



The Open Public Health Journal

Content list available at: <https://openpublichealthjournal.com>



RESEARCH ARTICLE

Semi-parametric Model to Study the Risk Factors of Tuberculosis among Adult men in South Africa

Muziwandile Nhlakanipho Mlondo^{1,*}, Sileshi Fanta Melesse¹ and Henry G. Mwambi¹

¹School of Mathematics, Statistics, and Computer Science, University of KwaZulu Natal, Pietermaritzburg Campus, Private Bag X01, Scottsville, 3209, South Africa

Abstract:

Background:

Understanding the relationship between tuberculosis and the risk factors of tuberculosis is vital to be able to address them. Even though tuberculosis is curable and preventable, it remains a public threat, especially in low- and middle-income countries. There are more cases of men infected with tuberculosis compared to women.

Methods:

This study determines the risk factors that influence TB infection among adult men in South Africa. The Generalized Additive Mixed Models that incorporate a nonparametric smooth additive function were used to analyze the 2016 South African Demographic Health Survey. The Generalized Additive Model and Generalized Additive Mixed Model were compared to find the most suitable model.

Results:

The finding reveals that the significant determinates associated with tuberculosis were: region, primary education compared to secondary, race, health, weight, and wealth index, which were modeled parametrically. The conclusion is that only the interaction effect of age and number of times away from home were significant in variables modeled non-parametrically.

Conclusion:

In conclusion, the South African government needs to intervene on men living in Western Cape, Eastern Cape, Northern Cape, Free State, and men with poor health - to reduce South Africa's infections. The government should also implement programs that will teach and discourage lower body mass index. Thus, targeting the factors that have a positive significant effect among adult men can help to reduce the risk of having tuberculosis.

Keywords: Tuberculosis, Semi-parametric models, Risk factors among men, Organs, Lungs, Brain.

Article History

Received: November 25, 2021

Revised: February 16, 2022

Accepted: March 4, 2022

1. INTRODUCTION

Tuberculosis remains a public health threat in all countries. Tuberculosis is an infection that typically affects the lungs. Nevertheless, it can also affect other organs like the brain and spine [1]. Tuberculosis is caused by an airborne bacterium named Mycobacterium. Mycobacterium is one of the top ten causes of death among infectious agents and is also one of the oldest recognized infections. Tuberculosis (TB) cannot be transmitted from surfaces where a TB patient has been and touched such as for COVID-19. However, Mycobacterium is

mainly transmitted when infected people exhale, thereby discharging small particles of e bacteria into the air, and nearby people breathe these bacteria and are consequently infected.

Getting infected by the TB bacteria does not necessarily result in sickness. There are two types of tuberculosis: Latent and Active TB. Latent infection occurs when a person's immune system prevents the bacteria from spreading. Active infection occurs when the bacteria multiply and infect the organs. Although TB can strike anyone, the literature suggests that most people who develop the disease are adults. TB is more common among men than women [2].

In this study, we examine the factors that affect TB among adult men and how these factors impact them, which we

* Address correspondence to this author at the School of Mathematics, Statistics, and Computer Science, University of KwaZulu Natal, Pietermaritzburg Campus, Private Bag X01, Scottsville, 3209, South Africa; Email: mlondo02@gmail.com

believe can help in addressing those factors. As a result, there could be a reduction in the number of TB cases and deaths. Ten million people contract tuberculosis each year, according to the World Health Organization (WHO) [3]. Despite being curable and preventable, TB still kills 1.5 million people each year, making it the leading infectious killer in the world. About two-thirds of new TB cases in 2019 are present in low-and middle-income countries [2].

A report on mortality and mortality causes in South Africa was released by Statistics South Africa (StatsSA) [4]. This report suggests that in 2010, there were 6,3% fewer deaths than in 2009. Data shows that more men than women die of tuberculosis, and 30-39-year-olds are at the highest risk. South Africa had the highest death rate from TB; about 12% of deaths in 2010 were caused by TB [4]. According to the data, there were more sick Black Africans than other race groups.

2. METHODS AND MATERIALS

2.1. Data Source

For this study, the 2016 South African Demographic Health Survey was used. The survey was carried out by Statistics South Africa and the South African Medical Research Council. The survey was conducted in 15292 households. In the sample of households, 8514 women, aged 15-49, completed interviews with individuals. 10336 men aged 15 and older completed interviews on adult health. The variable of interest is a binary variable indicating whether a health worker told a person about the status of having tuberculosis.

2.2. Statistical Methods

The relationship between the response and the predictors may not always be linear. When there is no such linearity, we need additive models. The Generalized Additive Models (GAM) follow additive models as the extension of the Generalized Linear Model (GLM) that incorporates an additive term in the linear predictor, and the response may belong to any exponential family distribution [5]. Suppose y_i is the response variable whose distribution belongs to the exponential family, then the generalized additive model is defined by:

$$g(\mu_i) = \eta_i = \mathbf{X}_i^* \boldsymbol{\theta} + \sum_j f_j(x_j)$$

$g(\mu)$ is a one-to-one function, \mathbf{X}_i^* is the i^{th} row matrix of the model, $\boldsymbol{\theta}$ is the parametric estimates vector. $f_j(\cdot)$ are the smooth functions. Smooth functions allow the flexible specification of the dependence of the response on the covariates, then a parametric relationship. P-spline estimates the inference of GAM.

The province's effect on the likelihood of developing tuberculosis is fixed for the Generalized Additive Mixed Model (GAMM). On the other hand, neighboring provinces may have a similar effect compared to non-neighboring due to spatial autocorrelation. We will study an extension of the Generalized Linear Mixed Model, which is the Generalized Additive Mixed Model, a model in which parametric fixed effects can be modeled non-parametrically by incorporating a smooth additive function as a random effect [6 - 10]. The GAMM explores the nonlinear relationship between the dependent variable and covariates. The GAMM has the following structure:

$$\text{logit}(\pi_{ijk}) = \mathbf{x}'_{ijk} \boldsymbol{\beta} + \sum_{r=1}^P f_r(z_{ijk}) + f_{\text{spat}}(s_i)$$

Where the $\boldsymbol{\beta}$ vector of the fixed effect of the predictors, $f_r(\cdot)$ is a smooth arbitrary function of the predictors and $f_{\text{spat}}(s_i)$ is the non-linear spatial effect. Using penalized splines (P-splines) with B-spline basis functions, smooth functions f_r are estimated [11]. A spline is defined as a linear combination of $M_r = n_r + v$ B-spline basis function B_{rm} , and the regression coefficient of α_{rm} as

$$f_r(z_r) = \sum_{m=1}^{M_r} \alpha_{rm} B_{rm}(z_r)$$

Approximating the smooth function involves choosing the number of knots (where the knots are $z_r^{(min)} < \xi_{r0} < \dots < \xi_{rnr} < z_r^{(max)}$ equally spaced). The choice of the number of knots is that too many knots can lead to curves that overfit the data, resulting in too rough functions. However, too few results do not always capture the variability in the data; between 20-40 equally spaced knots ensures flexibility [12]. Thus, the penalized likelihood estimation is given as:

$$P(\lambda_r) = \frac{1}{2} \lambda_r \sum_{m=v+1}^{M_r} (\Delta^v \alpha_{rm})^2$$

where λ_r is the smooth parameter and Δ^v is the differencing operator.

3. RESULTS AND DISCUSSION

Table 1 indicates that regions Western Cape, Eastern Cape, Northern Cape, Free State, KwaZulu-Natal, and Mpumalanga were positively significant with the risk of having tuberculosis when all these regions were compared to Limpopo. The odds ratios were 6.098, 3.808, 3.979, 3.881, 2.729, and 2.376, respectively. This implies that men from all provinces mentioned above are more likely to have TB than men from Limpopo [13].

Table 1. The parameter estimates for tuberculosis among adult men for the fixed effect of GAMM.

Parameter	Estimate	S.E	OR	P-Value
Intercept	-3.502	0.363	0.030	<2e-16***
Chronic disease (ref = NO)	-	-	-	-
Yes	0.129	0.168	1.138	0.443

(Table 1) contd.....

Parameter	Estimate	S.E	OR	P-Value
Region (ref = Limpopo)	-	-	-	-
Western Cape	1.808	0.385	6.098	2.70e-06***
Eastern Cape	1.337	0.289	3.808	3.78e-06***
Northern Cape	1.381	0.335	3.979	3.87e-05***
Free State	1.356	0.325	3.881	3.05e-05***
KwaZulu-Natal	1.004	0.307	2.729	0.001**
North West	0.279	0.337	1.322	0.407
Gauteng	-0.602	0.466	0.548	0.196
Mpumalanga	0.865	0.310	2.376	0.005**
Education Level (ref = Secondary)	-	-	-	-
No education	0.122	0.254	1.130	0.631
Primary	0.442	0.168	1.556	0.009**
Higher	0.437	0.283	1.548	0.122
Ethnicity (ref = Black/African)	-	-	-	-
White	-2.023	0.757	0.132	0.008**
Colored	-0.725	0.283	0.485	0.011*
Indian/Asian	-997.9	4840	0.000	0.984
Other	-997.8	7823	0.000	0.990
Wealth Index (ref = Middle)	-	-	-	-
Poorest	0.042	0.198	1.042	0.833
Poor	-0.051	0.195	0.950	0.987
Richer	-0.551	0.225	0.577	0.014*
Richest	-0.252	0.272	0.777	0.354
Marital status (ref = Never in a union)	-	-	-	-
Married	-0.242	0.197	0.785	0.220
Living with partner	-0.326	0.230	0.722	0.887
Widowed	0.729	0.324	2.073	0.024*
Divorced	0.973	0.425	2.645	0.022*
Separated	0.313	0.347	1.367	0.367
Health (ref = Good)	-	-	-	-
Poor	0.752	0.217	2.121	0.001***
Average	0.208	0.168	1.232	0.214
Excellent	-0.169	0.257	0.844	0.509
Weight (ref =Normal)	-	-	-	-
Underweight	-0.632	0.186	0.532	0.051
Overweight	-0.878	0.377	0.416	0.02*
Obese	1.216	0.847	3.374	0.151
Don't know	-0.321	0.660	0.723	0.623
Smoking Status (ref = Everyday)	-	-	-	-
Do not smoke	-0.0002	0.148	1.000	0.999
Sometimes	-0.130	0.294	0.878	0.658

Men with primary education (OR =1.556) are more likely to have tuberculosis than men with secondary education. Whites and Coloreds have a negative association with the risk of having tuberculosis compared to Black men. The corresponding odds ratios are 0.132 and 0.485, respectively. This implies that Whites are (1-0.132) %= 86.8%% less likely to have tuberculosis than Blacks, and Coloreds are (1-0.485) %= 51.5% less likely to have tuberculosis than Black men [14].

Furthermore, richer men from South Africa have a negatively significant effect on the risk of having TB compared to men with a middle wealth index. The corresponding odds ratio is .577. This implies that richer men are (1-0.577) %=42.3% times less likely to be at risk of having tuberculosis

than the middle wealth index. The parametric effect for divorced and widowed men is a positively significant effect on the risk of having TB compared to those who were never in a union, with an odds ratios of 2.645 and 2.073, respectively. This implies divorced men are 2.645 times more likely to have TB than men who were never in a union, and widowed men are 2.073 times more likely to be at risk of having tuberculosis than men who were never in a union. Men with poor health have a positively significant effect of tuberculosis compared to men with good health. The corresponding odds ratio is 2.121. This implies that the odds of having tuberculosis for men with poor health status are 2.121 times more likely than for men with good health. Overweight is a negatively significant effect

on the risk of having TB compared to underweight with OR =0.416. This implies that overweight men are (1-0.416) %=58.4% less likely to have TB than underweight men.

Table 2 indicates the approximate significance of smooth terms for the fitted GAMM. The value of the statistic test is 4.356 with 2.009 degrees of freedom (p-value=0.014) against the assumption that the number of household members is linearly associated with the risk of TB. The statistic test is 1.179 with 6.073 degrees of freedom with high significance (p-value=1.34e-08) against the assumption that the interaction effect of age and the number of times away from home is linearly associated with the risk of tuberculosis. Fig. (1) shows the smooth term and confidence interval. The number of household members has some quadratic effect; the risk of

having tuberculosis increases with eight household members.

The findings from Table 3 show some similarity in the results in terms of the p-value. All significant variables in GAM are also significant in GAMM. The standard error of White increased by 0.26%. The standard error for widowed and divorced was increased by 0.19% and 0.02%. Western Cape, Eastern Cape, Free State, Northern Cape, KwaZulu-Natal, and Mpumalanga regions standard errors were increased by 0.44%, 0.24%, 0.31%, 0.27%, 0.33% and 0.32% respectively. The standard errors of significant parameters in the GAMM are higher than those of the significant parameters in the GAM. This suggests that the GAM may lead to false precisions and estimates. Thus, GAMM with smooth terms as the random effect is better than GAM.

Table 2. Approximate significance of smooth terms for GAMM.

Smooth Terms	Edf	F-value	P-value
S(Number of household member)	2.009	4.356	0.014*
s(Surrent age)	1.141	0.367	0.486
s(Number of times away from home)	1.000	0.253	0.615
s(Age, number of times away from home)	6.073	1.179	1.34e-08***

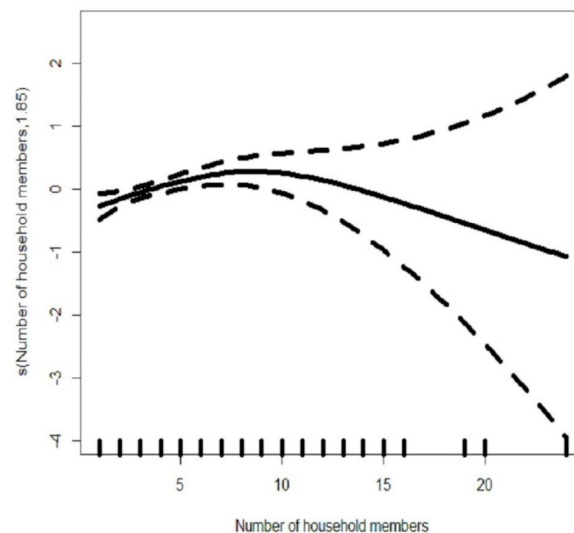


Fig. (1). Smoothing components for tuberculosis with number of household members.

Table 3. Comparison of GAM and GAMM.

Indicator	GAM			GAMM		
	Estimate	S.E	P-Value	Estimate	S.E	P-Value
Intercept	-3.490	0.362	<2e-16***	-3.502	0.363	<2e-16***
Chronic disease (ref = NO)	-	-	-	-	-	-
Yes	0.134	0.168	0.426	0.129	0.168	0.443
Region (ref = Limpopo)	-	-	-	-	-	-
Western Cape	1.824	0.383	2.02e-06***	1.808	0.385	2.70e-06***
Eastern Cape	1.337	0.288	3.47e-06***	1.337	0.289	3.78e-06***
Northern Cape	1.367	0.334	4.33e-05***	1.381	0.335	3.87e-05***
Free State	1.348	0.324	3.16e-05***	1.356	0.325	3.05e-05***
KwaZulu-Natal	1.013	0.306	0.005228**	1.004	0.307	0.001094**

(Table 3) contd.....

Indicator	GAM			GAMM		
	Estimate	S.E	P-Value	Estimate	S.E	P-Value
-						
North West	0.272	0.336	0.412	0.279	0.337	0.407
Gauteng	-0.599	0.464	0.196	-0.602	0.466	0.196
Mpumalanga	0.862	0.309	0.005228**	0.865	0.310	0.005241**
Education Level (ref = Secondary)	-	-	-	-	-	-
No education	0.100	0.253	0.693	0.122	0.254	0.631
Primary	0.432	0.168	0.009952**	0.442	0.168	0.008598**
Higher	0.462	0.282	0.101	0.437	0.283	0.122
Ethnicity (ref = Black/African)	-	-	-	-	-	-
White	-2.063	0.755	0.006279**	-2.023	0.757	0.007555**
Colored	-0.752	0.283	0.007887**	-0.725	0.283	0.010495*
Indian/Asian	-41.890	8523000.000	1.000	-997.900	4840.000	0.984
Other	-41.740	47454000.000	1.000	-997.800	7823.000	0.990
Wealth Index (ref = Middle)	-	-	-	-	-	-
Poorest	0.057	0.198	0.773	0.042	0.198	0.833
Poor	-0.023	0.195	0.908	-0.051	0.195	0.987
Richer	-0.562	0.224	0.012174*	-0.551	0.225	0.014212*
Richest	-0.276	0.272	0.310	-0.252	0.272	0.354
Marital status (ref = Never in union)	-	-	-	-	-	-
Married	-0.270	0.200	0.178	-0.242	0.197	0.220
Living with partner	-0.063	0.234	0.789	-0.326	0.230	0.887
Widowed	0.731	0.323	0.023568*	0.729	0.324	0.024309*
Divorced	0.977	0.425	0.021440*	0.973	0.425	0.022076*
Separated	0.311	0.347	0.370	0.313	0.347	0.367
Health (ref = Good)	-	-	-	-	-	-
Poor	0.739	0.216	0.000618***	0.752	0.217	0.000521***
Average	0.214	0.169	0.201	0.208	0.168	0.214
Excellent	-0.182	0.256	0.476	-0.170	0.257	0.509
Weight (ref = Normal)	-	-	-	-	-	-
Underweight	-0.358	0.185	0.054	-0.632	0.186	0.051
Overweight	-0.879	0.377	0.019644*	-0.878	0.377	0.020012*
Obese	1.288	0.840	0.125	1.216	0.847	0.151
Don't know	-0.346	0.659	0.600	-0.324	0.660	0.623
Smoking Status (ref = Everyday)	-	-	-	-	-	-
Do not smoke	-0.001	1472.000	0.936	0.000	0.148	0.999
Sometimes	-0.164	0.293	0.578	-0.130	0.294	0.657

CONCLUSION

The main purpose of this study was to examine tuberculosis risk factors using the 2016 South African Demographic Health Survey. The Generalized Additive Mixed Model was used to identify risk factors associated with tuberculosis. The response variable is the binary status of tuberculosis in men. The effect of age, number of household members, number of times away from home, and interaction effect of age and number of times away from home were analyzed non-parametrically while other covariates were modeled parametrically.

From the result, we can see that the only variables that are significant from both models are: region, race, marital status, weight, education and wealth index. We observe that the risk of having tuberculosis decreases for overweight men, as the higher body mass index (such as for overweight) negatively affects the risk of having tuberculosis compared to underweight. This result is supported by Buskin *et al.* (1994)

and Shetty *et al.* (2006).

The Western Cape, Eastern Cape, Northern Cape, Free State, KwaZulu-Natal, and Mpumalanga regions as compared to Limpopo, are highly affected by tuberculosis.

Education plays an important role as it is a significant effect on the risk of having tuberculosis. Men with primary education are most likely to be affected by tuberculosis when compared to men who have secondary education.

Whites and men of color have lower chances of having tuberculosis than Africans. Indians and other races have no significant effect of tuberculosis when compared to Blacks. Wealth index also affects the risk of having tuberculosis in South Africa. Richer men have a lower risk of having tuberculosis than men in the middle class. Furthermore, marital status affects the risk of having tuberculosis. Divorced men and widowed men are highly affected by tuberculosis when compared to men who have never been in a union.

In conclusion, the South African government needs to inspect men living in Western Cape, Eastern Cape, Northern Cape, Free State, and men with poor health - to reduce South Africa's infections. The government should also implement programs that will teach and discourage lower body mass index. Thus, targeting the factors that have a significant positive effect among adult men can help reduce the risk of having tuberculosis.

LIST OF ABBREVIATIONS

GAM	=	Generalized Additive Models
GLM	=	Generalized Linear Model
GAMM	=	Generalized Additive Mixed Model

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

HUMAN AND ANIMAL RIGHTS

Not applicable.

CONSENT FOR PUBLICATION

Not applicable.

AVAILABILITY OF DATA AND MATERIALS

The data supporting the findings of the article is available in the Demographic and Health Surveys at https://dh.sprogram.com/data/dataset_admin/login_main.cfm?CFID=35921003&CFTOKEN=eab0dc5e4d1182a9-994A35C3-A2C1-7748-F3B33EF5D3DBA210

FUNDING

None.

CONFLICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

Declared none.

REFERENCES

- [1] Hussain MS. Prevalence of tuberculosis 2020. <https://www.barnesandnoble.com/w/prevalence-rate-of-tuberculosis-muhammad-shabbir-hussain/1137783988?A=9786200431073>
- [2] World Health Organization. Tuberculosis in women. World Health Organisation 2018; p. 1.
- [3] World Health Organization. 2021. <https://www.who.int/news-room/fact-sheets/detail/tuberculosis>
- [4] What are South Africans dying of? Stats SA. 2013. <http://www.statssa.gov.za/?p=1023>
- [5] Hastie TJ, Tibshirani RJ. Generalized additive models. Boca Raton: FL CRC Press Sch 1990; Vol. 43.
- [6] Hastie Trevor J, Tibshirani Robert J. Gene Add Models. 1990; 43.
- [7] Hastie TJ, Tibshirani RJ. Generalized additive models 431986; : 1:297-310.
- [8] Chen C. Generalized additive mixed models. Commun Stat Theory Methods 2000; 29(5-6): 1257-71. [<http://dx.doi.org/10.1080/03610920008832543>]
- [9] Lin X, Zhang D. Inference in generalized additive mixed models by using smoothing splines. J R Stat Soc Series B Stat Methodol 1999; 61(2): 381-400. [<http://dx.doi.org/10.1111/1467-9868.00183>]
- [10] Breslow NE, Clayton DG. Approximate inference in generalized linear mixed models. J Am Stat Assoc 1993; 88: 9-25.
- [11] Eilers PHC, Marx BD. Flexible smoothing with B-splines and penalties. Stat Sci 1996; 11(2): 89-121. [<http://dx.doi.org/10.1214/ss/1038425655>]
- [12] Fahrmeir L, Kneib T, Lang S. Penalized structured additive regression for space-time data: a Bayesian perspective. Stat Sin 2004; 731-61.
- [13] Buskin SE, Gale JL, Weiss NS, Nolan CM. Tuberculosis risk factors in adults in King County, Washington, 1988 through 1990. Am J Public Health 1994; 84(11): 1750-6. [<http://dx.doi.org/10.2105/AJPH.84.11.1750>] [PMID: 7977912]
- [14] Shetty N, Shemko M, Vaz M, D'Souza G. An epidemiological evaluation of risk factors for tuberculosis in South India: a matched case control study. Int J Tuberc Lung Dis 2006; 10(1): 80-6. [PMID: 16466042]