ISSN- 0301-1216
Indian J. Prev. Soc. Med. Vol. 57, No.2, 2026

Identification of Predictors (Risk Factors) of Multidrug-Resistant Tuberculosis (MDR-TB) in the Amravati region using Logistic Regression Analysis

Rajesh Singh[1], Shaziya Farah Ahemed[2], Pritee Singh[3]

ABSTRACT

Multidrug-resistant tuberculosis (MDR-TB) is a global public health threat and burden on the health system. Facilities for its diagnosis and treatment are limited in many high-burden countries, including India. This study aims to determine predictors (risk factors) associated with MDR-TB among the new and old TB patients from the Amravati region to provide future guidance. A cross-sectional study was conducted on all new and old TB patients registered under DOTS in the Amravati region from May 2012 to December 2019. Complete information on 402 patients, which included treatment details and other relevant clinical data, was used in this study. Binary logistic regression analysis is employed to determine the predictors (risk factors) for MDR-TB. The significant predictors for MDR-TB found are 'Number of times TB treatment taken in past', 'Positive follow-up of at least 4 months treatment or Relapse/defaulter from previous DOTS', and MDR-contact, HIV positive, irregular TB patients, CBNAT-detected TB patients, whereas the factor 'number of past outcomes-type' did not predict MDR-TB. The study shows that the MDR contact and CBNAT test results are the strongest predictors of MDR-TB. These factors among newly diagnosed TB patients should be taken as an alarm for the screening of MDR-TB. Mantel-Haenszel approach is also discussed in the study.

Keywords: Multidrug-resistant tuberculosis (MDR-TB); Predictors; Amravati; CBNAT; Mantel-Haenszel test; odds ratio.

Authour(s) Details:

1. Department of Statistics, Sant Gadge Baba Amravati University, Amravati, India
2. Department of Statistics, Sant Gadge Baba, Amravati University, Amravati, India
3. Department of Statistics, Institute of Science, Nagpur, India

Corresponding Address: Pritee Singh, Department of Statistics, Institute of Science, Nagpur, India,
Email: priteesingh25@gmail.com

Citation: Singh R., Ahemed S Farah , Singh P. Identification of Predictors (Risk Factors) of Multidrug-Resistant Tuberculosis (MDR-TB) in the Amravati region using Logistic Regression Analysis. Indian J Prev Soc Med, 2026; 57 (2): **190-202**. **DOI:** https://doi.org/

Sequence of Article: **Submission: 11.02.2026** **Accepted: 11.04.2026** **Published: 30.06.2026**

Prior Publication: Nil; Source of Funding: Nil; Conflicts of Interest: None, Article # 104/105

Introduction

Tuberculosis (TB) remains the world's deadliest infectious disease, killing more than 1.4 million people in 2019, despite being curable[1]. Multidrug-resistant tuberculosis (MDR-TB) is defined as a form of TB infection caused by bacteria that is resistant to treatment with at least two of the most powerful first-line anti-TB drugs, isoniazid (H) and rifampicin (R)[2]. Drug-resistant TB continues to be a public health threat. Worldwide in 2019, close to half a million people developed rifampicin-resistant TB (RR-TB), of which 78% had multidrug-resistant TB (MDR-TB). The three countries with the largest share of the global burden were India (27%), China (14%) and the Russian Federation (8%).[3] A global total of 206 030 people with multidrug-or rifampicin-resistant TB (MDR/RR-TB) were detected and notified in 2019, a 10% increase from 186 883 in 2018.[4] MDR-TB is treatable and curable by using second-line drugs. However, second-line treatment options are limited and require extensive chemotherapy (up to 2 years of treatment) with medicines that are expensive and toxic. In some cases, more severe drug resistance can develop. TB caused by bacteria that do not respond to the most effective second-line anti-TB drugs can leave patients without any further treatment options[4].

The Revised National TB Control Program (RNTCP), based on the internationally recommended Directly Observed Treatment Short-course (DOTS) strategy, was launched in 1997 and expanded across the country in a phased manner with support from the World Bank and other development partners [5]. It has been acknowledged that good quality of treatment is a prerequisite to the prevention of the emergence of resistance. RNTCP recognises that the implementation of a good-quality DOTS program is

the priority for TB control in the country. Prevention of the emergence of MDR-TB in the community is more imperative than its treatment.[6]. From the perspective of public health policy, a study on the screening of risk factors linked to MDR-TB at the onset of therapy, among new cases, is important to identify patients vulnerable to getting infected with MDR-TB strains. This is because patients with MDR-TB strains respond poorly to first-line treatment (DOTS), which has been reported in a multi-country study [7]. The End TB Strategy milestones for 2020 and 2025 can only be achieved if TB diagnosis, treatment and prevention services are provided within the context of progress towards UHC, and if there is multi-sectoral action and accountability to address the broader determinants that influence TB epidemics and their socioeconomic impact. In the context of the COVID-19 pandemic, multi-sectoral action to address these and other determinants of TB and its consequences, including GDP per capita, poverty and social protection, is more important than ever [3]. For this, it is necessary to detect high-risk factors, which play an important role in developing MDR among TB patients.

Objective of the Study: The Present study is an attempt to develop a valid model to define the probability of emergence of MDR-TB, in such a way as to gain knowledge about the factors from *treatment details of patients* that play an important role in the emergence of MDR-TB among tuberculosis patients under DOTS. The main objective behind it is to provide a better understanding for the health agencies for future planning, about the risk factors for the emergence of MDR-TB patients from the Amravati region registered under DOTS. In this study, potential variables that are believed to have a significant impact on the emergence of MDR are identified.

What is new in the study: In most of the previous studies, the researchers mainly considered TB patients registered under different courses and tried to determine the effect of socio-economic and demographic factors on the emergence of TB or MDR. However, in the present study, an attempt is made to follow the details of the past treatment history of tuberculosis patients, in addition to their current status and other demographic factors. Mostly, previous studies consider merely the presence or absence of earlier TB history as a risk factor. In the present study first time, 'number of times the patient has undergone TB-treatment in past', and 'number of times a particular outcome type has been observed in the past' are included as explanatory variables for the prediction of MDR-TB.

Background

Risk Factors for MDR-TB: Most commonly, socioeconomic and demographic risk factors like age, gender, occupation, locality, HIV status, etc., are used by the researchers in their studies for predicting potential effects on TB/MDR. Some of the prominent key factors mentioned by a few researchers are the previous history of TB treatment and its duration. Risk factors suggested in some of the previous research are mentioned here. It is known that epidemiologic information on risk factors of MDR-TB is important for the prevention and control of the spread of the disease. An approximately linear increase was observed in the likelihood of having MDR-TB as the total time of prior anti-tuberculosis treatment, measured in months, increased. Multivariate analysis revealed that prior anti-tuberculosis treatment, but not HIV positivity, was associated with MDR-TB.[8] The other known risk factors for MDR-TB include contact with a drug-resistant TB patient, younger age, lower socioeconomic status, homelessness, male gender, and less than normal body mass index.[9,10] The intensive phase of drug-resistant TB treatment, the first 8–12 months, is when patients suffer from critical disease conditions and drug side effects.[11, 12] Patients who have received TB treatment in the past include those who relapsed (have TB disease again after being declared "cured" in the past) and those who defaulted (did not complete a course of TB treatment as recommended in the past). Close to one- third of all patients who default or relapse will have MDR-TB strains. It is therefore vital to obtain an accurate history of prior TB treatment, as this is a very strong risk factor for drug-resistant TB. Up to 90% of all patients who fail treatment, defined as a positive AFB sputum smear or culture after 5 months of therapy, will have MDR-TB strains[13]. Given this, patients who remain smear positive after 3 months of treatment should be investigated for possible drug-resistant TB.[14]

Prediction Modelling: Modelling any (phenomena) mainly has two goals, which are to obtain a valid estimate of the relationship existing between independent variables and the outcome variable and, secondly, to obtain a good predictive model to capture the future behaviour of phenomena[15]. For each unique combination of predictors, a prediction model provides an estimated probability that allows for risk stratification for individuals or groups. Hence, it can guide physicians in deciding upon further diagnostic tests or treatments. For example, patients with a high probability of having a disease might be suitable candidates for further testing, while in low-probability patients, it might be more effective to refrain from further testing[16]. Risk prediction models estimate the risk (absolute probability) of the presence or absence of an outcome or disease in individuals based on their clinical and non-clinical characteristics[17, 18].

Model: Most of the previous work done on TB and MDR-TB researchers used multivariate logistic and binary logistic regression for predictive modelling and to develop CPR (clinical prediction rule). Some of the previous related research to be mentioned here are: univariate and multivariate logistic regression were conducted by *Hamdouni* et al.[19] To determine factors associated with poor drug-resistant tuberculosis treatment outcomes in Morocco, modelling of the data was done after completion of the MDR-TB treatment course, and factors responsible for poor outcomes of the treatment have been discussed. *Ali et al* employed Univariate and multivariate logistic regression analysis in their work to determine the predictors of MDR-TB infection in Sudan using sociodemographic characteristics and other relevant clinical data.[20] *Tamirat et al* reported that adverse drug reactions were determinants of more extended hospital stays.[21] A binary logistic regression analysis was used in this study to identify factors associated with more extended hospital stays during the intensive phase of multidrug-resistant tuberculosis treatment. *Kolhari et al* have suggested a valid logistic regression model to predict the probability of fail outcome in the TB treatment DOTS course[22]. The risk factors adopted in this study are Sex, age, weight, nationality, imprisonment and history of TB cases as predictors of the given outcome. In a multicenter case–control study, *Ahmad* et al, with the help of Multivariable logistic regression models, showed that ATB patients in the house or a history of prior TB treatment are strongly associated with MDR-TB.[23] Cases were more likely to have had a TB patient in the house before the diagnosis of MDR-TB (adjusted odds ratio [OR adj] = 3.1, 95% confidence interval [CI]: 1.2, 8.3) or had a history of prior TB treatment (OR adj = 4.2, 95% CI: 1.1, 15.4). *Tangüis* et al detected factors predictive of non-completion of the TB treatment among HIV-infected patients 24. Variables showing a statistically significant association were analysed at the multivariate level by means of a logistic regression model in this study. *Melzer et al* developed A multivariable logistic regression model to examine the association between past treatment and drug resistance, adjusted for other factors and used rpoB genotyping to determine the utility of 'risk assessment' in selecting Mycobacterium tuberculosis isolates for Rifampicin resistance.[25]

Methodology

Variables and Model Selection: The binary logistic regression model is selected for this study. Previous work [19-25], is used as the main reference while selecting the model. References and previous work done[8-14, 23] are used to identify potential variables that are believed to have a significant impact on the emergence of MDR-TB. The case definitions are obtained from the World Health Organisation, Standard WHO definitions used by RNTCP to classify TB patients and TB treatment outcomes [26].

Since the main objective of the study is to predict risk factors for MDR-TB from treatment details of TB patients, under this objective, keeping previous studies as a source of reference, the following variables using treatment cards of the patients are shortlisted for the study. The demographic and clinical factors selected for this study are: Age, gender, weight, TB detection by three different tests (X-ray, sputum smear test, CBNAT). Patient's TB History details under various criteria are also considered. New patients who do not have any previous history of TB treatment are also selected. MDR suspect criteria are used to identify presumptive MDR-TB patients registered under DOTS, these patients are then referred to a MDR-TB test. To determine the effectiveness of 'MDR suspect criteria' and 'patients' status at the time of MDR diagnosis test' in the prediction of emergence of MDR, these two variables are also included in this study. At most six previous TB treatment details are considered for this study.

To make the study focused on new dimensions under the objective, a total of 19 variables were selected as main risk factors, which include treatment details of patients along with demographic and clinical factors. Socioeconomic factors, symptoms, information about other diseases, and medication details are not considered here to limit the study.

Variable Types and Categories: Variables selected for the study are: Binary variable MDR as a response variable, and 18 independent variables. Age, weight, 'Number of times patient took TB treatment in past', 'duration of the past TB treatment on totality (in days)', 'number of times patient's result was cured, relapsed, complete/unknowns, defaulter, or failure in the past' are taken as continuous explanatory variables. Among categorical variables there is a response variable MDR diagnosed by Culture/CBNAT (MDR present, MDR absent), and other categorical explanatory variables as-Gender(female, male),Type of disease (pulmonary, extra-pulmonary), Smear Sputum result (positive high, medium, low, negative), X-ray report (Stained/cavity present, clear), TB detection by CBNAT (Positive, negative), Patient status at the Time of MDR testing(relapse, failure, defaulter, Retreat , ongoing TB treatment suspended due to adverse drug reaction, new), Suspect criteria (Any failure in past, Sputum positive follow-up of at least 4 months of previous DOTS course/relapse/defaulter ,TB-patients having MDR contact, HIV positive TB patients, retreat extra- pulmonary, private referral, New TB patient tested positive by any/all of three test, New Sputum negative/ New extra pulmonary), Regularity status in past treatments (regular most of the time, irregular/frequent breaks / long gapes/ unknown),Transfer Code (history number and follow-up month when patients referred to MDR testing).

Population and sample size: All bacteriologically confirmed DR-TB patients who initiated treatment during the study are included. Parameters that are important inputs for sample size determination were not available. Hence, the estimation of sample size for this study is done by using the general principle of having >10 patients with unfavourable events per variable for predictive modelling to prevent the problem of overfitting [27]. When developing a risk model, a rule of thumb based on the events per variable (EPV) ratio is often used to determine the sample size. The EPV is the number of events in the data divided by the number of regression coefficients in the risk model. It has been suggested that an EPV of 10 or more is needed to avoid the problem of overfitting.[28, 29]

The Purposive Sampling is adopted at first, since this study is mainly on the emergence of MDR. So here, those patients are selected who have compulsorily undergone MDR-TB testing by any of the approved MDR-TB tests [2, 6], after routine TB-Test from among the patients who were registered under DOTS-TB for the years 2012 to 2019 in Amravati region. Out of these confirmed TB cases, 650 patients are selected initially for this study, and then only 402 complete data entries are retained.

Data collection: All the data are obtained from different DOTS centres under the Municipal Corporation Amravati region. **Primary** data is collected for the study. Patients' treatment cards are used to collect the data from the year 2012 to 2019. From the Municipal Corporation Head Office, data is gathered for all DOTS centers under 12 PHU in Amravati District. A total of 1190 presumptive MDR patients have undergone (CBNAT/culture) drug resistance test for this period, out of which 290 cases were confirmed MDR-TB patients and 900 non-MDR during this period. Initially, 360 non-MDR cases (40%) of the data from each PHU are collected by using stratified sampling, and all 290 MDR cases (100% selection) are considered for this study. From the initial 650 collected cases (MDR + non-MDR diagnosed patients), due to missing and incomplete data, 248 entries were dropped to avoid any kind of ill result. The remaining 402 complete entries are retained for the final analysis of this study. It consists of 195 non-MDR (22%) and 207 MDR (70%) patients of the total population under study. Even after dropping 248 (38 %) of incomplete entries, 402 patients' details constitute a large enough sample from the perspective of logistic regression modelling. This decision is supported by the previous studies [27,28].

Ethics: There was no requirement for ethical approval for this study; likewise, patients' data were kept strictly anonymous and confidential. Still, the permission is obtained from THO of the concerned DPT.

Mantel-Haenszel approach: The Cochran-Mantel-Haenszel method is a technique that generates an estimate of association between an exposure and an outcome after adjusting for or taking into account confounding. The method is used with a

dichotomous outcome variable and a dichotomous risk factor. The data is stratified into two or more levels of the confounding factor, which creates a series of two-by-two tables showing the association between the risk factor and outcome at two or more levels of the confounding factor; then a weighted average of the risk ratios or odds ratios across the strata is computed (i.e., across subgroups or levels of the confounder) [30]. The test provides a primary *p*-value test for the meeting of conditional independence and odds ratio with a 95% confidence interval associated with the main effect [31].

Limitations of Stratified Analysis and the Mantel-Haenszel approach: In Stratified Analysis as the number of strata increases, it keep whittling away at the number of people in each stratum, so sample size becomes a major problem, since many of the strata will contain few or no people. [30] Mantel Haenszel approach give the idea of association between selected predictors and the outcome, but its use is limited up to dichotomous categorical response and explanatory variables[32]. In the presence of more than one explanatory variable, separately testing each independent variable against the response variable introduces bias into the research[32]. The present study contains many continuous and multilevel categorical explanatory variables. Due to the above-mentioned limitations, for further analysis and modeling purpose, logistic regression is considered in the study.

Logistic Regression Model: After selecting the important predictors to develop the model, using the base of existing research and previous work, Binary logistic Regression model has been used to estimate the probability of emergence of MDR. In fact, the Logistic Regression Model is very flexible to use.

Logistic regression does not assume a linear relationship between the dependent and independent variables, the dependent variables do not need to be normally distributed, there is no assumption of homogeneity of variance, in other words, the variances do not have to be the same within categories, normally distributed error terms are not assumed and the independent variables do not have to be interval or unbounded[33]. In many application areas, such as epidemiologic and biomedical studies, logistic regression is the standard approach for the analysis of binary and categorical outcome data. Many researchers in these areas have reported estimated odds ratios obtained from fitting a logistic regression model as a standard component of the results [34].

The logistic regression model, or the logit model as it is often referred to, is a special case of a generalised linear model. Analysis for the logistic regression model assumes the outcome variable is categorical. It is common practice to assume the outcome variable, denoted by Y, as a dichotomous variable having either a success or a failure. If we suppose that there is a single quantitative explanatory variable (X), for a binary response variable (Y), we note that $\pi(x)$ denotes the success Probability at value x. The probability is the parameter for the binomial distribution [35]. The logistic regression model has a linear form for the logit of this probability as follows:

$$\textbf{Logit } [\pi(x)] = \log [\pi(x)/1- \pi(x)] = \alpha + \beta x$$

where α and β are the regression parameters estimated by the maximum likelihood method [35].

Model derivation: In this study, Binary logistic regression analysis is used to identify the factors and to derive a model for predicting the probability of MDR. A random variable Y_i (i=1,2…,n) is defined as a response variable for i^{th} TB patient having MDR. Here we have a dichotomous dependent variable Y_i ($Y_{i=1}$ if MDR is present in a patient and Y_i =0 ; if MDR is not present in the patient) with probability pi. Then Yi ~ b (n, pi) . P(Y) is the probability of a given outcome to be predicted, and in this study, it is the probability of the presence of MDR for X_{im} (m= 1,2,3,...18), explanatory variables. Then the logit link function is defined as

$$\text{Logit } (p_i) = \log[(p_i / (1-p_i)]$$

$$P(Y) = \log [(e^{\beta_o +\ldots+ \beta_m X_{im}} / (1 + e^{\beta_0+\ldots+ \beta_m X_{im}})) *(1+ e^{\beta_0 +\ldots+ \beta_m X_{im}})]$$

$$P(Y) = \beta_0+\ldots+ \beta_m X_{im} \quad \ldots\ldots\ldots\ldots(1)$$

Where β_o is a constant term and βm(m=1, 2, …,18) is the effect (regression) coefficient.

Statistical Analysis: Mantel-Haenszel test has been applied using SPSS version 22 to test the association between independent variables under study and the dependent outcome variable MDR, across 12 strata (PHU) of the confounding effects. Then the final odds ratio at the higher Corporation Level is calculated, considering odds ratios at PHU level.

For binary logistic regression analysis, data are analysed by using the statistical software, SPSS version 22 and JASP version 0.14.1. The inclusion of significant factors was determined by Forward stepwise (likelihood) logistic regression to compute the predicted risk of MDR-TB and to determine the predictors (risk factors) for MDR- tuberculosis. Here, MDR-TB is a response variable with a set of 18 mixed (categorical and continuous) explanatory variables. The lowest (last) level is used as the reference category for each variable. A p-value of <0.05 is regarded as a statistically significant predictor for the development of MDR-TB. Maximum likelihood estimators (coefficients) are calculated for each factor from which odds ratios are determined. The odds ratios are expressed as probabilities of developing MDR-TB. These probabilities for various baseline factors are calculated with their 95% confidence limits. So the maximum likelihood parameter estimates, diagnostic and goodness of fit statistics, residuals and odds ratios were obtained from the final fitted logistic regression model.

Results

Mantel-Haenszel approach result: Mantel-Haenszel test results, for the outcome MDR-TB and different predictors, stratified across 12 PHU's are obtained using SPSS Ver. 22. The 'Test of Homogeneity of the odds ratio' table provides a p-value above 0.05 for the following predictors- 'Type of disease', 'X-ray result', 'CBNAT test result', 'regularity status in past', and 'gender', indicating *no violation of the testing conditions* and to continue with the analysis. In the next table for 'Test of Conditional independence', the p-value is found to be below 0.05 except for 'age'. For the predictor 'age', the p-value under the table 'Test of Conditional independence' is above 0.05. For the rest of the predictors (risk factors), which are continuous in nature or categorical variables having more than two categories, tests of Homogeneity and conditional independence could not be calculated by SPSS, due to the small sample size per variable per strata or continuous type predictor variables. Odds Ratio for different predictors (risk factors * MDR test result by CBNAT or culture * PHU Cross tabulation) obtained from SPSS is given in table-1.

Table-1: Mantel-Haenszel Common Odds Ratio Estimate

Risk factors	Estimate (Odd ratio)	Ln (Estimate)	SD Error of Ln (Estimate)	Asymptomatic Significance (2-sided)	Asymptomatic 95% Confidence Interval (Common Odds Ratio)
X-ray TB test result	4.937	1.597	0.262	0.000	2.955 - 8.248
TB diagnosis by CBNAT	56.802	4.040	1.013	0.000	7.803 - 413.480
Patient's regularity status in past treatments.	1.845	0.612	0.209	0.003	1.226 - 2.777
Type of disease of a patient.	2.543	0.933	0.342	0.006	1.301 - 4.971
Gender	1.052	0.051	0.214	0.813	0.691 - 1.602

Logistic Regression analysis result: Here, two cutoff values are considered. The cut-off value 0.5 is by default in SPSS; eight out of 18 variables are found significant at this level. Then the significance is checked by reducing the cutoff level to 0.25, and the same result is obtained at this cutoff point also. The eight significant predictor variables (risk factors) obtained in the final step of Forward logistic regression are: type of disease, Sputum, X-ray and CBNAT tuberculosis test results, patients' status before MDR testing, suspect code for MDR testing, number of times TB treatment taken in the past, regularity status of the patients in history during DOTS treatment. The details are given in Table 2.

Table-2: Significant variables (predictors) for MDR

Variable	Model Log Likelihood	Change in -2 Log Likelihood	df	Sig. of the Change
TB disease	-192.756	20.236	1	0.000
Sputum result	-188.407	11.539	3	0.009
Xray-result	-197.696	30.116	1	0.000
CBNAT result	-202.656	40.037	1	0.000
Patient Status	-191.114	16.952	5	0.005
Suspect Code	-198.599	31.923	7	0.000
No of Time TB-Treatment- Taken in the past	-194.520	23.764	1	0.000
Regularity Status	-186.347	7.419	1	0.006

Table-2 shows the results of eight significant predictors (independent variables), out of 18 independent variables obtained in the final step, from the Forward Binary Logistic Regression (likelihood) Test method with MDR-TB as the dependent variable. It shows that -2loglikelihood has decreased, suggesting the rejection of the null hypothesis that the independent variables are not related to the dependent variable. And Omnibus Tests of Model Coefficients also verify the model improvement (χ^2=192, p<0.0001)[20].

Table-3: Significant categories of eight explanatory variables obtained by using SPSS and JASP analysis and their β coefficients.

Parameter	Estimate of Coeffi-cients-B	±SD Error	Odds Ratio Exp(B)	Wald Test		95% Confidence interval	
				df	P value	Lower bound	Upper bound
(Intercept)	-7.083	1.38	8.392e -4	1	<0.001	-9.788	-4.378
CBNAT result (TB detected by CBNAT)	4.087	1.055	59.564	1	<0.001	2.018	6.156
Suspect Code (any failure in past)	1.07	1.624	2.914	1	0.051	-2.113	4.252
Suspect Code (Sputum positive follow-up after at least 4 months of DOTS/ relapse/ defaulter)	2.536	1.255	12.634	1	0.043	0.076	4.997
Suspect Code (TB patient having MDR contact)	5.699	1.352	298.495	1	<0.001	3.049	8.348
Suspect Code (HIV pos TB)	2.535	1.167	12.619	1	0.03	0.247	4.823
Suspect Code (retreatment extra pulmonary)	2.331	1.508	10.29	1	0.122	-0.625	5.287
Suspect Code (Private referral patient)	3.133	1.112	22.939	1	0.005	0.953	
Suspect Code (New patient tested TB positive by any or all three tests)	3.522	1.101	33.841	1	0.001	1.363	5.68
No. TB treatment taken in past	1.556	0.334	4.738	1	<0.001	0.901	2.21
X-ray result (cavity /stained)	1.844	0.379	6.323	1	<0.001	1.101	2.587
Disease type (Pulmonary TB)	-2.796	0.717	0.061	1	<0.001	-4.202	-1.39
Sputum result (Low TB bacteria)	1.366	0.425	3.919	1	0.001	0.534	2.198
Sputum Test result (medium TB bacteria)	0.918	0.448	2.504	1	0.04	0.04	1.796
Sputum Test result (high TB bacteria)	0.773	0.454	2.166	1	0.089	-0.117	1.663
Patient Status (relapse)	-2.391	0.886	0.092	1	0.007	-4.127	-0.655
Patient Status (failure)	0.871	1.298	2.389	1	0.043	-1.672	3.414
Patient Status (defaulter)	-1.186	0.882	0.305	1	0.179	-2.914	0.542
Patient Status (retreated)	-1.492	1.17	0.225	1	0.202	-3.786	0.801
Patient Status (ongoing TB treatment suspended due to adverse drug reaction)	-2.16	0.858	0.115	1	0.012	-3.841	-0.479
Reg Status (patient did not take the drug regularly, frequent break/long gap in past)	0.838	0.358	2.311	1	0.019	0.136	1.539

Both the software SPSS version 22 and JASP 0.14.1 gave the same result, with slight presentation differences. The detail result for the above selected eight significant predictors for their categories are shown in Table-3.

In Table-3, the second column is the column of coefficients; it lists the constants of=β1, β2,....,β20 for each related predictor (category-wise) in which it should be applied in the logistic regression equation (Eq-1) to calculate P(Y). The value of the coefficient for the constant is β0 (intercept). The sixth column of this table presents p-values significant for $p < 0.05$. According to the value of the Wald test, 15 categories of eight independent variables retained at the last step of Forward (likelihood) binary logistic regression analysis are found to be highly significant predictors for the given outcome.

The column of Exp (B) shows the list of odds ratios, which indicates that for each unit increase in the independent variables, the odds of development of MDR-TB in a given patient will increase by a factor of the value mentioned in the column "Exp (B)" for each variable. The confidence interval gives the confidence that the relationship between all predictors and MDR-TB found in this sample will be found in 95% of samples from the same population.

Sensitivity and specificity: The Classification Table-4 shows that this rule correctly classifies 173 out of 208, i.e., 83% of the subjects, where the predicted event (presence of MDR-TB in the diagnosed patient) was observed. This is known as the sensitivity of prediction. It is also seen that this rule correctly classifies 146 out of 194 (75%) of the subjects, where the predicted event is not observed. This is known as the specificity of prediction. Overall, the predictions are correct 315 out of 402 times, for an overall success rate of the model of 79%.

Table-4: Classification of correct decisions

Observed		Predicted		
		MDR diagnosed by culture or CBNAT		Percentage Correct
		MDR not	MDR diagnosed	
MDR diagnosed by culture or CBNAT	MDR negative	146	48	75.3
	MDR positive	35	173	83.2
Overall Percentage				79.4

Model validation: To understand the proportion of true positive cases, model classified as being positive and the proportion of true negatives it classified as being negative, the sensitivity and specificity of the model are tested. The area under the Receiver Operating Characteristic (ROC) curve, which ranges from 0 to 1, is used to assess the model discrimination.

(a) ROC curve: The area under the receiver operating curve (AUROC) is calculated after computing the predicted probability for each patient using the logistic regression coefficients. (LB = 0.83,UB = 0.90) Presented in figure-1.

(b) Train and Test method: To check the model validity, the Train and Test method, which is composed of building a predictive model with the training set (50% of the whole data) and then validating it using test and validation samples (25% each), is carried out. Results are as shown in table -5.

Figure-1 : Model validation by ROC curve

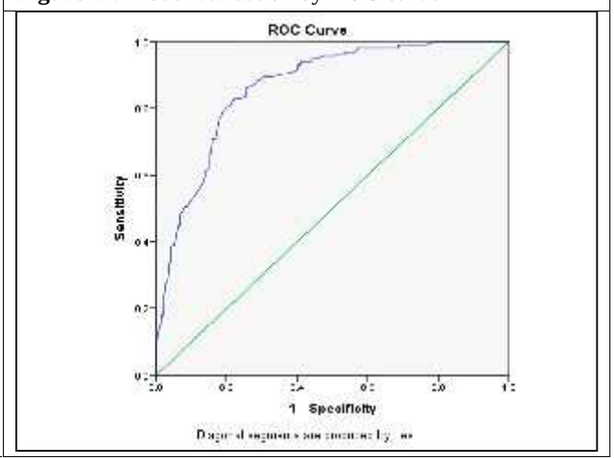

Table- 5: Train and test method results

Observed		Predicted					
		Training Sample			Testing Sample		
		Response recoded (1=Yes, 0=No		Percentage Correct	Response recoded (1=Yes, 0=No)		Percentage Correct
		No	Yes		No	Yes	
Response recoded (1=Yes, 0=No)	No	23	72	24.21	26	72	26.53
	Yes	0	87	100.0	3	112	97.39
Overall Percentage		0.0	54.72	60.44	10.34	60.87	64.79
Model Summary		**Training**			**Testing**		
Sum of Squares Error		48.373			25.641		
Percent Incorrect Predictions		26.1%			31.5%		

Model validation by R^2: Popular definition of R^2 uses the LR and −2 LL 0 as follows:

$$R^2 = (1 - \exp(-LR/n)) / (1 - \exp(- - 2\ LL0 / n)),$$

Where *n* is the number of patients. LL 0 refers to the log likelihood of the Null model, and LR is the likelihood ratio statistic. LR is the theoretically preferred quantity for testing significance in logistic models. This definition of *R* 2 was proposed by Nagelkerke and has the advantage of being scaled between 0 and 100%, making it a natural measure to indicate how close we are to our predictions to the observed 0 and 1 outcomes[36].

Discussion

This study is conducted in the Amravati region to determine the risk factors for MDR-TB infection among presumptive MDR-TB patients.

The Study developed a valid statistical model and identified risk factors (with significant categories, ref. Table-3, selected from patients' treatment details under DOTS, as the predictors for the emergence of MDR-TB. Multivariate binary logistic regression analysis was performed to ascertain the effects of age, weight, gender, and present and past treatment-related variables on the likelihood that participants have MDR-TB. The logistic regression model thus obtained is statistically significant; it is evident from the likelihood ratio (LR) test obtained from Omnibus Tests of Model Coefficients for the whole model as:

$\chi^2 = 192$, $p < 0.0001$).[20]

The model explained 46.0% (Nagelkerke R^2) of the variance in MDR-TB, indicating that predictions of the observed 0 and 1 outcomes are close to 0 and 46.0%, which is quite satisfactory.

The Error Sum of Squares for training samples (48.3) has reduced to (25.6) for testing samples. Overall percentage of correct classification has also improved from 60% to 65% for training and testing sets, respectively, revealing how correctly the model classified the given classes as MDR-Yes or MDR-not. AROC curve result is satisfactory when the area under the curve is under 87%.

The above suggested model correctly classified 79.0% of the cases. The sensitivity of the model, P (correct | event did occur), i.e. the percentage of occurrences of MDR-TB correctly predicted by the model, is 83%. The Specificity of the model, P (correct | event did not occur), i.e. the percentage of non-occurrence of MDR-TB correctly predicted by the model is 75%.

The Mantel Haenszel results table-1, for the predictors: 'Type of disease', 'X-ray result', 'CBNAT test result', 'regularity status in past' ($p<0.05$) indicates conditional independence between the groups (PHU strata) on the given outcome MDR-TB, it

means that association between these predictors and the outcome MDR-TB is significantly different, on different level of conditional variable PHU. For the predictor gender (p> 0.05), it indicates that groups (PHU strata) do not differ in their relationship with 'gender' and the outcome variable MDR-TB, which means the association between the predictor 'gender' and the outcome MDR-TB is not significantly different at different levels of the conditional variable PHU. The odds ratios obtained from the Mantel-Haenszel test do not vary much from the odds ratios of logistic regression analysis (table-3), for the same predictors.

From **logistic regression analysis**, the values of the odds ratio conclude that if the variable 'Number of times TB treatment taken in past' increases by one, then the odds of developing MDR- Tb would be 4.7 times higher. The risk of developing MDR-TB is 2.3 times higher in irregular TB patients than in regular ones.

The risk of MDR-TB development is highest among MDR contact patients. TB patient who has MDR-TB contacts is 298.4 times more likely to develop MDR-TB than a sputum-negative TB patient. Whereas a HIV-positive TB patient is 13 times more likely to develop MDR-TB than a sputum-negative TB patient. Failure patient from previous DOTS, is 3 times more likely to develop MDR-TB.

Positive follow-up of at least 4 months of treatment or Relapse/defaulter from previous DOTS has a 13 times higher risk of developing MDR than a sputum-negative TB patient. The risk of developing MDR-TB among private referral patients and new TB detected patients is 23 and 34 times higher than sputum negative TB patients. CBNAT-detected TB patients are 60 times more likely to develop MDR-TB than other TB patients. Whereas a positive X-ray patient is at 7 times higher risk than an X-ray negative. Inverted odds ratios for Pulmonary TB indicate that Pulmonary TB patients have 16 times more risk than Extra-pulmonary and sputum positive patient also have more risk of developing MDR-TB than negative TB patients.

This Study reveals MDR-TB contact as the strongest predictor (OR=298.4) for having MDR-TB. This result of the association between a history of close contact with MDR-TB patients and subsequent MDR-TB conforms to other studies [23, 37, 38]. In the present study, lung cavities are a risk factor for MDR-TB (OR= 6.3). This finding is in agreement with the previous study [38, 39]. This study also supports the results from earlier studies like 'History of TB treatment failure, relapse-TB, HIV-positive TB and irregular treatment as the strongest risk factors for development of MDR-TB [23, 37, 38]. It is not seen that age, weight and gender are significant predictors in the result of this study, as reported in earlier studies [23, 38, 39].

Conclusion

From this study, it can be concluded that the number of times a type of outcome in past TB treatments does not play any significant role in the prediction of MDR among old TB patients. Maybe this is due to low estimated power and lack of information for each outcome. The two important findings from this study are that an increase in the number of DOTS treatments for a patient increases his risk of being exposed to MDR-TB four times.

Most importantly, CBNAT TB-Test indicates the highest risk of developing MDR among TB-detected patients, so CBNAT can be considered the most powerful Test than the Sputum smear TB-Test and the X-ray TB-testing for early detection of MDR.

It may also be concluded from the study that for mixed-type data containing continuous and categorical types of variables, 'Logistic regression analysis' is more suitable in finding main effects of the significant predictors for a given outcome than the Mental Haenszel approach.

Since Multi Drug Resistant Tuberculosis testing is a high burden on poor Economic countries, there is a tool for early detection of the emergence of Multi Drug resistant among TB patients. Hence, this study helps in developing a valid model.

References

1. World Health Organization, WHO. Global TB Report 2020. https://apps.who.int/iris/ bitstream/handle/ 10665/ 336069/ 9789240013131-eng.pdf
2. Revised National Tuberculosis Control Programme. Guidelines on programmatic management of drug-resistant tuberculosis in India 2017. Central TB Division, Directorate General of Health Services, Ministry of Health & Family Welfare, Government of India , New Delhi. 2017; 1:216
3. World Health Organisation. Global Tuberculosis Report, Executive Summary. 2020; 1:2
4. WHO **fact** sheet on tuberculosis (TB). Oct 14, 2020. Available from: https://www.who.int/news-room/fact-sheets/detail/ tuberculosis.
5. Revised National Tuberculosis Control Programme. Annual status report: TB India 2012. Central TB Division, Directorate General of Health Services, Ministry of Health & Family Welfare, New Delhi. 2012; 1:2.
6. Revised National Tuberculosis Control Programme. DOTS-Plus Guidelines 2010. Central TB Division, Directorate General of Health Services, Ministry of Health & Family Welfare, New Delhi. 2010;1:1.
7. Dye C, Williams BG, Espinal MA, Raviglione MC. Erasing the World's Slow Stain: Strategies to beat multidrug-resistant tuberculosis. Science 2002; 295(5562):2042–6. [PubMed: 11896268]
8. Espinal MA, Laserson K, Camacho M, Fusheng Z, Kim SJ, Tlali RE. Determinants of drug-resistant tuberculosis: analysis of 11 countries. International Journal of Tuberculosis Lung Disease 2001; 5(10): 887-93.
9. Vaquero M, Gutierrez J, Casal MJ. Methodology of case–control studies in the epidemiology of multidrug-resistant tuberculosis. Rev Esp Quimioter 2000;13(1):20–30.
10. Lai HH, Lai YJ and Yen YF. Association of Body Mass Index with Timing of Death during Tuberculosis Treatment. Plos one 2017. Available from: https://doi.org/10.1371/ journal.pone. 0170104 .
11. Oladimeji O, Isaakidis P, Obasanya OJ, Eltayeb O, Khogali M, Bergh RV, Kumar AMV, Hinderaker SG, Abdurrahman ST, Lawson L, Cuevas LE. Intensive-phase treatment outcomes among hospitalised multidrug-resistant tuberculosis patients: results from a nationwide cohort in Nigeria. PLoS One 2014; 9(4):e94393.
12. Molie T, Teklemariam Z, Klinkenberg E, Dessie Y, Kumsa A, Mohammed H, Debebe A, Assefa D, Habte A, Bedru A, Fiseha D and Seyoum B. Intensive phase treatment outcome and associated factors among patients treated for multi-drug resistant tuberculosis in Ethiopia: a retrospective cohort study 2019. BMC Infect Dis; 19(1):818.
13. Lew W, Pai M, Oxlade O, Martin D, Menzies D. Initial drug resistance and tuberculosis treatment outcomes: systematic review and meta-analysis. Ann Intern Med 2008; 149(2):123-134.
14. Horne DJ , Royce SE, Gooze L , Narita M , Hopewell PC, Nahid P, Steingart KR. Sputum monitoring during tuberculosis treatment for predicting outcome, systematic review and meta-analysis. Lancet Infectious Diseases 2010; 10(6):387-394.
15. Singh GP, Tripathi A and Singh SK . Assessment of the effect of socioeconomic variables on child mortality through Mathematical Modelling. JP Journal of Biostatistics 2013; 9(1):818
16. Hendriksen JMT, Geersing GJ, Moons KGM, de Groot JAH. Diagnostic and prognostic prediction models. Journal of Thrombosis and Haemostasis 2013; 11(Suppl. 1):129–41.

17. Grobbee DE and Hoes AW. Clinical Epidemiology: Principles, Methods and Applications for Clinical Research. London Jones and Bartlett Publishers 2009.

18. Toll DB, Janssen KJ, Vergouwe Y, Moons KG. Validation, updating and impact of clinical prediction rules. Journal of Clinical Epidemiology 2008; 61(11): 1085–94.

19. Hamdouni ME, Bourkadi JE, Benamor J, Hassar M, Cherrah Y, Ahid S. Treatment outcomes of drug-resistant tuberculosis patients in Morocco: multicentric prospective study. BMC Infectious Diseases 2019; 19:316. Available from: https://doi.org/10.1186/s12879-019-3931-5

20. Ali MH , Al Rasheedy AA , Hassali MA, Dan Kibuule D and Godman B. Predictors of Multidrug-Resistant Tuberculosis (MDR-TB) in Sudan. Antibiotics 2019; 8(3):90. Available from: www.mdpi.com/ journal/ antibiotics,

21. Tamirat KS, Andargie G and Babel YA. Factors influencing the length of hospital stay during the intensive phase of multidrug-resistant tuberculosis treatment at Amhara regional state hospitals, Ethiopia: A retrospective follow-up study. BMC Public Health 2020; 20:1217. Available from: https://doi.org/10.1186/ s12889-020-09324-x

22. Sharareh R, Kalhori N, Nasehi M, Zeng XJ. A Logistic Regression Model to Predict High Risk Patients to Fail in Tuberculosis Treatment Course Completion. IAENG International Journal of Applied Mathematics 2010; 40:2

23. Ahsan M, Ahmad A, Akhtar S, Hasan R, Khan JA, Hussain SF, Rizvi N. Risk factors for multidrug-resistant tuberculosis in urban Pakistan: A multicenter case–control study. International Journal of Mycobacteriology, 2012; 1(1): 1 3 7 –1 4 2

24. Tangüis HG, Caylà JA, García PO, Jansà JM, Brugal MT. Factors predicting non-completion of tuberculosis treatment among HIV- infected patients in Barcelona. International Journal of Tuberculosis and Lung Disease 2000; 4(1): 55 –60.

25. Melzer M, Gupta N, Petersen I, Cook S, Hall B. Previous treatment in predicting drug-resistant tuberculosis in an area bordering East London, UK. International Journal of Infectious Diseases 2010;14(8):717-722.

26. Revised National Tuberculosis Control Programme. Technical and Operational Guidelines for Tuberculosis Control in India 2016, Central TB Division, Directorate General of Health Services, Ministry of Health & Family Welfare, Government of India, New Delhi. 2016.

27. Madan C, Chopra KK, Satyanarayana S, Surie D, Chadha V, Singh SK, Khanna A, Deshmukh R, Dutta L, Namdeo A, Shukla A, Sagili K, Chauhan SL. Developing a model to predict unfavourable treatment outcomes in patients with tuberculosis and human immunodeficiency virus co-infection in Delhi, India. PLOS ONE 2018; 13(10): e0204982. Available from: https://doi.org/10.1371/ journal. pone.0204982.

28. Pavlou M, Ambler G, Seaman SR, Guttmann O, Elliott P, King M, Omar RZ. How to develop a more accurate risk prediction model when there are few events. BMJ 2015; h3868. https://doi.org/ 10.1136/ bmj.h3868.

29. Peduzzi P, Concato J, Feinstein AR, Holford TR. Importance of events per independent variable in proportional hazards regression analysis. II: Accuracy and precision of regression estimates. J Clin Epidemiol 1995; 48: 1503-10

30. Wayne WL. Confounding and Effect Measure Modification. Boston University School of Public Health 2016. https://sphweb.bumc.bu.edu

31. Statistics. Cochran-Mantel-Haenszel, Conditional independence of a dichotomous categorical outcome. https:// www.scalestatistics.com/cochran-mantel-haenszel.html.

32. Sperandei S. Understanding logistic regression analysis. Biochemia Medica 2014; 24(1):12-8.

33. Maxwell Kori LH. Logistic Regression Analysis to Determine the Significant Factors Associated with Substance Abuse in School-Aged Children. Department of Mathematics and Statistics, Scholar Works, Georgia State University, Georgia. 2009.

34. O'Brien SM and Dunson DB. Bayesian Multivariate Logistic Regression. National Institute of Environmental Health Sciences, Biostatistics Branch 2003; MD A3-03.

35. Agresti A. An Introduction to Categorical Data Analysis (2nd ed). Wiley-Interscience, 2007.

36. Steyerberg EW. Clinical Prediction Models. Department of Public Health. Springer 2013; 4:60.

37. Mehari K, Asmelash T, Hailekiros H, Wubayehu T, Godefay H, Araya T, and Saravanan M. Prevalence and Factors Associated with Multidrug-Resistant Tuberculosis (MDR-TB) among Presumptive MDR-TB Patients in Tigray Region, Northern Ethiopia. Canadian Journal of Infectious Diseases and Medical Microbiology 2019. Available from: https://doi.org/10.1155/2019/2923549.

38. Mulu1 W, Daniel M and Mulat Y. Risk factors for multidrug-resistant tuberculosis patients in Amhara National Regional State. African Health Sciences 2015;15 (2).

39. Atre SR, D'Souza DTB , Vira TS, Chatterjee A, Mistry NF. Risk Factors Associated with MDR-TB at the Onset of Therapy among New Cases Registered with the RNTCP in Mumbai, India. Indian J Public Health 2011; 55(1):14-22.