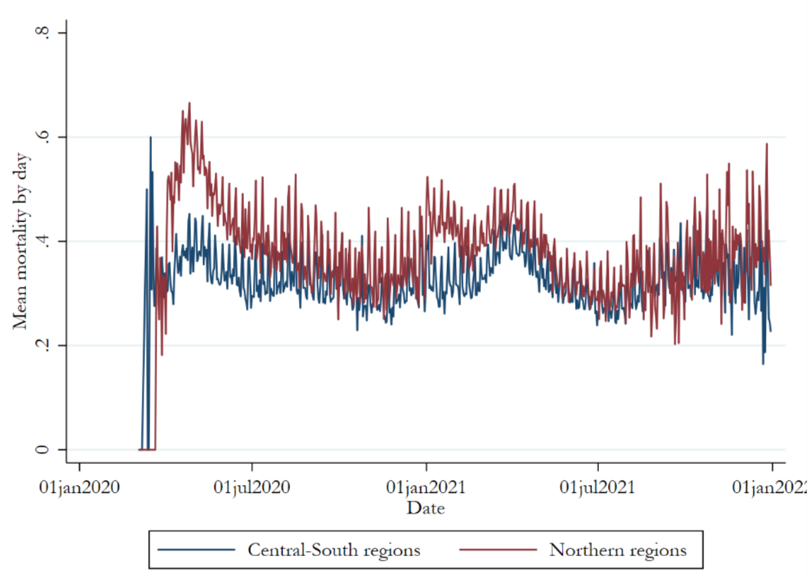
T-test for differences between Northern and Central-Southern Regions.

	N: Central-South	N: North	Mean Central-South	Mean North	difference	St Err	t value	p value
Delay in Admission: North and Central-South	1232594	403666	7.116	7.123	-.007	.008	-.9	.381
Distance to Healthcare facility: North and Central-South	1232331	403605	17.934	43.169	-25.236	.185	-135.95	0.000

Note. This table shows the t-test statistics comparing distance and delay between the Northern and Central-Southern regions of Brazil. Distance is calculated as the km distance between the municipality centroid of a patient's residence and treatment facility. Delay is measured as days between first symptoms and hospital admission.

Figure A1

Daily In-hospital Mortality Rates by Regions of Brazil in 2020 and 2021



Note: This figure shows the average in-hospital mortality rate in the Northern and Central-Southern regions of Brazil per day in 2020 and 2021, with blue indicating the Central-South and red indicating the North.

Table A3

Full Regression Results for Hypothesis 2, Table 3

VARIABLES	(1) Delay 2SLS regression: First Stage	(2) Death 2SLS regression: Second Stage
Log of km distance	0.0544***	
	(0.00911)	
North#distance	-0.0268	
	(0.0224)	
Admission delay in days		0.316***
		(0.0579)
North#delay		0.0599
		(0.112)
Urban	-0.00490	-0.00441
	(0.0203)	(0.00747)
Male	0.192***	-0.0259**
	(0.0108)	(0.0118)
Black	-0.291***	0.0995***
	(0.0240)	(0.0189)
East Asian	-0.163***	0.0299
	(0.0548)	(0.0220)
Parda	-0.110***	0.0237***
	(0.0147)	(0.00821)
Indigenous	-0.605***	0.223***
	(0.118)	(0.0567)
60-69	-0.103***	0.166***
	(0.0146)	(0.00819)
70-79	-0.667***	0.434***
	(0.0165)	(0.0386)
80-89	-1.285***	0.718***
	(0.0207)	(0.0740)
90-100	-1.738***	0.941***
	(0.0380)	(0.101)
Higher Educated	0.300***	-0.126***
	(0.0124)	(0.0182)
Risk Factors	-0.342***	0.212***
	(0.0120)	(0.0202)
2021	0.776***	-0.189***
	(0.0124)	(0.0449)
F-Statistic	16.533	
Observations	574,237	574,237
R-squared		-9.838
Hospital FE	YES	YES
Municipality FE	YES	YES

Note: This table shows the full regression results seen in Table 3, Section 5. It shows the IV regression of mortality on delay, instrumented using distance. Standard errors in parentheses. Significance levels indicated by *** p<0.01, ** p<0.05, * p<0.1.