Assessment of Socio-Demographic Factors of Tuberculosis Incidence in the Eastern Cape Province of South Africa

Davies Obaromi¹,²*, Qin Yongsong², James Ndege² and Akinwumi Odeyemi²

¹Biostatistics and Epidemiology Research Group, Department of Statistics, University of Fort Hare, PMB X1314, Alice 5700, South Africa.
²Department of Statistics, University of Fort Hare, PMB X1314, Alice 5700, South Africa.

* daviesobaromi@gmail.com

Abstract

Background: Socio-economic deprivation and some social vulnerability factors such as income and education level have been linked with tuberculosis incidence in some regions. This study aimed to extend and investigate the multimodel relationships between some socio-economic and demographic factors on the relative risk of Tuberculosis (TB) Incidence in the Eastern Cape Province of South Africa, using the Poisson regression model.

Methods: Ecological study of TB prevalence in the Eastern Cape Province of South Africa, in 2014 was investigated. We evaluated data from Electronic Tuberculosis Register (ETR), an electronic notification system for TB cases for the health sub-districts in the Province. Three socio-economic indicators and three demographic factors all calculated per sub-district were used. The data were analysed with a Bayesian regression model assuming a Poisson distribution for the observed new cases of TB in each health sub-district.

Results: The Bayesian regression model confirmed the relationship between demographic factors and TB distribution in the Eastern Cape Province. We demonstrated that the rate of TB incidence was positively correlated with average household size and the population density of each area. Socio-economic factors were not significant in any of the models.

Conclusion: This study signifies that the prevalence and incidence of TB in the Eastern Cape Province are more demographic dependent than socio-economic.

Keywords: Bayesian, Tuberculosis, Demographic, Multilevel, Epidemiology

Introduction

Tuberculosis is an infectious disease, generally chronic and caused by a group of bacteria: Mycobacterium tuberculosis, M. africanum, M. bovis, M. microti and M. canetti.¹ The main transmission is from person to person by microdrops generated by coughing or sneezing of a person with active TB.¹ It is still one of the oldest human diseases that still affect large population groups. Tuberculosis is closely linked to both overcrowding and malnutrition, making it one of the principal diseases of poverty. More people in the developing world contract tuberculosis because of a poor immune system, largely due to high rates of Human immunodeficiency virus (HIV) infection and the corresponding development of acquired immune deficiency syndrome (AIDS).²

South Africa ranks as the third highest in the world-burden of tuberculosis and for three consecutive years (2007, 2008 and 2009), tuberculosis ranked as the number one among the ten leading underlying natural causes of death as shown in Table 1.³ Tuberculosis in the Eastern Cape Province mainly affects the economically active age group. Within the age group of 25-34 years, the percentage distribution of reported...
TB cases was 15.9%, 0.7% and 23.1% for the years 2003, 2004 and 2005 respectively. A report on South African National Burden of Disease study 2000 Eastern Cape Province by MRC showed that tuberculosis was the second leading cause of death among women and the third leading cause of death among men aged 15-44 years. Eastern Cape Province ranks as the second highest burden of TB by Province after KwaZulu Natal as shown in Figure 2.

In people with normal immune systems, the lifetime risk of progressing from latent TB infection to active TB disease is 10%. HIV, by weakening the immune system, increases a person’s risk of progressing from latent TB infection to active TB disease by 10% per year.

In Figure 1 the Province has an extremely high burden of TB, co-infection with TB and HIV (TB/HIV). In 2008, there were more than 60 000 new TB cases in the

![Figure 1: HIV and TB Rates from 1980-2006](source)

Table 1: Number and Percentage of TB Deaths, South Africa, 2008–2013

<table>
<thead>
<tr>
<th>Year</th>
<th>Number of TB deaths</th>
<th>% of all deaths</th>
<th>Tb-specific rate per 100 000</th>
</tr>
</thead>
<tbody>
<tr>
<td>2008</td>
<td>75 281</td>
<td>12.6</td>
<td>153</td>
</tr>
<tr>
<td>2009</td>
<td>69 791</td>
<td>12.0</td>
<td>140</td>
</tr>
<tr>
<td>2010</td>
<td>63 281</td>
<td>11.6</td>
<td>125</td>
</tr>
<tr>
<td>2011</td>
<td>54 112</td>
<td>10.7</td>
<td>107</td>
</tr>
<tr>
<td>2012</td>
<td>48 409</td>
<td>8.4</td>
<td>92</td>
</tr>
<tr>
<td>2013</td>
<td>40 452</td>
<td>-</td>
<td>76</td>
</tr>
</tbody>
</table>

*Source: SANAC (2014) and Stats SA (2014)*
Province. Of these, there were 1,251 confirmed cases of MDR-TB and 385 confirmed cases of XDR-TB. In 2010, the total of new TB and re-treatment cases identified in the Province stood at 62,226.5

Household crowding, deficiency and access to health service are well-known factors linked with TB prevalence and should be integrated into health policies planning. An effective programme for TB control should include these features not only as individual risk factors but as population determinants, comprising the basis for a territorial approach to a surveillance system.6 Strong evidence for an association between TB and poverty are already available, expressed by higher TB incidence rates in crowded urban areas and amongst low income and illiterate populations.7 Health services access, in addition, is compromised in the same populations, both as a result of lack of sufficient health services in poor areas and because poorer and illiterate people are less aware of their own health status.

In a typical clinical practice and most epidemiological studies, socio-economic determinants of disease are included in terms of an individual person’s risk factors.8 In ecological studies, on the other hand, the focus is on the community as an entity in itself, an entity more complex than the sum of the individual persons who make it up.9 Although individual level factors such as behaviours and lifestyles, common to individuals across areas and subject to public health interventions, may also play a role in TB incidence. The question in ecological analysis is not about the causes of disease cases, but the causes of disease incidence.10

The connection between tuberculosis (TB) and socio-economic status is well acknowledged.6, 11 Besides, current studies succeeded in demonstrating at area level, an association between poverty and TB incidence rates.12 The objective of this paper is, therefore, to extend and model the relationship between TB incidence at the provincial level, the socio-

**Figure 2: TB Deaths per Province in South Africa**

DEATHS PER PROVINCE

- **NW:** 84,193 (7,7%)
- **GP:** 97,595 (21,3%)
- **MP:** 34,820 (7,6%)
- **KZN:** 84,193 (18,3%)
- **NC:** 13,699 (3%)
- **EC:** 63,935 (13,9%)
- **FS:** 33,382 (7,3%)
- **WC:** 46,007 (10%)

**Province population**

- **LP:** 5,518,000
- **NW:** 3,597,600
- **GP:** 12,728,400
- **KZN:** 10,456,900
- **MP:** 4,128,000
- **NC:** 1,162,900
- **FS:** 2,753,200
- **WC:** 6,016,900
- **EC:** 6,620,100

Source: Statistics South Africa (www.statssa.gov.za)
Assessment of Socio-Demographic Factors of Tuberculosis Incidence in the Eastern Cape Province of South Africa

economic and demographic measures, discussing the capability of various indicators in guiding preventive action and interventions. We used the indicators and covariate data from the Eastern Cape Socio-economic Consultative Council reports of 2014.

The detection of the different risks for tuberculosis in the Province will take into cognisance, the public health systems and deal with the underlying features of each risk factor that present higher incidences of the disease. This would also allow planning for the future, allocating scarce resources with priority and monitoring the impact of policy, political and economic changes in society and also showing the supposed direction that health policies and agencies should follow in enabling equal accessibility by all regardless of socio-economic status.

Methods

This study was carried out in the Eastern Cape Province of South Africa. The Province of the Eastern Cape is situated on the east coast of South Africa and lies between the Western Cape and KwaZulu-Natal Provinces. The Northern Cape and Free State Provinces, as well as Lesotho shares borders with this Province.

The Eastern Cape Province boasts of amazing natural diversity, stretching from the semi-arid Great Karoo to the forests of the Wild Coast. It also extends around the Keiskamma valley, the fertile Langkloof, and the mountainous southern Drakensberg region. The main feature of the Eastern Cape is its amazing coastline adjoining the Indian Ocean. The Province covers an area of 168 966 km² and with a population of 6 562 053 (Statistics South Africa, Censuses, 2011). The Province is situated at 32.2968°S and 26.4194°E of the country.

The Eastern Cape is the second-largest Province in South Africa by surface area and also the third-largest populated Province with its capital in Bhisho. Port Elizabeth, East London, Grahamstown, Mthatha (previously Umtata), Graaf Reinet, Cradock and Port St Johns are the major towns and cities in the Province. The Province is divided into two metropolitan municipalities, and they are Buffalo City Metropolitan Municipality and Nelson Mandela Bay Metropolitan Municipality. It has six district municipalities and which are further subdivided into 37 local municipalities.

The Eastern Cape is regarded as one of the poorest Provinces in South Africa. This is mostly as a result of the poverty found in the former homelands, where subsistence agriculture prevails.

Epidemiological Data Sources

This is a retrospective secondary data source from Eastern Cape Province TB notification and survey data. The data used in this study was extracted from the electronic tuberculosis register (ETR) records of TB cases from the twenty-four health sub-districts of the Province including the two metropolitan municipalities for the year 2014. The socio-economic and socio-demographic indicators and variables were obtained from publications of Eastern Cape Socio Economic Consultative Council (ECSECC, 2014) reports of all the local municipalities. Map of South Africa (Figure 3a) and map of the Eastern Cape Province, showing the twenty-four health sub-districts where data was collated is shown in Figure 3b.

Statistical Analyses

Bayesian Methods

We used a Bayesian regression model assuming a Poisson distribution for the observed new cases of TB in each area, averaged over the observed period and chosen covariates of socio-economic vulnerability and demographic factors, and it, therefore, takes the form:

\[
\log \lambda_i = \log(\text{pop}_i) + \beta_0 + \sum_{j=1}^{k} X_{ij} \beta_j + e_i
\]

Bayesian modelling depends on the ability to compute posterior distributions in order to provide estimates for all the corresponding model parameters. The majority of these posterior distributions are straightforward to calculate. Distributions with a conjugate prior typically have a posterior distribution which follows a standard distributional form.

In many cases, however, the computation required is more complex and a more advanced method is essential to calculate the posterior distribution. These advanced approaches usually make use of some form of numerical simulation, generally by drawing a sample of parameter values from an approximation of the posterior distribution to allow estimation of the distributions of the model parameters.
From the above model (equation 1), six covariates (socio-economic and demographic factors) were taken as explanatory variables for the relative risk of the disease. They are: $X_1 = \text{Gini coefficient}$ (a measurement of how income or poverty is equally distributed), $X_2 = \text{poverty rate}$ (% number of individuals living below the poverty line. Though, there is no official poverty line defined for South Africa), $X_3 = \text{unemployment rate}$ (%), $X_4 = \text{No schooling}$ (%; person aged 20+ years), $X_5 = \text{average household size}$ and $X_6 = \text{Population density}$ of the regions/municipalities. The Gini coefficient, poverty rate and unemployment are considered in this study as distal factors, while no schooling, average household size and population density are taken as proximal factors. An additional covariate, $\log \text{popi}$, was used as an offset variable in the equation and added in and out of the models.

In this study, four (4) separate multilevel models were developed and treated as non-independent Poisson random variables with means, to investigate whether the covariates influenced part of or all of the correlation in the TB relativity.

**Parameter Estimation by the Integrated Nested Laplace Approximation (INLA) Method.**

For the Bayesian estimation of the parameters of the model, we considered the use of INLA method, where the posterior marginal are not always presented in closed form as a result of the non-Gaussian response variables. For such models, Markov chain Monte Carlo methods can be applied, but they are not without some problems, both in terms of convergence and in computational time. In some practical uses, the extent of these problems is such that Markov chain Monte Carlo is simply not an appropriate tool for monotonous analysis.

It is shown in that by using an integrated nested Laplace approximation (INLA) and its simplified version, we can directly compute very precise approximations to the posterior marginal. The key advantage of these approximations is simply computational: where MCMC algorithms need hours and days to run, INLA provides more precise estimates in seconds and in minutes. Another benefit with INLA approach is its generality, which makes it possible to perform Bayesian analysis in a programmed, streamlined way, and to compute model comparison criteria and various predictive measures so that models can be compared and the model under study can be tested. Although in most cases, similar results will be obtained by MCMC and INLA inference. It should be noted that there are fundamental differences in the way that posterior distributions are estimated. MCMC can sample directly from a joint posterior distribution, while INLA uses a closed form expression to estimate the marginal posterior distributions.
For this study, we adopted the latter and the analysis was carried out in R. Models comparison were carried out using the Deviance Information Criterion (DIC), which combines a measure of fit and a measure of model complexity based on the effective number of parameters. Smaller values of DIC show a more fitting model.\textsuperscript{13}

**Results**

In Table 2 the estimates from the combination of Bayesian approach and GLMM to assess the TB relative risk for 2014 dataset was shown, by investigating its relationship with some socio-economic and demographic variables in the Eastern Cape Province of South Africa. For 2014 data, Eastern Cape had 37,365 notified TB cases from all the twenty-four (24) health sub-districts and with about 91.1% bacteriological coverage (ratio of the number of PTB patients diagnosed by bacteriological tests to the total PTB patients reported, excluding children 0-4 years).

Also in Table 2, the measures of association and their respective credibility intervals are presented for all covariates in all models. In the distal model (model 2), no socio-economic factor is found to be significant. Gini coefficient, unemployment, and no schooling, which are one of the most widely used socio-economic indicators in South Africa, were not significant in any of the models.

For the proximal model in model 3 and the combined model (model 4), average household size and population density were significant, indicating that an increase of one more person in the house increases the risk of TB and population density, which indirectly influences the average household number of persons living together. The most significant effect, however, is the average household size per house.

Although it is an indicator of poverty and population density, it is also an important covariate in the household transmission of TB. On the basis of DIC, model 2 gave the best fit but has no covariate that is significant for TB incidence. Models 1, 3 and 4 with higher DIC values than model 2, have positive indicators for the disease prevalence in the Province, and these three factors are more demographic than economic. Also, model 3 which has the most significant effects of population density and average household size, has the smallest random error value in all the models.

**Discussion**

The second most common cause of death from infectious disease (after HIV) is tuberculosis.\textsuperscript{14} The

| Table 2: Posterior Estimated Risk Factors Associated with TB by GLMM Model for Eastern Cape Province |
|-------------------------------------------------|---------------|---------------|---------------|---------------|
| Covariates | Model 1 | Model 2 | Model 3 | Model 4 |
| Intercept | 0.64 (1.40) | 0.03 (1.50) | 6.53 (0.96) | 7.38 (1.07) |
| Log pop | 1.24 (0.27) | 1.56 (0.28) | - | - |
| Gini coeff | - | -0.11 (0.54) | - | -0.65 (0.65) |
| Poverty | - | -1.68 (0.63) | - | -0.06 (0.68) |
| Unemploym. | - | -0.004 (0.01) | - | -0.02 (0.01) |
| Pop. Dens | - | - | 0.004 (0.001) | 0.003 (0.001) |
| No school | - | - | -7.22 (31.52) | -5.40 (31.51) |
| ave hhold size | - | - | 0.07 (0.26) | 0.10 (0.27) |
| Error term | 1.374 | 2.125 | 1.317 | 1.487 |
| DIC | 263.12 | 263.11 | 263.14 | 263.14 |
complete number of tuberculosis cases has been declining since 2005 and new cases since 2002. The Republic of South Africa has one of the most severe TB epidemics in the world. A major rise in TB prevalence in the pre-HIV period has been followed by a growing number of TB cases as a consequence of HIV and TB co-infection. Furthermore, there is now an increasing resistance to some of the anti-TB drugs.

The model by model approach used in this study to include variables in the models in connected blocks, enabled us to identify the significance of each covariate associated with the spread of the disease in the Province. Further, we established that socio-economic variables like Gini coefficient and unemployment, which are distal factors cannot on their own explain the incidence of TB: proximal variables classified as demographic variables, related to more exact extents (population density in the Province and average household size during the period under study) of the health-disease process were found to be vital.

The average household size, an information easily obtained in a demographic survey is a risk factor for TB, and neighborhoods with similar socio-economic features tend to converge together. More significantly is the population density of each census tract. Collectively, these variables (proximal and distal) show the role of different social and demographic practices, the first one signifying a structural dimension of the economic status of a population, while the second one points to the population structure of the unit under study.

Results from this study give two different substitutes to deal with observing TB in this Province. The use of socio-economic and demographic variables related to the prevalence of TB in the Province which points to the indicators to be examined and considered at the provincial level. In addition, the relationship either with population density and average household size, emphasizes the associations between demographic deprivation and TB in this setting.

**Strengths and Limitations of the Study**

Even with the well-known relationship between TB and some socio-economic indicators like poverty, this present study established that the TB rates in the Eastern Cape Province were more amplified with demographic factors such as population density and the average size of households (Table 2), which are more proximal than distal factors like socio-economic risk factors. Also, the link either with population density and average household size highlight the associations between demographic deprivation and TB in this setting. This also suggests that more multifaceted associations may be present between TB prevalence and a wide range of socio-demographic factors such as the individuals’ occupations, work conditions, poor nutrition, access to quality healthcare, living conditions and social behavior. Our study also utilised the strength of the Bayesian approach; a method which combines the likelihood of the data and a prior distribution to obtain the posterior distribution of the estimates. However, there are some limitations in this study.

First, a restraint usually faced in most studies of this methodology is that most TB identification surveys (especially, in high disease burden and low-income countries like South Africa), do not have specialists to perform such analyses satisfactorily and there is also scarcity of analytical statisticians in the Province.

Second, age and sex of the cases were not utilised in this study to capture the distribution and effects of such demographics on the disease. The Eastern Cape Socio Economic Consultative Council (ECSECC 2014) reports did not provide such information about the Province.

Finally, the data used for this study was the one reported for 2014 by the register (ETR). This may not suffice for the present state of tuberculosis in the Province.

**Findings in Context with Other Previous Studies**

The relationship between poverty and health is well acknowledged. Few studies have investigated the effects of socio-economic indicators on diseases prevalence. Many studies focused on risk factors to susceptibility and not prevalence of the disease in a selected region like our study. Some of these studies established that poverty, income and some vulnerability have links to infectious disease like pulmonary tuberculosis (PTB), HIV and TB/HIV co-infection. Also, they used the classical approach (Logistic regression) in establishing these associations. The studies found that socio-economic and some clinical factors influence susceptibility to tuberculosis (TB) disease.
Our study employed a more precise approach, the Bayesian approach to estimate the parameters of the model and it showed negative association between socio-economic factors and relative risk of the TB, although susceptibility and prevalence of a disease are two different things.

Several studies that examined socio-demographic factors on TB also existed, but they placed more emphasis on factors affecting knowledge levels of TB and not its prevalence or incidence is a selected area.\textsuperscript{20, 21, 22} Our present study did not establish any link between TB and any of the socio-economic factors used, but we found that the prevalence of TB in the Eastern Cape Province of South Africa has a more demographic and population dimension of prevalence than economic (Table 2).

\textbf{Conclusion}

By implementing a fully categorized or multilevel Bayesian model to assess the relative risk of the incidence of TB was a suitable technique for this study. It is known that ‘all models are incorrect, but some are useful’, the possibility to assess together in the same context, the impact of explanatory variables at ecological level should be encouraged and supported in epidemiology, in spite of the characteristic problems of this method.\textsuperscript{23} Moreover, identifying the social and demographic processes that lead to the circumstances of the collective risk of the disease are essential to the balanced planning of health and government mediations, and may aid with creating Epidemiological Surveillance Systems on a territorial basis.

\textbf{Acknowledgment:} We are grateful to data and research department, administrators, staff and members of Eastern Cape Department of Health, for the release of the data of the Province. Our profound gratitude also goes to GMRDC University of Fort Hare for their financial support.

\textbf{Authors’ Contributions:} Davies Obaromi designed and formulated the study. Davies Obaromi and Akinwumi Odeyemi drafted the manuscript, analyzed and interpreted the data. Qin Yongsong and James Ndege supervised the entire concept of the study and revised the manuscript critically for intellectual criticisms.

\textbf{Conflicts of Interest:} The authors declare no conflict of interest.

\textbf{Ethical Considerations:} This study was carried out under the authorization and permission of the Ethical committee of the University of Fort Hare, Alice, Eastern Cape, South Africa and approval of the Eastern Cape Department of Health, with ethical clearance reference number QIN041SOBA01 and EC_2015RP24_398 respectively.

\textbf{References}


