Identifying drivers of spatial-temporal heterogeneity in COVID-19 outcomes during the 2020-2021 outbreaks in Peru

Horacio Alfonso Chacon Torrico

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Public Health

University of Washington

2022

Reading Committee:

Robert Reiner

Abraham Flaxman

Jose Manuel Magallanes

Program Authorized to Offer Degree:

Global Health

University of Washington

Abstract

Identifying drivers of spatial-temporal heterogeneity in COVID-19 outcomes during the 2020-2021 outbreaks in Peru

Horacio Alfonso Chacon Torrico

Chair of the Supervisory Committee:
Robert Reiner, MS, PhD
Department of Health Metrics

According to the World Health Organization, it has been estimated that Peru had the highest per capita excess mortality rate for the 2020-2021 period among all countries in the world. Understanding the drivers of heterogeneity in the COVID-19 response is relevant from a public and global health perspective, as it could inform policymakers what interventions contribute to mitigating the outbreaks. To understand what province-level covariates determined better or worst outcomes in the Peruvian setting, we used vital registration data to fit time-varying mortality rates for the 196 provinces in Peru during the 2020-2021 period. Then, extracting transmission and disease features, we tested the relationship of those features with province-level covariates. We identified three main drivers of the variability of COVID-19 outcome, namely the overall wealth of the province by means of the human development index, the demographic

structure, and the internal migration. The results showed potential drivers of the epidemic in Peru and acted as an example of a failed COVID-19 response as a valuable and tragic lesson for Global Health.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I would like to thank my thesis committee members, Bobby Reiner, Abraham Flaxman and Jose Manuel Magallanes, whose guidance and patience were pivotal for the execution of this thesis work. Additionally, thanks to my friends, family, and the wonderful global health community I found at the University of Washington.

DEDICATION

I want to dedicate this thesis work to all families in Peru affected by the devastating effects of COVID-19, particularly those that lost a relative. We will continue working to ensure these avoidable tragic outcomes will not happen again.

1. INTRODUCTION

Peru is a South American country with a long Pacific Ocean coastline, home to more than 32 million inhabitants. This country has experienced significant macroeconomic growth for the past three decades; unfortunately, this prosperity has not necessarily been reflected in an improved social welfare status [1]. As a middle-income country, it mainly comprises an urbanized population (80%) in 25 administrative regions and 196 provinces [2]. The 2019 Global Burden of Disease report indicated Peru as the country with the lowest all-cause related Disability Adjusted Life Year rates in Latin America and the Caribbean, with 22,085 DALYs per 100,000 individuals [3]. Under this context, Peru was ranked as having a relatively low disease burden. However, during the pandemic, Peru suffered the worst COVID-19 outbreak globally. According to the World Health Organization, it has been estimated that Peru had the highest per capita excess mortality rate for the 2020-2021 period among all countries in the world [4]. This was translated into 437 excess COVID-19 associated deaths per 100,000 individuals, more than seven times the world median of 58 excess deaths per 100,000 people [4].

The World Health Organization declared the COVID-19 outbreak as a pandemic in March 2020. After and before that date, many countries worldwide applied distinct approaches to slow transmission progression, but similarities remained. Peru was not different from other countries when responding to COVID-19. The initial months of the Peruvian government's COVID-19 response included failures that evolved into an unfortunate number of deaths. On March 15, 2020, the Peruvian government dictated a two-week total country lockdown that was subsequently extended to June 26, 2022 [5]. Additionally, social support measures, such as food distribution and cash transfers, were implemented to counter the effect of lockdown on the informal sector, whose population relied mainly on daily income [6]. The government also put in place strict punishment for misinformation [7]. In Peru, the burden of COVID-19, as in most developing countries around the world, has been unequally distributed among different sectors, as poorer people suffered the most [6]. The underserved urban communities suffered a high COVID-19 burden, but so did native communities in the Amazonian jungle, both groups suffering a disproportionate number of deaths [8].

Although not discussed enough in the highly active and productive COVID-19 literature, some authors have proposed explanations for Peru's catastrophic and deadly COVID-19 experience. Theories include limited choices of testing supplies at the beginning of the pandemic, high economic and social inequality, the highly informal workforce and economy, as well as the lack of social protection and access to adequate health services [9]. Understanding the drivers of heterogeneity in the COVID-19 response is relevant from a public and global health perspective, as it could inform policymakers what interventions contribute to mitigating the outbreaks.

1.1 STUDY OBJECTIVE AND SPECIFIC AIMS

This study was a correlational and modeling analysis that explored the heterogeneity of COVID-19-related outcomes in Peru during the 2020 and 2021 Pandemic. Given the significant burden caused by the COVID-19 pandemic in terms of per-capita mortality, which was heterogeneously distributed along with the Peruvian geography, this study explored what drivers determined better or worst outcomes in the Peruvian setting particularly at the province-level (i.e., administrative level 2).

The specific aims of this study were the following:

1. To quantify and describe the spatial and temporal variations in COVID-19 related to outcomes during the 2020-2021 period in Peru.
2. To identify geographical and national level characteristics of COVID-19 related to outcome variability.

2. METHODS

This is an analytical, descriptive, correlational study. First, we described COVID-19 mortality and transmission patterns and features in Peru and then assessed potentially associated characteristics. To that end, we employed secondary and anonymized data to fit spatial-temporal models.

2.1 STUDY POPULATION AND SOURCES

The study utilized the national database of COVID-19 mortality, which is a subset of the Peruvian Vital Registration System containing only COVID-19 related deaths. This dataset is openly available with anonymized patient-level information [10]. The period that we chose for the mortality data was from March 2020 to June 2021, which corresponds to the date before the vaccination of the general population started. The final dataset included information about 189,943 deaths.

Peru has three administrative-political levels, starting from the largest, the department, the next level is the province, and finally, at the smallest level, the district. Therefore, we analyzed and modeled deaths at the national, departmental (n = 25) and province levels (n = 196). Additionally, to obtain standardized mortality rates, the country, departmental and province-level population numbers were obtained from the National Institute of Statistic and Informatics and 2017 census data [11]. Further province-level covariates were obtained from different sources such as United

Nations Development Programme, International Development Bank, IHME and other Peruvian governmental institutions (SUSALUD, CEPLAN).

2.2 STUDY DESIGN

We used a correlational and modeling analysis approach to times series data from the mortality datasets described previously. Three analytic stages comprised the study, described in detail in further sections:

- Mixed-effects Meta-Regression Bayesian Regularized and Trimmed (MR-BRT) model to fit time-varying province-specific mortality rates of COVID-19.
- COVID-19 province-level epidemic features extraction of the outbreak in Peru.
- Principal component and correlational analysis of epidemic features and province-level covariates.

2.3 VARIABLES

Dependent Variables

The dependent outcome for the study was the time-varying weekly per-capita COVID-19 mortality rate. This variable was obtained at the national, departmental, and provincial levels. It was estimated by fitting a Meta-Regression Bayesian, Regularized and Trimmed (MR-BRT) mixed-effects linear model.

Epidemic features

By obtaining province level-specific time series of COVID-19 mortality by means of the MR-BRT cascade spline model, we extracted outbreak features such as: time of the first peak, mortality rate of the first peak, time of the second peak, mortality rate of the second peak and highest initial effective reproductive number (Rt).

Covariates

During the correlational part of the analysis, we assessed the relationship of the epidemic features with several province-level covariates, such as:

- Rate of incoming internal migration per 100,000 individuals
- Socioeconomic status: Human development index
- Average years of education
- The proportion of population enrolled in social security insurance (proxy for informality)
- Physician to population ratio
- Demographic characteristics (population density, % 65 year or older, % female)
- Average province altitude over sea level.

2.4 DATA ANALYSIS

Step 1: Mixed-effects Meta-Regression Bayesian Regularized and Trimmed (MR-BRT) model to fit time-varying province-specific mortality rates of COVID-19.

For the first phase of the study, we aggregated the mortality data to report weekly death totals and per-capita mortality for every province. Once province-specific weekly death counts were obtained, a MR-BRT mixed effect cascade spline model was fitted as a time series for every province. We used the per-capita weekly mortality rates as the outcome and the number of weeks as the predictor. The cascading spline model first fits a mortality curve for the country, and that country estimate is used as a prior for the departmental level fit. Finally, the departmental model is used as a prior for the province-level estimates. This last step was vital for provinces with high standard errors and few data points, as it enables to borrow strength from the higher level and neighboring provinces. The MR-BRT mixed-effect linear model used cubic splines to ensure smoothness of the fitting time series and flexibility of the model. This modeling was processed with the mrtoolr R package (https://github.com/ihmeuw-msca/mrtoolr).

Step 2: COVID-19 province-level epidemic features extraction of the outbreak in Peru.

From the predicted mortality-province specific time series, we extracted different features that described the characteristics of the outbreak. These include mortality rate of the peaks, day until peak and effective reproduction number (Rt). To identify the highest moment of mortality (i.e., peak) for the two main COVID-19 waves, custom peak finder functions were developed, where the peak is defined by the highest value within a time series over a particular threshold. The minimal threshold for a peak was five deaths per week per 100,000 individuals. Also, to differentiate both waves, we divided two epidemic moments. Everything happening before day 300 of the epidemic was considered the first peak, and all the highest value after day 300 was considered the second peak.

The effective reproduction number Rt was estimated using the mortality rate and the time-varying incident fatality rates (IFR) to estimate incident cases; with the incident cases and a serial interval of 5.4 days [12], the effective reproductive number Rt could be modeled. We extracted the highest initial confident value of RT as the value that summarized that province's highest initial moment of transmissibility. Finally, Choropleth maps were developed to show the spatial distribution of the identified features.

Step 3: Principal component and correlational analysis of epidemic features and province-level covariates.

The last step was to use the extracted epidemic features at the province level to identify the relationship between individual and multiple drivers of transmission and COVID-19 outcomes. To guide the data exploration and covariate selection, we started with a principal component analysis to identify which covariates were colinear and explained the variability of the different features. Then, with the obtained epidemic features for every province, we performed a correlational analysis to identify if any province-specific characteristics such as internal migration, wealth, and age structure were associated with the COVID-19 features.

To ensure analysis reproducibility, all the code entailed for this thesis has been uploaded to a public GitHub repository (https://github.com/horaciochacon/MPH_thesis).

3. RESULTS

We modeled time-varying mortality rates and extracted the epidemic features for the 196 provinces during the specified time window. The summary of the epidemic features can be observed in **Table 1**. The province with the highest modeled mortality peak during the first wave was Cajatambo, with 94.26 weekly deaths per 100,000 population. Similarly, the highest mortality peak occurred in the Camana province for the second wave, with 38.47 weekly deaths per 100,000 population. Also, the day of the first and second peaks corresponded to the Ferreñafe and Huancabamba provinces at the days 70 and 307, respectively.

Table 1: Epidemic features and covariate summary for 196 provinces

Characteristic	N = 196[1]
Epidemic features	
Mortality rate first peak (weekly per 100,000)	10.70 (6.68, 19.01)
Mortality rate second peak (weekly per 100,000)	14.37 (9.95, 19.32)
Day of the first peak	169 (138, 182)
Day of the second peak	424 (400, 438)
Highest initial Rt value	1.34 (1.25, 1.52)
Covariates	
Human development index (HDI)	0.42 (0.35, 0.54)
Incoming migration (per 10,000)	220.1 (76.4, 398.0)
Population proportion over 65 years	0.073 (0.054, 0.088)
Population density (per km)	20 (8, 37)
Average years of education	6.93 (5.60, 8.33)
Percentage with EsSalud insurance	0.12 (0.08, 0.19)
Government density index	0.70 (0.66, 0.75)
Physician to population ratio	6.5 (4.9, 9.1)
Population proportion female	0.499 (0.490, 0.511)
Altitude in meters above sea level	2,593 (312, 3,283)

[1] Median (IQR)

To better describe the epidemic features, we mapped the distribution of the mortality peaks and day of first peaks to understand the spatial distribution and progression of the COVID-19 epidemic in Peru (**Figure 1**). We observed how the day of the peak for the first wave occurred predominantly in the northeast region of the country (i.e., the Amazon region) and the north coast. Similarly, the mortality rate of the first peak followed a similar spatial pattern. For the second wave, mortality rate peaks occurred first with a similar pattern in the northeast as compared with the first peak;

however, the intensity of the peaks followed a different distribution with less preponderance overall and more peaks occurring in the coastal, center and south of the country.

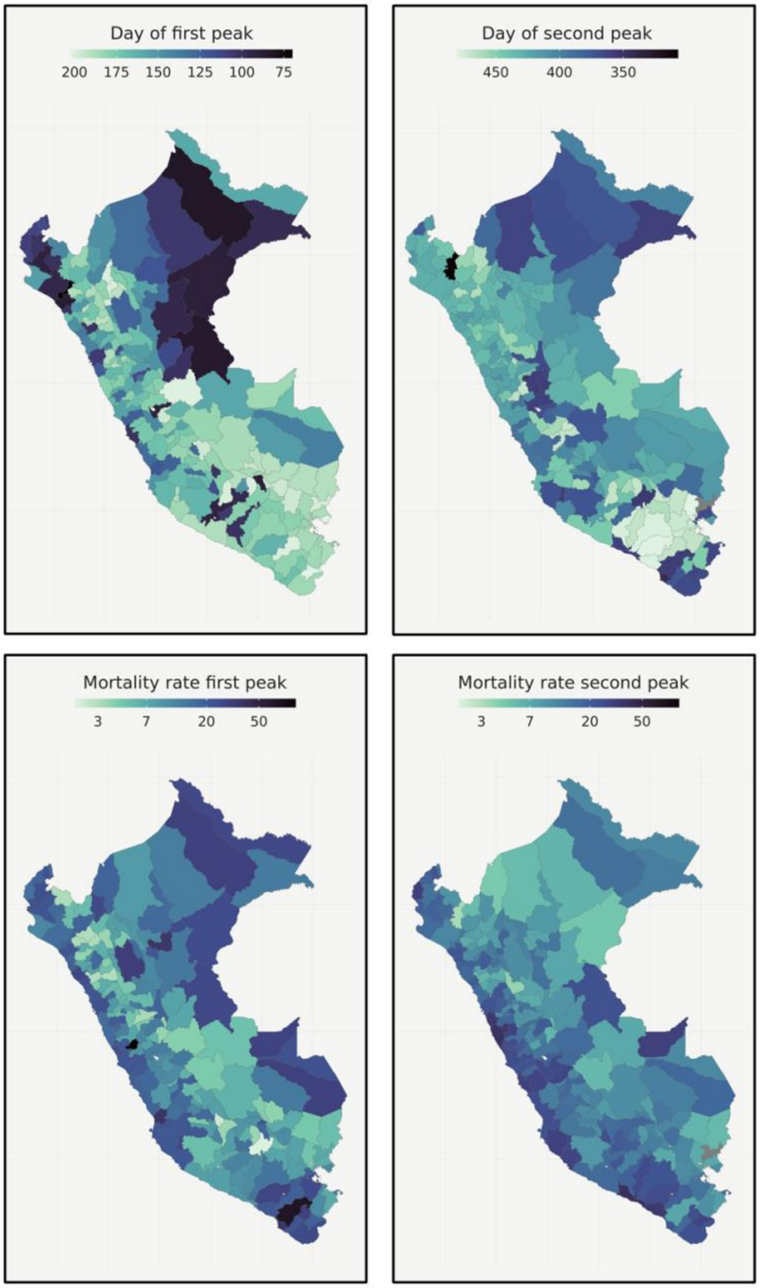

Figure 1: Spatial distribution of epidemic features

To guide the data exploration and covariate selection, we started with a principal component analysis to identify which covariates were colinear and explained the variability of the different features (**figures 2 and 3**). These plots show the first against the second principal components, both of which explain around 60% of the variability in the selected covariates. Every covariate is represented as a vector in this graph. Depending on the length and direction of the vector, the covariates are related to each other and the outcomes. Additionally, in both plots, we can see the size of the points as the mortality rate for the first (**figure 2)** and second (**figure 3**) peaks. We can see that the variables human development index, years of education and proportion of social security insurance (EsSalud) are highly colinear and run parallel to the first principal component (PC1). On the other side, variables of demographic characteristics such as the proportion of the population over 65 and the proportion of population females run parallel to the principal component number 2 (PC2).

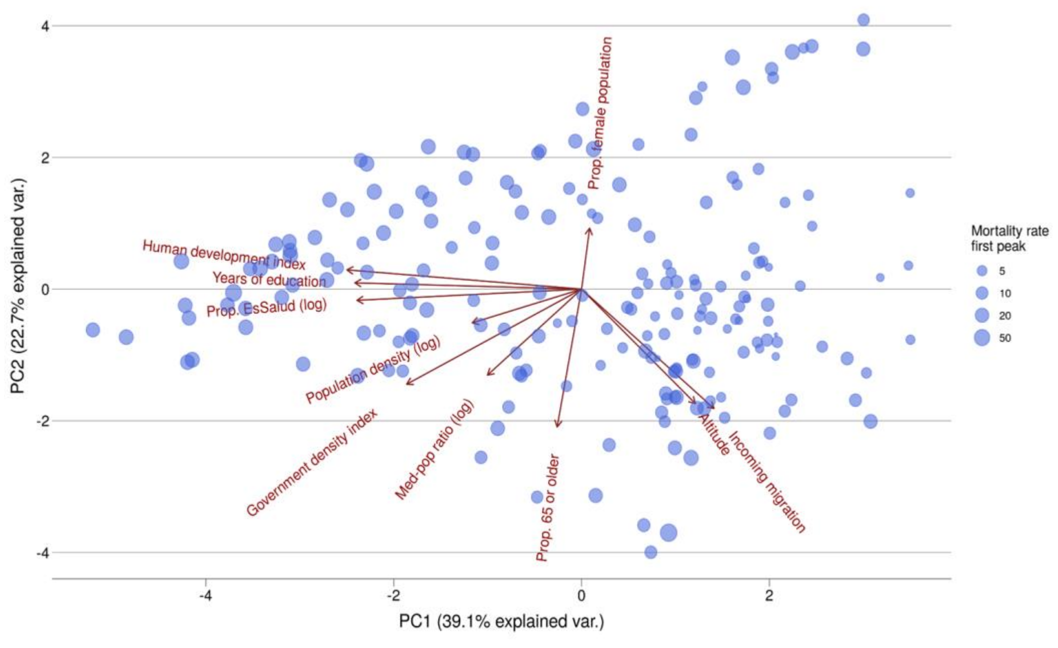

Figure 2: Principal component analysis (Mortality rate first peak)

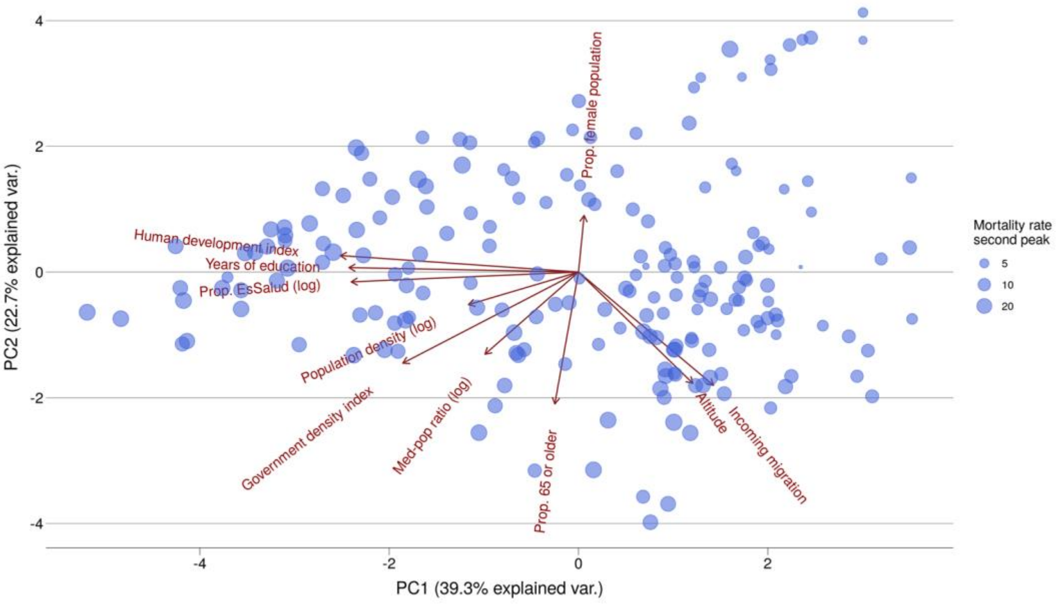

Figure 3: Principal component analysis (Mortality rate second peak)

As seen in the principal component analysis plot, the PC1 explains sociodemographic constructs, and in the following panel of plots (**figure 4**), we plot the PC1 against the peak mortality rates (first row of plots) during the first and second waves, shows a clear inverse linear relationship. That same relationship can be observed by plotting HDI against mortality peaks (bottom row of plots). Considering the human development index as an indicator of the overall wealth of the province, the plot shows that the wealthier the province, the higher the peak mortality rate that province suffered.

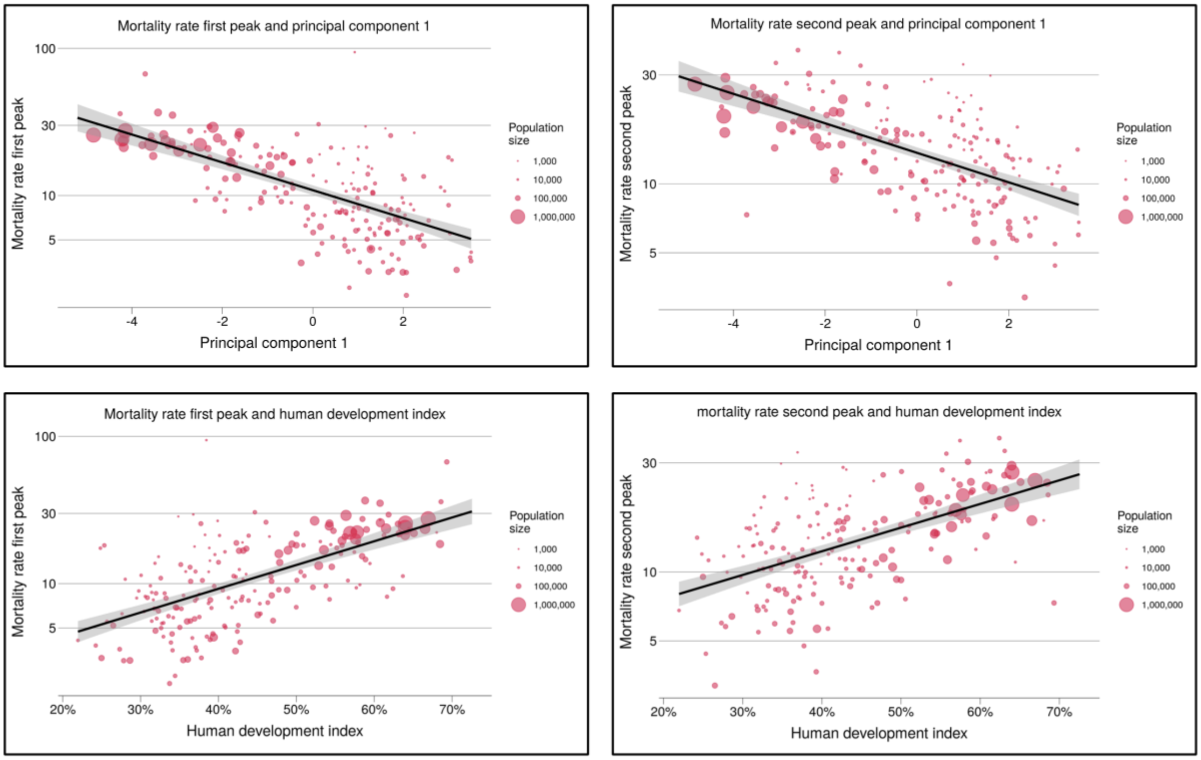

Figure 4: Scatter plots of mortality rates during the first and second peaks against principal component 1 and human development index

In **figure 5**, we plot the principal component 2 (PC2), which entailed the demographic structure of the provinces, against peak mortality rates. We observe no relationship for the first wave (top left plot) but an inverse relationship for the second mortality peak (top right plot). Similarly, in the two plots at the bottom, we see that mortality in the second wave worsened in provinces with a higher percentage of elderly, but this did not happen during the first wave.

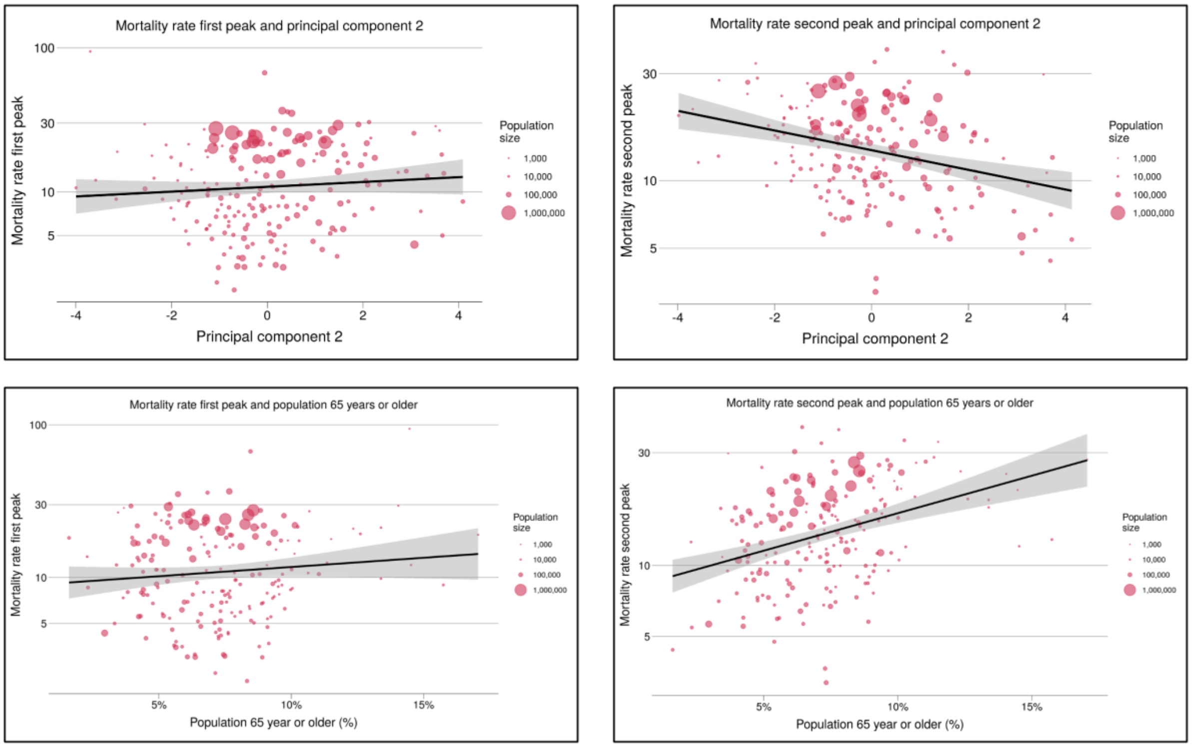

Figure 5: Scatter plots of mortality rates during the first and second peaks against principal component 2 and population over 65

Finally, we explored the relationship between incoming internal migration and mortality during the first and second waves. We also tested how this relationship interacted with the human development index as a dichotomous variable. We saw a linear relationship, a little bit stronger in the second wave, between internal migration and mortality only for the poorest half of the provinces. This means that the more incoming migration the province received, the highest the mortality, only for less wealthy provinces.

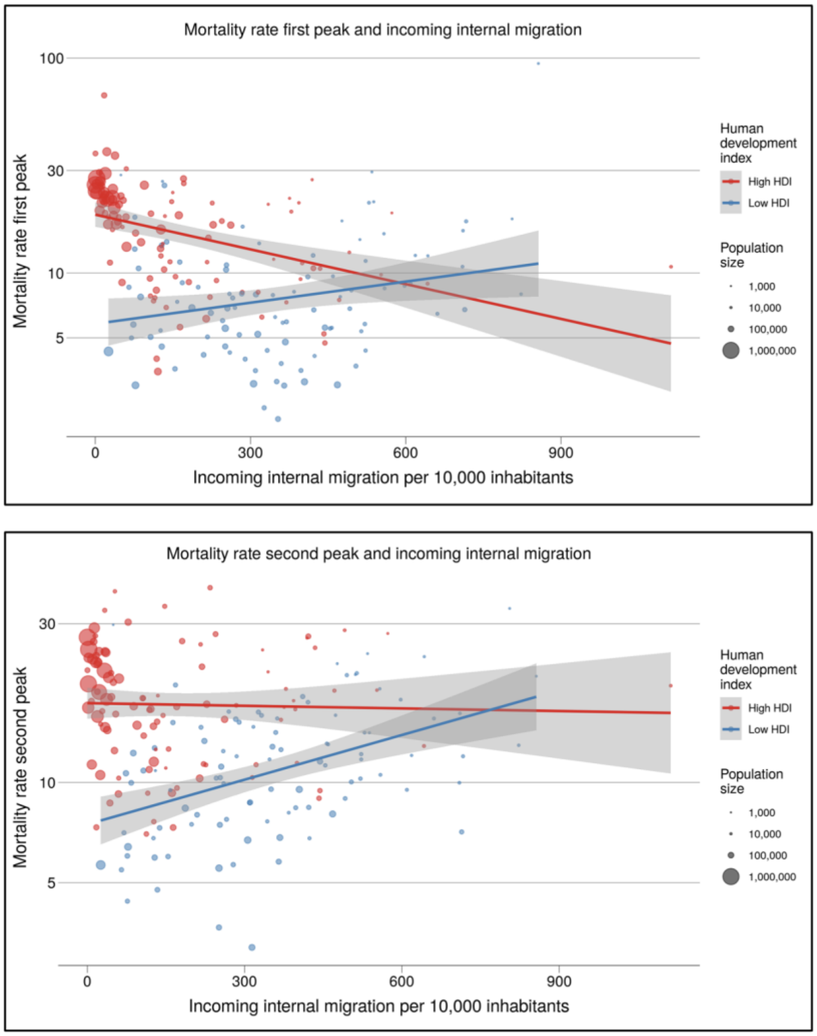

Figure 6: Scatter plots of mortality rate during first and second peak against incoming internal migration stratified by human development index

4. DISCUSSION

This study represents the first attempt to characterize the drivers of spatial-temporal variation of the heterogeneity in COVID-19 epidemic features in Peru. The main objective of the analysis was to understand how these provincial level covariates were related to the performance of the provinces in terms of mortality and transmission. We identified three main drivers of the variability of COVID-19 outcome, namely the overall wealth of the province by means of the human development index, the demographic structure, and the internal migration. Additionally, we described the spatial distribution and progress of the intensity of the outbreaks in space, where the first wave initially affected the Peruvian Amazon and the northern coastal regions, and the second wave highest peak affected the coast but was less intense and concentrated.

Regarding the human development index, some other studies have also found a positive correlation between its values and COVID-19 mortality rates [13, 14]. As other authors have suggested, at a national level ecological comparison, it would have been expected that administrative units with lower HDI values would have worse COVID-19 outcomes, given humanitarian and health system conditions that would have to predispose them to a nefarious response [14]. Several explanations for these results could arise because high-income provinces could have a higher prevalence of comorbidities associated with COVID-19 mortality.

The age structure of the population acted as a second important driver of COVID-19 epidemic features. Other global ecological studies also found the proportion of individuals age 60 or older positively associated with COVID-19 mortality [15]. This finding was an expected finding, as two years into the pandemic, we know how age is a significant risk factor for the severity and mortality of COVID-19 [16].

Finally, we also found a positive relationship between internal migration when stratified and observing only the poorest half of the provinces. In India, it has been described that states receiving a higher proportion of internal migration had more cases detected [17]. Peru had a big migration wave that started from the main cities to rural communities during the lockdown period [18], a movement which could have been a catalyzer for transmission within poorer, internal migration receiving provinces.

This study contains several limitations. First, the data quality could be potentially biased due to systematic underreporting of COVID-19 deaths in poorer and remote provinces. Second, there might be several unobserved confounders of the relationship that we present, which could better explain the variability of COVID-19 province-level outcomes heterogeneity. Finally, given that this is an ecological study, the results should not be extrapolated to the general population as we could incur in ecological fallacy. Despite the limitations, we consider this an important first attempt to characterize and model the heterogeneity and detect province-level covariates that could have influenced COVID-19 devastating effect in Peru.

This research did not pursue direct answers about the underlying reasons Peru presented the worst COVID-19 outbreak in the world. Instead, this assessment intended to describe what happened at the provincial level, using tools that let us observe epidemic features at a more granular level.

Nevertheless, the results show some potential drivers of the epidemic in Peru. Conclusively, understanding Peru as an example of a failed COVID-19 response is a valuable and tragic lesson for Global Health that we plan to dissect further and analyze.

BIBLIOGRAPHY

[1] Dargent E, Feldmann AE, Luna JP. Greater State Capacity, Lesser Stateness:: Lessons from the Peruvian Commodity Boom. *Polit Soc* 2017; 45: 3–34.

[2] OECD. *OECD Reviews of Health Systems: Peru 2017*. OECD Publishing, 2017.

[3] Institute for Health Metrics and Evaluation. Global Burden of Disease (GBD) 2019. GBD compare: Viz Hub, https://vizhub.healthdata.org/gbd-compare/ (2019, accessed May 2022).

[4] World Health Organization. *Global Excess Deaths Associated with COVID-19: January 2020–December 2021*.

[5] Lossio J. [Covid-19 in Peru: state and social responses]. *Hist Cienc Saude Manguinhos* 2021; 28: 581–585.

[6] Gianella C, Gideon J, Romero MJ. What does COVID-19 tell us about the Peruvian health system? *Rev Can Etudes Dev* 2021; 42: 55–67.

[7] Alvarez-Risco A, Mejia CR, Delgado-Zegarra J, et al. The Peru Approach against the COVID-19 Infodemic: Insights and Strategies. *Am J Trop Med Hyg* 2020; 103: 583–586.

[8] Montag D, Barboza M, Cauper L, et al. Healthcare of Indigenous Amazonian Peoples in response to COVID-19: marginality, discrimination and revaluation of ancestral knowledge in Ucayali, Peru. *BMJ Glob Health*; 6. Epub ahead of print January 2021. DOI: 10.1136/bmjgh-2020-004479.

[9] Rees GH, Peralta Quispe F, Scotter C. The implications of COVID-19 for health workforce planning and policy: the case of Peru. *Int J Health Plann Manage* 2021; 36: 190–197.

[10] Gobierno del Peru. Datos Abiertos de COVID-19, https://www.datosabiertos.gob.pe/group/datos-abiertos-de-covid-19 (2021, accessed December 6, 2021).

[11] National Institute of Statistics and Informatics. 2017 Peru National Census, https://censos2017.inei.gob.pe/bininei/RpWebEngine.exe/Portal?BASE=CPV2017DI&lang=esp (2018, accessed December 6, 2021).

[12] Rai B, Shukla A, Dwivedi LK. Estimates of serial interval for COVID-19: A systematic review and meta-analysis. *Clin Epidemiol Glob Health* 2021; 9: 157–161.

[13] Liu K, He M, Zhuang Z, et al. Unexpected positive correlation between human development index and risk of infections and deaths of COVID-19 in Italy. *One Health* 2020; 10: 100174.

[14] Buheji, AlDerazi, Ahmed, et al. The association between the initial outcomes of COVID-19 and the human development index: An ecological study. *Hum Syst Manage*, https://content.iospress.com/articles/human-systems-management/hsm210005.

[15] Alshogran OY, Altawalbeh SM, Al-Azzam SI, et al. Predictors of Covid-19 case fatality rate: An ecological study. *Ann Med Surg (Lond)* 2021; 65: 102319.

[16] Booth A, Reed AB, Ponzo S, et al. Population risk factors for severe disease and mortality in COVID-19: A global systematic review and meta-analysis. *PLoS One* 2021; 16: e0247461.

[17] Sarkar S. Significance of migration to the COVID 19 outbreaks in major states in India. *International Journal of Migration, Health and Social Care* 2021; 17: 402–413.

[18] Fort R, Espinoza M, Espinoza Á. COVID-19 y las migraciones de la ciudad al campo en el Perú: Identificación de amenazas y oportunidades para el uso sostenible del capital natural (Resumen …, https://publications.iadb.org/publications/spanish/document/COVID-19-y-las-migraciones-de-la-ciudad-al-campo-en-el-Peru-Identificacion-de-amenazas-y-oportunidades-para-el-uso-sostenible-del-capital-natural-Resumen.pdf (2021).