IDB WORKING PAPER SERIES Nº IDB-WP-01351

Access to water and COVID-19: a Regression Discontinuity Analysis for the Peri-urban Areas of Metropolitan Lima, Peru

Andrés Gómez-Lobo
Mauro Gutiérrez
Sandro Huamaní
Diego Marino
Tomás Serebrisky
Ben Solís

Inter-American Development Bank
Infrastructure and Energy Sector (INE)

July 2022

Access to water and COVID-19: a Regression Discontinuity Analysis for the Peri-urban Areas of Metropolitan Lima, Peru

Andrés Gómez-Lobo*
Mauro Gutiérrez**
Sandro Huamaní**
Diego Marino**
Tomás Serebrisky***
Ben Solís***

* Department of Economics, University of Chile
** National Superintendence of Water and Sanitation Services (Sunass)
*** Inter-American Development Bank

July 2022

Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library
Access to water and COVID-19: a regression discontinuity analysis for the peri-urban areas of Metropolitan Lima, Peru / Andrés Gómez-Lobo, Mauro Gutiérrez, Sandro Huamaní, Diego Marino, Tomás Serebrisky, Ben Solis.
p. cm. — (IDB Working Paper Series ; 1351)
Includes bibliographic references.
1. Water-supply-Peru-Regression analysis. 2. Infrastructure (Economics)-Peru-Regression analysis. 3. Water utilities-Peru-Regression analysis. 4. Coronavirus infections-Social aspects-Peru-Regression analysis. I. Gómez-Lobo, Andrés. II. Gutierrez, Mauro. III. Huamaní, Sandro. IV. Marino, Diego. V. Serebrisky, Tomás. VI. Solís, Ben. VII. Inter-American Development Bank. Infrastructure and Energy Sector. VIII. Series.

http://www.iadb.org

Address for correspondence: bsolissosa@iadb.org

Abstract

This paper presents the results of a quasi-experimental study using information collected through a survey conducted in peri-urban areas of Metropolitan Lima between October and November 2021. The survey was applied to households residing close to and on both sides of the geographic boundary of piped water supply. Our work finds that access to piped water was associated with a reduction in the probability of a COVID-19 infection. Furthermore, the model used shows heterogeneous effects that suggest that it is not enough for a household to be connected to the water supply network, but that a minimum consumption endowment must also be guaranteed. The results should be interpreted by taking into consideration the limitations of the information. These results highlight the need for investment in infrastructure to close access gaps and the importance of ensuring quality and affordable services for the population.

JEL codes: L95, I14, I15, I10, I18

Keywords: water, sanitation, COVID-19, health, regression discontinuity, Lima, Peru.

Table of Contents

Access to water and COVID-19: a Regression Discontinuity Analysis for the Peri-urban Areas of Metropolitan Lima, Peru[1]

1. Introduction

One of the multiple challenges facing Latin America and the Caribbean is closing the public service infrastructure gap to significantly improve its inhabitants' quality of life. Countries need to make additional public and private investment to guarantee access to basic services. The estimated investment gap in water, electricity, telecommunications, and transportation by 2030 is US$2.2 trillion (Brichetti et al., 2021), which is equivalent to 43% of the region's GDP in 2019. Specifically in the case of water and sanitation services, an estimated US$374 billion are needed to grant universal access to safely managed drinking water and sanitation and to ensure wastewater treatment in urban areas. That amount would allow building the main infrastructure components to achieve Sustainable Development Goal 6 and it would require the region to make an annual average investment of 0.5% of its GDP through 2030.

In the case of Peru, estimates show that by 2020 91.2% of the population had access to water through the public water system, with a significant disparity between urban (94.8%) and rural (77.6%) areas. In Metropolitan Lima[2], home to about one third of the country's population, coverage reaches 95.6% (INEI, 2021), with unserved communities located in the city's peri-urban areas.

Lack of access to water and sanitation services became even more critical in the context of the disease caused by the 2019 coronovairus (henceforth COVID-19), given that the World Health Organization (WHO) recommended frequent handwashing at the beginning of the pandemic as a strategy to prevent infection (WHO, 2020). However, considering that the recommendation calls for handwashing with soap and clean running water for at least 20 seconds, households not connected to the water supply network faced considerable challenge in implementing this recommendation. Furthermore, as further explained later in this paper, residents who are not connected to the water supply network must resort to alternative means for securing water

[1] The authors would like to thank Mateo Prochazka MD MSc, for his valuable contributions to a preliminary version of this document, Guillermo Cruces and Clara Pasman for their comments in the most recent versions of the document, and Beremiz Rojas for his support in systemizing statistical data. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the Inter -American Development Bank (IDB), its Board of Directors, or the countries it represents. The authors retain sole responsibility for any errors or omissions. Those interested in sending comments should direct them to bsolissosa@iadb.org.

[2] Throughout the present analysis, following the nomenclature of Peru's *Instituto Nacional de Estadística e Informática* ("National Institute of Statistics and Informatics"), the term Metropolitan Lima refers to the region consisting of the province of Lima and the constitutional province of Callao (2021 National Household Survey).

(primarily tanker trucks) and are more likely to engage in social interactions with members of other households (e.g., when queuing in long lines or sharing facilities). This in turn could increase the probability of COVID-19 infection among this population. In that regard, we are interested in analysing the association between a lack in access to water and increased exposure to the disease.

Peru is a particularly relevant case study, because the country reported the world's highest number of deaths due to COVID-19 for every 1,000 inhabitants by February 2022 (*The Economist*, 2022). In particular, the Metropolitan Lima region (which includes the capital of Peru) has remained one of the most heavily affected areas both in terms of cases and deaths. The high COVID-19 mortality rates observed in Peru are due in part to structural factors (Gianella et al., 2020; Schwalb and Seas, 2020). Notable among these are factors related to the healthcare system (lack of infrastructure and specialized personnel, system overload, and insufficient capacity for molecular testing) along with high levels of poverty and limited access to basic services—such as water and sanitation. Access to these services is particularly important in the face of a highly contagious infection that requires hygienic habits, such as handwashing, for its prevention.

This paper contributes to the growing literature that seeks to explain differences in the impact of COVID-19 on the health of different populations. To the authors' best knowledge, there are no studies that estimate the causal relationship between access to and quality of water services and the probability of COVID-19 infection, taking the individual as the unit of analysis. This study surveys households close to and on both sides of the geographic boundary of piped water supply in several of the city's peri-urban areas. This design reduces the observable and non-observable differences between households that are and are not connected to the water supply network, increasing the possibility of inferring causality between access to the public water system and the probability of infection. The survey collected information on access to water and sanitation, characteristics of the household and dwelling, personal hygiene, and COVID-19 morbidity, among other variables. This information was used to develop an assessment applying the Regression Discontinuity Design to establish the existing differences in COVID-19 morbidity attributable to households' access to water.

2. Literature Review

There is an abundance of literature documenting the relationship between water and sanitation (access, quality, hygiene), and human health. Waddington and Snilstveit (2009) carried out a systematic revision of evaluations conducted over the past three decades in 35 middle and low-income countries, establishing the relationship between water, sanitation, hygiene, and childhood diarrhea. Other works in the same line are Darvesh et al. (2017) and Wolf et al. (2018). Additionally, systematic reviews of studies analyze the relationship between water quality and lung cancer (Celik et al., 2008), and the impacts of water, sanitation, and hygiene interventions on cholera (Taylor et al., 2015). Cárdenas (2022) conducted an extensive revision of literature concerning the relationship between water and health.

Copious research shows that social, economic, and demographic factors contribute to disparities in the impacts of COVID-19. Lack of adequate infrastructure leads to social dynamics (such as overcrowding in public transportation or at water supply areas) that increase the risk of infection, while other socio-economic factors can severely limit the ability of households to implement preventive care (e.g., insufficient income and informal employment hinder compliance with confinement measures). Moreover, inequality in access to healthcare can be critical for patients requiring oxygen, hospitalization, or a bed in the Intensive Care Unit.

It would be remiss not to consider the scope and limitations of existing documents and publications when interpreting their authors' findings. Most papers analyzed the pandemic during the first months, meaning that initial transmission patterns were observed in a context where there was little information about the virus, no available vaccines, and one or more variants prevailing at the time[3]. Nearly all papers can be characterized as ecological studies – in other words, epidemiological assessments in which the unit of analysis is a population (at the country, state, county, or district levels) and not individuals. Due to the nature of these assessments, causality cannot be attributed to the relationships observed.

To measure the impact of COVID-19, the publications reviewed for our study mainly rely on incidence indicators (number of confirmed cases), mortality rate (number of deaths), and case fatality ratio (ratio between deaths and confirmed cases). For the purposes of this study, these publications can be divided into two groups. Those in the first group analyze the relationship between COVID-19 infections and the socio-economic and demographic characteristics of their location. Those in the second group explicitly include the lack of or

[3] Genomic surveillance varies in each country. In Peru, the most worrying variants were the Alfa, Lambda, Gamma (see Vargas-Herrera et al., 2022), and, most recently, Omicron.

inadequate access to water and sanitation services in their analysis of the geographic disparities in the impact of COVID-19.

Within the first group of publications, most authors identify risk factors that contribute to an increased COVID-19 infection rate or its severity among certain populations, such as age, gender, and comorbidities. These risk factors appear in publications in Brazil (Chauvin, 2021), China (You et al., 2020), Spain (Marí-Dell'Olmo' et al., 2021), United States (Ahmad et al., 2020; Clouston, Natale, and Link, 2021; Hyde, 2021; Kamis et al., 2021; Rozenfeld et al., 2020; Stokes et al., 2021; Strully et al., 2021; Xu et al., 2021), United Kingdom (Nicodemo et al., 2020), and an analysis of multiple countries (Chauvin et al., 2020). Several of these reports also include variables such as race, ethnicity, and level of education[4] (see **Table 1**). Similarly, a significant number of studies identify geographic factors that amplify the impact of the disease, including physical access to the location and the location's connection to roads or airports. Zhang et al. (2021) find that, at the beginning of the pandemic, proximity to Wuhan correlated with a higher infection rate in China. Scarpone et al. (2020) find that greater community interconnection can predict the disease's incidence level in Germany. Fortaleza et al. (2020) evaluate the time necessary for COVID-19 to be introduced into Sao Paulo and find that the disease's earliest introduction took place in districts with greater connectivity. They also find that the distance by land to the capital had a protective effect on Sao Paulo. In the same vein, Chauvin (2021) observes that in Brazil, distance to an international airport correlated with the number of deaths *per capita* during the first wave[5]. The results of these publications are consistent with expected transmission patterns of any outbreak or epidemic. With the exception of China, transmission began in all other countries with the arrival of the first cases from abroad. Transmission initially occurred between infected people and those with whom they had direct contact (focused infection), which offers explanation to the importance of connectivity and distance to airports.

As the process of community transmission progressed, other demographic variables such as population density (i.e. inhabitants per square kilometer), the presence of informal settlements, mobility patterns, and the use of public transportation increased the number and interpersonal proximity of social interactions, leading to a higher risk of infection. Fortaleza et al. (2020) find a positive correlation between Sao Paulo's population density and variables of the impact of COVID-19 (early introduction, incidence, and mortality); while Chauvin

[4] Stokes et al. (2021) find that the disproportionate impact of the pandemic on certain races and socio-economic characteristics is exacerbated when considering the excess of deaths not attributed to COVID-19. The interpretation of the results of ethnicity and race could be explained by cultural, behavioral, and attitudinal factors towards the pandemic.

[5] According to Chauvin (2021), the first wave peaked in July 2020.

(2021), also finds a positive correlation in Brazil between urban population density and the number of deaths *per capita*. You et al. (2020) in their analysis of the districts of Wuhan, observe a positive correlation between morbidity rate and variables such as population density and the building coverage ratio[6]. Similarly, Chauvin (2021) along with Brotherhood et. al (2020) observe that cities with a high percentage of homes in *favelas* suffered disproportionately higher deaths *per capita*, because of the difficulties residents of these neighborhoods encountered in following social distancing recommendations. Mobility patterns also play a significant role in the transmission of the virus. Chauvin (2021) observes higher rates of COVID-19 infection in higher income cities in Brazil[7], which typically experience greater levels of mobility as economic activities continued in spite of the pandemic. The author reaches this observation through a complementary analysis from a database with mobility data gathered from 60 million cellphones. A location's socio-economic status or GDP can act as a proxy variable, in certain cases, of its commercial activity and mobility levels (Zhang et al., 2021; Clouston et al., 2021)[8]. You et al. (2020) find a relationship between the impact of the pandemic and retail sales in China. Moreover, public transportation can contribute significantly to virus transmission, as was the case in New York during the pandemic's initial outbreak (Harris, 2020). The use and time spent on public transportation are also included in studies of Brazil (Chauvin, 2021; Freire de Souza et al., 2020) and the United States (Dasgupta et al., 2020; Hyde, 2021; Rozenfeld et al., 2020; Strully, 2021).

The literature also reveals that indicators in employment can be relevant to explain the impact of COVID-19. These indicators include variables such as employment in essential, manual, or in-person customer service, as well as unemployment (Hawkins, 2020; Lewis et al., 2020; Niedzwiedz et al., 2020; Scarpone et al., 2020; Strully et al., 2021). Chauvin et al. (2020) suggest that, among other factors, high rates of informal employment could contribute to young and middle-aged adults' higher risk of death by COVID-19 in developing countries.

Overcrowding of households (typically measured by the number of people per room) stands out as another variable in many reviewed publications. This is not only an indicator of proximity in indoor interactions but also on the ability to isolate a single member of the family in case of infection. Chauvin (2021) demonstrates a strong and statistically significant

[6] The assessment conducted in Zhang et al. (2021) for different cities in China has shown an inverse relationship—in other words, higher incidence—in areas with lower density. This pattern is both unique and counterintuitive, but the authors suggest it could be due to temporary local migration patterns in the context of local festivities, and the migration from larger to small and medium cities during the pandemic.

[7] Desmet and Wacziarg (2021) find that the opposite was true in the United States.

[8] Clouston et al. (2021) find that, at the beginning of the pandemic, higher socio-economic status was linked to higher virus transmission. As social distancing came into effect –a policy more easily adopted by wealthier households- the poorest counties began to exhibit higher incidence and mortality rates. This is consistent with the findings of De Groot and Lemanski (2021) for South Africa.

correlation of overcrowded homes with infection and deaths from COVID-19 in Brazil. Kamis et al. (2021) find that the disparity in mortality rates between counties in the US was initially low but became increasingly pronounced to the detriment of more heavily overcrowded counties. Dasgupta et al. (2020) also find that counties in the US with high social vulnerability, including a high rate of overcrowded homes, were more likely to be considered infection hotspots. Kamis et al. (2021) observe that the rate of overcrowded homes (controlling for poverty) could significantly help predict COVID-19 mortality. The findings of Ahmad et al. (2020), Lewis et al. (2020), Niedzwiedz et al. (2020), and Strully et al. (2021) also support this claim.

Table 1 summarizes the main studies reviewed that analyze the relationship of social and demographic indicators with COVID-19.

Table 1- Literature on the relationship of social and demographic indicators with COVID-19

#	Country	Author(s)	Year	Period analyzed	Main COVID-19 impact variable	Correlation, cause, and/or control variables	Main finding
1	Multiple countries (18 high-income countries, and 13 developing countries)	Chauvin, Fowler y Herrera	2020	Varies in every country (up to May or August 2020)	COVID-19 mortality rate	Age, gender, ethnic minorities, (pre-pandemic) public health conditions	Young and middle-aged adults in developing countries are more likely to die from COVID-19 as compared to their peers in developed countries.
2	Germany	Scarpone et al.	2020	January 2020 – March 2020	Confirmed COVID-19 cases	Connectivity, employment, church density, number of tourists in overnight accommodations.	The strongest predictors of COVID-19 incidence at the county level were associated with the community's interconnectedness, geographic location, transportation infrastructure, and structure of the labor market.
3	Brazil	Chauvin	2021	February 2020 – February 2021	Deaths *per capita*.	Population density, public transportation commute times, proximity to an international airport, higher income, elderly population, overcrowded homes, presence of *favelas*.	Cities with higher income (more mobility) were more affected by COVID-19. A high percentage of people living in *favelas* made the city particularly vulnerable to COVID-19.
4	Brazil	Fortaleza et al.	2020	Until May 2020	COVID-19 introduction time, and incidence, and mortality rates.	Population density, distance to the state capital, ratio of people in urban areas, Gini Index of income inequality.	Municipalities' levels of influence/connectivity, as well as population density displayed a positive correlation with early COVID-19 introduction, and higher incidence and mortality rates linked to the virus.
5	Brazil	Freire de Souza, Dornels et al.	2020	February 2020 – May 2020	Number of cases every 100,000 deaths per million, mortality rate.	17 indicators linked to COVID-19 infection: *per capita* income below half the minimum wage, work commutes over 1 hour long, school dropout rate, and abandonment of employment, among others.	Expansion begins in more developed municipalities, but it mainly affects the most socially vulnerable municipalities.

6	China	You et al.	2020	Until February 2020	COVID-19 mortality rate	Population density, building coverage ratio, retail sales per landmass, elderly population, GDP density, hospital density.	In an evaluation of Wuhan's 13 districts, variables such as population density, building coverage ratio, and age have shown a positive correlation with the mortality rates from COVID-19.
7	China	Zhang et al.	2021	Until March 2020	Number of confirmed cases	GDP, Population density, healthcare resources (staff, hospital beds, institutions), distance to Wuhan.	Higher risk of COVID-19 infections in cities with a higher GDP, limited health resources, and proximity to Wuhan. Inverse relationship between incidence and population density.
8	Spain	Marí-Dell'Olmo et al.	2021	March 2020 – November 2020	Cumulative COVID-19 cases	Age group, gender, personal income range.	Social inequalities affecting the incidence of COVID-19 were identified in Barcelona by age group, gender, geographic location, and income. Differences were noted between the first and second wave.
9	United States	Clouston, Natale, and Link	2021	January 2020 – May 2020	Confirmed daily cases and daily COVID-19 death toll	Socio-economic status, people over 65 years old, African Americans, the Hispanic community, urbanization.	At the beginning of the pandemic, higher socio-economic status correlated with early infection rates; however, lower-status counties ended up being the most affected.
10	United States	Dasgupta *el al.*	2020	June – July 2020	Classification of COVID-19 infection hotspots at the county level	Race, ethnicity, housing type, transportation	Socially vulnerable counties (especially, those of racial and ethnic minorities, and those with overcrowded households) were more likely to be deemed infection hotspots.
11	United States	Hawkins	2020	January 2020 – June 2020	Positive COVID-19 cases	Average income, percentage of inhabitants with no health insurance, poverty rate, unemployment, and percentage of workers employed in transportation services and in health and social care.	Higher COVID-19 rates in locations with more poverty, lower income, less health insurance coverage, higher unemployment, and higher percentage of the workforce in essential services.
12	United States	Kamis, Stolte et al.	2021	April 2020 – October 2020	COVID-19 death toll	Percentage of homes with more than one person per room, percentage of families under the poverty line, Afro community, Hispanic community, people over 65 years old, etc.	Controlling for poverty, overcrowded homes are a good predictor of deaths from COVID-19.
13	United States	Lewis et al.	2020	March 2020 – July 2020	Confirmed COVID-19 cases	Ethnicity, race, crowded homes, type of job, food insecurity, medical care.	The probability of COVID-19 infection in areas of high economic deprivation in Utah were three times higher than in those less isolated.

14	United States	Rozenfeld et al.	2020	February – April 2020	Positive test results (individual-based study)	Age, gender, race, ethnicity, language other than English, financial security of the neighborhood, air quality, home and transportation insecurity.	There was a higher risk of infection from COVID-19 linked to clinical variables, but also socio-demographic variables (race, ethnicity, home, neighborhood, and transportation conditions, among others).
15	United States	Stokes et al.	2021	January 2020 – December 2020	COVID-19 death toll, excess deaths (including those unrelated to COVID-19)	Population over 65 years old, Hispanics, African Americans, comorbidities (diabetes, obesity, smokers), rural populations, median income.	Total deaths exceed those officially recorded as due to COVID-19. When the total figure is considered, the impact on certain ethnicities and socio-economic characteristics is even more disproportionate.
16	United States	Strully et al.	2021	Until May 2020	Confirmed COVID-19 cases	Percentage of African American population, Asian population, Latinos, and foreigners. Population density, age and gender composition, unemployment rate, median household income, members in household, crowdedness, housing conditions (incomplete kitchen or plumbing installations).	Counties in the US with higher number of immigrants, as well as Central American or Black residents, presented more cases of COVID-19. There are other significant variables such as percentage of household with deficiencies, and the use of public transportation.
17	United States	Xu et al.	2021	January 2020 – December 2020	COVID-19 death toll	Gender, race, ethnicity, age group, underlying health conditions (diabetes, influenza, pneumonia).	Men displayed higher mortality rates in nearly all race and age groups. Certain disparities were noted between the Hispanics, African Americans, and Caucasians.
18	United Kingdom	Nicodemo et al.	2020	January 2020 – June 2020	COVID-19 death toll	Vulnerability to COVID-19 index (prevalence of high-risk diseases, urbanism, availability of resources in the health system, etc.).	The community vulnerability to COVID-19 index has a positive correlation with social isolation measures. The north of the United Kingdom (higher deprivation) was particularly vulnerable to the virus.
19	United Kingdom	Niedzwiedz et al.	2020	March 2020 – May 2020	Confirmed cases, cases that required hospitalization, and positive results (individual-based study)	Ethnicity, country of birth, socio-economic deprivation (unemployment, car and home ownership, crowdedness), education, type of job, urbanism.	Certain ethnic groups were considerably more likely to get COVID-19. Lower education levels and socio-economic deficiencies were linked to a higher risk of infection and hospitalization.

Source: Prepared by the authors.

A second, smaller group of studies analyzes the relationship between the impact of COVID-19 (incidence, mortality, fatality) and variables in access or quality of water and sanitation services (see **Table 2**).

For countries in Sub-Saharan Africa, Amankwaa and Fischer (2020) conduct a simple analysis of correlations, and find that fatalities from COVID-19 were negatively correlated with access to safe water and sanitation. Similarly, Silva et al. (2021) analyze morbidity and mortality data in Brazil and find that higher incidence rates were closely linked to low levels of coverage in water services, as well as to levels of fecal coliforms in drinking water exceeding safety ranges. Regarding COVID-19 mortality, the authors verified a significant correlation with low sanitation and wastewater treatment rates[9]. In India, Das et al. (2020) find that the incidence of COVID-19 was 90% higher than in districts of the megacity Chennai, where availability of water, sanitation, and hygiene services was limited.

In the United Stated, Ahmad et al. (2020) estimate the relative risks of COVID-19 incidence and mortality linked to poor living conditions, controlling for social and demographic variables at the county level. A home is considered to exhibit poor living conditions when, among other criteria, it lacks piped water, a flush toilet, or a bathtub/shower[10]. The authors learned that every 5% increase in the percentage of homes with poor living conditions led to a 50% increase in risk of infection and 42% increased risk of mortality due to COVID-19. Meanwhile, Hyde (2021) assesses the disease's fatality rate *vis-à-vis* violations of water quality regulations occurring in US counties. The author finds that the fatality rate from COVID-19 was higher in counties that registered water quality violations over the median—18% higher in counties more affected by serious quality infringements in terms of health, and 15% higher in counties more affected by violations involving pollutants associated with cardiovascular diseases[11].

Lastly, Revollo-Fernandez et al. (2022) in a recent study discover that the number of deaths from COVID-19 at the municipal level in Mexico were statistically linked to socio-economic and health indicators, including access to water. The authors find that a 20%

[9] Clusters were also found in the Northern and Northeastern regions, which are the poorest in Brazil, featuring low income, human settlements, and a poor sanitation system.

[10] Other criteria include overcrowding, high housing cost burden, and incomplete kitchen facilities. A dwelling is considered overcrowded when the ratio of inhabitants per room is more than 1. Households are considered to have a high cost of living when more than 50% of the monthly household income is allocated towards housing cost (including utilities). A home has incomplete kitchen facilities when it lacks a sink with running water, stove or range, or a refrigerator.

[11] Serious health violations pose an immediate threat to the health of those exposed, while violations involving lead, arsenic, cadmium, and copper are associated to a higher risk of cardiovascular disease.

increase in the percentage of people with no access to water is associated with an average increase of 12.3% in the cumulative number of deaths from COVID-19.

The papers presented analyze the relationship between COVID-19 and the characteristics of dwellings (including access to drinking water and sanitation) through correlations that cannot be considered causal. Our paper proposes to build upon on this aspect by making a methodological contribution to the existent literature.

Table 2- Papers on the relationships between water and COVID-19

N°	Country /region	Author(s)	Year	Period analyzed	Main COVID-19 impact variable	Correlation, cause, and/or control variables	Main finding
1	Sub-Saharan Africa	Amankwaa and Fischer	2020	January 2020 – May 2020	COVID-19 fatality rate	Safe access to water, safe access to sanitation.	The authors find a correlation between higher fatality rates and low rates of safe access to water and sanitation.
2	Brazil	Silva et al.	2021	February 2020 – May 2020	COVID-19 incidence and mortality	Water coverage, sanitation coverage, wastewater treatment, fecal coliforms in drinking water index.	High rates of COVID-19 incidence were linked to a significantly lower coverage of water services, and excessive fecal coliforms in drinking water. High COVID-19 mortality rates were linked to low sanitation and wastewater treatment coverage.
3	United States	Ahmad et al.	2020	Cumulative data up to March and April 2020	COVID-19 incidence and mortality	Percentage of dwellings in poor conditions (overcrowded, high cost, incomplete kitchen facilities, incomplete plumbing), comorbidities, ethnicity and race, gender, education, income.	Each 5% rise in the percentage of homes in poor conditions translated into a 50% higher risk of infection, and 42% growth in the COVID-19 mortality rate.
4	United States	Hyde	2021	January 2020 – September 2020	COVID-19 mortality	Water quality violations, gender, race, ethnicity, poverty rate, population density, population over 65 years old, use of public transportation, amount of people in the dwelling, education, air pollution.	Counties at higher risk of exposure to water pollutants (especially those that increase risk of cardiovascular disease) have a higher fatality rate than similar areas.
5	India	Das et al.	2020	May 2020	Confirmed COVID-19 cases	Poor dwelling conditions, water, sanitation, and hygiene services, gender disparity, low asset ownership.	Socio-economic deficiencies were linked to a significantly higher incidence of COVID-19 in the megacity of Chennai.
6	Mexico	Revollo-Fernandez et al.	2022	February 2020 – September 2021	COVID-19 mortality	Lack of access to water, poverty rate, social deprivation, vulnerable population according to income, male population, people over 60 years old, education, pneumonia mortality rate, diabetes mortality rate, hypertension mortality rate, cumulative number of COVID-19 infections, lack of access to health services, population density.	At a municipal level, COVID-19 mortality is associated with a higher percentage of people with no water, a higher percentage of people over 60 years old, and a higher mortality rate of diabetes.

Source: Prepared by the authors.

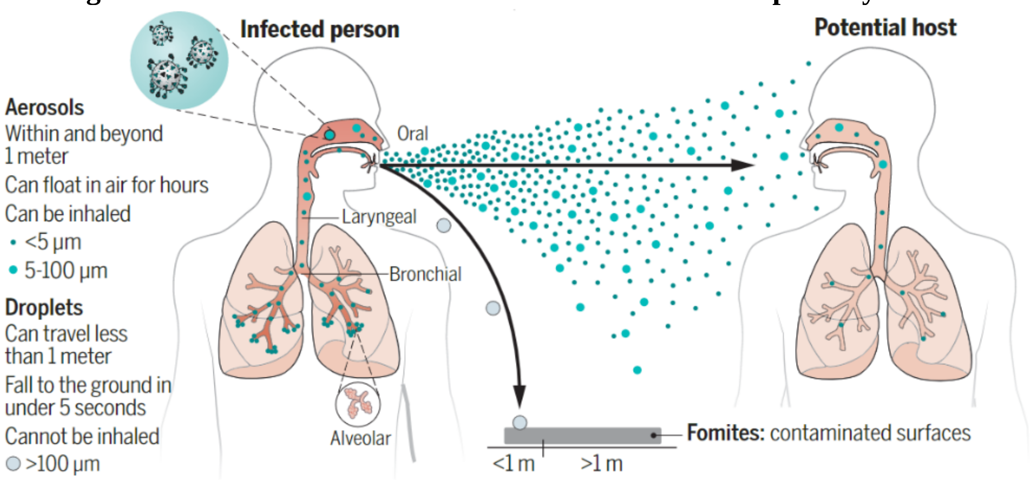
3. Context

3.1. COVID-19: transmission, risk, and prevention

COVID-19 is a communicable disease caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Patients may present symptoms (fever, cough, headaches, fatigue, difficulty breathing, loss of taste and smell), although a high percentage of patients present no symptoms.

The main form of transmission of SARS-CoV-2 is from the respiratory tract of an infected person to a potential host. When breathing, talking, or coughing, people who are infected exhale particles containing the virus that remain suspended in the air as aerosol particles. These can travel over a meter and remain suspended in the air for hours, facilitating their remote transmission (to over two meters), especially in rooms with little ventilation or little air exchange. The virus is also excreted in larger droplets (over 100 microns), which fall swiftly to the floor or surfaces, due to their weight, contributing to the potential transmission through fomites. Droplets can fall on hands, acting as fomites when people touch their face, nose, or mouth. Droplets reach a short range (less than a meter), so they only contribute to close contact transmissions between an infected patient and a potential host (see **Figure 1**). People with no symptoms or pre-symptomatic patients excrete the virus in aerosol particles and droplets inadvertently, contributing to spreading the virus even while having no symptoms.

Figure 1- Phases involved in airborne transmission of respiratory viruses



Phases involved in airborne transmission of respiratory viruses. Virus-laden aerosols (<100 μm) are first generated by an infected individual through expiratory activities, through which they are exhaled and transported in the environment. They may be inhaled by a potential host to initiate a new infection, provided that they remain infectious. In contrast to droplets (>100 μm), aerosols can linger in air for hours and travel beyond 1 to 2 m from the infected individual who exhales them, causing new infections at both short and long ranges.

Source: Wang et al. (2021).

Due to aerosol and droplet transmission, physical proximity and permanence in closed or poorly ventilated places for long periods of time increase the risk of transmission. Therefore, most countries agree in recommending social distancing, frequent hand washing, wearing facemasks, and ventilating closed rooms as primary prevention measures. The ability to follow these recommendations at home is highly limited to its socio-economic features (dwelling precariousness, crowdedness, informal employment, and lack of access to water and sanitation services, among others).

3.2. COVID-19 in Peru

The first case of COVID-19 in Peru was confirmed on March 6, 2020. Nine days later, Supreme Decree N. 044-2020-PCM declared a National State of Emergency on account of the dire consequences of the virus's transmission. The primary measure in the decree (subsequently extended in successive mandates) ordered social immobilization throughout the country except for essential activities such as the purchase of food and medicine. Other measures included the closing of the country's borders, banning transportation between regions, and cutting urban transportation by 50%. Every educational institution was closed, and public and private agencies adopted remote learning[12] and remote working as *modus operandi*. In addition, shops, including restaurants, were fully suspended from operations. Social gatherings and the use of public spaces were banned. Likewise, a curfew between 8 PM and 5 AM went into immediate effect. The armed forces and the police oversaw all of these restrictions[13]. To incentivize compliance with the restrictions, the government of Peru offered financial assistance to families whose income was affected[14]. However, implementation was difficult given the low percentage of the population owning bank accounts, high levels of informal employment, and deficiencies in the government's information systems. As such, the vulnerability of certain social groups to COVID-19 was exacerbated by shortcomings in the healthcare system and in targeting mechanisms of subsidies and transfers, as well as by the absence of assistance to migrants and refugees, among others (Vásquez-Rowe and Gandolfi, 2020). Particularly at the beginning of the pandemic—marked by restrictive measures of social isolation—socio-economic precariousness (informal employment paid in daily wages or the lack of a refrigerator at home) prevented a large share of the population from following

[12] To view the impact of COVID-19 on vulnerable higher education students in Peru, see the works of Elacqua et al. (2022).

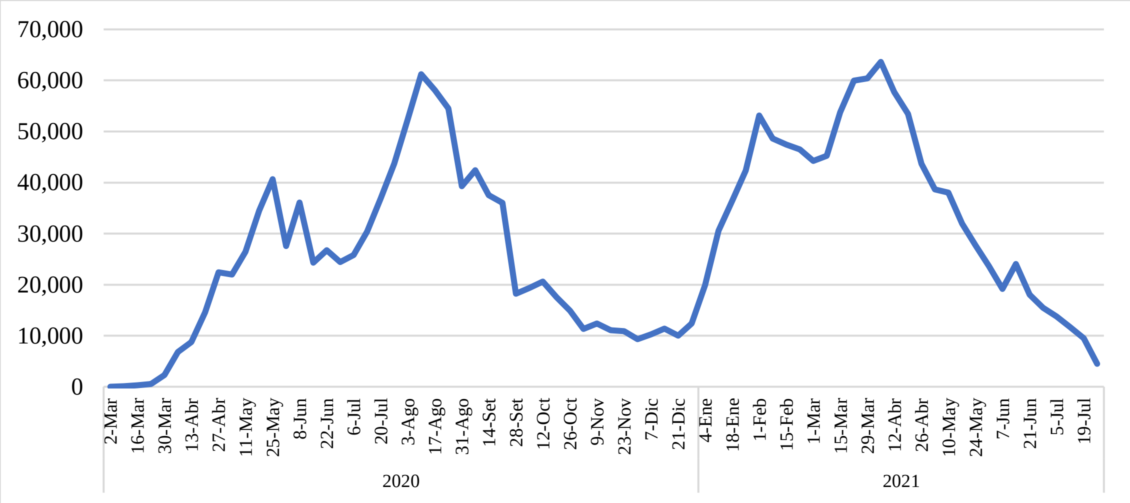
[13] The government of Peru temporarily implemented some other measures, such as the partial and total restriction of road traffic and the implementation of outings segregated by gender on different days, among others. Subsequent mandates progressively modified curfew hours, opened borders and shops, and permitted the reopening of public spaces, differentiating restrictions between regions.

[14] For an inventory and description of the main income-support programs at the onset of the COVID-19 pandemic in Latin America and the Caribbean, see Cejudo et al. (2021). According to Stampini et al. (2021), the *"Yo me quedo en casa", "Bono independiente", "Bono rural",* and *"Bono familiar universal"* benefits reached a joint coverage of 38% of the population.

isolation measures for long periods of time (Diaz-Cassou et al., 2020; Jaramillo and López, 2021). Thus, provisions implemented during the quarantine proved to be suboptimal (Brown et al., 2020).

The evolution in the number of cases during the period analyzed in this paper[15] was characterized by two transmission outbreaks, clearly noticeable in **Graph 1**[16]. During the first of these outbreaks, there was a particular generalized lack of diagnostic testing.

Graph 1- Number of new weekly COVID-19 infections in Peru



Source: Ministry of Health. Prepared by the authors.

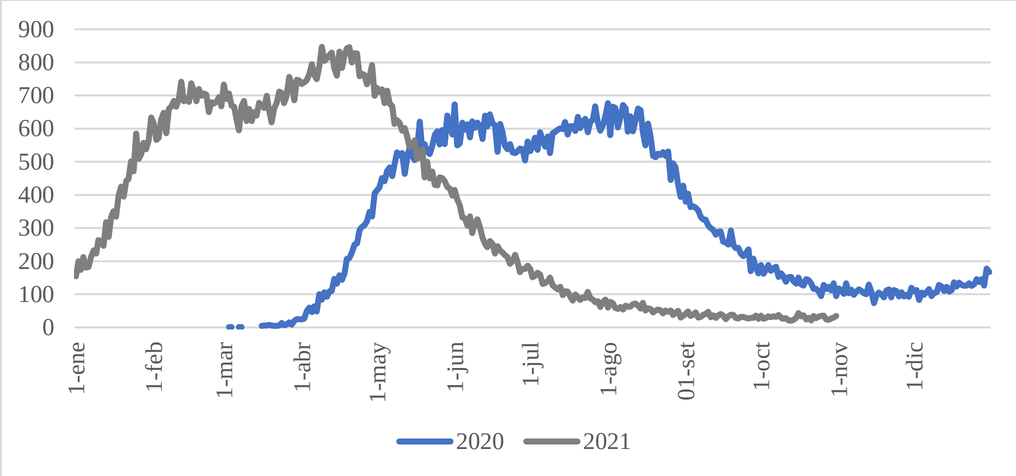
Uncertainty in the availability and frequency of diagnostic testing ultimately affects the total number of registed infections. Thus, total deaths from COVID-19 offers a better indicator of the virus's impact on the population and allows for more effective comparisons between countries. From March 2020 to October 2021, Peru recorded a total of 201,388 deaths from COVID-19. It is worth noting that these figures reflect improvements in the surveillance system of COVID-19 mortality in May 2021[17]. **Graph 2** presents the number of daily deaths from COVID-19, where both outbreaks are clearly marked.

[15] At the time the current document was drafted (early 2022), Peru was experiencing a third outbreak, with Omicron as the predominating variant, and a fatality rate considerably lower thanks to vaccination, and the variant's inherent lower fatality.

[16] To review dominating variants and their spatial variation, see Vargas-Herrera et al. (2022) and Romero et al. (2021).

[17] A temporary Technical Workgroup was created through Ministerial Resolution N. 095-2021-PCM to propose COVID-19 death toll updates criteria. The Technical Workgroup's final report can be found at https://cdn.www.gob.pe/uploads/document/file/1920118/Informe%20final%20del%20grupo%20de%20trabajo%20te%CC%81cnico%20con%20cifra%20de%20fallecidos%20por%20la%20COVID-19.pdf.pdf

Graph 2- Daily COVID-19 death toll in Peru



Source: National Open Data Platform. Prepared by the authors.

According to official country data, Peru ranks highest in the number of deaths from COVID-19 worldwide (618 per 100,000 inhabitants)[18]. Taking into consideration estimations of excess deaths—which help avoid underestimations in official figures—Peru also ranks highest in the number of deceased in Latin America (see **Table 3**)[19].

Table 3- Estimations of excess deaths in Latin America

Country	Excess deaths (per 100,000 inhabitants)	Difference with official records
Peru	640 - 670	+6%
Mexico	490-560	+100%
Bolivia	450-540	+200%
Honduras	240-530	+100%
El Salvador	180-500	+500%
Nicaragua	160-470	+12.000%
Argentina	320-470	+50%
Ecuador	390-430	+100%
Colombia	330-410	+50%
Brazil	330-380	+20%

Source: *The Economist* (2022).

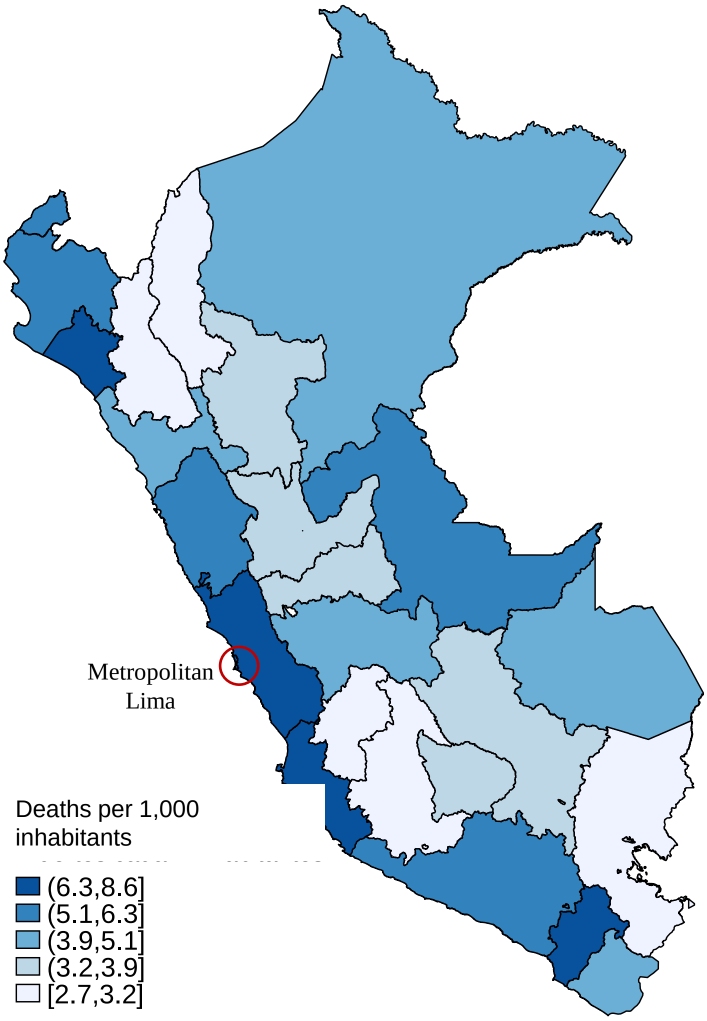
In terms of deaths from COVID-19 per 1,000 inhabitants at the departmental level in Peru (see **Figure 2**), the most affected departments are Callao (8.56), Ica (8.22), and Lima (8.10). This stresses the importance of conducting our study in Metropolitan Lima (which, as

[18] https://www.economist.com/graphic-detail/coronavirus-excess-deaths-tracker [checked on February 3, 2022]

[19] It is worth noting that, compared with other countries, Peru shows minimum discrepancies between the official number of deaths from COVID-19, and the indicator of excess deaths, because the government made an effort to disclose figures, improving the criteria to determine the number of deaths from COVID-19.

previously mentioned, is comprised by the province of Lima, and the constitutional province of Callao).

Figure 2- Deaths from COVID-19 every 1,000 inhabitants, by department



Note: Figures updated to 07-31-2021. Source: Ministry of Health. Prepared by the authors.

3.3. Current situation of access to water in Peru

As with other countries in Latin America and the Caribbean (LAC), Peru suffers from a significant infrastructure gap in terms of water and sanitation services. According to the National Infrastructure Plan for Competitiveness published by the Government of Peru (Ministry of Economy and Finance, 2019), gaps in short-term investment in water and sanitation are estimated at US$1.54 billion and US$7.39 billion[20], respectively. In the long term, investments of US$6.22 billion and US$18.35 billion will be necessary to reach access levels similar to those of developed

[20] Exchange rate employed: 3.9 PEN/USD.

countries[21]. In a more recent study, Brichetti et al. (2021) calculate that Peru must invest US$20.9 billion by 2030 to grant universal access to safely managed water and sanitation and to guarantee wastewater treatment in urban areas. This would allow the country to develop the main infrastructure components needed to comply with Sustainable Development Goal 6.

According to official information, in recent years there has been a slight increase in water coverage: in 2013, 86.1% of the population was served through the public water system, while in 2020, that percentage rose to 91.2% (see **Table 4**).

Table 4- Homes with access to piped water, 2013-2020 (%)

Sector	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**
National	**86.1**	**87.6**	**88.2**	**89.2**	**89.4**	**90.7**	**90.8**	**91.2**
Urban	93.4	93.6	93.9	94.5	94.4	95.3	94.9	94.8
Rural	63.2	68.3	69.5	71.2	72.2	74.4	75.6	77.6

Source: INEI (2021). Prepared by the authors.

According to the National Household Survey (NHS), by 2020, 90% of households nation-wide were connected to the public water system, greatly varying between geographic areas. As shown in **Table 5**, in the specific case of Metropolitan Lima, 5.7% did not have access to this service, resorting mostly to water supply from tanker trucks (4.6%). Considering that Metropolitan Lima concentrates nearly one third of the country's population, this means that over half a million people in this region lack access to drinking water.

[21] In the same sense, the gap estimated in 2030 by the Ministry of Housing, Construction and Sanitation is around US$25.64 billion (Ministry of Housing, Construction and Sanitation, 2021).

Table 5- Access to water services by type of supply and geographic domain in 2020 (%)

Source of water supply	Northern coast	Central coast	Southern coast	Northern hills	Central hills	Southern hills	Jungle	Met. Lima	Total
Tap water in dwelling	87.2	84.0	85.5	85.9	82.9	75.1	75.1	87.4	83.2
Tap water outside of dwelling (within building)	1.3	1.4	1.6	3.8	7.2	7.6	4.1	4.4	4.4
Public tap or standpipe	1.5	3.2	4.9	0.3	1.0	1.9	1.5	2.5	1.9
Tanker truck or similar	4.1	6.3	4.1	-	0.2	2.0	1.5	4.6	3.0
Borehole (groundwater)	1.4	1.5	1.2	0.5	0.4	6.2	3.6	0.4	1.8
Springs	0.1	0.1	0.1	5.3	4.3	2.3	2.6	0.0	1.6
Other	3.7	2.5	1.5	1.8	1.8	4.1	4.1	0.8	2.4
River, canal, lake, lagoon	0.9	0.9	1.1	2.4	2.1	0.9	7.5	-	1.6
Total	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**

Note: Metropolitan Lima includes the districts in the province of Lima, and the districts of the constitutional province of Callao. Source: 2020 NHS database. Prepared by the authors.

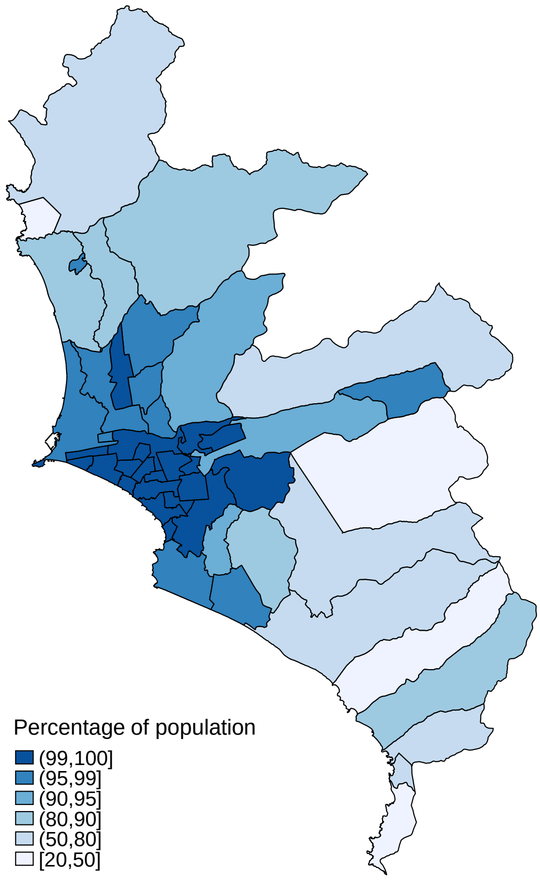
Homes that are not connected and are mainly supplied by tanker trucks ration their consumption due to the high costs. Also, because consumers are not guaranteed an uninterrupted supply, they must store water in containers. According to a survey conducted in Metropolitan Lima (Sunass, 2015), unconnected dwellings consume on average 4.8 cubic meters a month (a figure well below the average consumption of homes connected to the Sedapal's water supply network, which is around 15 cubic meters per month), and they pay on average 6 times more per cubic meter. In addition, this kind of water supply poses health risks due to the quality of the water received. In the context of COVID-19, water supply through tanker trucks could increase exposure to the virus because, to obtain water, people often must leave their homes, wait in line, or gather around a tanker truck, increasing close social interactions conducive to virus transmission. Additionally, access to water is strongly correlated with access to sanitation—a topic which will be discussed in further detail below. People who live in households that are not connected to the sewer network rely on systems like septic tanks, shared latrines, or could even practice open defecation.

In Metropolitan Lima, the company that provides water services is called *Agua Potable y Alcantarillado de Lima* (Sedapal). The company's coverage varies among the 43 districts of the province of Lima and the 7 districts of the constitutional province of Callao[22]. As **Figure 3** shows,

[22] Districts in the province of Lima: Ancon, Ate, Barranco, Breña, Carabayllo, Chaclacayo, Chorrillos, Cieneguilla, Comas, El Agustino, Independencia, Jesus Maria, La Molina, La Victoria, Lima, Lince, Los

the city's growth has led to major challenges in expanding drinking water coverage to the periphery. Our assessment mainly focuses on precisely those districts, where users with and without a connection to the water supply coexist. The goal of our study is to evaluate the association between the variable of lack of access to water with COVID-19 infection, controlling for other variables that could affect the results.

Figure 3 - Districts of Metropolitan Lima, by percentage of access to piped water



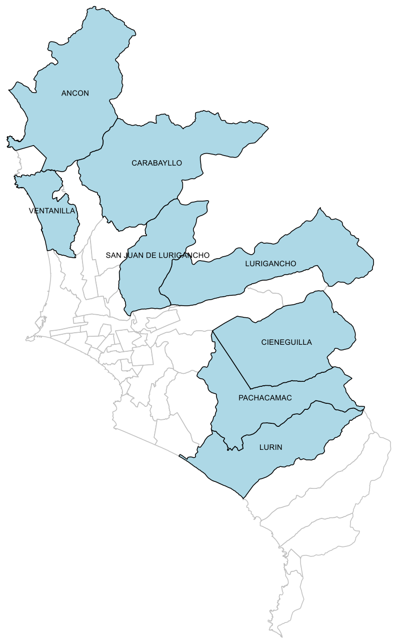
Source: INEI – 2017 Database Query of National Censuses.
Prepared by the authors.

Olivos, Lurigancho, Lurin, Magdalena Del Mar, Miraflores, Pachacamac, Pucusana, Pueblo Libre, Puente Piedra, Punta Hermosa, Punta Negra, Rimac, San Bartolo, San Borja, San Isidro, San Juan De Lurigancho, San Juan De Miraflores, San Luis, San Martin De Porres, San Miguel, Santa Anita, Santa Maria Del Mar, Santa Rosa, Santiago De Surco, Surquillo, Villa El Salvador, and Villa Maria Del Triunfo. Districts in the constitutional province of Callao: Bellavista, Callao, Carmen De La Legua Reynoso, La Perla, La Punta, Mi Peru, and Ventanilla.

4. Data specifications

Data was gathered from a survey specifically designed and administered for this study. Field research was conducted in person between October and November 2021 in peri-urban areas of the following districts of Metropolitan Lima: Ancon, Carabayllo, Cieneguilla, Lurigancho, Lurin, Pachacamac, San Juan de Lurigancho, and Ventanilla (see **Figure 4**). Information from a total 1,121 homes was gathered, and health-related questions were answered for 3,330 household members[23]. The sampling process was as follows: 634 homes with no access to water services were randomly selected (control group) in areas with an interrupted drinking water supply. Then, the closest houses with access to water were identified (treatment group). The survey's factsheet can be found in Annex 1.

Figure 4- Selected districts for the survey



Source: Prepared by the authors.

46.4% of the sample was connected to the public water system, while 53.6% relied on alternative supplies, mainly tanker trucks and public taps or standpipes (see **Table 6**). Although water extracted from public taps or standpipes originates from the water supply network, homes with this kind of access are not considered to be connected (treatment group) for the purposes of this study. As Howard, Bartram et al. (2020) argue, shared sources of water can be linked to a

[23] As will be further detailed, estimations are based on households located at most 2 kilometers away from either side of the boundary. Therefore, this section offers descriptive data from the 996 dwellings, and 3,015 individuals that fall under that criterion.

higher risk of virus transmission because of the difficulty in maintaining social isolation and social distancing, as well as the high number of people handling water faucets.

Table 6- Dwelling by water provision source in the sample (%)

	Supply source	%
Connected to public water system (Treatment group)	Piped water within the household	45.2
	Piped water that reaches outside the dwelling, but inside the building	1.3
Not connected to public water system (Control group)	Tanker truck	36.7
	Public taps or standpipes	9.6
	Requested from neighbor with access to piped water	4.3
	Groundwater borehole	2.7
	Others	0.3

Source: Survey conducted in peri-urban areas of Metropolitan Lima. Prepared by the authors, considering expansion factors.

Out of all household members in the sample, 28.7% were tested for COVID-19 (see **Table 7**).

Table 7- Percentage of members in a household tested for COVID-19 (%)

	Total	**Connected to public water system (Treatment group)**	**Not connected to public water system (Control group)**
Tested	28.7	30.1	27.5
Not tested	71.3	69.9	72.5

Source: Survey conducted in peri-urban areas of Metropolitan Lima. Prepared by the authors, considering expansion factors.

Among members who were tested, 34.2% were administered molecular tests, 33.9% rapid test (finger stick blood test), 27.2% antigen tests, and 4.7% the ELISA test (blood sample drawn from arm)[24].

[24] In order to help respondents with this question, they were shown visual aids (images) that included the main features of each type of test.

Table 8- Type of COVID-19 test taken (%)

	Total	Connected to public water system (Treatment group)	Not connected to public water system (Control group)
Molecular	34.2	34.7	33.8
Rapid	33.9	32.6	35.1
Antigen	27.2	28.0	26.4
ELISA	4.7	4.8	4.7

Note: Household members who were tested for COVID-19. Source: Survey conducted in peri-urban areas of Metropolitan Lima. Prepared by the authors, considering expansion factors.

According to our findings, the positivity rate was 33.6%, with a difference between respondents who were connected to the water supply network (treatment group) and respondents not connected to the water supply network (control group), as shown in **Table 9**. This difference of approximately ten percentage points in positivity between the two groups does not necessarily reflect the impact of access to water on the risk of infection. Although the control and the treatment groups should be otherwise similar, our methodology will help refine those estimations, considering possible differences found between the groups despite their geographic proximity.

Table 9- Results from COVID-19 tests

	Total	Connected to public water system	Not connected to public water system
Positive	33.6	28.0	38.9
Negative	66.4	72.0	61.1

Note: Household members who were tested for COVID-19. Source: Survey conducted in peri-urban areas of Metropolitan Lima. Prepared by the authors, considering expansion factors.

5. Methodology

This paper is a cross-sectional study with a quasi-experimental design and a comparator group, which aims to identify the relationship between access to piped water (independent, intervention, or treatment variable) and COVID-19 infection (result, dependent, or impact variable). The unit of analysis is the individual.

The study employs the Regression Discontinuity (RD) as its method. Studies with RD designs help determine the causal effect of an intervention or treatment, where a cut-off point or threshold has been established for a specific observable variable. This type of assessment is widely used in economics (Lee and Lemieux, 2010) and is increasingly recommended for epidemiological and public health research (Mosco et al., 2015) for its ability to quantify the effect of an intervention or treatment through comparison of similar observations at either side of the threshold. In a particular type of RD study, the discontinuity in treatment is geographic—often an administrative boundary that divides individuals into two groups (Keele and Titiunik, 2015).

Our study began by verifying the existence of marked geographic discontinuities in piped water provision in the peri-urban areas of Metropolitan Lima. These are clearly observable in national census maps. For example, the blue areas in **Figure 5** indicate city blocks with access to piped water, while those in red represent blocks that lack access in the district of Ventanilla, Callao. If we consider "access to piped water" as the treatment, the image shows a discontinuity in the treatment allocation, with a boundary that presumably reflects the extension of water distribution pipes. The red areas generally correspond to informal settlements, relatively close to blocks connected to the water supply network but are nonetheless located in hard-to-reach areas. These areas' topography typically prevents the water utility company from expanding the network there in the near future. It is furthermore not possible for those living in areas that lack access to move to other areas with access. This aspect is critical to the validity of our study's design—the allocation of treatment depends on households' geographic location, over which families have no influence.

Figure 5 – Geographic discontinuity in access to piped water (district of Ventanilla)

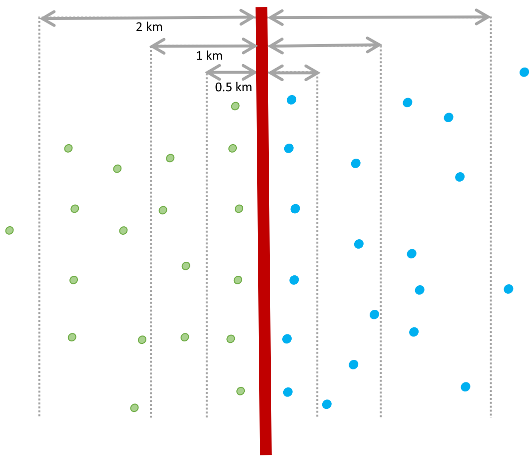
Source: 2017 National Censuses: Consultation System of Access to the Public Water System at block level, INEI.

For the purposes of this study, "geographic discontinuity" is defined as the boundary separating users with access to piped water from those who lack access[25] and was traced along equidistant points between both types of homes. Additionally, the assessment includes boundary or limit proximity bands, as shown in **Figure 6**. It is worth bearing in mind that estimates of the

[25] The pilot survey conducted prior to colleting the full sample's information allowed us to foresee that in some cases, homes have water and sanitation facilities, but are not served by Sedapal. They usually rely on communal solutions offering network services and water taps. To the effects of the current study, these observations are considered treatment observations, because they offer a relatively continued water supply, and do not require families to leave their homes to obtain water.

marginal effect of the treatment variable are more accurate for households in closer proximity to the geographic discontinuity. Moreover, most observations (approximately 82%) are located 0.5 km from the boundary at most.

Figure 6 – Defining geographic discontinuity



Households not connected to public water system
Households connected to public water system
Discontinuity

Source: Prepared by the authors.

Homes closer to the boundary share socio-economic indicators, except for their access to piped water. Therefore, the differences on the impact or result variable should theoretically be attributable to access to water. In practice, however, certain socio-economic indicators pertinent to this study may differ. As such, we have incorporated control variables.

This study utilizes an approach developed at Dell (2010) for calculating the estimation. It consists of adopting longitude and latitude variables as polynomial controls. Because the dependent variable is naturally dichotomous (1 if the person was infected with COVID-19, or 0 if they were not), we propose the following *logit* model for the main equation:

$$Pr[Y_i = 1] = \Phi(\beta + \beta T_i + \theta f(geographic\ location) + X'\gamma)$$

Where:

Y_i is a dichotomous variable equal to 1 if the individual "i" tested positive for COVID-19 between March 2020 and the date of data collection.

T_i is the dichotomous variable for treatment, indicating whether the individual "i" is connected to the public water system.

$f(\dots)$ is a second-degree polynomial, including each dwelling's latitude and longitude.

X' represents a vector of control variables.

Φ is the cumulative density function of standard normal distribution.

Because the dependent variable requires knowledge of respondents' history of COVID-19 infection, our sample is comprised of individuals who reported having been tested for COVID-19. Although this reduces the total sample, we consider this the most rigorous method to accurately account for respondents' history of COVID-19 infection[26]. In that vein, there are no statistically significant differences in the percentage of tested individuals between control and treatment observations (see median test in **Annex 2**). In other words, there were no meaningful differences in access to testing on either side of the geographic boundary of piped water supply [27]. In analyzing the main variables collected for this study (see median test in **Annex 3**), there were also no significant differences between individuals with access to testing and those without. Therefore, we do not consider that using the reduced sample (observations for tested individuals) would introduce a selection bias. Therefore our methodology is suitable to estimate the effect of the treatment (access to water) on the probability of COVID-19 infection[28].

6. Results

This section presents the results of our estimations derived from the aforementioned logit model. In addition, we include an extension of this model that incorporates heterogeneous effects related to the *per capita* endowment of water in households. A falsification or placebo test supplements our analysis to verify the estimations' validity. Clustered robust standard errors at the district level are included in parentheses.

[26] Other disregarded alternatives included making assumptions about the positivity of individuals who were not tested or using self-reported symptom information to infer positivity. Regarding the latter, it is worth keeping in mind that symptoms are neither a sufficient condition in themselves (symptoms may be similar to those of other respiratory diseases, for example) nor are they necessary (a large number of infected people may not have presented any symptoms).

[27] This can be attributed to a major deficit of testing, particularly at the beginning of the pandemic. Therefore, there were no significant differences in terms of access among different people. Subsequently, the Ministry of Health carried out screening campaigns in places like markets or transportation stations. In the following months, private labs offered testing, although their high costs could have been an access barrier to most (control and treatment) households considered in this study.

[28] As is discussed later in this paper, if the estimation was made considering a selection bias, the marginal effect would be higher. Nonetheless, the estimation would continue to be within the confidence interval of the preliminary results presented in this paper.

6.1. RD Logit Model: the role of water access

Table 10 presents the results of the estimation in four columns, each of which represents a range of distance to either side of the geographic boundary of piped water supply: 0.5, 1, 1.5, and 2 km, respectively. As previously mentioned, on average 82% of the observations are located 0.5 km or less away from the discontinuity, and the treatment variable's marginal effect can be better estimated in homes closer to the geographic discontinuity.

As evidenced in **Table 10**, the *treatment* variable is significant and its coefficient is negative. In other words, access to piped water is linked to a lower probability of reporting COVID-19 infection.

The *age* variable is also significant (for distances larger than 0.5 km) with a small coefficient that is positive, as expected. Although age itself does not demonstrably increase the risk of infection, this variable could signify greater exposure to the virus at places of employment, from activities like grocery shopping, and from the strict confinement of children and teenagers to the home, returning to face-to-face education only two years later. This variable could also reflect that the likelihood of respondents seeking testing increases with age.

The *first wave* variable –a dichotomous variable where 1 indicates that the person was tested between March and December 2020 and 0 indicates that they were tested later– is also significant. There are various possible explanations for this result. The first accounts for the limited availability of tests at the beginning of the pandemic in Peru. As such, testing prioritized patients reporting symptoms and/or close contact with infected individuals, making positive results more likely during the first wave. Other non-exclusive explanations include the more extreme restrictions of the first wave (increasing transmission through lock-downs), insufficient knowledge about virus transmission at the onset of the pandemic, and other aspects concerning variants of the virus predominant during the first wave.

The presence of *comorbidities*—defined as a dichotomous variable denoting conditions like obesity, diabetes, cardiac disease, lung disease, or a weakened immune system—is also significant, and the coefficient is positive, as expected. Comorbidities increase the severity of the disease, which could in turn increase the probability for patients with comorbidities to seek out testing.

Lastly, the *cash transfer* variable is significant and has a negative coefficient for the first three distance ranges. This dichotomous variable indicates whether the household obtained any benefits from the Peruvian Ministry of Development and Social Inclusion (MIDIS) such as *"Yo me Quedo en Casa"* or the *"Bono Familiar Universal"*. This would indicate that having received such benefits could reduce the risk of COVID-19 infection, presumably, because this financial assistance made social isolation at home more achievable, reducing the family members' exposure to COVID-19 infection.

Other variables included in the estimation—although not statistically relevant—are the indicator of *overcrowding* of the home (calculated as number of members per room), as well as dichotomous variables representing whether the *mother has completed secondary studies*, if the person who receives the primary income has a *face-to-face job in customer service*, if the family inhabits a *precarious house* (i.e. not made out of quality materials), and if the family has an average monthly *income higher than 1000 soles*. Although these variables were meaningful in much of the literature reviewed for this study, they likely increased the probability of transmission of SARS-CoV-2 only initially. Conversely, this paper's dependent variable includes cases of COVID-19 infection until October/November 2021—in other words, during the first *and* second waves.

Table 10- Logit Model Results

	<0.5 km	<1 km	<1.5 km	<2 km
	(1)	**(2)**	**(3)**	**(4)**
Treatment	-0.505***	-0.336*	-0.350**	-0.302*
	(0.1740)	(0.1890)	(0.1700)	(0.1590)
Age	0.004	0.010*	0.010*	0.009*
	(0.0060)	(0.0050)	(0.0050)	(0.0060)
Mother's incomplete	-0.265	-0.243	-0.248	-0.195
secondary studies	(0.1740)	(0.2150)	(0.2060)	(0.1970)
Face-to-face job	-0.091	-0.019	-0.062	-0.044
in customer service	(0.1750)	(0.1190)	(0.1290)	(0.1040)
Precarious house	0.081	0.213	0.195	0.306*
	(0.1930)	(0.2170)	(0.2040)	(0.1850)
Overcrowding	-0.063	-0.022	-0.015	-0.023
	(0.1740)	(0.1340)	(0.1280)	(0.1160)
Income >1000 soles	-0.454	-0.279	-0.219	-0.208
	(0.2830)	(0.2270)	(0.1830)	(0.1580)
First wave	0.915***	0.846***	0.900***	0.886***
	(0.1920)	(0.1060)	(0.1170)	(0.1210)
Comorbidities	0.754***	0.727***	0.752***	0.721***
	(0.1710)	(0.1900)	(0.2040)	(0.2230)
Cash transfer	-0.300*	-0.390**	-0.391**	-0.287
	(0.1730)	(0.1860)	(0.1880)	(0.2460)
Constant	-0.289	-0.830**	-0.833**	-0.910**
	(0.4620)	(0.3810)	(0.3830)	(0.3680)
Geo. controls	Yes	Yes	Yes	Yes
Observations	612	707	728	749
Log Likelihood	-362.645	-420.849	-430.603	-445.965
Akaike Inf. Crit.	765.289	881.698	901.205	931.929

Significance levels are: ***p<0.01, **p<0.05, *p<0.1

Note: Clustered robust standard errors at the district level are included in parentheses.
Source: Prepared by the authors.

Because this estimated model is a logit model, its coefficients cannot be directly interpreted, requiring the marginal effects to be calculated[29]. **Table 11** presents the marginal effects of the treatment variable, calculated for homes located at 0.5 km, 1 km, 1.5 km, and 2 km from the geographic discontinuity. According to the calculated marginal effects, access to piped water is associated with a reduction of between 6.2 and 9.9 percentage points in the probability

[29] If the estimations are made using a Linear Probability Model, the coefficient of the treatment variable is the same as the calculated marginal effect for the Logit Model.

of COVID-19 infection. Considering the level of positivity among residents not connected to the water supply network, it follows that the probability of COVID-19 infection could drop between 15% and 25% when the household is connected to the public water system[30].

Table 11- Marginal Effects of the Treatment Variable by Distance

Distance	AME	SE	z	p	lower	upper
0.5 km	-0.0999	0.0345	-2.8939	0.0038	-0.1676	-0.0323
1 km	-0.0646	0.0364	-1.7760	0.0757	-0.1359	0.0067
1.5 km	-0.673	0.0336	-2.0067	0.048	-0.1331	-0.0016
2 km	-0.0618	0.0329	-1.8805	0.0600	-0.1262	0.0026

Source: Prepared by the authors.

Annex 4 is a graphic representation of an individual's probability of COVID-19 infection according to their distance to the geographic discontinuity. The left panel shows control observations (not connected to the water supply network), while the right section displays treatment observations (connected to the water supply network). The graph clearly shows the difference in risk of COVID-19 infection[31].

6.2. RD Logit Model with heterogeneous effect by level of consumption: beyond access, understanding the importance of a minimum endowment

This section presents the results of an extension of the previously defined model—a model with heterogeneous effects, stratifying connected users by their level of consumption. This specification divides individuals who have access to piped water into two groups: those who consume less than 3 cubic meters *per capita* per month (or the equivalent of 150 liters per person per day)[32] and those who consume more than that[33].

---

[30] Annex 4 presents the estimation for a case of selection bias. The correction can be done by the methodology proposed by Heckman (1979). However, Greene (2006) points out that this methodology is not suitable when the main equation is not linear. For these cases, he proposes a general method to correct the problem of selection bias by maximizing the likelihood function through simulations.

This method consists of estimating two equations: the selection equation (which, in this case, calculates the probability of testing) and the main equation (which, in this case, estimates the probability of infection).

As can be observed, the marginal effect is lower, but it remains within the confidence interval of the primary results presented in this study.

[31] The graph presents second-degree polynomials at either side of the geographic boundary of piped water supply. This graph presents the probablity of COVID-19 infection solely according to distance. Therefore, attention should be focused on the "jump" at the cut-off point (geographic boundary).

[32] According to the World Health Organization (2017), to be considered "optimal access", the volume of water required is between 100 and 200 liters per person per day.

[33] To obtain the *per capita* consumption, the household's water consumption was calculated and divided by the number of residents in the household. For homes connected to the water supply network, the level

Table 12 presents the results for the estimation of the logit model with heterogeneous effects by level of water consumption. For simplicity, two ranges of distance were included (0.5 km and 1.5 km). For the low consumption group, control and treatment observations were compared where monthly consumption did not exceed $3m^3$. As can be observed, in this scenario the treatment variable was not significant. In the case of the high consumption group, the comparison takes place between control observations with a monthly *per capita* water consumption below $3m^3$ and treatment observations with a monthly *per capita* water consumption over $3m^3$. In this case, the coefficient associated to the treatment variable was significant. This would indicate that piped water connection is not enough in itself; a minimum supply of water must be guaranteed. This outcome could be explained by an unmet minimum amount of water required to comply with hygiene recommendations. Alternatively, homes connected to the water supply network (that receive poor-quality services) may need to supplement their supply with tanker trucks or public taps or standpipes, where they are exposed to a higher risk of COVID-19 infection. The marginal effects of this model, for the high consumption group, can be reviewed in **Table 13**.

of total consumption is infered from their monthly water and sanitation bill, considering the current tariff structure of Sedapal S.A. at the date of the field research. For homes not connected to the water supply network, total household consumption was calculted through questions included in the questionaire about recipients, content, and frequency of water storage.

Table 12- Results of the Logit Model with Heterogeneous Effects

	Distance <0,5 km		Distance <1,5 km	
	Low consumption (<3m³ *per capita*)	High consumption (> or = 3m³ *per capita*)	Low consumption (<3m³ *per capita*)	High consumption (> or = 3m³ *per capita*)
Treatment	-0.365	-0.754**	-0.14	-0.495*
	(0.4240)	(0.3110)	(0.3570)	(0.2590)
Age	0.005	0.003	0.011	0.010***
	(0.0090)	(0.0050)	(0.0080)	(0.0040)
Mother's incomplete secondary studies	-0.147	0.005	-0.25	-0.096
	(0.2150)	(0.2290)	(0.2060)	(0.2120)
Face-to-face job in customer service	0.369*	-0.018	0.241	0.06
	(0.2130)	(0.3420)	(0.1580)	(0.2190)
Precarious house	0.074	-0.325*	0.233	-0.11
	(0.2990)	(0.1910)	(0.2750)	(0.2160)
Overcrowding	-0.201	-0.116	-0.155	-0.067
	(0.2250)	(0.2010)	(0.1620)	(0.1420)
Income >1000 soles	-0.058	-0.582	0.044	-0.266
	(0.6660)	(0.4020)	(0.4410)	(0.2920)
First wave	0.687***	0.947***	0.766***	0.917***
	(0.1870)	(0.3250)	(0.1370)	(0.2030)
Comorbidities	0.763**	0.308	0.707***	0.332*
	(0.3660)	(0.2230)	(0.2360)	(0.1980)
Cash transfer	-0.054	-0.314	-0.146	-0.453**
	(0.3440)	(0.2870)	(0.3610)	(0.2260)
Constant	-0.398	0.029	-0.803*	-0.621
	(0.4480)	(0.6180)	(0.4480)	(0.4310)
Geo. controls	Yes	Yes	Yes	Yes
Observations	298	416	376	510
Log Likelihood	-180.26	-248.525	-227.233	-304.392
Akaike Inf. Crit.	400.52	537.05	494.466	648.785

Significance levels are: ***p<0.01, **p<0.05, *p<0,1

Note: The low water consumption group includes treatment and control homes with consumption below 3m³ *per capita*. The high water consumption group includes control homes that record consumption below 3m³ *per capita* and treatment homes with consumption greater than or equal to 3m³ *per capita*. Clustered robust standard errors at the district level are included in parentheses.

Table 13- Marginal Effect of the Logit Model with Heterogeneous Effects (high consumption)

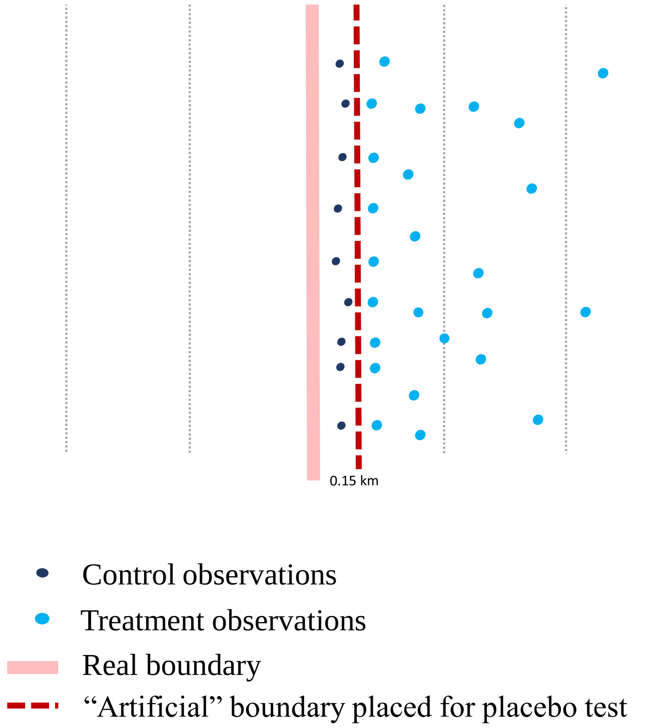
Distance	AME	SE	z	P	lower	upper
0.5 km	-0,1493	0.0618	-2.4160	0.0157	-0.2704	-0.0282
1.5 km	-0.1016	0.0538	-1.8871	0.0592	-0.2071	0.0039

Source: Prepared by the authors.

6.3. Falsification test

To verify the validity of our estimations, we conducted a falsification or placebo test by moving the boundary 0.15 km. For this falsification, we withdrew the sample's control (non-connected) households from the sample and set control observations 0.15 km further from the previous boundary that defined the discontinuity in piped water supply. We then set a new "artificial" boundary dividing the control and treatment homes at an equidistant point from each type of home (see **Figure 7**).

Figure 7 – Artificial boundary for placebo test



Source: Prepared by the authors.

By replicating the regression with these new control and treatment observations, we find that treatment (access to water) does not exhibit a significant coefficient (see **Table 14**). In other words, the null hypothesis, which states that the coefficient associated with the "artificial" treatment is not statistically different from zero, cannot be disregarded, supporting the validity of this paper's results.

Table 14- Logit Placebo Test Results (relocated boundary)

	<0.5 km (1)	<1 km (2)	<1.5 km (3)	<2 km (4)
Treatment	0.06	-0.02	0.028	0.005
	(0.4100)	(0.3350)	(0.3110)	(0.3100)
Age	0.01	0.009	0.008	0.008
	(0.0130)	(0.0110)	(0.0090)	(0.0090)
Mother's incomplete secondary studies	0.378	-0.029	-0.371	-0.342
	(0.4670)	(0.3720)	(0.3280)	(0.3260)
Face-to-face job in customer service	0.096	-0.17	-0.337	-0.313
	(0.4540)	(0.3510)	(0.3090)	(0.3070)
Precarious house	0.447	0.36	0.479	0.47
	(0.5290)	(0.4510)	(0.3920)	(0.3910)
Overcrowding	-0.553	-0.656*	-0.684**	-0.704**
	(0.4290)	(0.3400)	(0.3150)	(0.3130)
Income >1000 soles	1.868***	1.300***	1.009***	0.973***
	(0.4120)	(0.3170)	(0.2800)	(0.2780)
First wave	0.268	0.171	0.229	0.167
	(0.4380)	(0.3660)	(0.3400)	(0.3360)
Comorbidities	0.28	0.777**	0.742**	0.715**
	(0.5140)	(0.3920)	(0.3650)	(0.3640)
Cash transfer	-0.007	-0.06	-0.309	-0.316
	(0.4880)	(0.3910)	(0.3660)	(0.3640)
Constant	-2.091**	-1.292*	-0.703	-0.631
	(0.8620)	(0.6950)	(0.6080)	(0.6040)
Geo. controls	Yes	Yes	Yes	Yes
Observations	199	258	302	303
Log Likelihood	-94.859	-133.989	-163.821	-165.258
Akaike Inf. Crit.	229.719	307.978	367.642	370.516

Significance levels are: ***p<0.01, **p<0.05, *p<0.1

Note: Clustered robust standard errors at the district level are included in parentheses.
Source: Prepared by the authors.

7. Preliminary considerations to the discussion of results

This paper is a quasi-experimental study which demanded data collection with surveys where the individual is the unit of analysis. In this regard, we can confirm that, from a methodological perspective, this study's results go beyond an ecological report, making it possible to reach closer estimations of the effects of access to water on risk of infection. However, it is important to analyze certain limitations in this section, to explain how these limitations have been addressed, and their implications in interpreting the results.

A first group of limitations speaks to possible biases in the information used. Information bias is a systematic distortion of the phenomenon one is trying to measure with the collected information. In our study, data on every variable—including questions concerning COVID-19 testing and comorbidities—was collected through self-reporting and, therefore, influenced by respondents' memory or even pressure to provide what they consider to be socially acceptable answers. Meanwhile, our estimations consider that a single test could validly report whether a person had or had not been infected with COVID-19. However, people who tested negative could have been infected before or after having taken the test.

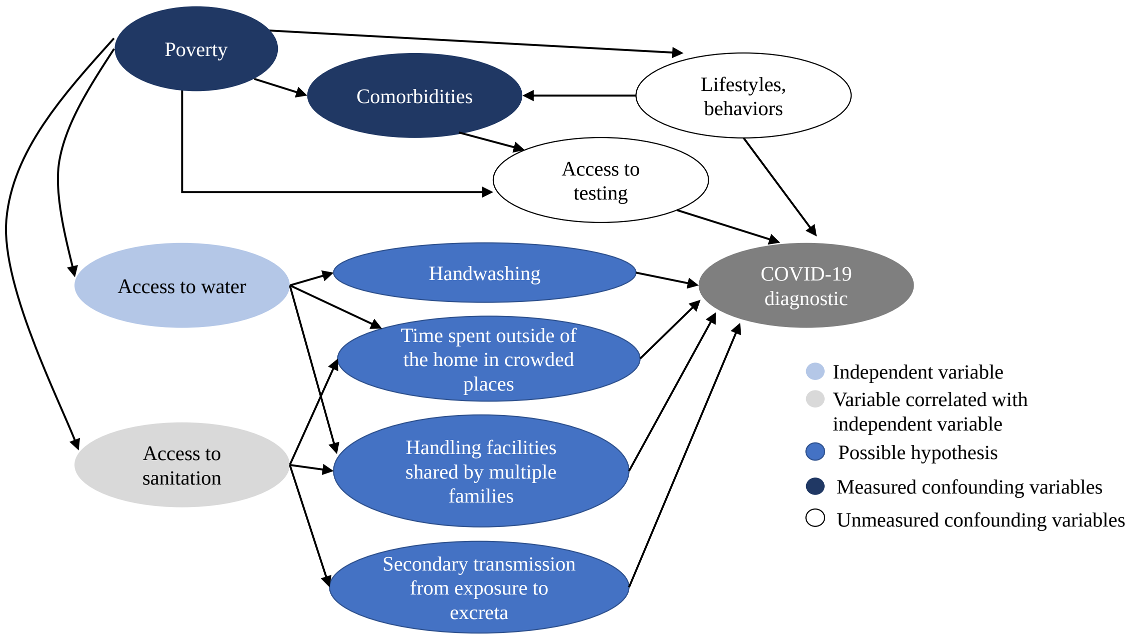
Another bias for consideration is selection bias—a systemic error by which individuals included in the sample differ from those not included, especially concerning variables of importance to the report. This paper includes means tests that revealed no significant differences in access to testing on either side of the geographic boundary of piped water supply. There are also no verifiable differences between individuals who were tested and those who were not. Given the absence of indicators of possible selection bias, we consider the estimation with the proposed methodology to be suitable. However, had we included corrections for selection bias in our estimation (see **Annex 4**), the estimation would remain within the confidence interval of the primary results presented in this study, albeit with a lower marginal effect.

A second group of limitations concerns confounding and its implications on inferring causality. First, the quasi-experimental design of this study took the *a priori* assumption that the control and treatment observations would be equivalent for a group of confounding variables, particularly in proximity to the geographic boundary of piped water supply. With this assumption, we aimed to emulating randomness. However, the fieldwork and subsequent results of the analyses of means have helped to identify differences in certain variables. For example, despite their geographic proximity, it is possible that dwellings lacking access to the water supply network also lack property deeds and are built on uneven ground. Despite their physical proximity, households also exhibited verified income disparities, among other differences. This has been partially addressed by including control variables in the regression.

A directed acyclic graph (DAG) is useful in explaining confounding. A DAG is a visual representation commonly employed in epidemiology, where causality assumptions are expressed to help identify the existence of confounding variables (Suttorp et al., 2015).

Socio-economic indicators (grouped under the concept of *Poverty* in **Figure 8**) are key confounding variables since they are linked both to access to water and COVID-19 infections. For example, people living in poverty tend to exhibit comorbidities contributing to increased COVID-19 severity, which in turn increases the probability of such patients seeking out testing. This is addressed through the comorbidities variable in the estimation model. Lifestyles and behaviors more commonly observed in people living in poverty may also increase the risk of infection. For example, diet can contribute to the prevalence of comorbidities, or individual behaviors like non-compliance with social isolation measures in pursuit of income can increase virus transmission. Structural risk factors (such as residential overcrowding) can further exacerbate these behavioral risks. This matter is addressed in this study by including socio-economic variables such as income or house overcrowding.

Figure 8 – Directed acyclic graph



Source: Prepared by the authors.

It is moreover necessary to clarify that there are multiple hypotheses to explain causation, though we find a significant statistical correlation between lack of access and COVID-19 infection. One hypothesis cites handwashing, in line with health authorities' recommendations, as the primary indicator in causation. However, this is less plausible according to epidemiological evidence, which shows that transmission is mostly airborne. A second hypothesis is that homes that are not connected to the public water system are forced into more frequent and closer social interactions. As previously mentioned, these people must often leave their homes, wait in lines, or gather near a tanker truck to obtain water, which increases the risk of COVID-19 infection. Thirdly, control households in this study include those that rely on public taps or standpipes, which increase risk of infection from time spent outside of the home in populated areas coupled with physical contact with facilities shared by multiple households.

In addition, access to water is strongly correlated with access to sanitation, as those who lack access to water share sanitation facilities with other homes and/or cannot safely dispose excreta. The SARS-CoV-2 virus has been detected in patient excreta (Amahmid et al., 2021) and its persistence for days in wastewater has also been documented (Godini et al, 2021), possibly accounting for a secondary form of transmission (Khumar Takur et al., 2021). Although the risk of infection from feces/urine to mouth/eye is low, risk could increase through person-to-person contact and in countries with poorly developed sanitation (Jones et al., 2020). Given that access to sanitation cannot be included in the regression to avoid multicollinearity issues, we cannot rule out that access to water is also picking up the aforementioned effects.

8. Discussion of results

This working paper contributes to the existing literature identifying the relationship between access to water and sanitation services and health indicators. In particular, the primary results of our study indicate that access to piped water was linked to a 6.2 to 9.9 percentage point drop in the probability of reporting COVID-19 infection. Considering the level of positivity among non-connected users, this would imply that the probability of COVID-19 infection decreases between 15% and 25% when the dwelling has access to piped water.

The results also suggest that it is not sufficient for the home only to possess a connection to the water supply network; a minimum supply must also be guaranteed. Our findings suggest that access to water is significantly ssociated to a lower probability of COVID-19 infection, as long as at least 3 monthly cubic meters of water *per capita* (or 150 liters per inhabitant per day) could be guaranteed. While service quality is demonstrably necessary to comply with hygiene recommendations, we also posit that households connected to the water supply network but with inadequate services must supplement their water supply with tanker trucks or public taps or standpipes, exposing them to a higher risk of COVID-19 infection.

We argue that a lack of access to basic services must be considered a public health issue. In addition to access to physical infrastructure, sufficient supply and affordability are essential to public health. Therefore, our findings stress the need for Peru and every other country in Latin America and the Caribbean to invest in infrastructure to close access gaps and implement policies to ensure minimum service consumption levels. Lack of access to piped water reduces people's quality of life, because, in addition to forcing them to resort to inadequate sources and to reduce their water consumption, lack of access also favors conditions that facilitate disease transmission particularly in the context of the COVID-19 pandemic.

9. Policy recommendations

Specialized papers point out that the high COVID-19 mortality rate in Peru is partially attributable to structural factors, such as a weak healthcare system, high levels of poverty, and deficiencies in basic services (Gianella et al., 2020; Schwalb and Seas, 2020). Our study observes that a lack in access to water services is linked to a higher risk of COVID-19 infection. In assessing alternative water sources for households lacking access to piped water, we identified different situations and behaviors that increase the risk of COVID-19 transmission. These results highlight the importance of boosting investments in the water and sanitation sector to achieve universal service coverage that meets safe management standards as established by the SDGs. This recommendation applies to Peru as well as to the rest of Latin America and the Caribbean, where there is still a significant gap in access to public services (Brichetti et al., 2021).

Our second policy recommendation stems from the results of the heterogeneous effects model. Although universal coverage is essential, it is also important to guarantee a minimum water supply to each dwelling. A household can be hindered or prevented from consuming water when there are significant deficiencies in service quality[34] or if the costs are too high. The absence of a minimum supply can compromise a family's hygiene habits or force them to resort to alternative water sources that, in the context of the pandemic, result in a higher risk of COVID-19 infection. Therefore, countries should implement operational initiatives and make investments aimed at improving the continuity and pressure of water services. Simultaneously, they should implement policies to guarantee their affordability for economically vulnerable households.

[34] When analyzing the information collected from the survey concerning satisfaction and continuity (in hours) of users connected to the public water system, both indicators increased with the distance from the geographic boundary of piped water supply.

References

Ahmad, K., Erqou, S., Shah, N., Nazir, U., Morrison, A. R., Choudhary, G., and Wu,W. C. (2020). *Association of poor housing conditions with COVID-19 incidence and mortality across US counties.* PLoS ONE, 15(11 November):1–13.

Amahmid, O., El Guamri, Y., Rakibi, Y., Ouizat, S., Yazidi, M., Razoki, B., Kaid Rassou, K., Asmama, S., Bouhoum, K., and D. Belghyti (2021). *Occurrence of SARS-CoV-2 in excreta, sewage, and environment: epidemiological significance and potential risks.* International Journal of Environmental Health Research. March 2021.

Amankwaa, Godfred, and Christian Fischer (2020). *Exploring the Correlation between COVID-19 Fatalities and Poor WASH (Water, Sanitation and Hygiene) Services*. MedRxiv N° June.

Brichetti, J.P; Mastronardi, L; Rivas Amiassorho, M.E., Serebrisky, T., and B. Solís (2021). *La brecha de infraestructura en América Latina y el Caribe. Estimación de las necesidades de inversión para progresar hacia el cumplimiento de los Objetivos de Desarrollo Sostenible al 2030.* Banco Interamericano de Desarrollo. Monografía 962.

Brotherhood, L., Cavalcanti, T., Da Mata, D., and Santos, C. (2020). *Slums and Pandemics.* Cambridge Working Papers in Economics: 2076.

Brown, Patrick, Greenwald, Zoë R., Salinas, Luis Ernesto, Aguirre Martens, Gabriel, Newcombe, Leslie, Gelband, Hellen, Veillard, Jeremy, and Prabhat Jha (2020). *Mortality from COVID in Colombia and Peru: Analyses of Mortality Data and Statistical Forecasts*. COVID-19 SARS-CoV-2 preprints from medRxiv and bioRxiv.

Cárdenas, Helena (2022). *Review on Urban Water & Sanitation on Three Topics: Subsidy Incidence, Price Elasticities, and Relations to Health: Framework and Research Paths for Future Studies*. Inter-American Development Bank. Technical note 2407.

Cejudo, Guillermo M., De los Cobos, Pablo, Michel, Cynthia L., and Diana L. Ramírez (2021). *Inventario y caracterización de los programas de apoyo al ingreso en América Latina y el Caribe frente a COVID-19*. Inter-American Development Bank. Technical note 2334.

Celik, Ismail, Lisa Gallicchio, Kristina Boyd, Tram K. Lam, Genevieve Matanoski, Xuguang Tao, Meredith Shiels, et al. (2008). *Arsenic in Drinking Water and Lung Cancer: A Systematic Review.* Environmental Research 108 (1): 48–55.

Chauvin, Juan Pablo (2021). *Why Does COVID-19 Affect Some Cities More than Others?: Evidence from the First Year of the Pandemic in Brazil*. Inter-American Development Bank. Working paper 1251.

Chauvin, Juan Pablo, Fowler, Annabelle, and Nicolás Herrera L (2020). *The Younger Age Profile of COVID-19 Deaths in Developing Countries.* Inter-American Development Bank. Working paper 1154.

Clouston, Sean A.P., Ginny Natale, and Bruce G. Link (2021). *Socioeconomic Inequalities in the Spread of Coronavirus-19 in the United States: A Examination of the Emergence of Social Inequalities*." Social Science and Medicine 268.

Darvesh, Nazia, Jai K. Das, Tyler Vaivada, Michelle F. Gaffey, Kumanan Rasanathan, and Zulfiqar A. Bhutta (2017). *Water, Sanitation and Hygiene Interventions for Acute Childhood Diarrhea: A Systematic Review to Provide Estimates for the Lives Saved Tool.* BMC Public Health 17 (Suppl 4).

Das, A., Ghosh, S., Das, K., Basu, T., Das, M., & Dutta, I. (2020). *Modeling the effect of area deprivation on COVID-19 incidences: a study of Chennai megacity, India*. Public Health, 185, 266-269.

Dasgupta, S., Bowen, V., Leidner, A., Fletcher, K., Musial, T., Rose, C., et al. (2020). *Association Between Social Vulnerability and a County's Risk for Becoming a COVID-19 Hotspot — United States, June 1–July 25, 2020*. US Department of Health and Human Services/Centers for Disease Control and Prevention. Morbidity and Mortality Weekly Report (MMWR). October 23, 2020 / Vol. 69 / No. 42.

De Groot, Jiska and Charlotte Lemanski (2021). *COVID-19 Responses: Infrastructure Inequality and Privileged Capacity to Transform Everyday Life in South Africa.* Environment and Urbanization 33 (1): 255–72.

Dell, M (2010). *The Persistent Effects of Peru's Mining Mita*. Econometrica, 78, no. 6, pp. 1863-1903.

Desmet, K. and Wacziarg, R. (2021). *JUE insight: Understanding Spatial Variation in COVID-19 across the United States*. Journal of Urban Economics.

Díaz-Cassou, Javier, Deza, María Cecilia, and Kenji Moreno (2020). *Peru: Desafíos del Desarrollo en el post COVID-19.* Inter-American Development Bank. Discussion paper 790.

Elacqua, G., López, K., Méndez, C., Novella, R., Prada, M. F., and D. Vásquez. *El impacto de la crisis del COVID-19 en estudiantes vulnerables de educación superior: El caso de Peru*. IDB Technical note 2449.

Fortaleza, C., Guimarães, R., De Almeida, G., Pronunciate, M., & Ferreira, C. (2020). *Taking the inner route: spatial and demographic factors affecting vulnerability to COVID-19 among 604 cities from inner São Paulo State, Brazil.* Epidemiology and Infection, 148, 1-5.

Freire de Souza, Carlos Dornels, Michael Ferreira Machado, and Rodrigo Feliciano Do Carmo. (2020). *Human Development, Social Vulnerability and COVID-19 in Brazil: A Study of the Social Determinants of Health.* Infectious Diseases of Poverty 9 (1): 1–11.

Friedrichs, M., Wagstaff, S., Sage, K., LaCross, N., Bui, D., et al. (2020). *Disparities in COVID-19 Incidence, Hospitalizations, and Testing, by Area-Level Deprivation — Utah, March 3–July 9, 2020*. Morbidity and Mortality Weekly Report, 69(38), 1369-1373

Gianella, C., Iguiñiz-Romero, R., Romero, M.J., and J. Gideon (2020). *Good Health Indicators are Not Enough: Lessons from COVID-19 in Peru*. Health and Human Rights Journal, 2020. 22(2): p. 317-319

Godini, H., Hoseinzadeh, E., and H. Hossini. *Water and wastewater as potential sources of SARS-CoV-2 transmission: a systematic review*. Reviews on Environmental Health, vol. 36, no. 3, 2021, pp. 309-317.

Greene, W. (2006). *A General Approach to Incorporating Selectivity in a Model*. Department of Economics, Stern School of Business, New York University.

Harris, Jeffrey (2020). *The subways seeded the massive Coronavirus epidemic in New York City*. NBER Working Paper Series.

Hawkins, D. (2020). *Social Determinants of COVID-19 in Massachusetts, United States: An Ecological Study*. Journal of Preventive Medicine & Public Health, 53(4), 220-227.

Heckman, J. (1979). *Sample Selection Bias as a Specification Error*. Econometrica, 47, pp. 153-161.

Howard, Guy, Bartram, Jamie et al. (2020). *COVID-19: Urgent Actions, Critical Reflections and Future Relevance of 'WaSH': Lessons for the Current and Future Pandemics*. Journal of Water and Health 18 (5): 613–30.

Hyde, K. (2020). *Residential Water Quality and the Spread of COVID-19 in the United States.* Social Science Research Network Web site.

Instituto Nacional de Estadística e Informática - INEI (2021). *Peru: Línea Base de los Principales Indicadores Disponibles de los Objetivos de Desarrollo Sostenible ODS 2021*.

Jaramillo, Miguel and Kristian López (2021). *Políticas para combatir la pandemia de COVID-19*. Grupo de Análisis para el Desarrollo - GRADE. Research paper 112.

Jones, D.L., Baluja, M.Q., Graham, D.W., Corbishley, A., McDonald, J.E., Malham, S.K., Hillary, L.S., Connor, T.R., Gaze, W.H., Moura, I.B., Wilcox, M.H. y K. Farkas. (2020). S*hedding of SARS-CoV-2 in feces and urine and its potential role in person-to-person transmission and the environment-based spread of COVID-19*. Science of Total Environment 749.

Kamis, Christina, Allison Stolte, Jessica S. West, Samuel H. Fishman, Taylor Brown, Tyson Brown, and Heather R. Farmer (2021). *Overcrowding and COVID-19 Mortality across U.S. Counties: Are Disparities Growing over Time?* SSM - Population Health 15: 100845.

Keele, Luke and Rocío Titiunik (2015). *Geographic Boundaries as Regression Discontinuities*. Political Analysis (2015) 23:127–155.

Kumar Thakur, A., Sathyamurthy, R., Velraj, R., Lynch, I., Saidur, R., Pandey, A.K., Sharshir, S.W., Elnaby Kabeel, A., Hwang, J., and P. GaneshKumar, (2021) *Secondary transmission of SARS-CoV-2 through wastewater: Concerns and tactics for treatment to effectively control the pandemic*. Journal of Environmental Management, Volume 290, 2021.

Lee, David S. and Thomas Lemieux. *Regression Discontinuity Designs in Economics.* Journal of Economic Literature. Vol. 48, No. 2, June 2010 (pp. 281-355)

Levin, A. T., Cochran, K. B., and Walsh, S. P. (2020). *Assessing the age specificity of infection fatality rates for COVID-19: Meta-analysis & public policy implications*. NBER Working Paper 27597.

Marí-Dell'Olmo, M., Gotsens, M., Pasarín, I., Rodríguez-Sanz, M., Artazcoz, L., García de Olalla, P. et al.. (2021). *Socioeconomic Inequalities in COVID-19 in a European Urban Area: Two Waves, Two Patterns*. International Journalof Environmental Research and Public Health, 18(3), 1256.

Ministry of Economy and Finance - MEF (2019). *Plan Nacional de Infraestructura para la Competitividad*.

Ministry of Housing, Construction and Sanitation (2021). *Plan Nacional de Saneamiento 2022-2026*.

Moscoe, Ellen, Bor, Jacob, and Till Bärnighausen. *Regression discontinuity designs are underutilized in medicine, epidemiology, and public health: a review of current and best practice*. Journal of Clinical Epidemiology 68 (2015) 132 – 143.

Nicodemo, C., Barzin, S., Cavalli, N., Lasserson, D., Moscone, F., Redding, S., & Shaikh, M. (2020). *Measuring geographical disparities in England at the time of COVID-19: results using a composite indicator of population vulnerability*. BMJ Open, 10(9), 1-8.

Niedzwiedz, C., O'Donnell, C., Dinesh, B., Demou, E., Ho, F., Celis-Morales, C., et al. (2020). *Ethnic and socioeconomic differences in SARS-CoV-2 infection: prospective cohort study using UK Biobank*. BMC Medicine, 18(160), 1-14.

World Health Organization (2017). *Guías para la calidad del agua de consumo humano: Cuarta edición que incorpora la primera adenda*.

World Health Organization (2020). *Handwashing an effective tool to prevent COVID-19, other diseases.* https://www.who.int/southeastasia/news/detail/15-10-2020-handwashing-an-effective-tool-to-prevent-covid-19-other-diseases [consulted on February 3, 2022]

World Health Organization (2022). *Tracking SARS-CoV-2 variants.* https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/ [consulted on February 9, 2022]

Revollo-Fernández, D., Rodríguez-Tapia, L., Medina-Rivas, C., and J.A. Morales-Novelo (2022). *Socioeconomic determinants of COVID-19 in Mexico.* Public Health.

Romero, P., Dávila-Barclay, A., Salvatierra, S., González, L., Cuicapuza, D., Solís, L., Marcos-Carbajal, P., Huancachoque, K., Maturrano, L., and P. Tsukayama (2021). *The Emergence of Sars-CoV-2 Variant Lambda (C.37) in South America*. Microbiology Spectrum. Volume 9 Issue 2.

Rozenfeld, Y., Beam, J., Maier, H., Haggerson, W., Boudreaud, K., Carlson, J., and Medows, R. (2020). *A model of disparities: risk factors associated with COVID-19 infection.* International Journal for Equity in Health.

Scarpone, C., Brinkmann, S., Große, T., Sonnenwald, D., Fuchs, M., and Byron, B. (2020). *A multimethod approach for county-scale geospatial analysis of emerging infectious diseases: a cross-sectional case study of COVID-19 incidence in Germany*. International Journal of Health Geographics, 19(32), 1-17.

Schwalb, A. and C. Seas (2021). *The COVID-19 Pandemic in Peru: What Went Wrong?* The American Journal of Tropical Medicine and Hygiene. 2021 Apr; 104(4): 1176–1178

Sheng, J., Malani, A., Goel, A., and Botla, P. (2021). *JUE Insight: Does Mobility Explain Why Slums Were Hit Harder by COVID-19 in Mumbai, India?* Journal of Urban Economics, page 103357.

Silva, R., Ribeiro, C., Moura, T., Santos, M., Santos, A., Tavares, D., & Santos, P. (2021). *Basic sanitation: a new indicator for the spread of COVID-19?* Transactions of The Royal Society of Tropical Medicine and Hygiene, 15(7), 832-840.

Stampini, Marco, Ibarrarán, Pablo, Rivas, Carolina, and Marcos Robles (2021). *Adaptive, but Not by Design: Cash Transfers in Latin America and the Caribbean Before, During and After the COVID-19 Pandemic*. Inter-American Development Bank. Technical note 2346.

Stokes, Andrew C., Dielle J. Lundberg, Irma T. Elo, Katherine Hempstead, Jacob Bor, and Samuel H. Preston (2021). *COVID-19 and Excess Mortality in the United States: A County-Level Analysis.* PLoS Medicine 18 (5): 1–19.

Strully, K., Yang, T.-C., & Liu, H. (2021). *Regional variation in COVID-19 disparities: connections with immigrant and Latinx communities in U.S. counties*. Annals of Epidemiology, 53, 56-62.

Sunass - National Superintendence of Sanitation Services (2015). *Comprar agua por camión cisterna les cuesta a las familias limeñas 72 soles mensuales*. Press release 161 – 2015.

Suttorp, M, Siegerink, B., Jager, K.J., Zoccali, C., and F. W. Dekker (2015). *Graphical presentation of cofounding in directed acyclic graphs*. Nephrol Dial Transplant (2015) 30: 1418-1423.

Taylor, Dawn L., Tanya M. Kahawita, Sandy Cairncross, and Jeroen H.J. Ensink. 2015. *The Impact of Water, Sanitation and Hygiene Interventions to Control Cholera: A Systematic Review*. PLoS ONE 10 (8): 1–19.

The Economist (2022). *The pandemic's true death toll. Our daily estimate of excess deaths around the world.* https://www-economist-com.access.idm.oclc.org/graphic-detail/coronavirus-excess-deaths-estimates [consultado el 3 de febrero de 2022]

Vargas-Herrera, N., Araujo-Castillo, R.V., Mestanza, O., Galarza, M., Rojas-Serrano, N. y L. Solari-Zerpa (2022). *SARS-CoV-2 Lambda and Gamma variants competition in Peru, a country with high seroprevalence*. The Lancet Regional Health – Americas, Volume 6, February 2022.

Vázquez-Rowe, I., and A. Gandolfi (2020). *Peruvian efforts to contain COVID-19 fail to protect vulnerable population groups*. Public Health in Practice 1 - 2020.

Waddington, Hugh, and Birte Snilstveit (2009). *Effectiveness and Sustainability of Water, Sanitation, and Hygiene Interventions in Combating Diarrhoea." Journal of Development Effectiveness 1 (3): 295–335.*

Wang, C.; Prather, K.; Sznitman, J.; Jimenez, J.; Lakdawala, S.; Tufekci, Z., and L. Marr. (2021). *Airborne transmission of respiratory viruses. Science 373, 981.*

Wolf, Jennyfer, Paul R. Hunter, Matthew C. Freeman, Oliver Cumming, Thomas Clasen, Jamie Bartram, Julian P.T. Higgins, et al. (*2018) "Impact of Drinking Water, Sanitation and Handwashing with Soap on Childhood Diarrhoeal Disease: Updated Meta-Analysis and Meta-Regression."* Tropical Medicine and International Health 23 (5): 508–25.

Xu, Jay J, Jarvis T Chen, Thomas R Belin, Ronald S Brookmeyer, Marc A Suchard, Christina M Ramirez, and Jay J Xu (2021). *National Population-Level Disparities in COVID-19 Mortality Across the Intersection of Race / Ethnicity and Sex in the United States.* MedRxiv, 1–18.

You, H., Wu, X., & Guo, X. (2020). *Distribution of COVID-19 Morbidity Rate in Association with Social and Economic Factors in Wuhan, China: Implications for Urban Development.* International Journal of Environmental Research and Public Health, 17(10), 1-14.

Zhang, H., Liu, Y., Chen, F., Baibing, M., Zeng, L., & Pei, L. (2021). *The effect of sociodemographic factors on COVID-19 incidence of 342 cities in China: a geographically weighted regression model analysis*. BMC Infectious Diseases, 21(428), 1-8

Annex 1 – Survey Factsheet

a) **Target population**

The study population is defined as the collection of every dwelling with and without access to drinking water services located in the peri-urban areas of the districts of Ancon, Carabayllo, Lurigancho, Cieneguilla, Pachacamac, Lurin, San Juan de Lurigancho, and Ventanilla, up to 2 km. away from the geographic boundary of piped water supply.

Districts were selected based on the results of the 12th National Population Census, the 7th National Housing Census, and the 3rd Census of Indigenous Communities conducted in 2017. Eight districts with the lowest access to drinking water were selected, excluding the districts in the southern coastline of Metropolitan Lima, covered by the scope of the Provisur project. Coverage levels could have suffered significant variations because of the Provisur project, and seasonal home occupation in those districts could hamper field research.

b) **Analyzed districts and total sample size**

<table>
<tr><th rowspan="2">Provinces</th><th rowspan="2">Districts</th><th colspan="2">Access to the public water system</th><th rowspan="2">Overall total</th></tr>
<tr><th>Without access</th><th>With access</th></tr>
<tr><td rowspan="7">Lima</td><td>Ancon</td><td>30</td><td>10</td><td>40</td></tr>
<tr><td>Carabayllo</td><td>80</td><td>61</td><td>141</td></tr>
<tr><td>Cieneguilla</td><td>20</td><td>10</td><td>30</td></tr>
<tr><td>Lurigancho</td><td>136</td><td>80</td><td>216</td></tr>
<tr><td>Lurin</td><td>57</td><td>30</td><td>87</td></tr>
<tr><td>Pachacamac</td><td>86</td><td>60</td><td>146</td></tr>
<tr><td>San Juan de Lurigancho</td><td>174</td><td>181</td><td>355</td></tr>
<tr><td>Callao</td><td>Ventanilla</td><td>51</td><td>55</td><td>106</td></tr>
<tr><td colspan="2">Total</td><td>634</td><td>487</td><td>1,121</td></tr>
</table>

Note: Includes some observations which where further than 2 km from the geographic boundary of piped water supply, which is why they were disregarded for the study.

c) **Margin of error and reliability:**

	Without access	With access
Margin of error	+/- 4.2%	+/- 4.3%
Reliability	95%	95%

d) **Field research timeline:** The field research was conducted by the company Impacto Directo / Directo Marketing between October 12 and November 5, 2021.

e) **Sample:** The study's regression discontinuity design required sampling homes near a geographic discontinuity. This has been defined as a boundary between dwellings that have access to piped water and those that do not. In other words, the sample did not have to be representative of all the homes in the analyzed districts. On the contrary, the sample should enable appropriate comparison between homes on either side of the boundary.

The sampling framework was designed based on the results of the 12th National Population Census, the 7th National Housing Census, and the 3rd Census of Indigenous Communities conducted in 2017. Blueprints were used to help identify conglomerations of city blocks with discontinuities in provision of drinking water services. Given how dated the information was, the boundary could have shifted (i.e. drinking water coverage could have increased). Thus, the field research validated and updated the sampling framework of users not connected to the public water system.

We developed a stratified multi-stage probability sample and a systematic final selection sample for homes that lacked access to piped water. We applied a criterion of proximity to households with no connection to homes that did have piped water, logging the dwellings' geolocations in both cases.

Selection of household procedure:

Stage	Sampling unit	Selection method
1	Population center	Sample proportionate to size
2	Blocks (urban), and compact segments (rural)	Sample proportionate to size
3	Individual homes	Systematic selection with random start

- The Primary Sampling Unit (PSU) is comprised of each district's population center, called conglomerates, which encompass the blocks.
- The Secondary Sampling Unit (SSU) is the peri-urban area of the population center made up of the blocks.
- The Tertiary Sampling Unit (TSU) is each individual household.

Annex 2 –Means test for the proportion of individuals who were administered a COVID-19 test

Variable	0.5 km from the boundary			1 km from the boundary			1.5 km from the boundary		
	Control	Treat.	P value t-test	Control	Treat.	P value t-test	Control	Treat.	P value t-test
	(1)	**(2)**	**(1) = (2)**	**(1)**	**(2)**	**(1) = (2)**	**(1)**	**(2)**	**(1) = (2)**
Proportion of individuals who were administered a COVID-19 test	0.259	0.300	0.191	0.270	0.292	0.456	0.269	0.290	0.432
	(0.017)	(0.017)		(0.015)	(0.019)		(0.014)	(0.018)	

Note: Clustered robust standard errors at bthe district level are included in parentheses.

Annex 3 –Means Tests Results (tested vs. not tested)

	0.5 km from the boundary			1 km from the boundary			1.5 km from the boundary			2 km from the boundary		
	Not tested	Tested	P val for t-test	Not tested	Tested	P val for t-test	Not tested	Tested	P val for t-test	Not tested	Tested	P val for t-test
	(1)	**(2)**	(1) = (2)	**(1)**	**(2)**	(1) = (2)	**(1)**	**(2)**	(1) = (2)	**(1)**	**(2)**	(1) = (2)
Overcrowding ratio	2.107	2.311	0.221	2.088	2.311	0.214	2.082	2.314	0.191	2.072	2.341	0.132
	(0.126)	(0.065)		(0.128)	(0.057)		(0.131)	(0.054)		(0.131)	(0.045)	
Precarious house	0.475	0.423	0.141	0.477	0.444	0.275	0.476	0.445	0.279	0.483	0.449	0.211
	(0.037)	(0.038)		(0.039)	(0.040)		(0.038)	(0.038)		(0.033)	(0.033)	
MIDIS Cash Transfer	0.228	0.227	0.992	0.224	0.227	0.956	0.224	0.230	0.909	0.220	0.235	0.809
	(0.023)	(0.041)		(0.021)	(0.034)		(0.021)	(0.037)		(0.021)	(0.039)	
Universal Cash Transfer	0.185	0.197	0.754	0.178	0.189	0.725	0.177	0.189	0.700	0.174	0.187	0.621
	(0.036)	(0.016)		(0.038)	(0.017)		(0.038)	(0.017)		(0.037)	(0.020)	
Total od received cash transfers	0.497	0.510	0.816	0.487	0.496	0.886	0.485	0.501	0.782	0.475	0.500	0.655
	(0.048)	(0.041)		(0.051)	(0.045)		(0.049)	(0.049)		(0.049)	(0.048)	
Comorbidities	0.262	0.357	0.141	0.248	0.347	0.081	0.243	0.343	0.068	0.245	0.339	0.071
	(0.028)	(0.035)		(0.023)	(0.033)		(0.024)	(0.031)		(0.023)	(0.030)	
Face-to-face job in customer service	0.690	0.613	0.094	0.688	0.602	0.090	0.692	0.606	0.098	0.693	0.612	0.101
	(0.048)	(0.036)		(0.057)	(0.029)		(0.055)	(0.024)		(0.053)	(0.023)	
Income < 750 soles	0.334	0.314	0.513	0.320	0.325	0.843	0.329	0.325	0.890	0.340	0.330	0.727
	(0.059)	(0.033)		(0.050)	(0.031)		(0.051)	(0.029)		(0.051)	(0.028)	
Income between 750 and 1000 soles	0.315	0.317	0.974	0.317	0.304	0.777	0.311	0.311	0.999	0.307	0.311	0.941
	(0.054)	(0.051)		(0.057)	(0.049)		(0.051)	(0.050)		(0.051)	(0.047)	
Income > 1000 soles	0.351	0.369	0.671	0.362	0.371	0.850	0.360	0.365	0.928	0.352	0.359	0.871
	(0.090)	(0.077)		(0.083)	(0.069)		(0.083)	(0.068)		(0.079)	(0.065)	
Household members under 17 years old	1.152	1.323	0.513	1.102	1.347	0.309	1.106	1.361	0.262	1.088	1.381	0.189
	(0.257)	(0.080)		(0.240)	(0.081)		(0.230)	(0.081)		(0.230)	(0.082)	
Household members over 65 years old	0.145	0.246	0.085	0.139	0.244	0.058	0.157	0.241	0.122	0.152	0.242	0.118
	(0.026)	(0.036)		(0.023)	(0.032)		(0.026)	(0.032)		(0.027)	(0.032)	
Mother secondary studies' completion	0.661	0.684	0.741	0.654	0.688	0.643	0.656	0.680	0.760	0.661	0.667	0.927
	(0.056)	(0.032)		(0.054)	(0.031)		(0.055)	(0.033)		(0.053)	(0.032)	
Access to drinking water services	0.486	0.498	0.781	0.466	0.463	0.943	0.463	0.459	0.918	0.449	0.443	0.857
	(0.061)	(0.040)		(0.073)	(0.052)		(0.072)	(0.054)		(0.065)	(0.052)	

Note: Clustered robust standard errors at the district level are included in parentheses.

Annex 4 – Model correction for potential selection bias

Following the methodology proposed by Greene (2006) for non-linear main equations, this model was estimated using probit models both for the selection equation (which, in this case, estimates the probability of being tested) and the main equation (which, in this case, calculates the probability of COVID-19 infection).

Apart from *treatment*, the following variables were included in the main equation: *overcrowding*, *use of public transportation by the head of household, income of at least 1000 soles*, and the existence of *comorbidities*. For the selection equation, in addition to the variables considered in the main equation, we include indicators such as *symptoms*, *age*, employment involving *face-to-face job in customer service*, and if the individual work as a *dependent employee*. The marginal effects and results of the estimation are presented below. That being said, it is worth mentioning that the estimation includes variables that could differ from those included in this paper's primary results. We included variables that ensure the estimations converge, considering that they were conducted by maximizing the likelihood function through simulations.

Marginal Effects of the Treatment Variable by Distance

Distance	AME	SE	z	p	lower	upper
0.5 km	-0.043	0.017	-2.490	0.013	-0.076	-0.009
1 km	-0.036	0.013	-2.800	0.005	-0.061	-0.011
1.5 km	-0.036	0.013	-2.790	0.005	-0.061	-0.011
2 km	-0.038	0.012	-3.140	0.002	-0.062	-0.014

Source: Prepared by the authors.

Estimation Results (Heckman Probit)

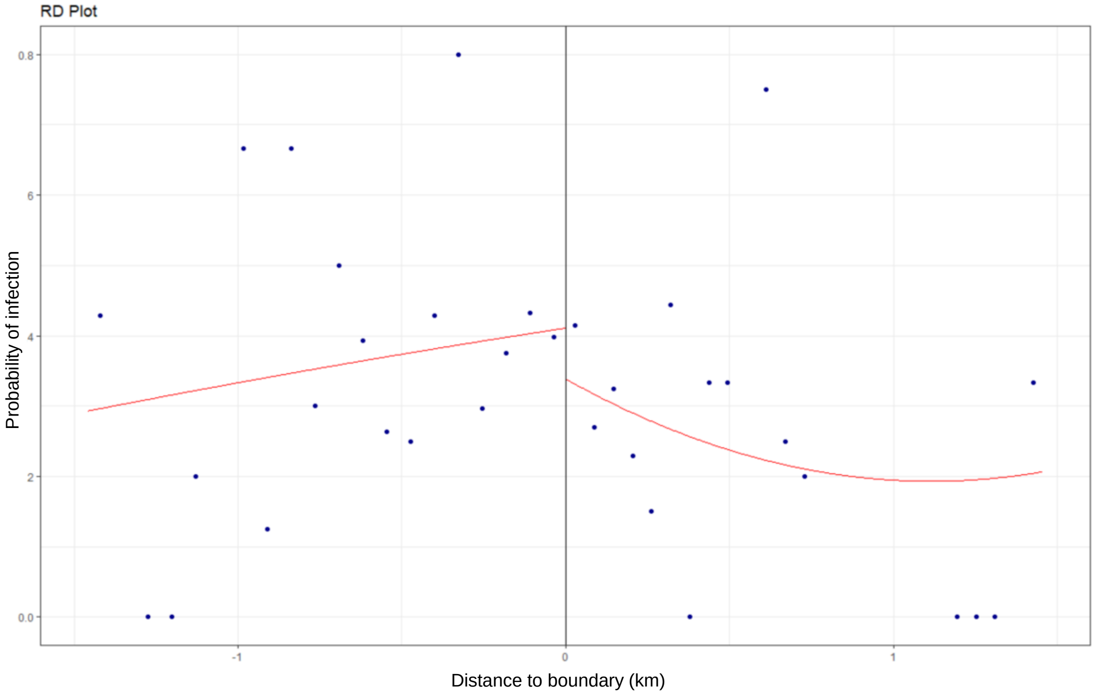
	Distance to boundary			
	<0.5 km	<1 km	<1.5 km	<2 km
Main Equation				
Dependent variable: Tested positive for COVID-19 (yes=1)				
Treatment (access to piped water)	-0.146***	-0.124***	-0.122***	-0.131***
	-(0.054)	-(0.042)	-(0.042)	-(0.039)
Overcrowding	0.0791	0.0781**	0.0759**	0.0674*
	-(0.051)	-(0.040)	-(0.038)	-(0.036)
Uses public transportation	0.0873	0.0920	0.102	0.096
	-(0.099)	-(0.086)	-(0.082)	-(0.081)
Household income > 1000 soles	-0.204*	-0.117	-0.100	-0.098
	-(0.122)	-(0.100)	-(0.096)	-(0.086)
Comorbidities	0.0268	0.00923	0.030	0.012
	-(0.094)	-(0.094)	-(0.096)	-(0.100)
Constant	0.727***	0.707***	0.688***	0.724***
	-(0.130)	-(0.074)	-(0.072)	-(0.070)
Selection equation				
Dependent variable: Covid_test				
Has COVID-19 symptoms (yes=1)	1.196***	1.180***	1.181***	1.184***
	-(0.107)	-(0.089)	-(0.084)	-(0.090)
Age	0.00823**	0.00767***	0.00798***	0.00841***
	-(0.003)	-(0.003)	-(0.002)	-(0.003)
Face-to-face job in customer service	-0.0484***	-0.0745***	-0.0760***	-0.0748***
	-(0.018)	-(0.028)	-(0.026)	-(0.026)
Dependent work	0.157**	0.110***	0.111***	0.111**
	-(0.063)	-(0.040)	-(0.037)	-(0.045)
Overcrowding	-0.104***	-0.103***	-0.101***	-0.0857***
	-(0.027)	-(0.032)	-(0.031)	-(0.031)
Uses public transportation	0.120	0.066	0.051	0.041
	-(0.090)	-(0.079)	-(0.074)	-(0.073)
Household income > 1000 soles	0.179**	0.0961*	0.076	0.0903*
	-(0.076)	-(0.054)	-(0.056)	-(0.048)
Comorbidities	0.062	0.062	0.040	0.022
	-(0.128)	-(0.122)	-(0.120)	-(0.103)
Constant	-1.061***	-0.956***	-0.949***	-1.001***
	-(0.134)	-(0.093)	-(0.087)	-(0.083)
athrho	-2.744***	-2.871***	-2.821***	-2.795***
	-(0.550)	-(0.465)	-(0.385)	-(0.401)
Selected observations	612	707	728	749
Non-selected observations	1991	1991	1991	1991
Total observations	2603	2698	2719	2740

Significance levels are: ***p<0.01, **p<0.05, *p<0.1

Note: Note: Clustered robust standard errors at the district level are included in parentheses. Includes geographic location variables (altitude, latitude).

Source: Prepared by the authors.

Annex 5 – Probability of COVID-19 infection by distance to boundary



Note: The left panel represents users lacking connection to the water supply network; the right panel reflects users connected to the water supply network.
Source: Prepared by the authors.