



The Open Public Health Journal

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



RESEARCH ARTICLE

Risk Factors Associated with Tuberculosis Among Men; A Study of South Africa

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

¹School of Mathematics, Statistics and Computer Science, University of KwaZulu Natal, Berea, Durban, South Africa

Abstract:

Background:

One of the public health problems all over the world is tuberculosis. An important factor for human well-being is good health. Worldwide, there are more cases of men with tuberculosis than women. Therefore, identifying risk factors associated with tuberculosis among men is essential. This study uses a survey logistic regression model to identify risk factors associated with tuberculosis in South Africa using the 2016 South African Demographic Health Survey data.

Methods:

Based on the fact that tuberculosis status is a binary variable, logistic regression and survey logistics were used for analysis.

Results and Conclusion:

The findings using the survey logistic model are presented. The results suggest that a survey logistic model that accounts for complex sampling design is better than logistic regression. The findings from the study show that the risk factors associated with tuberculosis are: chronic disease, current age, region, race, number of times away from home, marital status, weight, smoking status, the interaction effect of chronic disease and age, and the interaction effect of smoking status and number of household members. These factors can be used to implement strategies for reducing the risk of having tuberculosis.

Keywords: Tuberculosis, Risk factors, Survey logistic regression, Binary response, Chronic disease, Men.

Article History

Received: September 17, 2021

Revised: November 9, 2021

Accepted: December 15, 2021

1. INTRODUCTION

One of the public health threats that remains in all countries is tuberculosis. An essential factor for human well-being is good health. At the same time, tuberculosis is a contagious infection that usually attacks the lungs. However, it can also spread to other parts of the body, such as the brain and spine [1]. A type of bacteria called Mycobacterium causes tuberculosis, which spreads through the air. However, being infected by the tuberculosis bacteria does not always mean a person will get sick. The diseases have two different forms, which are: Latent Tuberculosis and Active Tuberculosis. Latent TB is when someone has the bacteria, but their immune system prevents the bacteria from spreading. Active TB is when someone has bacteria that can multiply (spread) and attack their organs. TB can affect anyone anywhere, but the literature shows that most people who develop the disease are adults.

There are more cases of men than women [2]. In this study, we looked at the factors that affect the spread of tuberculosis among adult men and how they affect it so that those factors can be addressed. The number of persons developing TB can be reduced, and thus the number of deaths.

According to the World Health Organization [3], ten million people develop tuberculosis (TB). Even though the disease is preventable and curable, 1.5 million people die from TB each year, "making it the world's top infectious killer." About two-thirds of new TB cases in 2019 are in these eight countries: Bangladesh, Indonesia, China, India, Nigeria, Pakistan, Philippines, and South Africa, which are low-and middle-income countries [4]. Insights Statistics South Africa [5] has released a report dealing with mortality and the cause of death in South Africa. This is predicated on data collected from deaths in 2010 and was registered at the Department of Home Affairs. The total number of deaths decreased by 6.3% in 2010 compared to 2009. Data shows that more males than females died due to tuberculosis, and the highest number of deaths

* Address correspondence to this author at the School of Mathematics, Statistics and Computer Science, University of KwaZulu Natal, Berea, Durban, South Africa; Tel: 0849725211; E-mail: mlondo02@gmail.com

recorded was among the age group 30-39 years. TB was the leading cause of death in South Africa, accounting for about 12% of deaths which occurred in 2010 [5].

General household survey (GHS) results showed that 2.9% of tuberculosis sufferers said they were sick [6]. The most common age groups affected by TB were 25-34, 35-44, 45-54, and 55-64. The data showed that there were more Black Africans compared to the other race groups who were sick. Most sick or injured people included in the survey and who suffered from TB resided in KwaZulu-Natal.

The critical problems with tuberculosis are that it is not easy to diagnose the patient fast enough and treat them before spreading the germs to the communities. Another problem is to control the spread in public areas and public transport. In this study, we help determine the relationship of the factors affecting the spread and determine the risk factors for tuberculosis.

2. METHODS

2.2. Data Source

South Africa is one of the sub-Saharan Africa countries located in the southernmost region of Africa. South African Demographic and Health Surveys were conducted in 1998, 2003 and 2016. The 2016 South African Demographic and Health Survey (SADHS) was used in this study. The survey was implemented by Statistics South Africa (StatsSA) with the South African Medical Research Council (SAMRC). The sampling technique was performed in two-stage stratified sampling. In the first stage, samples were selected with a probability proportional to the sampling size of the primary sampling units. Secondly, systematic sampling was used in all dwelling units. Since SA is divided into nine provinces, primary sampling units were used to ensure survey precision across regions. Each region was stratified into urban, farm, and traditional areas [7].

2.2. Study Variable

The dependent or response variable in this study is when a male person was told by a health worker that he has tuberculosis, which is a binary variable. The explanatory variables utilised in this study are current age, region, ethnicity, smoking status, chronic disease, health, weight, education level, marital status, number of times away from home, wealth index, and the number of household members. Numerous researchers suggested these variables.

2.3. Statistical Methods

An outcome with two categories is called a binary outcome. The dichotomous response variable used in this research lends itself to logistic regression models as an obvious choice [8 - 11]. The binary logistic model is intended to depict a probability number between zero and one [12]. The logistic regression model is used to determine the association between a dichotomous response and a set of explanatory variables.

This model assumes that the data is collected by using a simple random sample. The logistic regression model can be defined mathematically as:

$$\text{logit}(\pi(x)) = \ln \left[\frac{\pi(x)}{1 - \pi(x)} \right] = \alpha + \sum_{i=1}^p \beta_i x_i$$

where $\pi(x)$ = probability of having tuberculosis, α = the intercept, β s = slope parameters, and X s = the independent or explanatory variables for the model.

However, a complex survey design was used due to the DHS sampling technique [13, 14]. Survey logistic regression is an extension of logistic regression, which includes the effect of sampling design to adjust estimates of standard errors and variability [15 - 20].

The survey logistic regression for a dichotomous dependent variable Y_{ijh} , $i = 1, \dots, n_{ij}$; $j = 1, \dots, n_j$; $h = 1, \dots, H$, where h is the stratum, j is the cluster and I is the household and denotes the sampling weight for ijh^{th} observation as w_{ijh} and x_{ijh} . The row vector of the matrix corresponding to the i^{th} adult man within the j^{th} primary sampling unit, nested in h^{th} cluster. Suppose that $\pi_{ijh} = P(Y_{ijh} = 1 | X_{ijh})$ is the probability of having tuberculosis.

Survey logistic regression model is then defined as:

$$\text{logit}(\pi_{ijh}) = \mathbf{x}_{ijh}^T \boldsymbol{\beta}$$

or

$$\pi_{ijh} = \frac{\exp(\mathbf{x}_{ijh}^T \boldsymbol{\beta})}{1 + \exp(\mathbf{x}_{ijh}^T \boldsymbol{\beta})}$$

where \mathbf{x}_{ijh} is covariate matrix, and $\boldsymbol{\beta}$ are an unknown vector of regression coefficients to be estimated. Pseudo-maximum likelihood was applied to achieve the estimation of unknown parameters of the model. The pseudo-maximum likelihood approach incorporates the sample weights and sample design to estimate unknown parameters [21, 22]. PROC SURVEYLOGISTIC in SAS 9.4 was used to fit a model to the data.

3. RESULTS

Table 1 shows that the interaction effect of smoking status and the number of household members has a negative association. As the number of household members increases, the odds of having tuberculosis for men who do not smoke and those who smoke sometimes decreases compared to those who smoke every day with an odds ratio of 0.827 and 0.756, respectively. The main effect of men with chronic disease was a statistically significantly higher risk of having tuberculosis than those without a chronic disease (OR=24.989, p -value<0.0001). Age has a positive association with the risk of having tuberculosis. This implies that with a one-unit increase in age, the risk of having tuberculosis increases by (1.067-1) %=6.7%.

Table 1. Survey logistic regression analysis of maximum likelihood.

Indicator	Estimate	S.E	P-value	OR
Intercept	-6.3473	0.6317	<.0001	0.002 (0.001;0.006)
Chronic disease (ref = NO)				
Yes	3.2184	0.6471	<.0001	24.989 (7.029;88.829)
Current age	0.0646	0.0109	<.0001	1.067 (1.044;1.090)
Region (ref = Limpopo)				
Western Cape	1.8474	0.5281	0.0005	6.343 (2.253;17.858)
Eastern Cape	1.3801	0.4258	0.0012	3.975 (1.726;9.158)
Northern Cape	1.6928	0.465	0.0003	5.434 (2.185;13.520)
Free State	1.3921	0.4464	0.0019	4.023 (1.677;9.651)
KwaZulu-Natal	1.1000	0.4278	0.0103	3.004 (1.299;6.948)
North West	0.3787	0.4787	0.4292	1.460 (0.571;3.732)
Gauteng	-0.8337	0.5851	0.1547	0.434 (0.138;1.368)
Mpumalanga	0.8692	0.4297	0.0435	2.385 (1.027;5.537)
Education Level (ref = Secondary)				
No education	0.4745	0.3401	0.1634	1.607 (0.825;3.130)
Primary	0.3808	0.2312	0.1000	1.464 (0.930;2.302)
Higher	0.6558	0.3409	0.0548	1.927 (0.988;3.758)
Ethnicity (ref = Black/African)				
White	-2.6057	0.9553	0.0065	0.074 (0.011;0.480)
Colored	-1.1217	0.4245	0.0084	0.326 (0.142;0.749)
Indian/Asian	-14.3155	0.4362	<.0001	0(2.58e-7;1.43e-6)
Other	-15.0754	1.1363	<.0001	0(3.05e-8;2.63e-6)
Times away from home	0.0437	0.0170	0.0105	1.045 (1.010;1.080)
Wealth Index (ref = Middle)				
Poorest	-0.0291	0.2970	0.9221	0.971 (0.543;1.738)
Poor	-0.0459	0.2871	0.8731	0.955 (0.544;1.677)
Richer	-0.1743	0.3044	0.5672	0.840 (0.463;1.526)
Richest	0.2703	0.3345	0.4194	1.310 (0.680;2.524)
Marital status (ref = Never in a union)				
Married	-0.00109	0.2480	0.9965	0.999 (0.614;1.624)
Living with partner	0.3022	0.3045	0.3213	1.353 (0.745;2.457)
Widowed	1.3966	0.6509	0.0323	4.041 (1.128;14.474)
Divorced	1.2704	0.4762	0.0078	3.562 (1.401;9.059)
Separated	0.6590	0.4233	0.1200	1.933 (0.843;4.431)
Health (ref = Good)				
Poor	0.4489	0.3173	0.1577	1.567 (0.841;2.918)
Average	0.0624	0.2177	0.7744	1.064 (0.695;1.631)
Excellent	-0.1349	0.3280	0.6809	0.874 (0.843;4.431)
Weight (ref =Underweight)				
Normal	-0.5296	0.2776	0.0568	0.589 (0.342;1.015)
Overweight	-1.1039	0.4789	0.0215	0.332 (0.130;0.848)
Obese	-0.5436	1.2255	0.6575	0.581 (0.053;6.413)
Don't know	-0.4199	0.9872	0.6708	0.657 (0.095;4.549)
Smoking Status (ref = Everyday)				
Do not smoke	0.8552	0.3173	0.0072	2.352 (1.263;4.380)
Sometimes	1.3504	0.6731	0.0452	3.859 (1.032;14.435)
Number of household members	0.1076	0.0430	0.0126	1.114 (0.095;4.549)
Interaction effect				
Chronic disease and age (ref = No)				
Having chronic disease and current age	-0.0654	0.0156	<.0001	0.937 (0.908;0.966)
No. of household members and smoking status(ref=Everyday)				
No. of household members and do not smoke	-0.1900	0.0638	0.0030	0.827 (0.730;0.937)

(Table 1) contd.....

Indicator	Estimate	S.E	P-value	OR
No. of household members and sometimes smokes	-0.2798	0.1356	0.0394	0.756 (0.579;0.986)
Current age and times away from home	-0.00103	0.000466	0.0280	0.999 (0.998;1.000)

All the regions that are statistically significant to tuberculosis have a positively associated risk of having tuberculosis. Men from Western Cape have higher odds than Limpopo (OR=6.343, p-value=0.0005), followed by men from Northern Cape compared to men from Limpopo (OR=5.434, p-value=0.0012). The risk of having tuberculosis for men from the Eastern Cape is 3.975 times higher compared to men from Limpopo (p-value =0.0012, for men from KwaZulu-Natal, it is 3.004 times higher compared to men from Limpopo with p-value= 0.0103. Furthermore, the risk of having tuberculosis for men from Mpumalanga is 2.385 times higher than in Limpopo men. Whites, Colored, Indians, and others are all negatively associated with the risk of having tuberculosis compared to Blacks/ Africans.

Table 1 also suggests that the increase in the number of times away from home for adult men increases the risk of having TB by (1.045-1) %=4.5%. The risk of having tuberculosis for widowers is 4.041 times higher than for men who were never in a union, followed by men who are divorced, which is 3.562 times higher than those men never in a union. The risk of having tuberculosis for overweight men is (1-0.332)%=66.8% less likely than underweight men; furthermore, as the number of household members increases, the risk of tuberculosis increases by (1.114-1)%= 11.4%.

Fig. (1) suggests that up to the approximate age of 50, the risk of tuberculosis is smaller for those with no chronic disease. After the approximate age of 50, the risk is higher for those with no chronic disease. Fig. (2) shows that for up to approximately five household members, the risk of tuberculosis

is lesser for daily smokers than for a non-smoker and those who smoke sometimes. After approximately five household members, the risk of tuberculosis is higher for daily smokers.

Table 2 provides the results of the two models, the fitted classical logistic regression model and survey logistic regression. The estimated coefficients and standard error from survey logistic regression were different from the logistic regression model under simple random sampling. The survey logistic estimates of the parameters of current age, chronic disease, Northern Cape region, White and Colored ethnicity, widowed and divorced were increased and significant for both models. The estimates of Western Cape, Eastern Cape, Free State, KwaZulu-Natal, and Mpumalanga regions, and the number of household members was decreased and significant in survey logistic regression compared to logistic. The coefficient of chronic disease in the survey logistic model increased by 53.84%, and standard errors were increased by 48.66%. The coefficient of White and Colored increased by 9.69% and 57.39%, respectively, with an increase in standard error by 25.58% and 50.11%, respectively. The coefficient of widowed and divorced was increased by 72.91% and 7.67%, with an increase in standard error by 98.63% and 13.49%, respectively. Northern Cape and KwaZulu-Natal coefficient were increased by 15.87% and 3.935%, respectively, increasing standard error by 38.56% and 38.72%, respectively. The coefficient of the Western Cape, Eastern Cape, Free State and Mpumalanga regions were decreased by 1.78%, 3.21%, 3.78%, and 18%, respectively. The standard error was increased by 36.85%, 46.68%, 37.1%, and 37.9%, respectively.

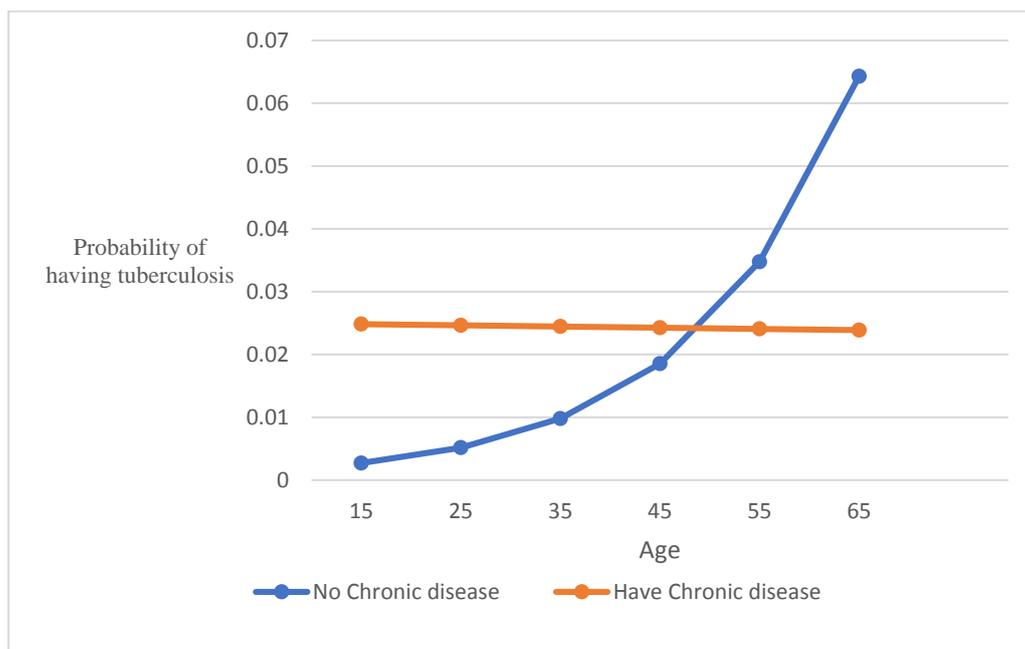


Fig. (1). Interaction effect for the current age and chronic disease for survey logistic regression.

The standard errors of significant parameters in the survey logistic regression were higher than the corresponding standard errors of the significant parameters in the logistic regression model. In the survey logistic model, the variables weight (overweight), smoking status and the interaction effect of the number of household members and those who smoke

sometimes, and the interaction effect of current age and number of times away from home were significant, while in logistic regression these were not. This suggests that excluding the complex design may lead to false precision and estimates. Thus, survey logistic regression model was suitable for this study.

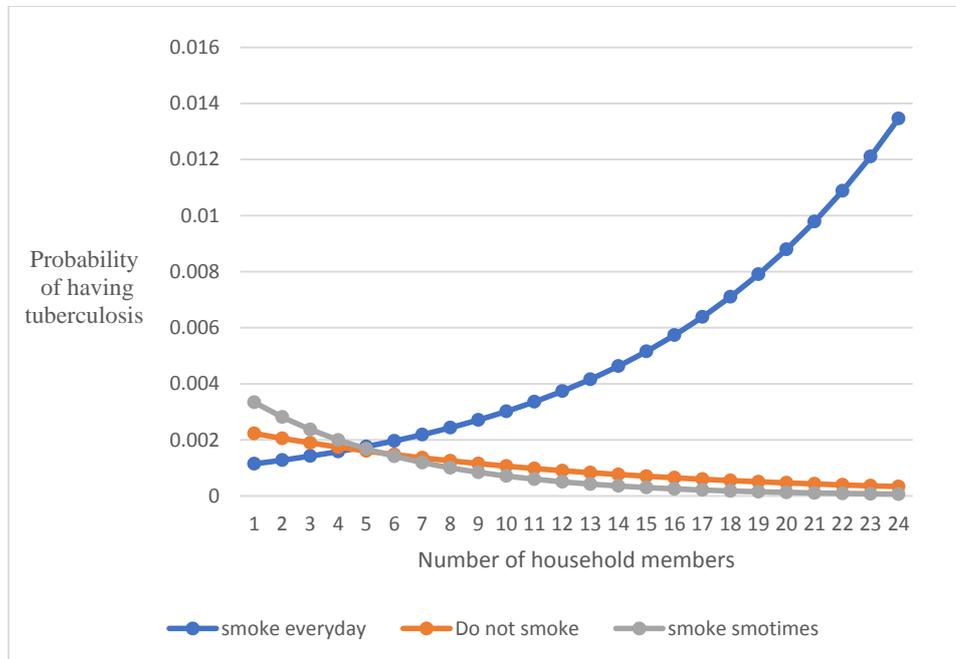


Fig. (2). Interaction for number of household members and smoking status for SLR.

Table 2. Comparison of the survey logistic regression and logistic regression.

Indicator	Survey Logistic Regression			Logistic Regression		
	Estimate	S.E	P-value	Estimate	SE	P-value
Intercept	-6.3473	0.6317	<.0001	-5.6284	0.4296	<.0001
Chronic disease (ref = NO)						
Yes	3.2184	0.6471	<.0001	2.0921	0.4353	<.0001
Current age	0.0646	0.0109	<.0001	0.0318	0.00609	<.0001
Region (ref = Limpopo)						
Western Cape	1.8474	0.5281	0.0005	1.8809	0.3859	<.0001
Eastern Cape	1.3801	0.4258	0.0012	1.4259	0.2903	<.0001
Northern Cape	1.6928	0.465	0.0003	1.4610	0.3356	<.0001
Free State	1.3921	0.4464	0.0019	1.4468	0.3256	<.0001
KwaZulu-Natal	1.1000	0.4278	0.0103	1.0584	0.3084	0.0006
North West	0.3787	0.4787	0.4292	0.3574	0.3373	0.2894
Gauteng	-0.8337	0.5851	0.1547	-0.5263	0.4658	0.2585
Mpumalanga	0.8692	0.4297	0.0435	1.0601	0.3114	0.0007
Education Level (ref = Secondary)						
No education	0.4745	0.3401	0.1634	0.0788	0.2581	0.7602
Primary	0.3808	0.2312	0.1000	0.43080	0.1675	0.0101
Higher	0.6558	0.3409	0.0548	0.48990	0.2807	0.0809
Ethnicity (ref = Black/African)						
White	-2.6057	0.9553	0.0065	-2.3755	0.7607	0.0018
Colored	-1.1217	0.4245	0.0084	-0.7127	0.2828	0.0117

(Table 2) contd.....

Indicator	Survey Logistic Regression			Logistic Regression		
	Estimate	S.E	P-value	Estimate	SE	P-value
Indian/Asian	-14.3155	0.4362	<.0001	-13.9226	464.7	0.9761
Other	-15.0754	1.1363	<.0001	-13.9182	2936.8	0.9962
Times away from home	0.0437	0.017	0.0105	0.0289	0.0136	0.0340
Wealth Index (ref = Middle)						
Poorest	-0.0291	0.297	0.9221	0.0415	0.1965	0.8328
Poor	-0.0459	0.2871	0.8731	-0.0174	0.1944	0.9285
Richer	-0.1743	0.3044	0.5672	-0.5068	0.2229	0.0230
Richest	0.2703	0.3345	0.4194	-0.1481	0.2705	0.5840
Marital status (ref = Never in a union)						
Married	-0.00109	0.248	0.9965	0.0257	0.1981	0.8967
Living with partner	0.3022	0.3045	0.3213	0.2416	0.2232	0.2791
Widowed	1.3966	0.6509	0.0323	0.8077	0.3277	0.0137
Divorced	1.2704	0.4762	0.0078	1.1799	0.4196	0.0049
Separated	0.6590	0.4233	0.1200	0.4508	0.3488	0.1962
Health (ref = Good)						
Poor	0.4489	0.3173	0.1577	0.7799	0.2157	0.0003
Average	0.0624	0.2177	0.7744	0.219	0.1668	0.1892
Excellent	-0.1349	0.328	0.6809	-0.233	0.2541	0.3592
Weight (ref =Underweight)						
Normal	-0.5296	0.2776	0.0568	0.3377	0.1849	0.0678
Overweight	-1.1039	0.4789	0.0215	-0.5149	0.3463	0.1371
Obese	-0.5436	1.2225	0.6575	1.5828	0.8265	0.0555
Don't know	-0.4199	0.9872	0.6708	0.0318	0.6415	0.9604
Smoking Status (ref = Everyday)						
Do not smoke	0.8552	0.3173	0.0072	0.4582	0.2468	0.0634
Sometimes	1.3504	0.6731	0.0452	0.5350	0.5107	0.2949
Number of household members	0.1076	0.043	0.0126	0.1132	0.0331	0.0006
Significant interaction effects in both models						
Chronic disease and age (ref = No)						
Having chronic disease and current age	-0.0654	0.0156	<.0001	-0.0395	0.0084	<.0001
No. of household members and smoking status(ref=Everyday)						
No. of household members and do not smoke	-0.1900	0.0638	0.0030	-0.1306	0.0447	0.0035

4. DISCUSSION

The likelihood ratio, score tests, and Wald test are statistically significant at a 5% level of significance. This means that there is a significant contribution of covariates in

the prediction of having tuberculosis. Table 3 shows that 79.8% of the probabilities are predicted correctly, suggesting a perfect association between the predicted and actual probabilities. The concordant is 79.4%, Gamma is 60%, and Somers' D is 59.5%.

Table 3. SLR model evaluation.

Model evaluation				
Overall Significance	F-value	Num DF	Den DF	Pr>F
Likelihood Ratio	6.96	34.7474	23772	<.0001
Score	6.22	41	625	<.0001
Wald	60.80	41	6250	<.0001
Association of predicted probability with observed				
Percent Concordant	79.4	Somers'D	0.595	
Percent Discordant	19.9	Gamma	0.600	
Percent Tied	0.8	Tau-a	0.069	
Pairs	752402	c	0.798	

CONCLUSION

Therefore, policymakers need to focus on the significant factors to develop strategies to reduce the risk of having tuberculosis. This study suggests that improving health by targeting chronic diseases will reduce the risk of tuberculosis. Targeting adult men to keep track of their tuberculosis status will help to reduce the risk of tuberculosis. Reducing the number of times away from home among men will reduce cases of tuberculosis. Implementing programs that will teach Blacks or Africans will reduce the risk of tuberculosis. The government of South Africa needs to implement programs targeted to the regions of Western Cape, Eastern Cape, Northern Cape, Free State, KwaZulu-Natal, and Mpumalanga to develop strategies of reducing the risk of having tuberculosis among men in South Africa.

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://dhsprogram.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] M S H Prevalence of tuberculosis 2020. Available from: <https://www.barnesandnoble.com/w/prevalence-rate-of-tuberculosis-m-uhammad-shabbir-hussain/1137783988?A=9786200431073>
- [2] World Health Organization. Tuberculosis in women. World Health Organisation 2018; p. 1.
- [3] Zaidi HA, Wells CD. Digital health technologies and adherence to tuberculosis treatment. *Bull World Health Organ* 2021; 99(5): 323-323A. [<http://dx.doi.org/10.2471/BLT.21.286021>] [PMID: 33958817]
- [4] World Health Organization. Tuberculosis World Health Organisation 2021. Available from: <https://www.who.int/news-room/fact-sheets/detail/tuberculosis>
- [5] What are South Africans dying of? Stats SA 2013. Available from: <http://www.statssa.gov.za/?p=1023>
- [6] Lehohla P. Use of health facilities and levels of selected health conditions in South Africa: Findings from the General Household Survey, 2011. Statistics South Africa Pretoria 2013.
- [7] South Africa - Demographic and Health. 2019. Available from: <https://microdata.worldbank.org/index.php/catalog/3408>
- [8] Agresti A. Introduction to categorical data analysis. 2007. [<http://dx.doi.org/10.1002/0470114754>]
- [9] Agresti A. Categorical data analysis. 2nd ed. New York: Wiley-Interscience 2002. [<http://dx.doi.org/10.1002/0471249688>]
- [10] Park H-A. An introduction to logistic regression: From basic concepts to interpretation with particular attention to nursing domain. *J Korean Acad Nurs* 2013; 43(2): 154-64. [<http://dx.doi.org/10.4040/jkan.2013.43.2.154>] [PMID: 23703593]
- [11] Queen JP, Quinn GP, Keough MJ. Experimental design and data analysis for biologists. Cambridge University Press 2002. [<http://dx.doi.org/10.1017/CBO9780511806384>]
- [12] Melesse S, Sobratee N, Workneh T. Application of logistic regression statistical technique to evaluate tomato quality subjected to different pre-and post-harvest treatments. *Biol Agric Hort* 2016; 32(4): 277-87. [<http://dx.doi.org/10.1080/01448765.2016.1184587>]
- [13] BA. A. Performing logistic regression on survey data with the new SURVEYLOGISTIC procedure. In Proceedings 27th Annu SAS Users Group Int Conf SUGI. 14-7.
- [14] Liu X, Koirala H. Fitting proportional odds models to educational data with complex sampling designs in ordinal logistic regression. *J Mod Appl Stat Methods* 2013; 12(1): 26. [<http://dx.doi.org/10.22237/jmasm/1367382300>]
- [15] Berglund PA. Applied survey data analysis. Chapman and Hall/CRC 2017.
- [16] Kish L. Survey sampling. New York: John Wesley & Sons 1965.
- [17] Moeti A. Factors affecting the health status of the people of Lesotho. PhD Thesis. 2007.
- [18] Skinner CJ, Holt D, Smith TF. Analysis of complex surveys. Wiley 1989.
- [19] Rao JN, Scott AJ. The analysis of categorical data from complex sample surveys: Chi-squared tests for goodness of fit and independence in two-way tables. *J Am Stat Assoc* 1981; 76(374): 221-30. [<http://dx.doi.org/10.1080/01621459.1981.10477633>]
- [20] Lu M, Yang W. Multivariate logistic regression analysis of complex survey data with application to BRFSS data. *J Data Sci* 2012; 10(2): 157-73. [[http://dx.doi.org/10.6339/JDS.201204_10\(2\).0001](http://dx.doi.org/10.6339/JDS.201204_10(2).0001)]
- [21] Hosmer DW, Lemeshow S. Special topics. *Appl Logist Regres* 2000; pp. 260-351.
- [22] Pfeffermann D. The role of sampling weights when modeling survey data. *Int Stat Rev Int Stat* 1993; pp. 317-37.