Isworo, *et al. Kesmas: Jurnal Kesehatan Masyarakat Nasional (National Public Health Journal)*. 2023; 18 (4): 252-257
DOI: 10.21109/kesmas.v18i4.7357

Kesmas: Jurnal Kesehatan Masyarakat Nasional
(Kesmas: National Public Health Journal)

Development of an Evidence-Based Tool to Assess the Relative Vulnerability of Different Communities to Tuberculosis

Slamet Isworo[1], Sri Handayani[2*], Reece Hinchcliff[3], Zainal A Hasibuan[4]

[1]Department of Environmental Health, Faculty of Health Sciences, Universitas Dian Nuswantoro, Semarang, Indonesia, [2]Department of of Public Health, Faculty of Health Sciences, Universitas Dian Nuswantoro, Semarang, Indonesia, [3]School of Applied Psychology, Griffith Health Group, Griffith University, Queensland, Australia, [4]Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang, Indonesia

Abstract
Identifying specific tuberculosis (TB) vulnerabilities in populations based on their geographical, demographic, and epidemiological characteristics is an essential yet challenging requirement to help reduce and eliminate TB. Assessment tools that can accurately quantify the risks associated with key factors could be used to measure TB vulnerability efficiently and indicate the most appropriate range of interventions. This study aimed to develop TB vulnerability assessment tools based on a TB vulnerability assessment conceptual framework developed with Leximancer. Three steps to produce the tools were facet analysis, interpreting the facet to create a list of questions, and expert judgment to confirm the suitability of the questionnaire. The "everything is data" principle was used to identify the data sources and build the tools. The data came from multiple primary data sources, with a questionnaire survey and observational form, and secondary data from various governmental statistical departments in Indonesia to collect data related to demography, health indicators, climate, temperature, and air quality. These tools will be optimized at scale next year to evaluate their utility for prioritizing and prescribing health system responses to TB in different communities in Central Java Province.

Keywords: big data, Leximancer, tuberculosis, vulnerability assessment

Introduction

The World Health Organization released the End TB strategy to eliminate tuberculosis (TB) in 2035. The strategy aims to reduce TB incidence to 90%, reduce TB death-related diseases by 95%, and protect families from negative impacts.[1] As of 2022, the target is still far from achieved as the reduction of TB incidence has only reached 10%, death-related TB has only reduced by 5.9%, and 48% of people with TB face catastrophic health expenditure.[2] At the same time, countries with a high burden of TB, such as Indonesia (second highest TB incidence globally), also face additional challenges related to high rates of TB-HIV and TB multidrug resistance.[2]

Indonesia's TB control strategy for 2020-2024 focuses on finding and curing cases through integrated public–private mixed programs and strengthening diagnostic tools.[3] The TB control strategy has been implemented in every city/district in Indonesia, producing variable outcomes.[3] Previous studies have found significant variations in TB risk and outcomes based on the characteristics of communities residing in geographically defined areas.[4-6] The implication is that to accelerate the End TB goal in Indonesia, specific interventions need to be developed and delivered based on each community's unique level and type of TB vulnerability.[7] Even though the vulnerability measurement based on an individual was developed, identifying specific TB vulnerabilities based on area is challenging as no specific guidance is available. Other factors include lacking practical tools and methods to support these efforts and variable access to the required data types.[8]

Developing TB vulnerability assessment tools that quantify the risks associated with unique factors within different communities may facilitate the accurate and efficient measurement of TB vulnerability and guide more effective health and social welfare interventions. Previous studies have proposed strategies to measure TB vulnerability only based on individual factors.[8,9] Another study has measured TB vulnerability by solely relying on social factors analysis.[10] Based on a prior evidence synthesis project,[11] a framework to measure TB vulnerability based on geographical area was developed. This study aimed to extend that body of work by developing specific tools to measure TB vulnerability based on the previous-

Correspondence*: *Sri Handayani, Department of Public Health, Faculty of Health Sciences, Universitas Dian Nuswantoro, Nakula 1 Street No. 5-11, Pendrikan Kidul, Semarang City, Central Java 50131, Indonesia, Email: sri.handayani@dsn.dinus.ac.id, Phone: +62 811-2680-990*

Received : August 20, 2023
Accepted : November 20, 2023
Published : November 30, 2023

ly-created TB vulnerability framework.[11] This study would provide tools to measure the vulnerability of TB in the community and a guide to prioritize and prescribe health system responses to TB based on different risk factors in communities. The use of several concept components in the concept framework allows the comprehensive exploration of the risk factors and helps to establish robust tools for measuring vulnerability based on geographical area.

Method

The TB vulnerability assessment conceptual framework has five components: risk of TB transmission, damage caused by TB, available health facilities, TB burden, and TB awareness.[11] These components required data from individual and health facilities–based data. The individual-based data were the risk of TB transmission (including the level of knowledge, environmental condition, and individual susceptibility), the damage caused by TB (economic impact and social support), and TB awareness (TB literacy). The health facilities–based data were TB burden (the number of TB cases) and the number of health facilities available in the area. The framework was built with Leximancer, a software under the license of Queensland University of Technology (QUT) Australia that uses machine-learning techniques to perform quantitative content analysis.[12] The use of Leximancer as automatic content analysis helps data processing visualize conceptual maps by generating main concepts within the text and determining how they are related. Leximancer enables the analysis of more data more frequently.[13] While the output of the Leximancer analysis showed a collection of the most relevant concepts related to TB risk and reduction, along with their interconnections, these insights could not initially be directly converted into a questionnaire or any other tool that could be used practically to guide TB vulnerability assessment and reduction activities.

To extend this work, three further steps were undertaken to develop methodologically rigorous and practically useful tools: facet analysis, formulation of tools (a questionnaire, an observational form, and a secondary data list), and expert judgment.[14] Facet analysis was undertaken to interpret concepts and translate them into tools (a questionnaire, an observational form, and a secondary data list) to measure each framework component. The resulting concepts were then organized into a logical classification to build a hierarchical structure. Several alternative methods were also undertaken to develop a facet: drawing from Leximancer's topic guide as it is (clear description), structuring several concepts into a make-sense facet (need analysis), and digging deeper into sub-concepts to attain the meaning (need deeper analysis). The three alternatives were used when the concept was built with a more complex meaning of the sub-concept.

A questionnaire was developed in the second stage by interpreting the context created in the facet. One or more assertions were related to each context. The facet not only performed on the theme but also analyzed the quotation result. The questionnaire was created with a series of assertions, a Likert scale asking respondents how much they agree with each statement, and "yes" or "no" answers to knowledge questions. The type of respondent target for each setting and survey topic was likewise carefully examined. This process was repeated until every potential context for each candidate dimension had been discussed. A second analysis examined each question representing each prospective dimension to avoid making repetitive claims. The end outcome was a distributed array of statements for each potential dimension.[15]

This procedure was repeated until each theme's potential contexts had been examined. Further analysis was done to eliminate the likelihood of redundant responses by comparing each theme's questionnaire statements. Based on the component's candidates (risk of TB transmission, damage caused by TB, health facility, TB burden, and TB awareness) created in the previous stage, the resulting statements were automatically sorted into groups. This study employed additional clustering techniques to identify new representable groupings with synchronized and explicit concepts or dimensions. These results aligned with grouping questions with the least amount of redundancy.

In the third step, five experts were invited to evaluate the critical components that resulted from the facet analysis. These experts consist of three lecturers with at least four years of TB research experience until 2022; the lecturers were experts on TB study related to management and intervention. Two worked in the TB field in the Semarang City (Central Java Province, Indonesia) health office, with at least six years of experience until 2022. These two experts were the head of the TB program and the field chief of the TB intervention program in the Semarang City health office. The experts were invited to a meeting, informed of the results of the measurement tools (questionnaire, observation form, and secondary data form), and asked to select each list of questions represented in each TB vulnerability assessment tool component (risk of TB transmission, damage caused by TB, health facility, burden of TB, and awareness of TB).

Results

Figure 1 shows the sample of facet analysis results, for example, how the facet regarding health facilities built the context of services, facilities, systems, and policy.

Besides creating the theme for the analysis output, Leximancer produced quotations based on the paper analyzed. Table 1 outlines the quotation analysis used to build a context conclusion. For example, the quotation ". . . Treatment completion is understood to depend, in part, on immigrant knowledge and attitudes toward latent TB infection (LTBI) . . ." concludes as level of knowledge context. This result was part of the facet analysis using quotation analysis.

Table 2 shows how the context conclusion is defined (Table 1), and the list of questions and data sources are built. For example, the level of knowledge has three list topics for questions and can be gathered through the survey. The final result of the tools produced three instruments: a questionnaire, an observational form, and a secondary data form (Table 3).

Table 4 presents the observational form for collecting primary data on tuberculosis vulnerability assessment, while Table 5 provides a list of secondary data sources required to measure tuberculosis vulnerability assessment (see below). Table 5 lists the secondary data sources that will be used to supplement the primary data and provide additional insights into tuberculosis vulnerability.

Discussion

This study showed how to measure TB vulnerability with a big data approach to enable the prioritization of specific interventions. The assessment tools produced in this study were based on the big data concept (data gathered from multiple sources) with the principle that "everything is data."[16] While novel to Indonesia, this approach has been applied commonly for disease prevention and control in China and various other countries,

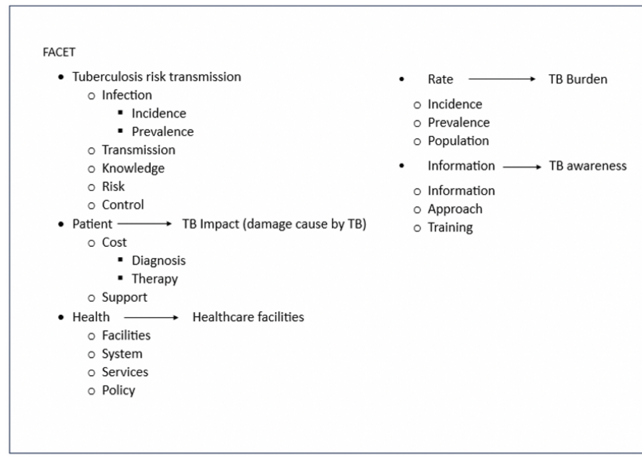

Figure 1. Facet Analysis of Tuberculosis Vulnerability Component

Table 1. Putting Context into Tuberculosis Vulnerability Facets Using Quotation Analysis

Facet	Quotation Example	Context Conclusion
TB risk transmission	. . . Treatment completion is understood to depend, in part, on immigrant knowledge and attitudes toward LTBI . . . . . . Participants with good knowledge of TBIC were three times more likely to practice TBIC than those with poor knowledge . . .	Level of knowledge
TB impact (damage caused by TB)	. . . He did not receive financial support from his family but did mention that the doctor gave him money sometimes . . . . . . Outpatient clinic costs constituted the highest proportion (38.7%) of the total cost, followed by medicine costs (21.6%) . . .	Social support, family support, economic impact

Notes: TB = Tuberculosis, LTBI = Latent Tuberculosis Infection, TBIC = Tuberculosis Infection Control

Table 2. How Facets Developed into the Tuberculosis Vulnerability Assessment Questionnaire

Facet	Context	Definition	Topic for the Question	Source
TB risk transmission	Level of knowledge	Measuring the level of knowledge about TB, including the cause, prevention, and treatment	TB is caused by *Mycobacterium tuberculosis*, which spreads through sputum when people cough How TB can be prevented How TB can be treated	Survey
TB risk transmission	Environmental condition	Measuring the environmental conditions, including humidity, temperature, and housing conditions	House humidity House temperature House condition: floor condition, building condition (permanent, semi-permanent)	Observational
TB risk transmission	Geographical condition	Measuring the geographical condition based on economic status, poverty level, and living density level	Poverty level based on city Population density based on city	Secondary data

Note: TB = Tuberculosis

Table 3. Questionnaire: Tuberculosis Vulnerability Assessment

Variable	No.	Question	Answer Choice	Analytic Expression
Knowledge	1	TB is caused by Mycobacterium tuberculosis, which spreads through sputum when people cough.	Yes / No	Yes: 1, No: 0
	2	If your family member gets a TB infection, you are also at risk of getting an infection.	Yes / No	Yes: 1, No: 0
	3	TB can be prevented when a person closes their mouth when they cough.	Yes / No	Yes: 1, No: 0
	4	TB can be prevented when a person closes their mouth when sneezing.	Yes / No	Yes: 1, No: 0
	5	TB can be prevented by having a BCG vaccine.	Yes / No	Yes: 1, No: 0
	6	TB can be treated by consuming OAT for at least six months.	Yes / No	Yes: 1, No: 0
	7	TB can be resistant if the patient's adherence to taking medicine is low.	Yes / No	Yes: 1, No: 0
Individual susceptibiliy	1	Have you been diagnosed with diabetes mellitus?	Yes / No	Yes: 1, No: 0
	2	Have you been diagnosed with HIV?	Yes / No	Yes: 1, No: 0
	3	Do you have someone in your house diagnosed with diabetes mellitus?	Yes / No	Yes: 1, No: 0
	4	Do you have someone in your house diagnosed with HIV?	Yes / No	Yes: 1, No: 0
	5	Do you have someone in your house diagnosed with TB?	Yes / No	Yes: 1, No: 0
Social support	1	In my community, TB patients are excluded from community gatherings.	Scale from 1 to 7 1: Unlikely 7: Most likely	Score from 1 to 7
	2	In my community, TB patients are more likely to get financial support from the neighborhood.	Scale from 1 to 7 1: Unlikely 7: Most likely	Score from 1 to 7
	3	In my community, TB patients are more likely to get mental support from the neighborhood.	Scale from 1 to 7 1: Unlikely 7: Most likely	Score from 1 to 7
	4	In my community, TB patients are more likely to get social support from the neighborhood.	Scale from 1 to 7 1: Unlikely 7: Most likely	Score from 1 to 7
Family support	1	If my family member gets a TB infection, I will help bring them to a doctor.	Scale from 1 to 7 1: Unlikely 7: Most likely	Score from 1 to 7
	2	If my family member gets a TB infection, I will help do the daily chores.	Scale from 1 to 7 1: Unlikely 7: Most likely	Score from 1 to 7
	3	If my family member gets a TB infection, I will listen to their private worries.	Scale from 1 to 7 1: Unlikely 7: Most likely	Score from 1 to 7
	4	If my family member gets a TB infection, I will show love and affection to them.	Scale from 1 to 7 1: Unlikely 7: Most likely	Score from 1 to 7
Economic impact	1	If my family gets TB, it will affect the family's income.	Scale from 1 to 7 1: Unlikely 7: Most likely	Score from 1 to 7
	2	If my family gets TB, it will affect the family's food consumption.	Scale from 1 to 7 1: Unlikely 7: Most likely	Score from 1 to 7
	3	If I get TB, it will affect my work.	Scale from 1 to 7 1: Unlikely 7: Most likely	Score from 1 to 7
TB literacy	1	How easy is it for you to get information about TB?	Scale from 1 to 7 1: Not easy 7: Very easy	Score from 1 to 7
	2	How easily can you understand the information you get about TB?	Scale from 1 to 7 1: Not easy 7: Very easy	Score from 1 to 7
	3	How often do you use the information to decide the next act?	Scale from 1 to 7 1: Never 7: Always	Score from 1 to 7
	4	How easy is it to process the information into action?	Scale from 1 to 7 1: Not easy 7: Very easy	Score from 1 to 7

Notes: TB = Tuberculosis, BCG = Bacillus Calmette Guerin, OAT = *Obat Anti Tuberculosis* (TB drugs), HIV = Human Immunodeficiency Virus

demonstrating its utility for efficiently and effectively detecting infectious and chronic diseases.[17] The study's findings revealed that big data analytics applications have been beneficial in managing chronic diseases in different stages and could potentially ease the burden of chronic illnesses.

Big data analytics have demonstrated the ability to extract insight from massive data sets and improve out-

Table 4. Observational Form of Tuberculosis Vulnerability Assessment

Subject	Answer
Humidity	Numeric
Temperature	Numeric
Ventilation condition	Bad
	Good
Floor condition	Soil
	Tile
	Carpet
Building condition	Permanent
	Semi-permanent
	Wood
House density	Numeric

Table 5. Secondary Data List for Tuberculosis Vulnerability Assessment

Data	Source
Poverty level based on city	Statistics Indonesia
Population density based on city	Statistics Indonesia
Number of clinics based on city	Health Office
Number of doctor practices based on city	Health Office
Number of pharmacies based on city	Health Office
Number of other health facilities based on city	Health Office
Incidence of tuberculosis based on city	Health Office

comes while minimizing expenses.[15] A previous study identified that the big data approach can be applied to improve chronic disease management, resulting in better outcomes, a lower disease burden, and lower treatment costs.[18] The tools outlined in this study will be able to measure disease burden and, through subsequent action, lower the burden and reduce the cost of prevention and treatment.

Compared to existing instruments, the tools created in this study focused not only on specific factors but also on measuring the complex parameters that can increase the chance of TB transmission. These tools will be able to provide a broader view of risk factors related to TB transmission and combine the analysis with machine learning to give more insight into how all the factors in combination increase the TB vulnerability of communities. A previous tool was provided only for measuring vulnerability in specific conditions related to individual or social vulnerability.[8,10] That has a high chance of failing to acknowledge and account for other factors that may contribute to vulnerability, such as environmental conditions.

Concerning TB prevention at both the global and national levels, one size does not fit all.[19] The mapping of geographical vulnerability, enabled through the tools developed in this study, will help TB program managers identify specific risk factors and build a bespoke array of interventions to decrease TB transmission in specific communities. For example, an education program could be the priority in areas with low TB knowledge, and community engagement may need to be increased in communities with high TB stigma. The same approach has been successfully applied to the education field, resulting in seven clusters of students and providing different interventions based on the risk of each cluster.[16]

One potential bias that might occur in this study was cognitive bias, where the experts' prior knowledge and experience in TB study and intervention might influence their evaluation of the measurement tools. There could also be confirmation bias if the experts only select questions or components that align with their preexisting beliefs or assumptions about TB vulnerability. To minimize the bias, the experts were provided with clear instructions for evaluating the measurement tools and encouraged to base their decisions on available evidence rather than personal opinion or experience. Additionally, a diverse group of experts from various backgrounds was consulted to ensure a broader perspective. Future study is required to explore the validity of the instruments and the framework component in measuring the degree of vulnerability of TB in communities. The big data approach and end-to-end methodology will be used for future investigation.

Conclusion

The application of big data, data linkage, and machine learning has the potential to greatly strengthen current approaches to identifying and reducing TB in vulnerable communities. While the project findings contribute to the increasing body of evidence and practical tools that can facilitate this goal, there are several limitations: there are few expert participants, and the tools have not yet been utilized for policy development purposes, impeding the ability (at this stage) to rigorously determine their validity and reliability, in practice. Future studies are underway to address this limitation.

Abbreviations

TB: Tuberculosis; LTBI: Latent Tuberculosis Infection.

Ethics Approval and Consent to Participate

This study was approved by the Institutional Review Board of Universitas Negeri Semarang (No. 315/KEPK/EC/2023) and performed following the principles of the Declaration of Helsinki. Informed consent was waived because of the retrospective nature of this study.

Competing Interest

The authors declared that there are no significant competing financial, professional, or personal interests that might have affected the performance or presentation of the work described in this manuscript.

Availability of Data and Materials

The datasets are not publicly available but are available from the corresponding author upon reasonable request.

Authors' Contribution

Conceptualization: SI, ZAH, SH; Data curation: SH, ZAH, SI; Formal analysis: all authors; Funding acquisition: SI, SH; Investigation: SI, SH; Methodology: SH, ZAH, RH; Project administration: SH; Resources: RH; Software: RH; Supervision: RH; Validation: ZAH; Visualization: SH; Writing–original draft: SH, SI; Writing–review and editing: all authors.

Acknowledgment

This study was supported by the Research Grant "Program Hibah Penelitian Fundamental Reguler DIKTI," Ministry of Research, Technology and Higher Education, of the Republic of Indonesia. Contract number 065/A38-04/UDN-09/VII/2023.

References

1. World Health Organization. The End TB Strategy. Geneva: World Health Organization; 2015.
2. World Health Organization. Global Tuberculosis Report 2022. Geneva: World Health Organization; 2022.
3. Kementerian Kesehatan Republik Indonesia. Strategi Nasional Penanggulangan Tuberkulosis di Indonesia 2020–2024. Jakarta: Kementerian Kesehatan Republik Indonesia; 2020.
4. Rahayu SR, Katsuyama H, Demura M, et al. Factors associated with tuberculosis cases in Semarang District, Indonesia: Case-control study performed in the area where case detection rate was extremely low. Environ Health Prev Med. 2015; 20 (4): 253–261. DOI: 10.1007/s12199-015-0443-9
5. Zhang Y, Liu M, Wu SS, et al. Spatial distribution of tuberculosis and its association with meteorological factors in mainland China. BMC Infect Dis. 2019; 19 (1): 379. DOI: 10.1186/s12879-019-4008-1
6. Beiranvand R, Karimi A, Delpisheh A, et al. Correlation Assessment of Climate and Geographic Distribution of Tuberculosis Using Geographical Information System (GIS). Iran J Public Health. 2016; 45 (1): 86–93.
7. The Republic of Indonesia Joint External Monitoring Mission for Tuberculosis. Geneva: World Health Organization; 2020.
8. Balakrishnan SK, Suseela RP, Mrithyunjayan S, et al. Individuals' Vulnerability Based Active Surveillance for TB: Experiences from India. Trop Med Infect Dis. 2022; 7 (12): 441. DOI: 10.3390/tropicalmed7120441
9. Vanleeuw L, Zembe-Mkabile W, Atkins S. Falling through the cracks: Increased vulnerability and limited social assistance for TB patients and their households during COVID-19 in Cape Town, South Africa. PLoS Glob Public Health. 2022; 2 (7): e0000708. DOI: 10.1371/journal.pgph.0000708
10. Trivianita N, Marsisno W, Wilantika N. Social Vulnerability Index to Tuberculosis of Provinces in Indonesia. Adv Soc Sci Educ Hum Res. Atlantis Press. 2019; 216: 170–180.
11. Handayani S, Hinchcliff R, Hasibuan ZA. Development of Tuberculosis Vulnerability Assessment Conceptual Framework Using Automatic Content Analysis. Healthcare Transformation with Informatics and Artificial Intelligence. Amsterdam: IOS Press; 2023. p. 220–223. DOI: 10.3233/SHTI230467
12. Ward V, West R, Smith S, et al. The role of informal networks in creating knowledge among healthcare managers: A prospective case study. Health Serv Deliv Res. 2014; 2 (12). DOI: 10.3310/hsdr02120
13. Engstrom T, Strong J, Sullivan C, et al. A Comparison of Leximancer Semi-automated Content Analysis to Manual Content Analysis: A Healthcare Exemplar Using Emotive Transcripts of COVID-19 Hospital Staff Interactive Webcasts. Int J Qual Methods 2022; 21: 16094069221118993. DOI: 10.1177/16094069221118993
14. Rahmah A, Santoso HB, Hasibuan ZA. Critical Review of Technology-Enhanced Learning using Automatic Content Analysis Case Study of TEL Maturity Assessment Formulation. Int J Adv Comput Sci Appl. 2022; 13 (1). DOI: 10.14569/IJACSA.2022.0130148
15. Raghupathi W, Raghupathi V. Big data analytics in healthcare: Promise and potential. Health Inf Sci Syst. 2014; 2: 3. DOI: 10.1186/2047-2501-2-3
16. Yunita A, Santoso HB, Hasibuan ZA. 'Everything is data': Towards one big data ecosystem using multiple sources of data on higher education in Indonesia. J Big Data. 2022; 9: 91. DOI: 10.1186/s40537-022-00639-7
17. Zhou X, Lee EWJ, Wang X, et al. Infectious diseases prevention and control using an integrated health big data system in China. BMC Infect Dis. 2022; 22: 344. DOI: 10.1186/s12879-022-07316-3
18. Bhardwaj N, Wodajo B, Spano A, et al. The Impact of Big Data on Chronic Disease Management. Health Care Manag. 2018; 37 (1): 90–98. DOI: 10.1097/HCM.0000000000000194
19. Bhargava S, Jain M, Jain M. In tuberculosis, "One size does not fit all." Lung India. 2019; 36 (1): 1–2. DOI: 10.4103/lungindia.lungindia_489_18