# A MULTILEVEL AND SPATIAL ANALYSIS OF GEOGRAPHICAL VARIATION AND DETERMINANTS IN TUBERCULOSIS INCIDENCE IN BOLIVIA

**NICOLAS SARAVIA**

**SEPTEMBER 2015**

Presented as part of, and in accordance with, the requirements for the Final Degree of MSc in Human Geography: Society and Space at the University of Bristol, School of Geographical Sciences, September 2015.

University of BRISTOL School of Geographical Sciences

**CERTIFICATION OF OWNERSHIP OF THE COPYRIGHT**

**IN A TYPESCRIPT OR MANUSCRIPT**

***Dissertation presented as part of, and in accordance with, the requirements for the Final Degree of MSc in Human Geography: Society and Space at the University of Bristol, School of Geographical Sciences.***

I hereby assert that I own exclusive copyright in the item named below. I give permission to the University of Bristol Library to add this item to its stock and to make it available for use and re-copying by its readers.

| | |
|---|---|
| **AUTHOR** | NICOLAS SARAVIA |
| **TITLE** | A MULTILEVEL AND SPATIAL ANALYSIS OF GEOGRAPHICAL VARIATION AND DETERMINANTS IN TUBERCULOSIS INCIDENCE IN BOLIVIA |
| **DATE OF SUBMISSION** | 3rd of September of 2015 |

## Abstract

*Tuberculosis remains a public health issue of importance in Bolivia, with an incidence total of 5487 in 2014 (PNCTB); however, there is a large variation on where these new cases are concentrated. Previous research on the subject in Bolivia has focused almost exclusively at the department level and in hospital case studies. Thus, this dissertation studies the situation using multilevel modelling and spatial analysis to account for differences in geographical divisions, and determine the social and infrastructural determinants impacting the dissemination of the disease. The results confirm some of the theories previous literature on the subject in regards to the effects of poverty on incidence. However, this study further contributes by analysing sanitary condition predictors, and more importantly, by suggesting a positive correlation between incidence and primary school coverage, thus proposing these centres as major points of contagion in need of greater control.*

## Word Count: 10,453

## Acknowledgements

I would like to thank my adviser, Dr. Winnie Wang, for her time and feedback. I would also like to thank Ms. Silvia Padilla and Dr. Anna Volz of the Pan-American Health Organization for their assistance in obtaining data, and Dr. Dennis Mosqueira of the Bolivian National TB Control Programme, for providing me with the data for incidence cases.

I would like to thank my family for their support during the months of writing this dissertation.

# Author’s Declaration

*I declare that the work in this dissertation was carried out in accordance with the requirements of the University’s Regulations and Code of Practice for Taught Postgraduate Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, this work is my own work. Work done in collaboration with, or with the assistance of others, is indicated as such. I have identified all material in this dissertation which is not my own work through appropriate referencing and acknowledgement. Where I have quoted from the work of others, I have included the source in the references/bibliography. Any views expressed in the dissertation are those of the author.*

SIGNED: ………Nicolas Saravia………….     DATE: ……2nd of September of 2015……….

# Table of Contents

## List of Tables



## List of Figures

# Chapter 1. Introduction

Tuberculosis (TB) remains as the second cause of death from a preventable or curable infectious disease, with an estimate of 1.5 million deaths and 9 million new cases worldwide in 2013, of which 95 percent take place in developing countries (WHO, 2015). Bolivia has the second most incidence of TB in the Americas, with 5487 new cases in 2014 (PNCTB, 2014). Currently, most scholarly articles and NGO reports studying the issue, do so considering only the status of the problem at the national level, with only a handful further analysing the variance and its factors occurring at the regional and local levels within countries. As of yet, there have been no published studies regarding the disease that have delved into the different geographical levels of Bolivia; thus, the purpose of this analysis is to provide a multilevel and spatial analysis to more accurately understand and determine the social and infrastructural factors that influence the variance in incidence across the municipalities, provinces, and regional departments of the chosen country of study.

## 1.1 Background on the disease

Approximately one-third of the population of the world has the *mycobacterium tuberculosis* in their system (Russell, 2010). In most cases, the immune system creates a granuloma or wall around the infection, which prevents it from developing into the disease and spreading; however, in 5 to 10 percent the illness develops, becoming infectious (CDC, 2014; Russell, 2007). The spread of TB occurs via air through droplets expelled through the mouth by saliva and mucus, which can remain airborne for hours (Huaroto, 2009). The bacteria require oxygen to survive, therefore it is usually found in the lungs; although, it can also spread through the bloodstream to different parts of the body (González-Marín, 2010). The disease is treatable with antibiotics; however, due to inadequate, incomplete or abandoned treatments, mutations in the genome of the bacteria develop resistance to medications such as rifampicin and isoniazid, resulting in a multi-drug resistant form (MDR-TB) (Álvarez-Galviria, 2013). In more recent cases, the bacteria has developed “resistance to at least three of the six classes of second-line anti-TB drugs”, a form known as extremely drug-resistant TB (XDR-TB), which has complicated even further the efforts to halt incidence (Migliori, 2007).

Based on the medical assessments suggesting the airborne nature of the disease by saliva and mucus, it may be worth to analyse where and how households discharge their sewage, in addition to variables that suggest agglomeration, such as population density.

## 1.2 Country Background

Bolivia is divided into three geo-administrative levels, 9 departments, 112 provinces and 327 municipalities (of which five have indigenous autonomy) (UNISDR, 2012). In 2014, the country had a total of 9,595,277 inhabitants, of which 1,544,315 live in its most populous city, Santa Cruz de La Sierra, located in the eastern lowlands (PNCTB, 2014). Bolivia's capital, La Paz, and its neighbouring city, El Alto, form a major population cluster in the western area of the country, with 813,509 and 900,566 people respectively (UNISDR, 2012). Along with Cochabamba in the centre of the country, these areas form an axis of interconnectivity in regards to transportation and migration (Ibid.). The life expectancy in Bolivia is of 66.93 years, and its total mean poverty percentage reaches 70.2 percent (World Bank, 2012; INE, 2012). There is great variation within the department of La Paz, as it has the highest mean poverty at the department level, at 78.12 percent of its population, yet the lowest at the municipality level, as the city of La Paz reaches 14.3 percent (INE, 2012). The department with the second most poverty is Oruro (75.76), while the lowest percentages are found in Cochabamba (66.87), Santa Cruz (57.11), and Tarija (48) (Ibid.). The municipality with the highest poverty is El Choro, in the department of Oruro, at 97.9 percent (Ibid.). The highest mean income per capita is observed in Pando, at 5240 Bolivianos (£494), which is also among the highest in poverty, with 72.57 percent (INE, 2012, PAHO, 2012). The other higher earning departments are Tarija (1476 Bolivianos) and Beni (1190), while the lowest earnings are in Cochabamba (703 Bolivianos), La Paz (684), and Santa Cruz (641) (PAHO, 2012).

Figure 1 below shows the map of Bolivia with the administrative division in 9 departments, 112 provinces and 327 municipalities distributed within the provinces.

This map also denotes the geographical boundaries of Bolivia with five countries, Argentina, Brazil, Chile, Paraguay and Peru.


Fig. 1. Map of Bolivia denoting department, province, and municipality boundaries, produced using ESRI ArcGIS, utilising PNCTB (2014) data and Hijmans (2015) Shapefile boundaries.

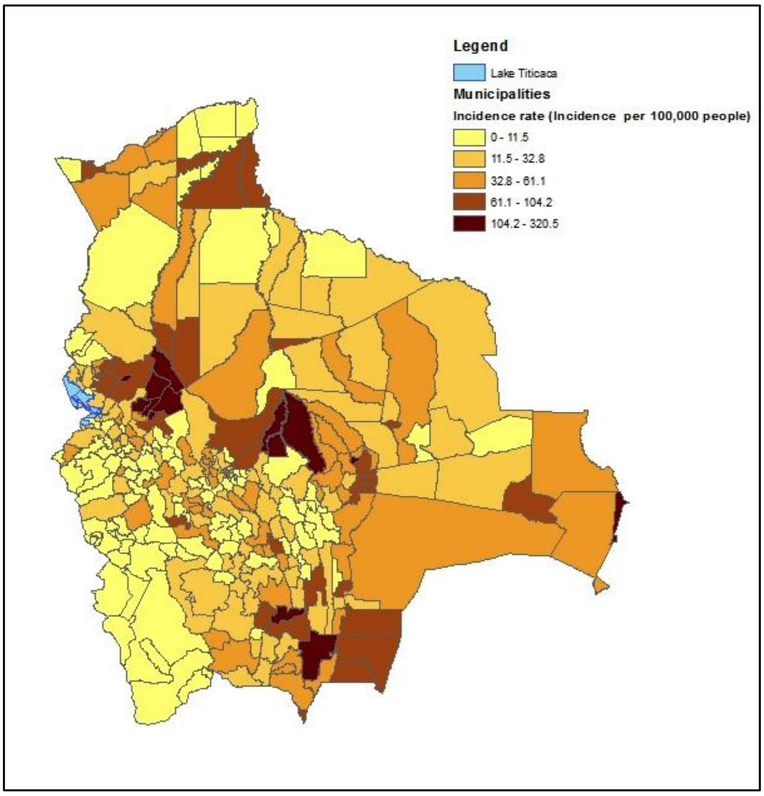

Fig. 2 Map of incidence rates (per 100,000) in Bolivia, produced using ESRI ArcGIS utilising PNCTB (2014) data and Hijmans (2015) Shapefile boundaries.

## 1.3 Tuberculosis in Bolivia

The highest TB incidence rate in municipalities occurs in Tipuani, Larecaja province, in the department of La Paz, with 320.5 cases per 100,000 people (PNCTB, 2014). On the other hand, municipalities such as Lagunillas and Moro Moro in Santa Cruz, Llica and Tahua in Potosí, and Coipasa in Oruro, among others, have zero new cases of TB (Ibid.) In Tipuani, the incidence rate has ameliorated since 2011, when it reached 608.9. Its worst year was in 2012, with 705 cases per 100,000 individuals (Ibid.). More information on the issue and causality theories of tuberculosis in Bolivia is explained in the literature review in the next Chapter.

# Chapter 2. Literature Review

## 2.1 Introduction

The available literature regarding the study of TB and its social and infrastructural causality factors offers a diverse range of risk determinants. Social factors such as income, poverty, homelessness, and level of education tend to overlap in studies concerning developing, as well as developed countries, with a variety of results in relation to risk causality. In the global context, numerous papers have proposed overcrowding as a major determinant of risk, as it increases the likelihood of exposure due to poor air quality and ventilation, while poverty and malnutrition increase susceptibility to the bacteria (Hargreaves, 2011; Murray, 2011) An ecological study, using non-linear regression, proposed a direct relation between countries with low Human Development Index and high TB incidence rates (Castañeda-Hernández, 2013).

## 2.2 Literature on developed countries

A descriptive statistics study of TB incidence across Europe, found an association with socioeconomic inequality, while it did not associate net migration, smoking, and social trust as significant factors (Ploubidis, 2012). In another study, using a logistic regression of a sample of 248 adults in Estonia proposed "education level below higher", unemployment and not having a residence, as important predictors in pulmonary TB incidence (Tekkel, 2002). However, in an analysis focusing on Spain at the national level, Kriging interpolation was implemented to estimate incidence across the population, identifying education level variables as not statistically significant, while having determined overcrowding, unemployment, illiteracy, immigration and immunodeficiency as significant factors (Gómez-Barroso, 2009). The methodology presented its limitations due to the lack of data for municipalities, as spatial shading may alter the accuracy of incidence levels based only on the distance to provincial capitals.

More specifically, in the case of Barcelona, a uni-variate analysis performed by chi-square, based on a sample of 892 cases of native and foreign-born persons with TB, determined risk factors associated to alcohol abuse, “residence in the old city district”, human immunodeficiency virus, diabetes, and homelessness (Borrelli, 2010). A case-control study of 112 individuals with TB residing within different postcodes in Liverpool, England, found that they were “four times less likely to have additional bathrooms” (Tocque, 2001).

In the United States, a Center for Disease Control study identified homelessness in six percent of the total reported cases between 1994 and 2010 at the national level (Bemrah, 2013). More specifically, in a correlational study at the states level, income inequality, social capital, and poverty were proposed as the main factors influencing new cases (Holtgrave, 2004). A multilevel analysis of the state of Washington, focusing on 2161 cases of individuals within neighbourhoods delineated by ZIP codes, associated “socio-economic disadvantage” to higher incidence (Oren, 2012).

A multivariate regression analysis of notification rates from 1988 to 1997 in Tokyo, Japan, propose associations of TB incidence with population density, lack of housing space, among other factors (Nishiura, 2003). On the other hand, a spatial analysis of Hong Kong suggested density as a non-significant predictor, and associated lower income and low educational achievement as the most significant variables (Chan-Yeung, 2005).

## 2.3 Literature on developing countries

Poverty has been suggested as the main determinant in risk of TB in studies such as a logistic regression of rural China and a spatial analysis of Hermosillo, in northwestern Mexico (Jackson, 2006; Álvarez-Hernández, 2010). However, contrary to most studies suggesting poverty as a significant predictor in connection to TB, a control study using logistic regression of 265 cases, of which 174 were of individuals with TB in Zambia, propose an association between prevalence and relative wealth in households (Boccia, 2011). Similar results have appeared in studies of cases in Malawi and Peru (Ibid.). Additionally, low education has been proposed as a main risk determinant in a case-control study of 189 patients in Bangalore, South India (Shetty, 2006). Lower education was also a factor of importance in a study using logistic regression, focusing on 202 controlled cases in Cartagena, Colombia, which also determined alcohol consumption, illiteracy as causality for TB prevalence[1] (Castillo, 2013).

### 2.4 Literature on neighbouring countries

In the case of Brazil, a multilevel, case-controlled study of 1452 individuals with tuberculosis within 52 areas in Recife, Pernambuco, suggested a significant risk of TB incidence in high illiteracy areas and in unemployed and low-income individuals (Ximenes, 2009). Through a spatial clustering and Poisson models analysis, lower socioeconomic status was found to be the main cause of TB incidence in Vitoria, Espiritu Santo (Maciel, 2010). A hybrid ecological and time-tendency study focusing on Ribeirão Preto, São Paulo, argued a higher risk among low-income families, in addition to individuals with less than three years of education (Hino, 2011).

A study of Lima, Peru, using descriptive statistics for a random sample of 150 public transportation commuters presenting pulmonary TB symptoms found that 39 percent of the individuals lived in overcrowding conditions, while an approximated 46 percent travelled daily on crowded public transportation (Horna-Campos, 2007). A descriptive analysis of a total of 106 TB patients at a hospital in Cordoba, Argentina, found unemployment as the most significant socio-economic variable, in half of the cases (Rivera, 2014). While these two focused studies may provide a glimpse into the issue within specific areas, they do not offer answers in terms of causality or variation.

### 2.5 Literature on Bolivia

As mentioned in the introduction section of this dissertation, most works discussing tuberculosis in Bolivia are limited to overviews of the situation at the cross-national, national and departmental levels, with a handful involving case-control studies in

[1] Prevalence refers to the total number of cases.

specific hospitals. Among the cross-national studies of the subject, Sobrero and Peabody (2006) argued that the main factor for the continuing spread of TB in Bolivia and its neighbouring countries is the absence of a directly observed treatment strategy, or DOTS (based on political commitment, quality detection, supervised treatment, drug management system, and monitoring), rather than a socio-economic factor (WHO, 2015b). More specifically, a case-control study of 34 pulmonary tuberculosis cases in the Modelo Corea Municipal Hospital of El Alto, suggests overcrowding in housing, temporary migration outside the country, and nutrition deficiency, as the main factors in new cases (Zubieta, 2012). The small number of individuals considered limits the study, however, it benefits from the controlled method by following up with the cases and observing trends over time.

Another study, using descriptive statistics of 1444 cases at a children's hospital in Cochabamba, suggests a connection between rural and lower socio-economic status with TB incidence (Barrios, 2012). However, the method of analysis overlooked the complexity of the issue, by not considering the interaction or dependence between predictor variables and the disease.

## 2.6 Contributions to literature

Most of the literature on the issue of tuberculosis and its determinants has focused on the hospital and local levels, with few examples using multilevel modelling or other techniques to account for the geographical variations that may occur in relation to the predictors' association with TB incidence. Thus, the aim of this dissertation is to determine whether the factors mentioned throughout the literature, such as poverty, population density, and income, have an influence in new cases at different levels in Bolivia. In addition, this study contributes to the literature by analysing infrastructural variables such as those pertaining to sewage discharge, source of water consumption, and primary school coverage.

# Chapter 3. Methodology, Data Preparation and Evaluation

## 3.1 Methodology

The objective of this study is the analysis of the high levels of incidence of TB, searching for the factors that are related to this incidence, based on the available data and the hypothetical models that have been proposed to explain this complex problem. The method consists firstly on selecting a comprehensive set of variables accounting for social factors such as poverty, income per capita, education; together with the inclusion of variables relative to availability of hospitals and to the number of health personnel in level 1 units

Due to the extremely high levels of contamination of TB, a set of relevant variables associated with sanitary conditions of common practice, particularly in areas of extreme poverty. It is to be determined whether these variables have or not a significant impact on the outcomes of the dependent variable. Some of these variables account for sewage discharges on streets, rivers, and lakes, as well as variables that account for the origin of the drinking water, obtained also from rivers, wells, public faucets, etc. The Section that follows provides a list of all the proposed variables, along with the structure of the sample data in a form suitable to be used in the SPSS Mixed Model Analysis software.

In the first phase of the data analysis, which is presented in Chapter 4, a simple level multi-regression model is applied to the dependent variable versus the independent variables with the goal to determine the degree of variance of the intercepts and slopes of the linear regressions, to conclude the validity or justification of using a multilevel analysis on this study

The Mixed Linear Models routine of the IBM SPSS software, is used to run multilevel analysis. The statistical model is described in Chapter 5, indicating the level 1 and level 2 structure of the model. It is also essential to centre the data points of the predictor variables to obtain a statistically significance for the intercepts as the mean of means of the dependent variable. The multilevel analysis is firstly applied to each predictor one by one into the model, to estimate its degree of influence on the outcome of the dependent variable, which is given by the relative value of the Schwarz's Bayesian Criterion (BIC). The combined effect of all the relevant predictors is analysed in Section 5.4 by including them into the model in a sequential manner. The results of this model provide the overall effect of the predictors on the research variable, considering the interaction within level 1 units and across level 2 units.

As a conclusion of the data analysis, a spatial vision of the interaction between the dependent variable and the predictors, as well as to determine the degree of influence that these predictors may have, Chapter 6 presents a set of maps obtained with ArcGIS showing the values of the variables within the level 1 units, nested within the level 2 boundaries. By using the method of Boolean intersection of these maps, the combined effect of these predictors to the research variable can be clearly visualised.

## 3.2 Data Preparation and Evaluation

The data sets in this study are obtained from different sources. The incidence of TB and Population, obtained from *Programa Nacional de Control de Tuberculosis*, Bolivia, 2014 (PNCT). The health care personnel and income per capita from the Panamerican Health Organization (PAHO), 2012; Poverty and water origin and sewage discharge are drawn from the *Censo de Población y Vivienda, 2012* (Census of Population and Housing) of the *Instituto Nacional de Estadística*, Bolivia. The number of hospitals is obtained from the *Unidad de Análisis de Políticas Sociales y Económicas* (UDAPE), data set *Establecimientos de Salud según Municipios*, 2013. Primary school coverage data set drawn from Indicadores de *Educación Alineados a las Metas del Milenio según Municipio*, UDAPE, 2012.

As a result of gathering data from different sources, distinct file types were converted into a single format. Moreover, the names of municipalities in these different datasets were often mismatched with formal and alternative names or spellings. Therefore, in order to bring together the data for usage, municipality names and IDs were consolidated into a single standardised form.

### 3.2.1 Variables

The dependent variable of the study consists of the tuberculosis incidence rate (Incidence), which is the total number of new cases of per year per 100,000 people.

Due to the contagious nature of the disease, the independent variables were chosen to reflect possible infrastructure and social factors that could have an effect in the geographical variance of new cases of the disease. The independent variables listed below have data values obtained at the municipality level, and are normalized to thousands of peoples to allow for the appropriate comparison across municipalities and higher level structures. The source of this data for each of the variables is also indicated in this list.

**Health care and infrastructure variables:**

**Healthcare personnel rate:** Number of doctors, nurses and/or other qualified personnel working in a municipality per 1,000 people, (PAHO, 2012).

**Hospital rate:** Number of hospitals in a municipality per 100,000 inhabitants, (UDAPE, 2013).

**Socio-economic variables:**

**Income per capita:** Mean income, in Bolivianos, of the people in a municipality. The gross domestic product divided by the population per municipality, (PAHO, 2012).

**Poverty:** Percentage of the total population of a municipality under the poverty line, defined as monthly income per household of 693.20 Bolivianos (£ 65.00), in 2012 (INE, 2012)

**Primary School Coverage:** Percentage of the net coverage of primary schools per municipalities, (UDAPE, 2012).

**Water Supply Infrastructure variables (INE, 2012):**

**Origin of Water- Tap rate:** Number of households receiving tap water via a supply network per 100,000 inhabitants.

**Origin of Water- Faucet rate:** Number of households receiving water via a neighbourhood public faucet per 100,000 inhabitants.

**Origin of Water-Truck rate:** Number of households receiving water via delivery trucks per 100,000 people.

**Origin of Water-Well rate:** Number of households receiving water via a well per 100,000 people.

**Origin of Water-River rate:** Number of households collecting water for consumption via rain, rivers, slopes, and/or ditches per 100,000 inhabitants.

**Origin of Water-Lake rate:** Number of households collecting water for consumption via lakes, ponds, and/or pools per 100,000 inhabitants.

**Sewage discharge variables (INE, 2012):**

**Sewerage rate:** Number of households using a sewerage system for sewage discharge per 100,000 people.

**Discharge-Septic tank rate:** Number of households discharging sewage into septic tanks per 100,000 inhabitants.

**Discharge-Cesspool rate:** Number of households discharging sewage into cesspools per 100,000 inhabitants.

**Discharge-Street rate:** Number of households discharging sewage onto streets per 100,000 people.

**Discharge-River rate:** Number of households discharging sewage into rivers and ravines per 100,000 inhabitants.

**Discharge-Lake rate:** Number of households discharging sewage into lakes, ponds and water pools per 100,000 inhabitants.

### 3.2.2 Data structure

For the purpose of enabling the comparison of different data sources, the chosen levels for this study are based on the geo-administrative divisions in Bolivia of the year 2009 (prior to the addition of seven new municipalities), which consist of 327 municipalities, nested into 112 provinces, within 9 departments.  Since the sampled data is for each of the municipalities, it is reasonable to consider these as level 1. It can be supposed that the municipalities have certain common characteristics within a given department; however, it is expected that there would be a differentiation of these characteristics between departments. The variance across departments will be analysed in a first approach using linear regressions of the Incidence versus the predictors per department, presented in the next chapter.  Based on the fact that the number of departments is too few, and the requirements of the validity of the multilevel analysis, it is anticipated that the department would not be a good choice as a second level in a multilevel analysis.

Another alternative approach to conform a multilevel structure of the data is based on the fact that the 327 municipalities are indeed distributed into 112 provinces, resulting in three to four municipality units per province, which again are too few to be considered a representative sample of Level 1 within Level 2 layer. However, by analyzing the geographical locations of the provinces within the departments it seems reasonable to come up with a third alternative for structuring of the data. As shown in Fig.1 in the Introduction, the majority of the provinces are located at the boundaries with Argentina, Brazil, Paraguay or Chile. Therefore, the departments are subdivided in interior-provinces and boundary-provinces, grouping them into Divisions. The municipalities will be nested into 21 divisions, with an average of 17 municipalities per division. Appendix A includes a Table with the number of municipalities nested within divisions, and departments. This re-structuring of the data increases the size of the level 1 units per division and the number of level 2 divisions, without altering the data at all, preserving the nesting of the municipalities within the divisions, which better represents the grouping of the level 1 units in smaller division clusters than within the department structures. It is expected that this assumption will be confirmed in Ch. 4 below, by estimating the intra-class coefficient (ICC) between the variance of Incidence between divisions and the variance of Incidence within divisions.

## Chapter 4. Single-level Multiple Regression Analysis

With the objective of obtaining a preliminary idea of the relation of the independent variables, or predictors, and the dependent variable, Incidence, a linear regression analysis was conducted using the IBM-SPSS software. The validity of this model, is based on the supposition that the statistical samples have been obtained through simple random sampling and such also assumes that the subjects of study, municipalities, are independent from each other. The objective of this dissertation is to demonstrate that this second assumption is not sustained in this case study and the characteristics of the municipalities grouped into divisions are not independent units due to the relations between the population that reside in these geographical units, with similar economic, social, and structural situations in education and public health. The single-level model does not consider the common or similar aspects between the samples of the municipalities, or the fact that these units are nested into higher levels of hierarchy, and thus, it is expected that the coefficient estimations and their residual errors were biased in their predictions.

Despite this limitation, the single-level analysis results useful in obtaining an idea of the degree of variation of the linear regressions when these are done separately for each division, taking into account the samples of Incidence and the independent variables of the municipalities within the corresponding division. Fig. 3 below shows the scattered plots with the estimated linear fit for each of the regressions for each division between Incidence and the predictor variable Healthcare personnel rate (HPR). These plots show the line fitting corresponding to each division, with a large range of values of intercepts from near zero to 35, slopes between -17 to 19 and $R^2$ coefficients between 0.002 to 0.30 across divisions. This could suggest a significant distribution of Incidence within the 21 divisions. In some of the divisions, negative slopes are observed, which would indicate that the existence of health personnel in such divisions contribute to the decrease of the number of incidences of TB, while the divisions with positive slopes indicate that a presence of health personnel is not sufficient to reduce the degree of incidence.

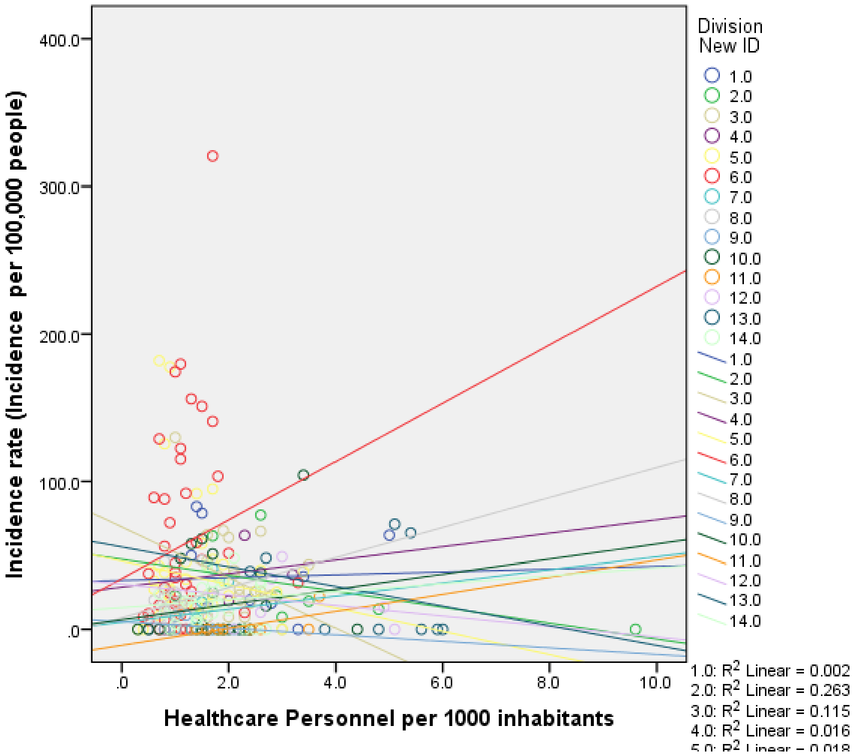


Fig. 3. Linear regressions of Incidence versus number of health personnel per 1,000 habitants for some of the 21 divisions

Similarly, Fig 4 illustrates the variability in the dependence of Incidence on the independent variable in relation to the quality of water consumed by the population in the municipalities within each of the divisions. In this case, the slopes vary between -0.02 a 0.07, the intercepts from near zero to 53, and the $R^2$ coefficients in the range of near zero to 0.9. The former suggests that there are possible divisions in which there is very little variance in Incidence in respect to the mean value, and others in which the variance is very large in respect to the mean of the municipalities within those divisions. The great variation in slopes could imply that in the divisions with steeper slopes, discrete increases in the number of people with access to this type of water could have a greater impact in the Incidence of TB than in the cases of divisions with flatter slopes. In those in which the slope is near zero, it is expected that the Incidence is not sensitive to the number of people who consume this type of water, which would imply that the source is healthy and not contaminated with TB.

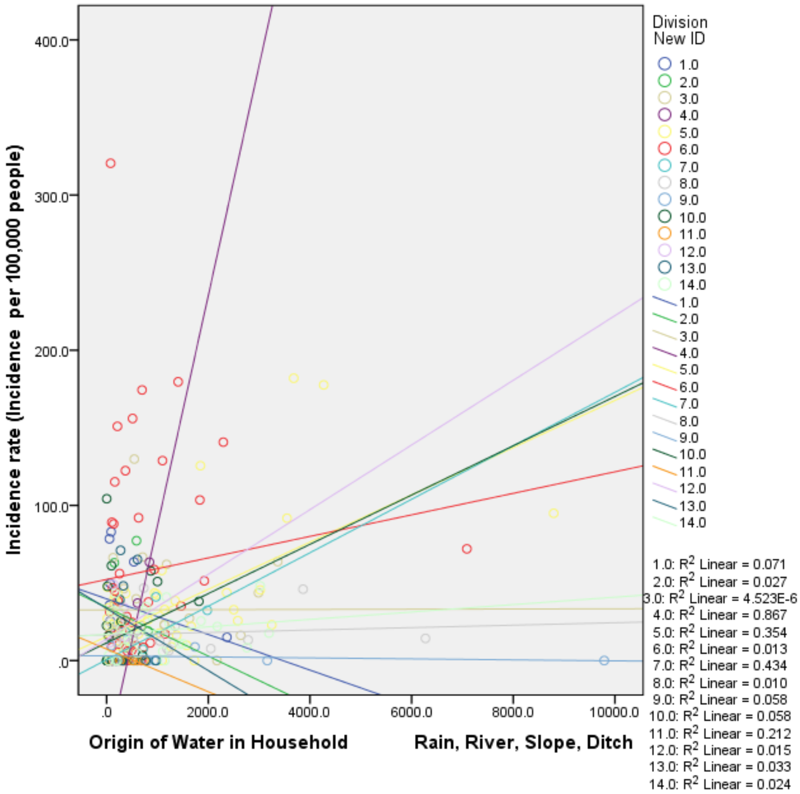


Fig. 4 Linear regressions of Incidence versus number of people using non-treated water for some of the 21 divisions

These types of single level multi regressions were conducted for most of the predictors of the independent variables listed in Chapter 2, with very similar patterns in the variability of the intercepts and slopes of the linear regressions across the divisions.

# Chapter 5. Two Level Regression Model

According to the theories of multilevel statistical analysis, the focus of this study complies with the necessary conditions that justify their application to the chosen research for looking for the relationships between TB incidence and a variety of independent variables including public health, economics, education and those related to the source of water that people use for their everyday needs, and the sanitary conditions where they lived.

## 5.1 Null Model

The first multilevel model, known as Null model in the literature, does not include any predictors. the Null model allows to partition the variance in Incidence into its components within the divisions and between the divisions. Indexes (i) samples will be associated with level 1 variables and (j) samples with level 2 variables. It is assumed that level 1 units are nested within level 2 unit, which is a good model for this research subject, where level 1 can be identified with municipalities and level 2 with divisions. The equation for this model can be written as

$$\mathbf{I_{ij} = \alpha_{0j} + \varepsilon_{ij}}\,, \qquad \textbf{(1)}$$

where $\mathbf{I_{ij}}$ is the Incidence in municipality (i) within the division ( j), $\boldsymbol{\alpha_{0j}}$ is the intercept of the municipalities within the division (j), and $\boldsymbol{\varepsilon_{ij}}$ is the variation in estimating the Incidence in the municipality ( i) within division (j). Since there are no independent explanatory variables included in this model, the level of Incidence is due to the variance of level 1 and level 2 components. Between the divisions, variation in intercepts can be represented as

$$\boldsymbol{\alpha_{0j} = \delta_{00} + u_{0j}}\,, \qquad \textbf{(2)}$$

where $\boldsymbol{\delta_{00}}$ represents the mean value of the intercepts of municipalities at level 1 and $\mathbf{u_{0j}}$ is the variation in intercepts around its mean value for division (j) at level 2.

And by substituing Eq (2) in (1) a single equation combining these two relations can be written as:

$$\mathbf{I_{ij} = \delta_{00} + u_{0j} + \varepsilon_{ij}} \; . \qquad \textbf{(3)}$$

This equation corresponds to the model including the estimates for the fixed mean value $\boldsymbol{\delta_{00}}$, also known as the mean of means (for centered data) for all the divisions, and the estimates of the partition of the variance between the random parameters, $\mathbf{u_{0j}}$ for level 2 and $\boldsymbol{\varepsilon_{ij}}$ for level 1.

Furthermore, this model allows for the calculation of the proportion of the variance between divisions with respect to the variance within the divisions. This proportion, which is known as intraclass correlation (ICC), can be calculated in IBM SPSS via the Variance Components or MIXED procedures. The expression for ICC, according to Goldstein (2011) and Heck and Thomas (2015), is defined as:

$$\mathrm{ICC} = \frac{\sigma_{\mathrm{B}}^2}{(\sigma_{\mathrm{B}}^2 + \sigma_{\mathrm{W}}^2)} \qquad \textbf{(4)}$$

where $\sigma_B^2$ represents the variance between divisions and $\sigma_W^2$ represents the variance within divisions, and ICC corresponds to the ratio of between divisions to the total variance. Therefore, high values of ICC imply that there is significant variability between divisions and a multiple level analysis would provide valuable information about nested model. Values of ICC lower than 5% would indicate that such a model would not provide any more information than a single regression model. (Nezlek, 2011; Goldstein, 2011; Heck and Thomas, 2015).

## 5.2 Data Centring

Mathematically, the Intercept of Incidence corresponds to the Incidence value when the predictor variable is zero. In this study, this is true for certain independent variables, such as those related to the sanitary conditions of sewage disposal and the source of water for consumption. However, for variables such as poverty, income per capita and many others, the intercept does not correspond to a zero for the predictor variable. Moreover, the intercept does not have any statistical significance or interpretation unless the intercept corresponds to the point where the mean value of the predictor is zero.

To centre the data of the independent variable $X_{ij}$ around a zero mean, it is necessary to translate the sample data respect to its mean value $\bar{X}$, that is,

$$X_{ij}\,(\text{grand-centered}) = X_{ij} - \bar{X}\ , \qquad (5)$$

where $\bar{X}$ represents the grand mean of the sample. All the data for the predictors used in the analysis of the following sections are grand centred, and therefore, the intercepts correspond to the Incidence mean values in each of the models.

## 5.3 Null Model Analysis

Using the SPSS linear mixed models option, the null model analysis was performed considering the 327 municipalities in level 1, nested into 21 divisions in level 2 and Incidence as the dependent research variable, with no predictors being included in this case.
The model considers only random effects on intercept parameter and defining divisions as the subject grouping. The software runs provide the estimate for the means of means ($\boldsymbol{\delta_{00}}$), and the variances between and within divisions, respectively. Table 1 below shows the results of the null model for this case.

Table 1 Null model for Incidence

Table 1 (a) Model Dimension

| | | Number of Levels | Covariance Structure | Number of Parameters | Subject Variables |
|---|---|---|---|---|---|
| Fixed Effects | Intercept | 1 | | 1 | |
| Random Effects | Intercept | 1 | Variance Components | 1 | DivisionNew |
| Residual | | | | 1 | |
| Total | | 2 | | 3 | |

Table 1 (b) Estimates of Fixed Effects

| Parameter | Estimate | Std. Error | df | t | Sig. | 95% Confidence Interval | |
|---|---|---|---|---|---|---|---|
| | | | | | | Lower Bound | Upper Bound |
| Intercept | 28.197638 | 3.809604 | 18.801 | 7.402 | .000 | 20.218324 | 36.176952 |

Table 1 (c) Estimates of Covariance Parameters

| Parameter | | Estimate | Std. Error | Wald Z | Sig. | 95% Confidence Interval | |
|---|---|---|---|---|---|---|---|
| | | | | | | Lower Bound | Upper Bound |
| Residual | | 1345.9672 | 108.129434 | 12.448 | .000 | 1149.879 | 1575.49305 |
| Intercept [subject = DivisionNewID] | Variance | 169.15806 | 87.331725 | 1.937 | .053 | 61.49509 | 465.312798 |

Schwarz's Bayesian Criterion (BIC) : 3309.686

From these Tables, we obtain the estimates of the means of means $\delta_{00} = 28.197$ , and the variances $\sigma_B^2$ = 169.158 and $\sigma_W^2$ = 1345.967. Using equation (4), the estimated ICC value for this case is close to 11.2%, which means that the intercepts vary significantly across divisions, with about 11.2% of the total variability in Incidence due to between division effects, and 88.8% due to within division effects. These results justify the applicability of multilevel analysis to the chosen research model for TBC in Bolivia using the available data for municipalities within a division hierarchy structure.

For comparison purposes, and to insure that the division grouping of the samples was a better approach than the one using departments, the null analysis was performed for Incidence considered the municipalities, nested within the 9 departments. The null analysis provided a mean of mean value of $\delta_{00} = 31.139$, and variances of $\sigma_B^2$ = 87.81 and $\sigma_W^2$ = 1447.46, which gives an ICC of only 5.72%, considered to be too low for a multilevel analysis to be applicable or useful.

This confirms that the approach of selecting a better grouping structure based on divisions with a larger sample size at level 2, nesting the municipalities also on sizable sample subgroups within the divisions, is the appropriate one to expect meaningful results from the two-level analysis.

## 5.4. Level 1 Random Intercept Model and Analysis

### 5.4.1 Random Intercept Models

The interest of using the multilevel analysis in this study is to be able to determine the relationships between Incidence and independents variables or predictors, with the added complexity that the samples at the municipality level are not considered independent and that these municipalities are nested into divisions.

The simplest model is to analyze the effect of a particular predictor on Incidence, assuming that it is the only independent variable acting on it and that the relationship is linear. It is also assumed that this interaction will be the cause of a random variation on the intercepts. This model is similar to the Null model in the sense that it allows for random variations of the intercepts, but a new term proportional to the predictor variable $X_{ij}$ is added to it. Equation 5 shows the modified model

$$\mathbf{I_{ij}} = \boldsymbol{\delta_{00}} + \mathbf{u_{0j}} + \boldsymbol{\varepsilon_{ij}} + \boldsymbol{\alpha_{1j}}\, X_{ij} \qquad \textbf{(5)}$$

where the new term $\alpha_{1j}$ is the slope of the linear regression of the municipality (i) nested into division (j). In this model it is assumed that the intercept changes in a random manner and that the slope is fixed, of the form $\boldsymbol{\alpha_{1j}} = \boldsymbol{\delta_{10}}$, that is, a constant value for the sample, with no random component. Then the equation for Incidence can be written as

$$\mathbf{I_{ij}} = \boldsymbol{\delta_{00}} + \boldsymbol{\delta_{10}}\, X_{ij} + \mathbf{u_{0j}} + \boldsymbol{\varepsilon_{ij}}\ , \qquad \textbf{(6)}$$

The first two terms on the right side of this equation can be interpreted as the fixed component prediction of Incidence outcomes, and the last two terms correspond to the variances of the slope and the variance at the municipality level.

The model will provide the values of fixed parameters of the mean of means and slope as well as the variances $\sigma_B^2$ of $\mathbf{u_{0j}}$ and $\sigma_W^2$ of $\boldsymbol{\varepsilon_{ij}}$, for a given predictor. Deviations of the value of the intercept $\boldsymbol{\delta_{00}}$, with respect to the mean value of the Null model and the OLS models, for different predictors, provide an estimate of the degree of variability that a particular predictor may have on the outcome of Incidence. However, the SPSS MLM provides a model deviance statistics, among others, the Akaike Information Criterion (AIC), and the Schwarz Bayesian Criteria (BIC) which are used to determine how well the model fits the data and even more important, to examine the improvement of the model fit by comparing two sequences of models that is, in particular, when an additional variable is added to the previous model. It is well established in the literature that there is no substantial improvement between two models when the difference of their BIC values is less than 2, and that there is a significant difference between two models when the BIC difference is greater than 10, In this case, the model with the lower BIC is selected (Raftery, 1966).

### 5.4.2 Analysis of Random Intercept Model with predictors included individually

In this section, several random intercept models are obtained by using the SPSS Linear Mixed Model, starting from the Null model as a base line. One predictor at the municipal level is added to the Null model at a time. The 15 models correspond to the 15 independent variables $X_{ij}$ described previously in the methodology section. First, social and economic variables are introduced in the model, followed by variables related to health, and finally, variables related to the sanitary conditions in the individual municipalities.

The results of these models are presented in Table 2 below. The coefficients obtained for the Null model in previous section are shown on the first column of the Table, indicating the intercept value of 28.197, the ICC of 11.2% and BIC of 3309.686. The BIC for the Null is considered as the reference model to select those models with a BIC value below BIC for the Null model. Following the criteria mentioned above, the model with the lowest BIC value, and under that of the Null model, is regarded as the best model to fit the data, with the predictors that most effect have on the Incidence. The last column in Table 2 shows the difference of the BIC number of each of the models with respect to the BIC of the Null Model.

The models are presented in descendent order from the best BIC values on the top rows of Table 2. Model 1 corresponds to predictor Primary school coverage with BIC difference of 29.599 under the value of BIC for the Null model, and intercept of 29.694. Then, it follows Model 2 with predictor Poverty with BIC difference of 24.430 also below the Null model, and intercept de 28.724. Model 3, for the predictor Health personnel rate, has a BIC 0.214 above the BIC of Null and intercept of 29.176. All the intercept values show standard error of the order of 10% and p-values <0.001. Models 4, 5 , 6 and 7 have BIC values greater than in the first three models, when compare one-to-one independently. Nevertheless, these predictors may still be relevant and influential on the Incidence of TB when they are considered in aggregate models, in which several predictors are included together in the multilevel model. These multi-predictor models are analysed in the following section. Additionally, it is worth to notice that the observed ICC values for all models are smaller than the one of the Null Model but greater than 5, which is indicative that the cross-division effect is relevant for most of the models with individual predictors.

From these results it can be said that the interactions of the majority of the predictors related to sanitary conditions, such as sewage disposals and water sources for consumption, do not have an statistically significant effect on Incidence of TB in Bolivia, with the only possible exemption of the Street discharge of household rate, shown as Model 6 on Table 2. On the other hand, models 3, 4 and 5 clearly represent effects of both individual and structural influence of the level of poverty, hospital infrastructure on the degree of Incidence of TB. Besides these results, that are somehow expected in this research study when compared to the existing literature, Model 1 with predictor Primary school coverage appears to be very determinant on the outcome of Incidence in Bolivia. The multilevel analysis including all the relevant predictors considered in an aggregated manner, will provide a more conclusive argument to understand the role of this predictor in the most reliable model.

## Table 2

Multilevel Regression Models for Incidence versus several Predictors

| Model # | Predictor ($X_{ij}$) | Intercept $\delta_{00}$ | Intercept SE | Intercept p-value | Slope $\delta_{00}$ | Slope SE | Slope p-value | ICC (%) | BIC | BIC difference |
|---|---|---|---|---|---|---|---|---|---|---|
| **1** | **Null Model** | **28.197** | **3.8** | **<0.001** | **---------** | | | **11.16** | **3309.686** | **0.00** |
| 2 | Primary school coverage | 29.694 | 3.1 | <0.001 | 0.54 | 0.09 | <0.001 | 6.50 | 3280.087 | **-29.599** |
| 3 | Poverty | 28.724 | 3.3 | <0.001 | - 0.59 | 0.11 | <0.001 | 8.42 | 3285.256 | **-24.430** |
| 4 | Heath personnel/1,000 | 29.176 | 3.9 | <0.001 | -3.16 | 1.96 | <0.011 | 11.2 | 3309.90 | **0.214** |
| 5 | Hospital rates (x100,000) | 28.629 | 3.7 | <0.001 | -0.03 | 0.025 | 0.129 | 9.90 | 3313.569 | **3.883** |
| 6 | Street discharge household rate | 28.245 | 3.8 | <0.001 | $1.7x10^{-3}$ | $3.8x10^{-4}$ | 0.65 | 10.94 | 3318.79 | **9.104** |
| 7 | Income per capita | 28.64 | 3.9 | <0.001 | -0.002 | $3.1x10^{-3}$ | 0.448 | 11.51 | 3318.815 | **9.129** |
| 8 | Sewage river discharge rate | 28.266 | 3.8 | <0.001 | $3.1x10^{-4}$ | $6.5x10^{-4}$ | 0.64 | 10.10 | 3322.30 | 12.614 |
| 9 | Septic tank household rate | 28.189 | 3.8 | <0.001 | $2.0x10^{-4}$ | $4.6x10^{-4}$ | 0.651 | 10.60 | 3323.00 | 13.314 |
| 10 | Water source from lake/pond rate/100,000 | 28.460 | 3.7 | <0.001 | $2.4x10^{-4}$ | $2.0x10^{-4}$ | 0.215 | 10.58 | 3323.37 | 13.684 |
| 11 | Water source from well rate/100,000 | 28.534 | 3.7 | <0.001 | $1.1x10^{-4}$ | $6.5x10^{-5}$ | 0.093 | 11.30 | 3324.38 | 14.694 |
| 12 | Public faucet water rate/100,000 | 28.406 | 3.7 | <0.001 | $1.2x10^{-4}$ | $1.1x10^{-4}$ | 0.273 | 10.24 | 3324.9 | 15.214 |
| 13 | Population density | 28.512 | 3.7 | <0.001 | $3.6x10^{-5}$ | $1.7x10^{-5}$ | 0.042 | 10.40 | 3325.621 | 15.935 |
| 14 | Water source from rain/river rate/100,000 | 28.298 | 3.8 | <0.001 | $-1.8x10^{-4}$ | $1.8x10^{-5}$ | 0.530 | 10.79 | 3329.3 | 19.614 |
| 15 | Sewage house rate | 28.189 | 3.8 | <0.001 | $-2.0x10^{-6}$ | $1.0x10^{-5}$ | 0.84 | 11.20 | 3330.882 | 21.196 |
| 16 | Tap water household rate/100,000 | 28.206 | 3.8 | <0.001 | $-1.0x10^{-6}$ | $8.0x10^{-6}$ | 0.927 | 11.10 | 3331.279 | 21.593 |

### 5.4.3 Analysis of Random Intercept Model with multiple predictors

Based on the results of the previous section, several models were implemented to analyse the combined effect of predictors included sequentially into the basic Null model. Table 3 below summarizes the results of nine models, which result from various combinations of predictors that provide the best fitting to the data and the most significant set of variables that may affect the outcome of the Incidence. The coefficients obtained for the Null model of Section 4.1 are shown on the first row of Table 3 as a reference to more complex models. The last column corresponds to the difference of the BIC of a new model minus the BIC of a previous model. A negative difference number indicates a drop of the BIC value, and therefore an improved model.

Model 17 represents the effect of Poverty on Incidence that, when compared to the Null model, shows a slight increase of intercept (SE 3.8, and $p<0.001$), a reduction of the ICC to 8.4 and a drop of the BIC of 24.43. In Model 18, predictor Primary school coverage is added to Model 17 resulting in an additional drop of 13.97 in the BIC showing an improvement on Model 17. The addition of predictor Hospital rate to Model 18 results in an increase 5.44 of BIC of Model 19 with respect to Model 18. As the next step in Model 20, predictor Health personnel is added to Model 18, causing a drop of 9.8 of the BIC of Model 18, with an increase of the means of means of Incidence to 30.857, and maintaining a reasonable ICC value of 6.1. None of the Models 21 and beyond show a reduction of the BIC and therefore can be considered as statistically non-significant to cause any impact to Incidence of TB.

Besides confirming the effect of Poverty as an important variable impacting Incidence, these results also show that Primary school coverage seems to have a strong relation with Incidence as well, more than other variables such as Income per capita, Population density, among others cited in the literature reviewed in Chapter 2.

Firstly, the Primary school predictor is a variable that takes into account the number of primary (or elementary) schools with respect to the school age students living in a specific municipality. The results indicate that areas with high incidence rate are also areas of high attendance of children of ages 5 to 11 years old. Given the epidemiological known facts about germ dissemination of TB, it is plausible to propose that there may be a strong correlation between Primary school attendance and Incidence in certain municipalities in Bolivia.

Chapter 6 presents the analysis of the data in a graphical form, using ESRI ArcGIS programme to determine the consistency with the results obtained in this Section.

**Table 3**

Multilevel Regression Models for Incidence versus Multi – Predictors

| Model # | Predictors ($X_{ij}$) | Intercept | | | ICC (%) | BIC | Model comparison | BIC difference |
|---|---|---|---|---|---|---|---|---|
| | | $\delta_{00}$ | SE | p-value | | | | |
| **1** | **Null Model** | **28.197** | **3.8** | **<0.001** | **11.16** | **3309.686** | **--------** | **0.00** |
| 17 | Poverty | 28.724 | 3.3 | <0.001 | 8.4 | 3285.256 | **17 vs 1** | **- 24.43** |
| 18 | Model 17 + Prim. School coverage | 29.638 | 3.05 | <0.001 | 6.6 | 3271.286 | **18 vs 17** | **-13.97** |
| 19 | Model 18 + Hospitals rate | 29.725 | 3.04 | <0.001 | 6.4 | 3276.726 | **19 vs 18** | **+5.44** |
| 20 | Model 18 + Health personnel/1,000 | 30.857 | 2.99 | <0.001 | 6.1 | 3261.486 | **20 vs 18** | **-9.80** |
| 21 | Model 20 + Hospitals rate | 30.837 | 3.01 | <0.001 | 6.2 | 3266.990 | **21 vs 20** | **+5.50** |
| 22 | Model 20 + Income per capita | 30.62 | 2.99 | <0.001 | 6.0 | 3270.157 | **22 vs 20** | **+8.67** |
| 23 | Model 20 + Population density | 30.852 | 2.99 | <0.001 | 6.1 | 3269.964 | 23 vs 20 | +8.48 |
| 24 | Model 20 + Street Sew. discharge | 30.852 | 3.0 | <0.001 | 6.1 | 3270.913 | 24 vs 20 | +9.43 |
| 25 | Model 20 + Public water source | 30.85 | 3.0 | <0.001 | 6.2 | 3277.98 | 25 vs 20 | +16.50 |

## Chapter 6. Spatial Analysis of Incidence versus Predictors

The objective of this analysis is to represent, in a geographical form, the data for Incidence and the most relevant predictors found in previous Sections. The maps of Incidence and the predictors are generated by using the ESRI ArcGIS 10.3.1. Fig. 5 below shows the maps for the municipalities in Division 1, in the Beni department, for Incidence, Primary schools and Poverty, indicating the ranges in graduated colours; being the darker colour for the larger values. Map (d) represents the Boolean intersection of areas of Incidence, Primary school and Poverty with values greater than 15, 80, and 40%, respectively. These limits are considered as relatively high when compared to the values of the country overall. It is observed that most of the municipalities in this department show an overlap of these three most affecting predictors, which is consistent with the results of the multilevel analysis in the previous chapter.

Fig. 6 below also illustrates the maps for Division 16 in the Department of Potosí. The maps are displayed in the same order as in Fig 5. In this case, the Boolean intersection is obtained for values of Incidence, Primary school and Poverty greater than 13, 76, and 20%, respectively. Additional Intersection maps are presented in Fig. 8 Appendix B for other divisions where the shaded areas also indicate that the municipalities with higher levels of Poverty, and with the lager coverage of Primary schools, show the highest values of Incidence.

Besides proving consistency of the results obtained by the multilevel analysis, the maps also provide detailed information on how the Incidence and the predictors relate to the spatial dimension of the data. The identification of geographical areas or clusters of municipalities where the most affecting predictors concur with the highest values may be particularly in designing health control programmes and in developing new processes to control the disease.

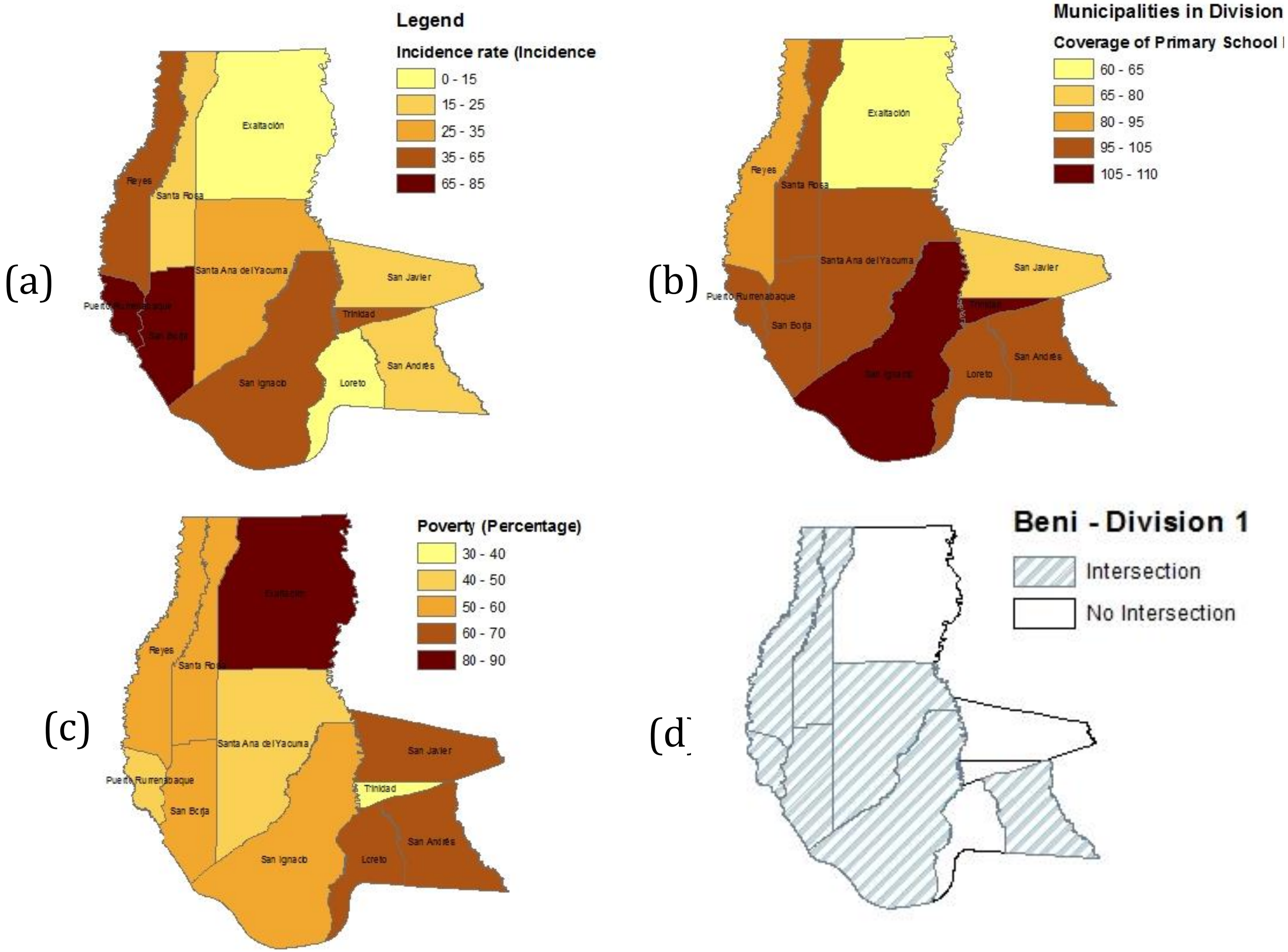


Figure 5. Maps of (a) Incidence, (b) Prim. School Cov., (c) Poverty. (d) Map pf Intersection , in Division 1, Beni department.

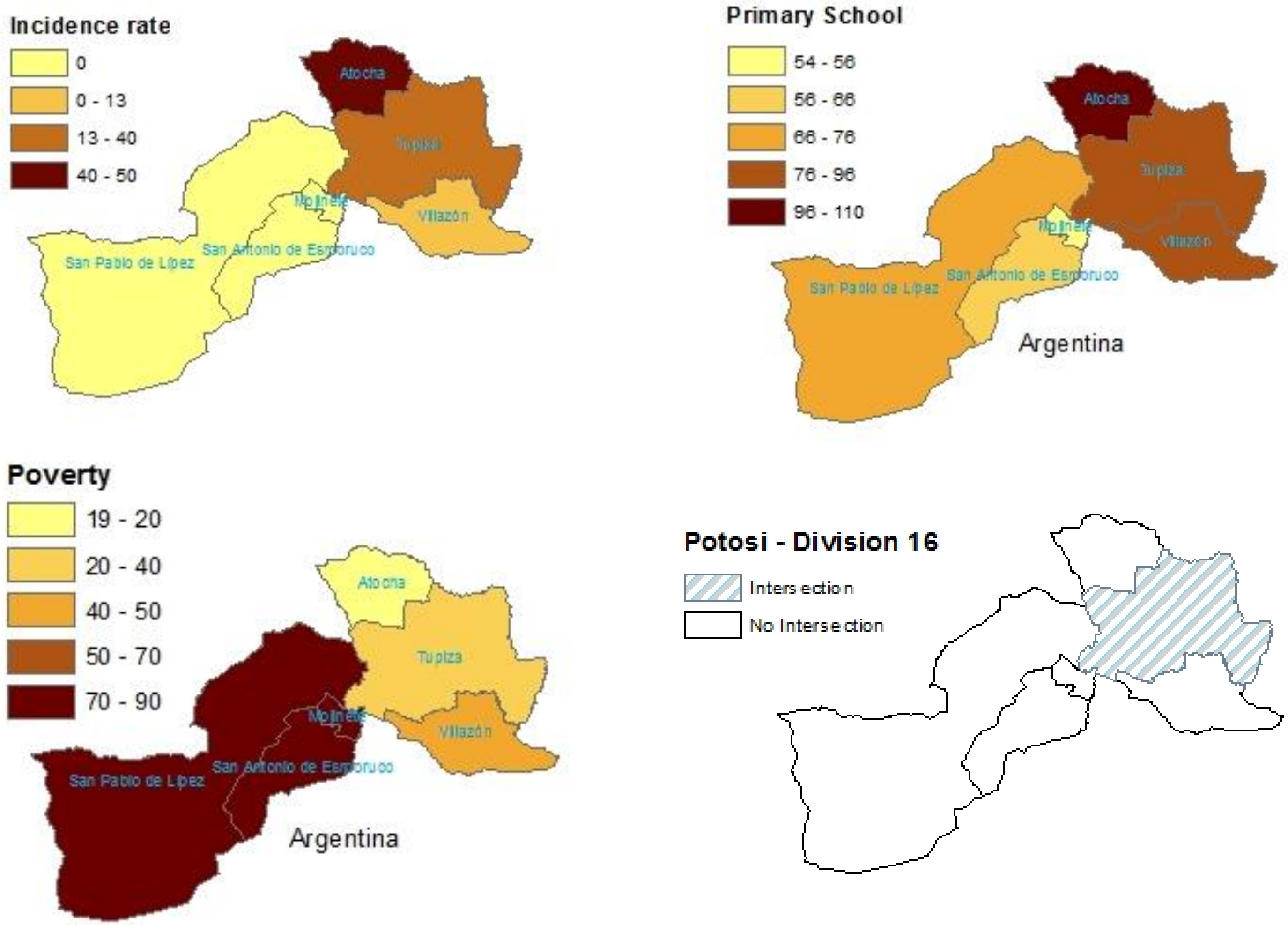


Figure 6. Maps of (a) Incidence, (b) Prim. School Cov., (c) Poverty. (d) Map pf Intersection , in Division 16, Potosí

# Chapter 7. Conclusions

## 7.1 Main Findings

The results obtained from this study based on the multilevel regression analysis, for the most part, support the published literature on the causes for the increase of TB incidence, in particular those related to the levels of poverty, availability of hospitals, and healthcare personnel. The positive correlation between the number of hospitals and health personnel with Incidence, is a consequence of the initiatives of the public health entities in Bolivia, which have focused their efforts on attacking the problem of incidence of the disease by the treatment of patients and follow-ups in health centres with specialised personnel.

The multilevel analysis in this study is very clear in determining the main effect of poverty, the most statistically significant predictor in the impact on Incidence. Also, the results of this dissertation indicate that, with the exception of Street sewage disposal, the variables related to the sanitary conditions have a negligible contribution to the incidence as indicated in the BIC parameters, with great error dispersion and large p-values. This is mainly because the effect is very localised in a few municipalities with a great quantity of null values of the majority of the rest of the municipalities.

Additionally, based on the multilevel analysis of this study, it can be concluded that the Primary school coverage has an important role in the impact on incidence, as it provides a good model in combination with Poverty and Healthcare personnel, with statistical parameters that assure a high degree of confidentiality and significant data fitting. Due to the efforts and assignation of special resources by the entities of the country and fund by international agencies, there has been a significant increase in the number of primary schools in isolated and low resource areas in Bolivia, especially in recent years; however, the extreme poverty levels remain very high. It is possible to conceive the situation in which school coverage and education are very high in these areas, yet the rest of the conditions associated to sources of contagion, overcrowding in housing, and insufficient sanitary conditions in the prevention of TB still persist, and could even be amplified by more contact in these educational centres. This could not be the case if sanitary control measures and reduction of contagion sources were implemented.

The spatial analysis of the data of this study based on the search for the conjunction of the effects of the main effect predictors on the incidence, provides results similar to those of the multilevel analysis, by obtaining geographic areas in which there is a positive concurrence in the increase of the level of incidence forming which could be denoted as “TB hot spots”. In this sense, the map diagrams with the Boolean intersections of the maps for each predictor and those showing Incidence clearly demonstrate this point.

Nevertheless, it is necessary to conduct additional analyses that would permit to prove in a convincing manner the causal relation between the primary school coverage predictor and incidence.

### 7.2 Limitations of the study and future research

As per the previous discussion, further investigation is necessary to establish causality and effect of the predictor primary school coverage on Incidence, by looking into additional health variables that could link the results obtained in this study. The study may be expanded to include variables pertaining to other diseases, such as influenza, chronic lung disease, diabetes, immunodeficiency virus, and Chagas, that compromise the immune systems of individuals, and thus, also could become susceptible to TB. Chagas is endemic to Bolivia, with an approximated prevalence reaching 1.1 to 1.8 million cases, and areas such as Santa Cruz possess high incidence of both diseases, thus a study analysing co-infection could provide additional answers (Medrano-Mercado, 2008). Data on health insurance adherence and types of healthcare facilities may also be worth examining in relation to TB prevalence and mortality. In addition, this study has been limited to considering all documented cases without distinction between pulmonary and extra-pulmonary TB. In this study, no differentiation was made between non-resistant, multi-drug resistant, and extremely drug-resistant TB, which could perhaps provide additional insight into the understanding of social and structural determinants of the different forms of the disease.

A longitudinal analysis could provide answers in regards to the impact of the factors on incidence over time, and to what could have ameliorated or worsened the situation in the different areas. The study may be expanded to include variables pertaining to other diseases, such as influenza, chronic lung disease, diabetes, immunodeficiency virus, and Chagas, that compromise the immune systems of individuals, and thus, also could become susceptible to TB. Chagas is endemic to Bolivia, with an approximated prevalence reaching 1.1 to 1.8 million cases, and areas such as Santa Cruz possess high incidence of both diseases, thus a study analysing co-infection could provide additional answers (Medrano-Mercado, 2008).

Another social aspect that could result useful to explore is migration, including internal, emigration, and immigration, and how this could affect incidence in new areas, and the increase or decrease of new cases in certain areas. Other variables that could be included in future studies include those related to alcohol and tobacco consumption, and malnutrition, and sleep deficiency.

# APPENDIX A
## Municipalities, Divisions and Departments

| Department | Division ID | Number of Municipalities per Division |
|---|---|---|
| Beni | 1 | 11 |
| | 2 | 8 |
| Chuquisaca | 3 | 25 |
| | 4 | 3 |
| Cochabamba | 5 | 45 |
| La Paz | 6 | 45 |
| | 7 | 8 |
| | 8 | 19 |
| | 9 | 8 |
| Oruro | 10 | 27 |
| | 11 | 8 |
| Pando | 12 | 6 |
| | 13 | 9 |
| Potosi | 14 | 28 |
| | 15 | 4 |
| | 16 | 6 |
| Santa Cruz | 17 | 42 |
| | 18 | 7 |
| | 19 | 7 |
| Tarija | 20 | 4 |
| | 21 | 7 |

Summary: 9 Departments, 21 Divisions and 327 Municipalities.

# APPENDIX B

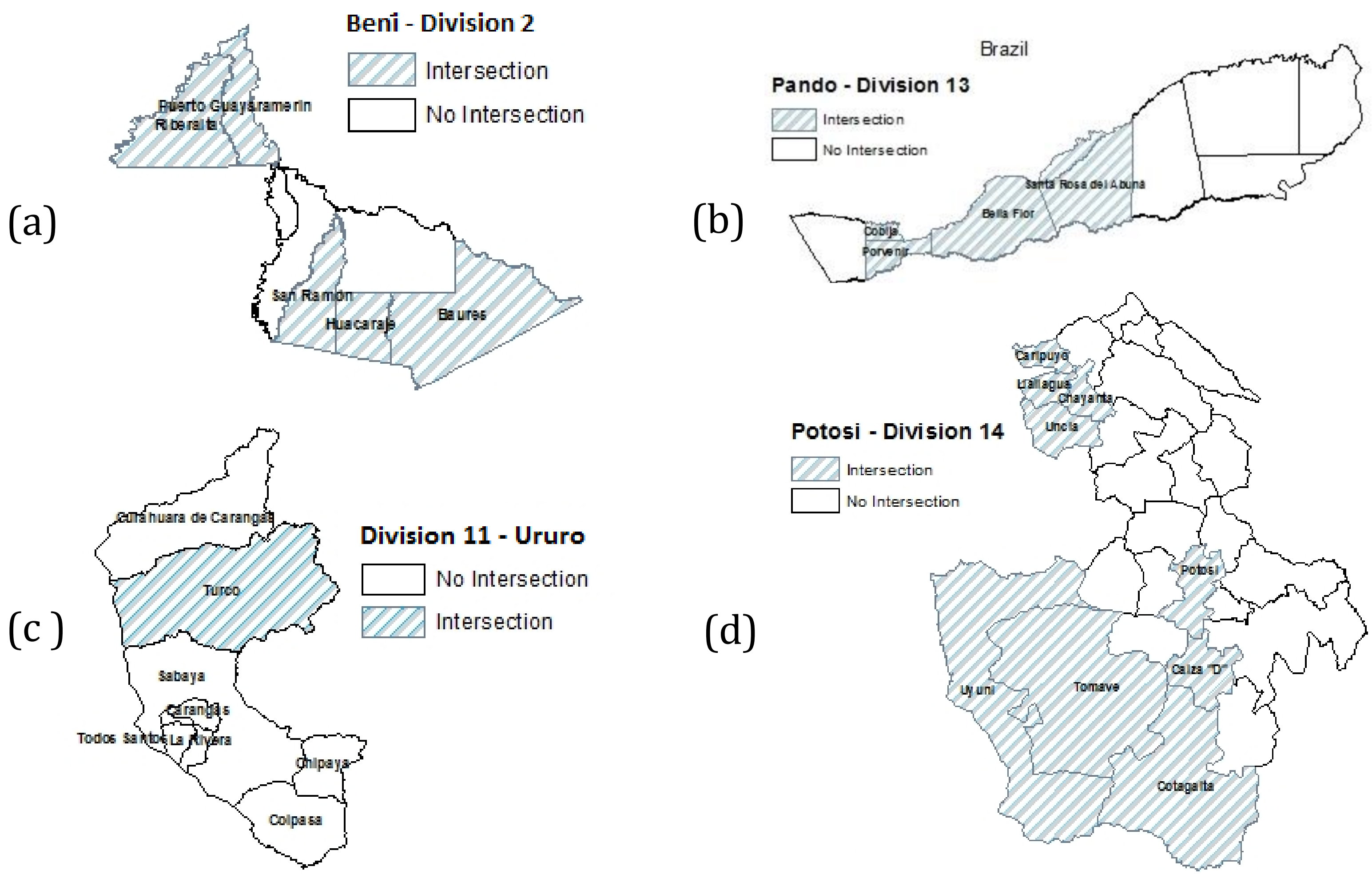


Figure 8. Maps of Intersections for (a) Division 2, (b ) Division 13, (c) Division 11 and (d) Division 14, in Beni, Pando, Ururo and Potosí departments, respectively

## References

Álvarez-Gaviria, M. et al. (2013) ‘Multi-drug resistant tuberculosis’, *Revista Ciencias Biomédicas,* vol. 4, no. 1, pp.159-164.

Álvarez-Hernández, G. et al. (2010) ‘An analysis of spatial and socio-economic determinants of tuberculosis in Hermosillo, Mexico, 2000–2006,’ *International Journal of Tuberculosis and Lung Disease,* vol. 14, no. 6, pp. 708-713.

Barrios, M.E. (2012) ‘Características de la tuberculosis en niños menores de 15 años durante las gestiones 2009-2010 en el Hospital del Niño Manuel Ascencio Villaroel’, *Revista Científica Ciencia Médica,* vol. 15, no. 1, pp. 22-25.

Bemrah, S. et al. (2013) ‘Tuberculosis among the homeless, United States, 1994-2010’, *International Journal of Tuberculosis and Lung Disease,* vol. 17, no. 11, pp. 1414-1419.

Boccia, D. et al. (2009) ‘Tuberculosis infection in Zambia: the association with relative wealth’, *The American Journal of Tropical Medicine and Hygiene,* vol. 80, no. 6, pp. 1004-1011.

Borelli, S. et al. (2010) ‘Tuberculosis transmission patterns among Spanish-born and foreign-born populations in the city of Barcelona’, *Clinical Microbiology and Infection,* vol. 16, no. 6, pp. 568-574.

Castañeda-Hernández, D. et al. (2013) ‘Association between incidence of tuberculosis and Human Development Index in 165 countries of the world’, *Revista Peruana de Medicina Experimental y Salud Pública,* vol. 30, no. 4, pp. 560-568.

Castillo, I.Y. et al. (2013) ‘Factors associated with the prevalence of tuberculosis in Cartagena’, *Revista de la Universidad Industrial de Santander, Salud,* vol. 45, no. 2, pp. 21-27.

Center for Disease Control (2014) *TB Elimination: The Difference between TB Infection and TB Disease* [Online]. Available at http://www.cdc.gov/tb/publications/factsheets/general/ltbiandactivetb.pdf (Accessed 8 July 2015).

Chan-Yeung, M. et al. (2005) 'Socio-demographic and geographic indicators and distribution of tuberculosis in Hong Kong: a spatial analysis', *International Journal of Tuberculosis and Lung Disease,* vol. 9, no. 12, pp. 1320 – 1326.

Goldstein, H. (2011) *Multilevel Statistical Models,* 4th edn., Chichester, John Wiley & Sons Ltd. Gómez-Barroso, D. (2009) 'Space distribution of Tuberculosis in Spain by geostatistical methods', *Revista Española de Salud Pública,* vol. 83, no. 5, pp. 737-744.

González-Martín, J. et al. (2010) 'Documento de consenso sobre diagnóstico, tratamiento y prevención de la tuberculosis', *Archivos de Bronchoneumología,* vol. 46, no. 5, pp. 255-274.

Hargreaves, J., et al. (2011) 'The social determinants of tuberculosis: from evidence to action', *American Journal of Public Health,* vol. 101, no. 4, pp. 654-662.

Heck, R.H. and Thomas, S. L. (2015) *An introduction to Multilevel Modeling Techniques,* 3rd Edn., Routledge.

Hijmans, R. et al. (2015) *Global Administrative Areas: Bolivia* [Shapefile].1:200, GADM, Using: ESRI ArcMap.

Hino, P. et al. (2011) 'Spatial patterns of tuberculosis and its association with living conditions in the city of Ribeirão Preto in the state of São Paulo', *Ciência & Saúde Coletiva,* vol. 16, no. 12, pp. 4795-4802.

Holtgrave, D.R. and Crosby, R.A. (2004) 'Social determinants of tuberculosis case rates in the United States', *American Journal of Preventive Medicine,* vol. 26, no. 2, pp. 159-162.

Horna-Campos, O. et al. (2007) ‘Public Transportation and Pulmonary Tuberculosis, Lima, Peru’, *Emerging Infectious Diseases,* vol. 13, no. 10, pp. 1491–1493.

Huaroto, L. et al. (2009) ‘Recomendaciones para el control de la transmisión de la tuberculosis en los hospitales’, *Revista Peruana de Medicina Experimental y Salud Pública,* vol. 26, no. 3, pp. 364-369.

Instituto Nacional de Estadística (2012) *Resultados Censo de Población y Vivienda,* La Paz, INE.

Jackson, S. et al. (2006) ‘Poverty and the economic effects of TB in rural China’, *International Journal of Tuberculosis and Lung Disease*, vol. 10, no. 10, pp. 1104-1110.

Maciel, E. (2010) ‘Spatial patterns of pulmonary tuberculosis incidence and their relationship to socio-economic status in Vitoria, Brazil’, *International Journal of Tuberculosis and Lung Disease,* vol. 14, no. 11, pp. 1395-1402.

Medrano-Mercado, N. et al. (2008) ‘Urban transmission of Chagas disease in Cochabamba, Bolivia’, *Memórias do Instituto Oswaldo Cruz,* vol. 103, no. 5, pp. 423-430.

Migliori, G.B. et al. (2007) ‘125 years after Robert Koch's discovery of the tubercle bacillus: the new XDR-TB threat. Is “science” enough to tackle the epidemic?’, *European Respiratory Journal,* vol. 29, no. 3, pp. 423-427.

Murray, M. et al. (2011) ‘Modeling social, environmental and biological determinants of tuberculosis’, *International Journal of Tuberculosis and Lung Disease,* vol. 15, no. 6, pp. 564-570.

Nezlek, J.B. (2011) *Multilevel Modeling for Social and Personality Psychology,* London, SAGE Publications Ltd.

Nishiura, H. (2003) ‘Socioeconomic factors for tuberculosis in Tokyo, Japan-unemployment, overcrowding, poverty, and migrants’, *Kekkaku: Tuberculosis,* vol. 78, no. 6, pp. 419-426.

Oren, E., et al. (2012) ‘Area-based socio-economic disadvantage and tuberculosis incidence’, *The International Journal of Tuberculosis and Lung Disease*, vol. 16, no. 7, pp. 880-885.

Pan-American Health Organization (2012) *Map of Municipal Health Index, Data* [CSV Dataset] Pan-American Health Organization-World Health Organization.

Ploubidis, G. et al. (2012) ‘Social determinants of tuberculosis in Europe: a prospective ecological study’, *European Respiratory Journal,* vol. 40, no. 4, pp. 925-930.

Raftery, A.E. (1966) ‘Bayesian model selection in social research’, In Marsden, P.V., (ed.), *Sociological Methodology,* Oxford, Basil Blackwell, pp.111-163.

Rivera, L.T. et al. (2014) ‘Características epidemiológicas de pacientes con tuberculosis en el Hospital Tránsito Cáceres de Allende’, *Revista Americana de Medicina Respiratoria,* vol. 14, no.4, pp. 404-411.

Russell, D. et al. (2010) ‘Tuberculosis: What we don’t know can, and does, hurt us’, *Science,* vol. 328, no. 5980, pp. 852-856.

Russell, D. (2007) ‘Who puts the tubercle in tuberculosis?’, *Nature Reviews Microbiology,* vol. 5, no. 1, pp. 39-47.

Shetty, N. et al. (2006) ‘An epidemiological evaluation of risk factors for tuberculosis in South India: a matched case control study’, *International Journal of Tuberculosis and Lung Disease,* vol. 10, no. 1, pp. 80 – 86.

Sobrero, R. and Peabody, J. (2006) ‘Tuberculosis control in Bolivia, Chile, Colombia, Peru: Why does incidence vary so much between neighbors’, *International Journal of Tuberculosis and Lung Disease,* vol. 10, no. 11, pp. 1292-1295.

Tekkel, M. et al. (2002) ‘Risk factors for pulmonary tuberculosis in Estonia’, *International Journal of Tuberculosis and Lung Disease*, vol. 6, no. 10, pp. 887-94..

Tocque, K. et al. (2001) ‘A case-control study of lifestyle risk factors associated with tuberculosis in Liverpool, North-West England’, *European Respiratory Journal,* vol. 18, no. 6, pp. 959-964.

Unidad de Análisis de Políticas Sociales y Económicas (2013) *Establecimientos de Salud, según Municipio* [XLS Dataset].

Unidad de Análisis de Políticas Sociales y Económicas (2012) *Indicadores de Educación alineados a las metas del milenio, según Municipio* [XLS Dataset].

United Nations Office for Disaster Risk Reduction (2012) *Documento País: Bolivia,* Geneva, Switzerland, UNISDR, DIPECHO Report no. 7.

Ximenes, R. et al. (2009) ‘Is it better to be rich in a poor area or poor in a rich area? A multilevel analysis of a case–control study of social determinants of tuberculosis’, *International Journal of Epidemiology,* vol. 38, no. 5, pp. 1295-1296.

World Bank (2012) *Life expectancy at birth (total), in years* [Online] Available at http://data.worldbank.org/indicator/SP.DYN.LE00.IN (Accessed 10 July 2015)

World Health Organization (2015a) *Tuberculosis Fact Sheet* [Online] Available at http://www.who.int/mediacentre/factsheets/fs104/en/ (Accessed 10 July 2015).

World Health Organization (2015b) *Tuberculosis: The five elements of DOTS* [Online]. Available at http://www.who.int/tb/dots/whatisdots/en/index4.html (Accessed 10 July 2015).

Zubieta, A. and Cruz, J. (2012) ‘Factores de riesgo socioeconómicos asociados a tuberculosis pulmonar en pacientes de 15 y más años en el Hospital Municipal Modelo Corea, ciudad de El Alto. Gestión 2012’, *Cuadernos Hospital de Clínicas,* vol. 55, no. 2, p. 17.