Closing the Health Gap in Brazil: Data-led Infrastructure Expansion

December 2022

Shreya Chaturvedi   Kwang Jun Lee   Luisa Leite
Bharath Ram   Masato Takahashi   Carlos Víquez

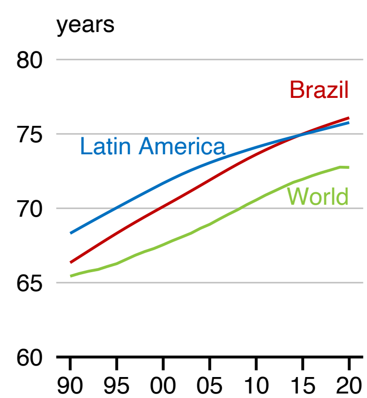
Executive Summary

- XX
- XX
- XX
- XX

1. Introduction

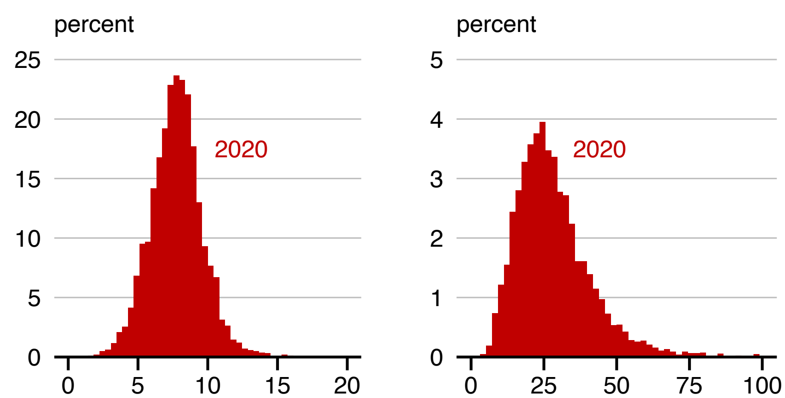
- **Over the past several decades, Brazil has made tremendous improvements in terms of life expectancy and infant mortality, outpacing the world and Latin America (Chart 1).** This is a stark achievement to develop a robust healthcare system to address the needs of its population.
- **Not all Brazilians, however, have equal access to high quality medical services.** Some pieces of evidence can be found in the unique distributions of the number of deaths and hospitalization relative to population across municipalities (Chart 2)
- **One of the solutions towards achieving equal access to medical services could be to expand the medical infrastructure in municipalities that have not been able to meet medical needs, i.e., those with high burden of disease.**
- **This is easier said than done.** One of the challenges is that the burden of disease cannot be measured by a single indicator, e.g., the mortality rate, and, hence, must be evaluated in a multi-faceted manner. Another challenge is that, even if the burden of disease is assessed at the municipality level, it does not say what medical infrastructure should be expanded.
- **Against this backdrop, our analysis aims to address the followings:**
  - **First, we develop a new comprehensive measure of burden of disease to inform policymakers a new perspective on municipalities that are falling behind in terms of medical provision.** The features of this measures, which we call the Municipal Public Health Index, are that it incorporates multiple factors including future medical needs and that can be constructed with relative ease compared to other measures.
  - **Second, we analyze the correlation between the burden of disease according to the Municipal Public Health Index and various indicators of health infrastructure, based on which we make a policy recommendation of how to expand the health infrastructure in certain municipalities.** While the lack of time-series data does not allow us to conduct a full-fledged analysis regarding the effect of expanding health infrastructure on health status over time, we believe this will be a good starting point.
- **As the first step, we provide a quick overview of the current health indicators in Brazil.**

Chart 1:
Life expectancy



Source: World Bank

Chart 2: Distributions by municipalities (per 1,000 people)
(i) Deaths (ii) Hospitalization



Sources: Ministry of Health; Instituto Brasileiro de Geografia e Estatística

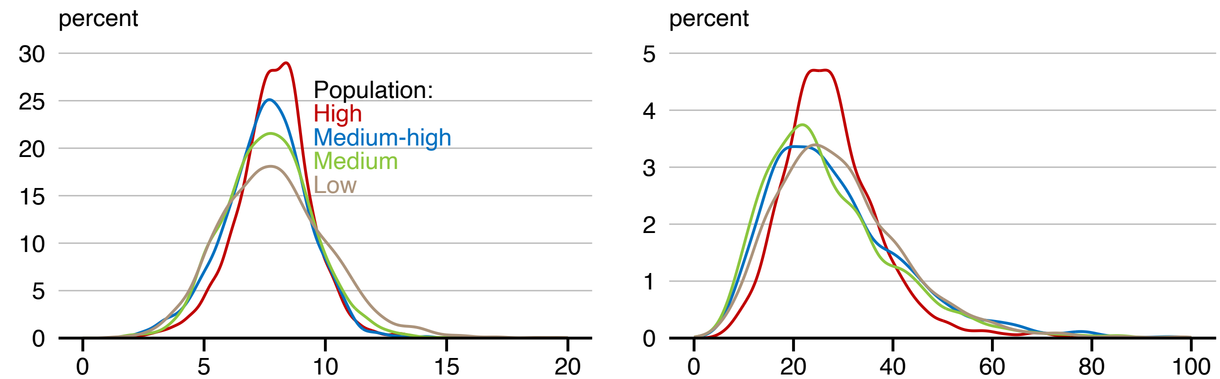
2. Characterizing health care in Brazil

Deaths and hospitalizations by population group

- We divide the municipalities into four groups by quartile of population to see whether the difference in the number of deaths or hospitalizations across municipalities comes from urban and rural settings (Chart 3). While the distributions do not differ substantially across population groups, we notice some slight differences which include that the share of municipalities with more than 10 deaths per 1,000 people is higher in the low population group and that the share of municipalities with more than 40 hospitalizations per 1,000 people is higher in bottom three population groups.

Chart 3: Distributions by quartile of population (per 1,000 people; 2020)

(i) Deaths (ii) Hospitalization

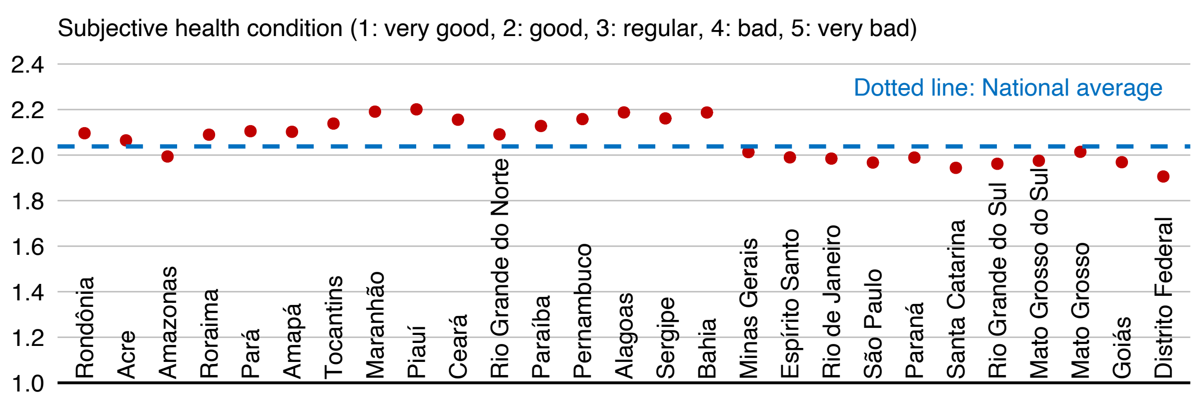


Note: "High": >75 percentile; "Medium-high": 50–75 percentile; "Medium": 25–50 percentile; "Low": <25 percentile.
Sources: Ministry of Health; Instituto Brasileiro de Geografia e Estatística (IBGE)

Subjective health conditions at the state level

- The National Health Survey asks respondents, among other things, to describe their state of health on a five-point scale from "very good" (1) to "very bad" (5). We consider the responses to this question as a measure of the subjective health conditions. Using the publicly available dataset, we can aggregate the subjective health conditions at the state level. Chart 4, which plots the measure for all 27 states, reveals that there is some degree of variation. We will further explore this point in Section 3.2.

Chart 4: Subjective health conditions (2019)

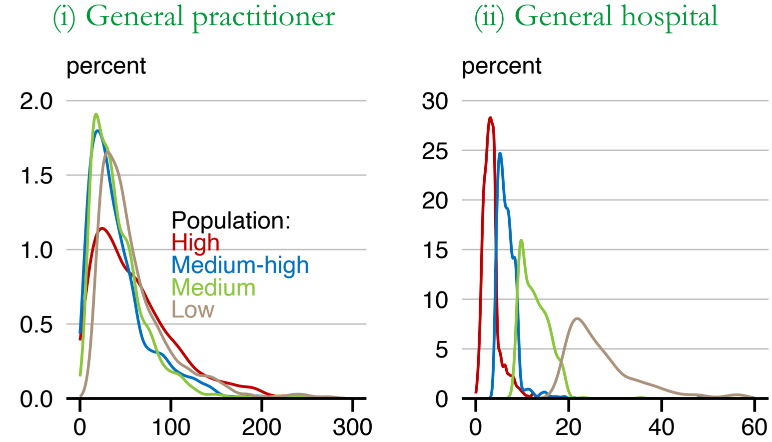


Source: IBGE

Health care access

- Finally, we compare health care access in terms of medical infrastructure by municipality group (the same classification as in Chart 3). Surprisingly, we do not observe any clear relationship between the size of the population and the medical infrastructure (Chart 5). Thus, our solution needs to be more than just expanding medical infrastructure in areas with population.

Chart 5: Medical infrastructure (per 100,000 people)



Sources: Health Infrastructure Database; IBGE

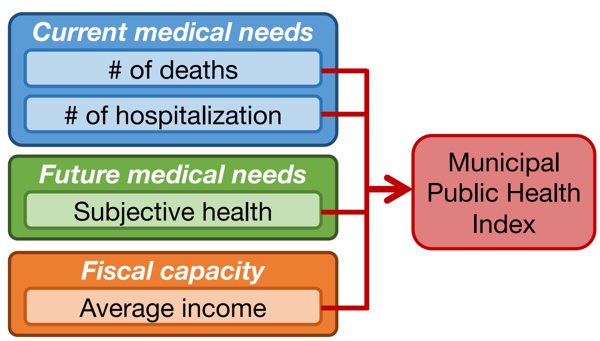
3. Measuring the disease of burden at the municipality level

3.1. Defining the burden of disease

- **We recognize the burden of disease to be determined by current and future medical needs as well as fiscal capacity.** We think this way because our ultimate goal is to link the burden of disease with the necessary investment decision, which must be made in a forward-looking way. Every investment decision, however, has to be financed in one way or the other. In this regard, it is also important to take fiscal capacity into account.

- **Under this definition, we construct the Municipal Public Health Index (MHPI) by synthesizing the indicators of both current and future medical needs.** We use the numbers of deaths and hospitalization per capita as the indicators of current medical needs. We intend to use the subjective health conditions as the indicator of future medical needs. Finally, we use income as a proxy for tax revenue and, hence, fiscal capacity (Chart 6).

Chart 6: Our measure of disease of burden



3.2. Predicting the subjective health conditions at the municipality level

- **One of the challenges is that the data on the subjective health conditions, which we introduced earlier, is only available at the state level.**

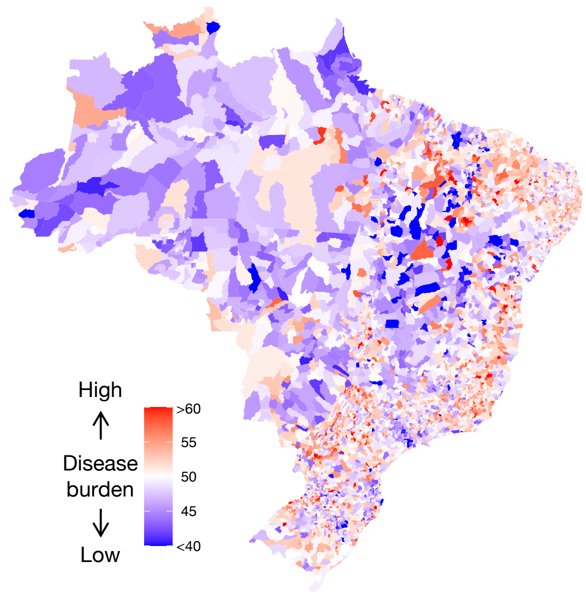
- **We assume that the degree of variation in this measure across states reflects the difference in various socioeconomic and hygiene conditions.** They include the average age and the share of households with access to sewage system. They are available at the state and municipality level.

- **We estimate the subjective health conditions in municipalities using the relationship between subjective health conditions and socioeconomic and hygiene conditions at the state level.** To do so, we use a machine learning approach (see Technical Appendix for details).

3.3. Municipal Public Health Index

- **We calculate the MPHI in the following two steps. First, we standardize all four indicators:** deaths per capita; hospitalizations per capita; the subjective health conditions; and average income (inverted). A higher value implies a greater burden of disease for that municipality. This first step is necessary to make the indicators measured in different units comparable.
- **Second, we aggregate the three (standardized) indicators by taking a simple average.**
- **Through these steps, we obtain the MPHI as shown in Chart 7.** According to this analysis, the burden of disease is identified to be higher in municipalities that are located in the north-east region or in the coastal region.

Chart 7: Municipal Public Health Index



Source: Authors' calculation

4. Relationship with the existing medical infrastructure

- **We compare the medical capital stock between the groups of municipalities with low burden of disease (frontier municipalities) and those with high burden of disease (catching-up municipalities).** The former group is municipalities whose MPHIs are equal to or higher than the 75th percentile, and the latter is equal to or lower than the 25th percentile.
- **We notice that catching-up municipalities generally have fewer health professionals (general practitioners and nurses) and lower amount of medical equipment (X-ray and ultrasound), whereas the number of general hospitals is higher (Chart 8).**

Chart 8: Comparison of the medical capital stock (per 100,000 people)

	Medical professionals				Medical facilities				Medical equipment	
Municipalities	Pharmacist	General practitioner	Nurse	Family physician	Health center	General hospital	Speciality care clinic	Health management	X-ray	Ultrasound
Frontier (Low burden)	24	58	110	17	28	4	22	7	36	22
Catching-up (High burden)	16	30	94	15	42	7	13	10	17	13

Note: Each red square indicates the greater of the two.

4. Relationship with the existing medical infrastructure

- **Based on the above, we recommend the government increase the number of medical workers and the amount of medical equipment in municipalities with high burden of disease as identified by our MPHI, so as to promote the full utilization of existing medical facilities and, hence, reduce the numbers of deaths and hospitalization.**

Technical Appendix

Chart 2

- Mortality and hospitalization data are obtained from the Mortality Information System and the Hospitalizations Information System, respectively. The number of deaths is calculated as the sum of deaths from causes classified in Chapters 1–18 and 20 in the ICD-10 (Chapters 19 and 21 are not included due to data unavailability). The number of hospitalizations is calculated as the sum of hospitalizations from causes classified in all Chapters 1–21.[1] The missing observations are treated as zero.
- For each municipality, the numbers of deaths and hospitalizations are divided by the population according to the Brazilian Census in 2010 and then multiplied by 1,000 to obtain the numbers of deaths and hospitalizations per 1,000 people which are used in the charts.

Details on predicting the subjective health conditions at the municipality level

- **We predict the subjective health conditions at the municipality level in the following three steps.** See Chart A1 (p. 9) for a visual explanation.

Step 1: Aggregate the Census data at the state level

We begin by aggregating the National Health Survey (PNS) dataset at the state level. This survey was conducted by the Instituto Brasileiro de Geografia e Estatística in 2019 in an effort to evaluate health conditions of the people in Brazil. The variable of our interest is the subjective health score (J00101) which comes with a five-point scale from "very good" (1) to "very bad" (5). This variable will be our dependent variable. For aggregation, we use the weights (V00281) which are adjusted for likelihood of selection and rate of non-response by sex and age category.

Next, we aggregate the 2010 Census at the state level. The Census consists of two datasets: those from the household survey and the individual survey. First, the household survey dataset contains the following pieces of information for each municipality: the number of households; average room density; average household size; share of households by types of wall materials; share of households with toilets, sewage, water, garbage collection system, and electricity. We aggregate the household survey at the state level using the number of households as weights. Second, the individual survey dataset provides the following variables for each municipality: population; share of male population; average age; share of population by race (white, black, Asian, mixed, and indigenous). We aggregate the individual survey at the state level using the population as weights.

[1] Chapter 1: Certain Infectious and Parasitic Diseases; Chapter 2: Neoplasms; Chapter 3: Disease of the blood and blood-forming organs and certain disorders involving the immune mechanism; Chapter 4: Endocrine, Nutritional, and Metabolic Diseases; Chapter 5: Mental, Behavioral and Neurodevelopmental disorders; Chapter 6: Diseases of the Nervous System; Chapter 7: Diseases of the Eye and Adnexa; Chapter 8: Diseases of the Ear and Mastoid Process; Chapter 9: Diseases of the Circulatory System; Chapter 10: Diseases of the Respiratory System; Chapter 11: Diseases of the Digestive System; Chapter 12: Diseases of the Skin and Subcutaneous Tissue; Chapter 13: Diseases of the Musculoskeletal System and Connective Tissue; Chapter 14: Diseases of Genitourinary System; Chapter 15: Pregnancy, Childbirth, and the Puerperium; Chapter 16: Certain Conditions Originating in the Perinatal Period; Chapter 17: Congenital malformations, deformations, and chromosomal abnormalities; Chapter 18: Symptoms, signs, and abnormal clinical and laboratory findings, not elsewhere classified; Chapter 19: Injury, poisoning, and certain other consequences of external causes; Chapter 20: External Causes of Morbidity; Chapter 21: Factors influencing health status and contact with health services.

Step 2: Regress PNS subjective health conditions on the Census data

As shown in Chart 4 in the main text, there is a variability in the subjective health conditions across states. We assume that this reflects differences in various socioeconomic and hygiene conditions as documented in the Census in a comprehensive manner. Based on this assumption, as the second step, we empirically examine the relationship between the subjective health conditions and demographic characteristics at the state level.

Our empirical strategy is to estimate the three different models using ordinary least squares. As mentioned earlier, the dependent variable is the subjective health conditions which we calculated in the first step. We have the following 21 potential independent variables: average room density; average household size; share of households by types of wall materials (masonry, rigged wood, Taipa, used wood, straw, other materials, or no walls); share of households with toilets, sewage, water, garbage collection system, or electricity; share of male population; average age; share of population by race (white, black, Asian, mixed, or indigenous). The number of observations is 27, which is the number of states in Brazil.

The three regression models are the simple model (short model), the kitchen-sink (long model), and the LASSO model. The brief explanation of each model is as follows:

- ✓ **Simple model:** We use the arbitrarily chosen following four variables as independent variables: room density, sewage, water, average age.
- ✓ **Kitchen-sink model:** We use almost all the potential independent variables in this model. Given the number of potential independent variables is close to the observations, we decide to include the following 14 variables on the right-hand side of the regression equation: room density. In this regard, this model should be referred to as a sort of kitchen-sink model strictly speaking, but for simplicity we refer to this simply as a kitchen-sink model.
- ✓ **LASSO model:** The third model is a regression-based machine learning model, and appropriate independent variables will be automatically chosen from the pool of 21 potential independent variables.[2]

We separate the sample into two parts: 70 percent goes to the training sample and 30 percent goes to the test sample. We estimate the models using the training sample and evaluate the performance based on the root mean squared error (RMSE) in the test sample, i.e., out-of-sample RMSE. The results of the performance check are shown in Chart A2. We confirm that the LASSO model has a better performance than both the simple and kitchen-sink models.

[2] The least absolute shrinkage and selection operator, or LASSO, is a method for estimating the regression coefficients by minimizing the following equations:

$$\hat{\beta}_{\lambda}^{LASSO} = \arg\min_{\beta}\left[\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 + \lambda\sum_{j=1}^{p}\left|\beta_j\right|\right]$$

where λ is a penalty parameter and p is the number of coefficients in the model.
In the case of positive penalty parameter, the method penalizes the non-zero values of coefficients, which results in assigning zero to the coefficients of the variables that are less important. We choose λ based on a cross-validation procedure.

Step 3: Predict municipal health scores based on the relationship at the state level

Finally, we use the relationship identified in the second step to predict the subjective health score at the municipality level.

- Finally, we address some caveats. The biggest shortcoming of this analysis is the small sample size. It would be an interesting extension of this analysis if the PNS-based subjective health conditions become available even for some selected municipalities. Second, given the small number of observations, we only sought potential independent variables from the Census, it would also be worth further exploring the deep determinants of the subjective health conditions.

Details on the Municipal Public Health Index

- Each of the four indicators are standardized to mean 50 and standard deviation of 10. The choices of 50 for the mean and 10 for the standard deviation are arbitrary, and they can be set at different values because the MPHI seen to be ordinal instead of cardinal.
- In the main text, we put the map of the MPHI to get the big picture. But in this technical appendix, we provide a list of top 20 and bottom 20 municipalities just for an illustrative purpose. We do not include this list in the main text because we consider it best to use the MPHI to get the overall idea about the public health status in small groups of municipalities, and not in a way to compare two adjacent municipalities on which is doing better. This is because each of the three indicators can vary year to year, especially in small municipalities.
- For this analysis, we took a simple average of the four standardized indicators to integrate the indices into one indicator. We did so because we take an agnostic approach about which indicator is has the highest importance and which does not. This does not, however, limit the potential users of the MPHI to stick to this particular construction method. Rather, this is one illustration that the index can be easily modified to suit the different needs of policymakers. For example, if a policymaker wants to put more emphasis on mortality, then the weight placed on the number of deaths per capital can be raised.

Other notes on future considerations

- One of the things that this analysis did not explicitly take into account is the fact that the major causes of deaths or hospitalizations vary a lot among municipalities. For instance, pregnancy, childbirth, and puerperium accounts for 18 percent of hospitalization at the median, but it only accounts for a few percent in some municipalities, whereas it accounts for around 50 percent in some municipalities (Chart A4). Another example is the diseases of respiratory system. It accounts for 13 percent of hospitalization in a typical municipality, but this is a more serious cause of hospitalization in some municipalities and a less so in others. These pieces of evidence might imply that municipalities have different medical needs, and such needs are better met with a tailor-made strategy for health infrastructure expansion.
- It is also worth considering whether expanding infrastructure in each municipality will be effective, or it is more cost effective if the government makes a major investment in regional hubs across the country. While the availability of data does not allow us to do a deep dive into developing an investment plan that incorporates this aspect, i.e., the accessibility to other municipalities which heavily depends on whether a municipality is in an urban or rural area, we think this is an interesting avenue for future analysis.

Chart A1: Visual illustration of the method to predicting subjective health conditions

Step 1

Code		PNS	Census data							
State	Municipality	Health score	Room density	···	Sewage	Water	···	Male ratio	Average age	···
11	110001	?	1.63	···	0.02	0.94	···	0.511	41.0	···
11	110002	?	1.76	···	0.08	0.99	···	0.507	43.1	···
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
11	110180	?	1.56	···	0.01	0.97	···	0.511	46.2	···
12	120001	?	1.99	···	0.12	0.93	···	0.529	43.7	···
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Aggregate at the state level

Code		PNS	Census data							
State		Health score	Room density	···	Sewage	Water	···	Male ratio	Average age	···
11		2.096	1.76	···	0.23	0.95	···	0.509	43.5	···
12		2.064	2.19	···	0.39	0.86	···	0.502	44.0	···
⋮		⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Step 2

Code		PNS	Census data							
State	Municipality	Health score	Room density	···	Sewage	Water	···	Male ratio	Average age	···
11	110001	?	1.63	···	0.02	0.94	···	0.511	41.0	···
11	110002	?	1.76	···	0.08	0.99	···	0.507	43.1	···
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
11	110180	?	1.56	···	0.01	0.97	···	0.511	46.2	···
12	120001	?	1.99	···	0.12	0.93	···	0.529	43.7	···
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Regress Census data on health score from PNS

Code		PNS	Census data							
State		Health score	Room density	···	Sewage	Water	···	Male ratio	Average age	···
11		2.096	1.76	···	0.23	0.95	···	0.509	43.5	···
12		2.064	2.19	···	0.39	0.86	···	0.502	44.0	···
⋮		⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Step 3

Code		PNS	Census data							
State	Municipality	Health score	Room density	···	Sewage	Water	···	Male ratio	Average age	···
11	110001	?	1.63	···	0.02	0.94	···	0.511	41.0	···
11	110002	?	1.76	···	0.08	0.99	···	0.507	43.1	···
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
11	110180	?	1.56	···	0.01	0.97	···	0.511	46.2	···
12	120001	?	1.99	···	0.12	0.93	···	0.529	43.7	···
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Predict based on the relationship at the state level

Code		PNS	Census data							
State		Health score	Room density	···	Sewage	Water	···	Male ratio	Average age	···
11		2.096	1.76	···	0.23	0.95	···	0.509	43.5	···
12		2.064	2.19	···	0.39	0.86	···	0.502	44.0	···
⋮		⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Chart A2: Performance check

Model	In-sample RMSE	Out-of-sample RMSE
Simple	0.039	0.079
Kitchen-sink (sort of)	0.016	0.092
LASSO	0.024	**0.062**

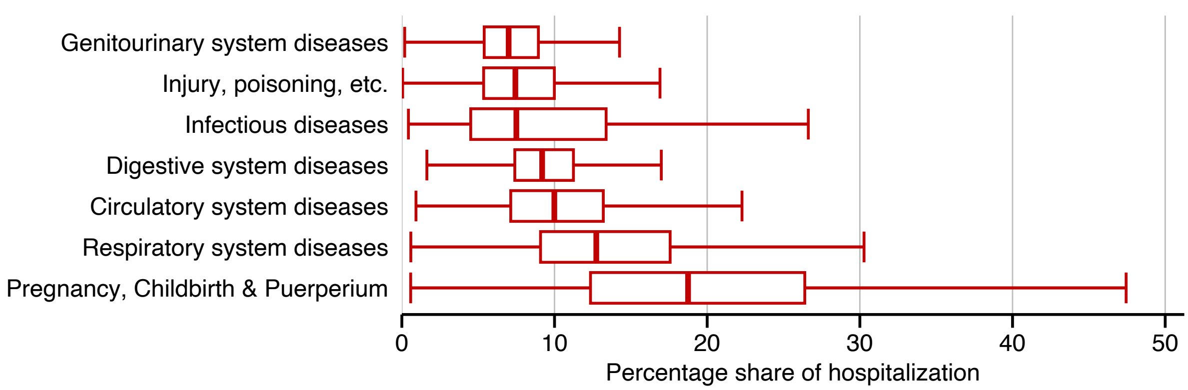
Note: See Technical Appendix for details.
Source: Authors' calculation

Chart A3: Municipalities with top and bottom 20 MPHIs

Top	State	Municipality	Worst	State	Municipality
1	Santa Catarina	Florianópolis	1	Maranhão	Passagem Franca
2	São Paulo	Santana de Parnaíba	2	Piauí	São Raimundo Nonato
3	Distrito Federal	Brasília	3	Santa Catarina	Treze de Maio
4	Espírito Santo	Vitória	4	Piauí	Fronteiras
5	Santa Catarina	Balneário Camboriú	5	Bahia	São Miguel das Matas
6	São Paulo	Santa Cruz da Conceição	6	Bahia	Wanderley
7	Santa Catarina	Treze Tílias	7	Santa Catarina	Ponte Serrada
8	Paraná	Curitiba	8	Paraíba	Santo André
9	São Paulo	Valinhos	9	Bahia	Cristópolis
10	Pernambuco	Fernando de Noronha	10	Goiás	São Miguel do Passa Quatro
11	Rio Grande do Sul	Carlos Barbosa	11	Piauí	São Braz do Piauí
12	Minas Gerais	Nova Lima	12	Piauí	Wall Ferraz
13	Rio de Janeiro	Niterói	13	Piauí	Parnaguá
14	São Paulo	São Paulo	14	Bahia	Itiruçu
15	Rio Grande do Sul	Chuí	15	Paraíba	Uiraúna
16	Minas Gerais	Belo Horizonte	16	Paraíba	Brejo dos Santos
17	Rio Grande do Sul	Porto Alegre	17	Piauí	Simões
18	Rio Grande do Sul	Caxias do Sul	18	Paraná	Santa Mariana
19	Santa Catarina	Joaçaba	19	Maranhão	São Félix de Balsas
20	Rio Grande do Sul	Garibaldi	20	Goiás	Indiara

Source: Authors' calculation

Chart A4: Major causes of hospitalization in Brazil



Note: The chart plots only the subset of diseases for which the median percentage share of hospitalization is higher than 5 percent. Outliers are not shown. The data are from 2010.
Sources: Ministry of Health; Instituto Brasileiro de Geografia e Estatística