Identification of Risk Factors of Multidrug-Resistant Tuberculosis by using Classification Tree Method

Dixin Tan,^{1,2} Bin Wang,¹ Xuhui Li,^{1,2} Xiaonan Cai,¹ Dandan Zhang,¹ Mengyu Li,^{1,2} Cong Tang,¹ Yaqiong Yan,³ Songlin Yu,¹ Qian Chu,⁴ and Yihua Xu^{1,2*}

¹Department of Epidemiology and Biostatistics, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China; ²The Ministry of Education (MOE) Key Lab of Environment and Health, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China; ³Wuhan Centers for Disease Control and Prevention, Wuhan, Hubei, China; ⁴Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China

Abstract. Multidrug-resistant tuberculosis (MDR-TB) has become a major public health problem. We tried to apply the classification tree model in building and evaluating a risk prediction model for MDR-TB. In this case-control study, 74 newly diagnosed MDR-TB patients served as the case group, and 95 patients without TB from the same medical institution served as the control group. The classification tree model was built using Chi-square Automatic Interaction Detector method and evaluated by income diagram, index map, risk statistic, and the area under receiver operating characteristic (ROC) curve. Four explanatory variables (history of exposure to TB patients, family with financial difficulties, history of other chronic respiratory diseases, and history of smoking) were included in the prediction model. The risk statistic of misclassification probability of the model was 0.160, and the area under ROC curve was 0.838 ($P < 0.01$). These suggest that the classification tree model works well for predicting MDR-TB. Classification tree model can not only predict the risk of MDR-TB effectively but also can reveal the interactions among variables.

INTRODUCTION

Multiple drug-resistant tuberculosis (MDR-TB) is caused by strains of *Mycobacterium tuberculosis* (MTB) that are resistant to isoniazid (INH) and rifampicin (RFP) at least. It is linked to the high morbidity and mortality¹ of TB and threatens the achievements of TB control at present. In 2015, MDR-TB cases in India, China, and the Russian Federation accounted for 45% of the total 580,000 cases.² Although the detection rate of MDR-TB among smear positive TB patients was high,³ the diversity, and complexity of drug resistant spectrum poses threat to TB control in China. Abundant literature demonstrated many risk factors,^{4,5} (demographics, environment, behavior, and genetic susceptibility^{6,7}) impact on the incidence of MDR-TB complicatedly. However, the correlation between the mechanism of these risk factors and the incidence of MDR-TB still remains unclear. Some researchers have attempted to screen and identify a set of optimal predictors potentially for MDR-TB in terms of pathogenesis through a logistic regression model.^{8,9} While the techniques of the logistic regression model are relatively mature, there are still several weaknesses (co-linearity of variables, interactions between variables, and judgment of high-risk groups¹⁰). By contrast, as outstanding data mining technology, the classification tree method becomes an alternative strategy to make up for shortcomings of traditional parametric tests, and identify some main factors affecting the occurrence of disease effectively. It has been used to predict a variety of diseases.^{11–13} Chi-square Automatic Interaction Detector (CHAID) decision tree has been used to formulate pathways for the early detection of metabolic syndrome in young adults.¹⁴ It also contributes to identify clients for methadone

treatment who experienced poorer treatment outcomes. Besides, it worked for investigators to build a prediction model for the incidence risk of ischemic stroke.¹³

Therefore, we hypothesized that classification tree model based on simple and reasonable decision rules can be established to predict the incidence risk of MDR-TB. Thus, we conducted a case-control study to assess the risk factors of MDR-TB and how these factors interact, and to evaluate the model, thus providing evidence for the early prevention of MDR-TB in China.

MATERIALS AND METHODS

Study subjects. From January to June in 2013, a non-matching case-control study was conducted to identify risk factors of MDR-TB. Seventy four newly diagnosed MDR-TB patients served as the case group. To ensure the representation and comparability of the control, 95 patients without TB from the same medical institution were randomly selected as the control group. Sputum culture and drug susceptibility testing with the proportion method for INH, RFP, ethambutol, and streptomycin were used to identify MDR-TB according to the guidelines for drug-resistant surveillance in TB (4th edition) published by the World Health Organization.¹⁵

Data collection. We used a structured questionnaire, including sociodemographic characteristics, living environment, dietary patterns, daily life, mental status, past medical history, and other potential risk factors, to explore risk factors of MDR-TB. All data were collected by self-reporting. “Town/city” and “Rural area” were two options of “Residence.” Exposure to patients with TB and with/without a mask when talking was also recorded. If a participant was in line with the description, record “Yes”; otherwise, record “No.” History of other chronic respiratory diseases was identified through asking participants “Has a doctor ever told you that you have chronic obstructive pulmonary disease, bronchial asthma, lung cancer, or etc.?” “Cigarette smoking” was defined as people who have smoked more than 100 cigarettes in their lifetime. Family economic condition was assessed according

* Address correspondence to Yihua Xu, Department of Epidemiology and Biostatistics and The Ministry of Education (MOE) Key Lab of Environment and Health, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Hongkong road no. 13, Wuhan, Hubei, China 430030. E-mail: xuyihua_6@hotmail.com

to the selection of “below median” and “equal to or over median,” with median income of 5,000 RMB/month. Physical exercise was investigated by two alternative terms: “less than three times a week” and “equal to or over three times a week.” The questionnaire was pretested. All interviews were conducted by unified-training interviewers under protection of the privacy of subjects.

Ethical approval. The study protocol was approved by the Ethics Review Committee of School of Public Health, Tongji Medical College, Huazhong University of Science and Technology. Before interviews, written informed consents were obtained.

Statistical analysis. First, we conducted a univariate analysis to identify significant variables primarily. Secondly, the multivariate unconditional logistic model was established to identify potential risk factors. Finally, the classification tree model with CHAID technique was applied to analyze the relationship between MDR-TB and risk factors discovered. We used stepwise analysis with the most significant predictor (the largest χ^2 value) to divide the entire cases into two or more mutually exclusive subgroups. Same as the first step, the cases in subgroups are further separated by the second significant predictor of the original outcome. The analysis continued until there were no more significant predictors. A two-sided P value ≤ 0.05 was considered with statistical significance.

All statistical analyses were performed using the SPSS statistical package, version 21.0.

RESULTS

Characteristics of subjects. All participants were Han, and the male-to-female sex ratio was 88:81 (52.1% were men). The average age of the case group (38.37 ± 14.55 years) was higher than that of the control group (36.86 ± 13.03 years), but there was no significant difference between two groups on age distribution (Table 1).

The univariate analysis demonstrated that the differences of the characteristics and risk factors between two groups were

statistically significant: occupation ($\chi^2 = 4.334$, $P = 0.037$), residence ($\chi^2 = 4.164$, $P = 0.041$), whether belongs to the floating population ($\chi^2 = 6.917$, $P = 0.009$), family economic condition ($\chi^2 = 16.761$, $P < 0.001$), vaccination of *Bacillus Calmette Guérin* ($\chi^2 = 12.161$, $P = 0.002$), history of other chronic respiratory diseases ($\chi^2 = 4.594$, $P = 0.032$), history of exposure to TB patients ($\chi^2 = 56.273$, $P < 0.001$), history of smoking ($\chi^2 = 12.325$, $P < 0.001$), exercise ($\chi^2 = 9.309$, $P = 0.002$), whether mask nose when talking with others ($\chi^2 = 4.509$, $P = 0.034$), and whether avoid others when cough ($\chi^2 = 8.216$, $P = 0.004$).

The unconditional logistic regression model was built to explore risk factors of MDR-TB. The results are presented in Table 1, which shows that some risk factors are found to be significantly associated with MDR-TB. These factors are residence, family with financial difficulties, suffering from other chronic respiratory diseases, exposure to TB patients, smoking history, physical exercise, and wore a mask when talking.

Results from classification tree method and rules for predicting the incidence risk of MDR-TB. Figure 1 revealed that the classification tree included four major predictor variables and nine nodes (including five terminal nodes) with a growing depth of three. Exposure to TB patients was the most important predictor as it split the first level of the tree into two branches. Subjects with exposure to TB patients had a significantly higher risk of MDR-TB than those without (81.4% to 17.2%, $P < 0.05$). For the group of subjects with exposure to TB patients, subjects whose families had financial difficulty would have a significantly higher MDR-TB risk (91.1% to 64.0%, $P < 0.05$). For the group of subjects who were not exposed to TB patients, the risk of MDR-TB between the subjects with or without other chronic respiratory diseases was significantly different (80.0% to 13.8%, $P < 0.05$). History of smoking was identified as the third prominent variable. Subjects with history of smoking had a higher risk of MDR-TB.

Table 2 demonstrates the decision rules of the model. The terminal nodes were ranked by the probability of MDR-TB (from 9.3% in node 7 to 91.1% in node 4). The decision rules

TABLE 1
 P value and estimated OR-value of risk factors from unconditional logistic regression model

Variables	Wald χ^2 value	P value	OR (95% CI)
Residence			
Town/city	—	—	—
Rural area	5.488	0.019	3.664(1.236–10.860)
Family financial condition			
Equal to or over median	—	—	—
Below median	7.166	0.007	3.881(1.538–10.476)
With other respiratory diseases			
No	—	—	—
Yes	12.324	0.000	83.522(7.061–987.995)
Exposure to TB patients			
No	—	—	—
Yes	41.407	0.000	30.593(10.792–86.719)
History of smoking			
No	—	—	—
Yes	6.146	0.013	4.069(1.342–12.339)
Physical exercise			
Less than three times a week	—	—	—
$\geq$ three times a week	4.028	0.045	0.342(0.120–0.975)
Mask his nose when talking			
No	—	—	—
Yes	6.502	0.011	0.262(0.094–0.734)

CI - confidence interval; OR = odds ratio.

FIGURE 1. The classification tree diagram by chi-square automatic interaction detector algorithm for predicting the incidence risk of multidrug-resistant tuberculosis. This figure appears in color at www.ajtmh.org.

included four significant predictor variables of MDR-TB: exposure to TB patients, family with financial difficulties, with other chronic respiratory diseases, and history of smoking. Subjects who were exposed directly to drug-resistant TB (cases of the second generation of TB patients) were those who had exposure to TB patients. Subjects who could not afford liver protection drugs were perceived as family with financial difficulties.

Calculation of the gain and index are presented in Table 3. The gain percent equaled to the gain number divided by the total cases. The method for calculating the response percentage was similar to that of the probability mentioned previously. The nodes 4, 5, and 3 had more cases than other cases (with index values greater than 100%), but the nodes 8 and 7 (with index values less than 100%) showed opposite results.

TABLE 2
Decision rules for the classification tree model of MDR-TB

Node number	Exposure to TB patients	Family with financial difficulties	With other chronic respiratory diseases	History of smoking	Probability of MDR-TB (%)
No.4	Yes	Yes	—	—	91.1
No.5	No	—	Yes	—	80.0
No.3	Yes	No	—	—	64.0
No.8	No	—	No	Yes	31.6
No.7	No	—	No	No	9.3

MDR-TB = multidrug-resistant tuberculosis.

The CHAID classification tree analysis showed that all participants were initially split based on the exposure to TB patients; hence, exposure to TB patients was the most important factor in regards to MDR-TB (Figure 2). Subjects with exposure to TB patients were at a higher risk of MDR-TB. Family with financial difficulties was another important factor. When Subjects had low income (node 4), 91.1% of the subjects developed MDR-TB. History of other chronic respiratory diseases was also an important variable for the development of MDR-TB. Without other chronic respiratory diseases (node 6), few subjects (13.8%) became MDR-TB, but with other chronic respiratory diseases (node 5), 80.0% became MDR-TB. The CHAID classification tree also showed that subjects reported of smoking were more likely to suffer from MDR-TB. As nodes 7 and 8 showed, if the subjects suffered to other chronic respiratory diseases, the rate of MDR-TB in those smoked was more than three times as high as in subjects without history of smoking (31.6% versus 9.3%, respectively).

The evaluation of exhaustive CHAID prediction model.

The prediction model was assessed by the misclassification risk estimate. Quantitatively, 84.0% of the subjects were correctly classified through the decision rules of this model (with 0.160 risk estimate and 0.028 standard errors). Graphically, the index value and the gains chart were both compliant with the standards: the index value started above 100%, remained on a high plateau as it moved along and then declined rapidly toward 100%; the gains chart rose steadily toward 100% (Figure 2). Furthermore, the results of the receiver operating characteristic curve (ROC) were presented in Figure 3. Specificity, sensitivity and the area under ROC curve were 82.4%, 85.3%, and 0.838, respectively. The area under ROC curve was statistically significant.

DISCUSSION

Classification tree is an effective data mining method in selecting risk factors and predicting the risk of multifactorial association diseases. CHAID of classification tree method

has been widely used in various fields of research.^{13,14,16–18} Compared with traditional model methods, such as multiple linear regressions, Cox proportional hazards model, and logistic regression models, CHAID has distinctive superiority in dealing with problems such as risk factors screening and risk prediction, and can particularly demonstrates the complicated multifactorial interactions.¹⁰ Classification tree method with CHAID technique would be completely unaffected by collinearity, outliers, or distribution errors. In addition, it can discover and expose the interactions between the selected variables.¹³

On the basis of extensive literature review, we found that the classification tree model suited to studies of chronic diseases (hypertension, diabetes). Because TB is one of the major public health problems worldwide, we attempt to use this method to analyze its incidence risk for individuals. In our research, we identified four variables (exposure to TB patients, family with financial difficulties, history of other chronic respiratory diseases, and history of smoking) for the prediction of MDR-TB incidence risk using the classification tree model with simple and reasonable decision rules. Individuals who are exposed directly to cases of the second generation of TB patients may become primary MDR-TB, not acquired MDR-TB during or after treatment.^{19–23} It is the most important factor that influences the risk of developing MDR-TB.²⁴ Individuals who cannot afford liver protection drugs may result in bad therapy outcome, thereby becoming more financial difficult family.^{23,25–36} Individuals with other chronic respiratory diseases have low resistance to diseases and are more likely to get MDR-TB.^{28,29} And individuals who smoke or smoked before are regarded as population easier to be infected with diseases, including MDR-TB.^{30,31} Compared with logistic regression models, this classification tree model generated by CHAID algorithm reveals the multifactorial interactions among risk factors and determines the individuals who are at high risk of MDR-TB.

Contact with TB, family income, chronic respiratory diseases, and smoking, which were in turn selected by CHAID algorithm of classification tree method, may play more cardinal roles in the incidence risk of MDR-TB than other factors. These selected variables could be considered as the most fundamental factors of MDR-TB and be targeted as the major aspects in the primary prevention strategies of MDR-TB. The assessment of the prediction model demonstrated that it could determine the major risk factors of MDR-TB and reveal their hidden interactions reliably. Therefore, the results may help us screen individuals who are at high risk of MDR-TB and predict the incidence risk for particular group of people on the basis of the decision rules of classification tree. However, previous treatment of TB, the most significant risk factor associated with MDR-TB found in many studies^{6,8,23,25,27,30–36} was not included in this analysis. The reason was that we

TABLE 3
Calculation of gain and index for nodes (node by node)

Node number	Node		Gain		Response, %	Index, %
	N*	Percent†, %	N	Percent, %		
No.4	45	26.6	41	55.4	91.1	208.1
No.5	5	3.0	4	5.4	80.0	182.7
No.3	25	14.8	16	21.6	64.0	146.2
No.8	19	11.2	6	8.1	31.6	72.1
No.7	75	44.4	7	9.5	9.3	21.3

Growing method: chi-square automatic interaction detector; dependent variable; multidrug-resistant tuberculosis.

* Represents the number of cases in each node.

† Represents the percentage of cases in the total number of subjects for each node.

FIGURE 2. The gains and index charts of the classification tree by chi-square automatic interaction detector (CHAID) algorithm for predicting the incidence risk of multidrug-resistant tuberculosis (MDR-TB) (Growing method: CHAID; dependent variable; MDR-TB; Target category: cases of MDR-TB).

selected 95 healthy people without history of TB as controls in this study.

Some limitations in the process of this research should be acknowledged. First, the sample size was comparatively small, which might have a great impact on the results when the parameters of classification tree model were changed. To some extent, this would affect the accuracy of prediction and the causal association between risk factors and MDR-TB. Second, although this study was conducted under a case-control design that 95 healthy people without history of TB from the same medical institution as the control group. History of TB treatment, which was found as the most important risk factor of MDR-TB in many studies, was not included in the analysis. And the information biases might be inevitable when data were collected. Finally, we designed the risk model aimed at validating the risk factors identified from the former meta-analysis and found that the risk factors have the high degree to match the identified risk factors. Although this study is a validated study, overfitting cannot be ruled out either. At that

time, individuals who are at high risk of MDR-TB will be more easily identified and health promotion or preventive treatment can be taken immediately to prevent the occurrence of MDR-TB.

CONCLUSION

In summary, we identified four significant predictor variables (exposure to TB patients, family with financial difficulties, history of other chronic respiratory diseases, and history of smoking) of MDR-TB through a case-control study and established a prediction model by means of CHAID algorithm of classification tree method. It explained the main risk factors and their latent interactions to predict the risk of MDR-TB. Parameters of ROC curve suggested that the classification tree model worked well for predicting MDR-TB.

Received January 12, 2017. Accepted for publication August 23, 2017.

Published online September 25, 2017.

Acknowledgment: We thank Hong Xie and Qionghong Duan for their guidance and assistance in the process of field data collection during our study.

Financial support: This work was supported by a grant from The National Social Science Fund of China (No. 15BSH118) and Innovation Research Fund of Huazhong University of Science and Technology (No. 2013TS004).

Authors' addresses: Dixin Tan, Xuhui Li, Mengyu Li, and Yihua Xu, Department of Epidemiology and Biostatistics and The Ministry of Education (MOE) Key Lab of Environment and Health, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China, E-mails: tandixin0907@outlook.com, 562073458@qq.com, 271754582@qq.com, and xuyihua_6@hotmail.com. Bin Wang, Xiaonan Cai, Dandan Zhang, Cong Tang, and Songlin Yu, Department of Epidemiology and Biostatistics, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China, E-mails: gswangbin@hust.edu.cn, caixiaonan0415@hotmail.com, 634247983@qq.com, 2771114282@qq.com, and 9994636@hust.edu.cn. Yaqiong Yan, Department of Tuberculosis Control, Wuhan Centers for Disease Control and Prevention, Wuhan, Hubei, China, E-mail: 89519752@qq.com. Qian Chu, Department of Chest Oncology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China, E-mail: qianchu@tj.hust.edu.cn.

FIGURE 3. The receiver operating characteristic charts of the classification tree by chi-square automatic interaction detector algorithm for predicting the incidence risk of multidrug-resistant tuberculosis. This figure appears in color at www.ajtmh.org.

REFERENCES

1. Stellah GM, Isaack AL, Alexander WM, Riziki MK, Scott KH, 2015. The influence of mining and human immunodeficiency virus infection among patients admitted for retreatment of tuberculosis in northern Tanzania. *Am J Trop Med Hyg* 93: 212-215.

2. World Health Organization, 2016. *Global Tuberculosis Report*. Available at: http://www.who.int/tb/publications/global_report/en/. Accessed February 2, 2017.
3. Zhou ML et al., 2012. Analysis of the case detection and short-term effect of the Wuhan MDR-TB project. *Chin J Antituberculosis* 34: 299–303.
4. Zhang GL et al., 2013. Application of a hybrid model for predicting the incidence of tuberculosis in Hubei, China. *PLoS One* 8: e80969.
5. Chen W, Shu W, Wang M, Hou YC, Xia YY, Xu WG, Bai LQ, Nie SF, Cheng SM, Xu YH, 2013. Pulmonary tuberculosis incidence and risk factors in rural areas of China: a Cohort Study. *PLoS One* 8: e58171.
6. Gunther G et al., 2015. Multidrug-resistant tuberculosis in Europe, 2010–2011. *Emerg Infect Dis* 21: 409–416.
7. Zetola NM, Modongo C, Kip EC, Gross R, Bisson GP, Collman RG, 2012. Alcohol use and abuse among patients with multidrug-resistant tuberculosis in Botswana. *Int J Tuberc Lung Dis* 16: 1529–1534.
8. Liang L et al., 2012. Factors contributing to the high prevalence of multidrug-resistant tuberculosis: a study from China. *Thorax* 67: 632–638.
9. Jenkins HE, Gegia M, Furin J, Kalandadze I, Nanava U, Chakhaia T, Cohen T, 2014. Geographical heterogeneity of multidrug-resistant tuberculosis in Georgia, January 2009 to June 2011. *Euro Surveill* 19: 29–38.
10. Maroco J, Silva D, Rodrigues A, Guerreiro M, Santana I, de Mendonça A, 2011. Data mining methods in the prediction of Dementia: a real-data comparison of the accuracy, sensitivity and specificity of linear discriminant analysis, logistic regression, neural networks, support vector machines, classification trees and random forests. *BMC Res Notes* 4: 299.
11. Zhang JF, Goode KM, Rigby A, Balk AHMM, Cleland JG, 2013. Identifying patients at risk of death or hospitalization due to worsening heart failure using decision tree analysis: evidence from the Trans-European Network-Home-Care Management System (TEN-HMS) study. *Int J Cardiol* 163: 149–156.
12. Baltzer PAT, Dietzel M, Gröschel T, Kaiser WA, 2012. A simple and robust classification tree for differentiation between benign and malignant lesions in MR-mammography. *Eur J Radiol* 81 (Suppl 1): S4–S5.
13. Gan XM, Xu YH, Liu L, Huang SQ, Xie DS, Wang XH, Liu JP, Nie SF, 2011. Predicting the incidence risk of ischemic stroke in a hospital population of southern China: a classification tree analysis. *J Neurol Sci* 306: 108–114.
14. Miller B, Fridline M, Liu PY, Marino D, 2014. Use of CHAID decision trees to formulate pathways for the early detection of metabolic syndrome in young adults. *Comput Math Methods Med* 2014: 242717.
15. World Health Organization, 2010. *Multidrug and Extensively Drug-Resistant Tuberculosis (MXDR-TB) Global Report on Surveillance and Response*. Available at: http://www.who.int/tb/publications/global_report/en/. Accessed November 16, 2016.
16. Thein TL, Leo YS, Lee VJ, Sun Y, Lye DC, 2011. Validation of probability equation and decision tree in predicting subsequent dengue hemorrhagic fever in adult dengue inpatients in Singapore. *Am J Trop Med Hyg* 85: 942–945.
17. Horner SB, Fireman GD, Wang EW, 2010. The relation of student behavior, peer status, race, and gender to decisions about school discipline using CHAID decision trees and regression modeling. *J Sch Psychol* 48: 135–161.
18. Lahmann NA, Tannen A, Dassen T, Kottner J, 2011. Friction and shear highly associated with pressure ulcers of residents in long-term care—classification tree analysis (CHAID) of Braden items. *J Eval Clin Pract* 17: 168–173.
19. Becerra MC, Franke MF, Appleton SC, Joseph JK, Bayona J, Atwood SS, Mitnick CD, 2013. Tuberculosis in children exposed at home to multidrug-resistant tuberculosis. *Pediatr Infect Dis J* 32: 115–119.
20. Seddon JA, Hesselning AC, Godfrey-Faussett P, Fielding K, Schaaf HS, 2013. Risk factors for infection and disease in child contacts of multidrug-resistant tuberculosis: a cross-sectional study. *BMC Infect Dis* 13: 392.
21. Furukawa NW, Haider MZ, Allen SJ, Carlson SL, Lindquist SW, 2017. Resistance to first-line antituberculosis drugs in Washington state by region of birth and implications for latent tuberculosis treatment among foreign-born individuals. *Am J Trop Med Hyg* 96: 543–549.
22. Vashakidze L et al., 2009. Prevalence and risk factors for drug resistance among hospitalized tuberculosis patients in Georgia. *Int J Tuberc Lung Dis* 13: 1148–1153.
23. Li XX et al., 2015. Comparing risk factors for primary multidrug-resistant tuberculosis and primary drug-susceptible tuberculosis in Jiangsu province, China: a Matched-Pairs Case-Control Study. *Am J Trop Med Hyg* 92: 280–285.
24. Wang K et al., 2014. Factors contributing to the high prevalence of multidrug-resistant tuberculosis among previously treated patients: a case-control study from China. *Microb Drug Resist* 20: 294–300.
25. Yang XJ, Yuan YL, Pang Y, Wang B, Bai YL, Wang YH, Yu BZ, Zhang ZY, Fan M, Zhao YL, 2015. The burden of MDR/XDR tuberculosis in coastal plains population of China. *PLoS One* 10: e117361.
26. Chen S et al., 2013. Risk factors for multidrug resistance among previously treated patients with tuberculosis in eastern China: a case-control study. *Int J Infect Dis* 17: e1116–e1120.
27. Zhao P, Li XJ, Zhang SF, Wang XS, Liu CY, 2012. Social behaviour risk factors for drug resistant tuberculosis in mainland China: a meta-analysis. *J Int Med Res* 40: 436–445.
28. Liu CH, Li L, Chen Z, Wang Q, Hu YL, Zhu BL, Woo PCY, 2011. Characteristics and treatment outcomes of patients with MDR and XDR tuberculosis in a TB referral hospital in Beijing: a 13-year experience. *PLoS One* 6: e19399.
29. Bartu V, Kopecka E, Havelkova M, 2010. Factors associated with multidrug-resistant tuberculosis: comparison of patients born inside and outside of the Czech Republic. *J Int Med Res* 38: 1156–1163.
30. Rifat M, Milton AH, Hall J, Oldmeadow C, Islam MA, Husain A, Akhanda MW, Siddiquea BN, 2014. Development of multidrug resistant tuberculosis in Bangladesh: a case-control study on risk factors. *PLoS One* 9: e105214.
31. Skrahina A et al., 2013. Multidrug-resistant tuberculosis in Belarus: the size of the problem and associated risk factors. *Bull World Health Organ* 91: 36–45.
32. Liu Q, Zhu LM, Shao Y, Song HH, Li GL, Zhou Y, Shi JY, Zhong CQ, Chen C, Lu W, 2013. Rates and risk factors for drug resistance tuberculosis in northeastern China. *BMC Public Health* 13: 1171.
33. Ricks PM, Mavhunga F, Modi S, Indongo R, Zezai A, Lambert LA, DeLuca N, Krashin JS, Nakashima AK, Holtz TH, 2012. Characteristics of multidrug-resistant tuberculosis in Namibia. *BMC Infect Dis* 12: 385.
34. Daniel O, Osman E, 2011. Prevalence and risk factors associated with drug resistant TB in south west, Nigeria. *Asian Pac J Trop Med* 4: 148–151.
35. Caminero JA, 2010. Multidrug-resistant tuberculosis: epidemiology, risk factors and case finding. *Int J Tuberc Lung Dis* 14: 382–390.
36. Lomtadze N, Aspindzelashvili R, Janigava M, Mirtskhulava V, Wright A, Blumberg HM, Salakaia A, 2009. Prevalence and risk factors for multidrug-resistant tuberculosis in the Republic of Georgia: a population-based study. *Int J Tuberc Lung Dis* 13: 68–73.