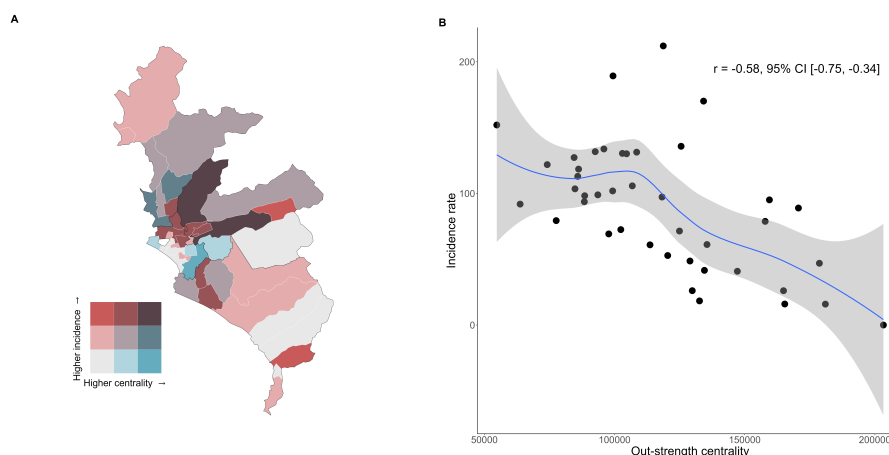


# The association of commuting patterns and the incidence of Tuberculosis in Lima, Peru: A network analysis

Anónimo

**Abstract Background.** Peru is one of the countries with the highest burden of Tuberculosis (TB) in the region of the Americas. The provinces of Lima, where the capital is located, and Callao report more than half of the cases in the country and nearly half of the deaths. These are the most important metropolitan areas of the country, with millions of people commuting every day mainly to study or work. Most people commute for many hours on public transportation in overcrowded conditions, which can lead to an increased risk of contracting TB. This study leverages the use of network analysis techniques to characterize the commuting patterns and investigates its association with the incidence of TB in the province of Lima. **Methods.** District-to-district commuting flow data was extracted from the 2017 Peru Census for commuting to school, commuting to work, and commuting to school and/or work. Each commuting flow dataset was represented as a weighted graph and analyzed using network analysis techniques in R version 4.3.0. We described the geographical subregions of Lima by the origin and destination of their commuting flows and by their burden of TB in 2017. From each graph, we computed indicators of centrality in the graph for each district and evaluated the association of these indicators with the incidence of TB in the year of study. Finally, we built generalized linear and additive models to model the incidence rate using poverty indices and pollutant concentration measurements as predictors, and for each centrality measure we assessed whether adding them to the models improved the fit. **Results.** We found that the subregions of East, North and South Lima had a higher percentage of people that commute to other subregions and within the same subregion than the subregion of Center Lima. Concurrently, the former subregions had the highest incidence rates (IR - per 100,000 population). Looking at the flow rates and centrality measures by districts, a moderate positive linear correlation between the intra-district flow rate to schools and the IR ( $r = 0.55$ , 95% CI [0.25, 0.74]), and a moderate negative linear correlation between the out-strength centrality (i.e. out-district flow rate) to schools and the IR were found ( $r = -0.58$ , 95% CI [-0.75, -0.34]) (Figure 1). A better fit to the data was found using Generalized Additive Models (GAM) with a Poisson distribution for the counts of cases of TB and a log-link function. The best model without a centrality measure included the percentage of people living with at least one unsatisfied basic need and the concentration of PM2.5 (AIC = 382.51). On the other hand, the best model with a centrality measure included UBN, the concentration of NO2, and the in-closeness centrality measure from the commuting to work network (AIC = 351.16), meaning a moderate quality of fit improvement of 8.2%. **Conclusions.** An association was found between the commuting patterns of the Lima subregions and the burden of TB. At a district level, an association was found between some centrality indicators that measure the outward and intra flow rate, and the incidence rate. It was also evidenced that some centrality indicators together with socio-demographic and environmental predictors can be used to model the incidence of TB. We found that the effects of these predictors may be better explained using non-linear terms. These results further show that human mobility is an important social factor when it comes to understanding the dynamics of TB transmission.

**Palabras clave:** Tuberculosis - Commuting patterns - Network analysis



**Figure 1.** A: Bivariate choropleth map of the incidence rate per 100,000 people and the out-strength centrality. B: Scatterplot for the incidence rate and the out-strength centrality showing a moderate linear correlation.