

RESEARCH

Determinants of TB data quality in completeness, accuracy, and timeliness for TB programme decision-making in Tshwane District : A protocol for a cross-sectional analytical study

L Motimele,^{1,2} MB BCh, Diploma in Public Health; **T Maimela**,^{3,4} MB BCh, MMed, PHM, FCPHM (SA)(CMSA);
D Basu,^{1,2} PhD, MMed, FCPHM (SA)(CMSA), MPH, MSc (Human Genetics), MBBS

¹ Department of Public Health Medicine, University of Pretoria, South Africa

² Steve Biko Academic Hospital, Pretoria, South Africa

³ Charlotte Maxeke Johannesburg Academic Hospital, Johannesburg, South Africa

⁴ Wits School of Public Health, Department of Community Health, Johannesburg, South Africa

Corresponding author: L Motimele (lerato.motimele@up.ac.za)

Background. The quality of health information is crucial for TB programme monitoring and decision-making. South Africa faces challenges in the quality of reported TB data as evidenced in discrepancies reported between different sources. This poses a need to systematically assess the quality of TB data from districts to understand factors that contribute to TB data quality for use in decision-making.

Objective. To determine factors influencing TB data quality for decision-making in the TB programme at a district level in South Africa.

Methods. The study will follow a cross-sectional mixed methods design using an explanatory sequential method to achieve triangulation between the two research methods. The study setting is the Tshwane Health District (THD). Data will be collected from multiple TB data sources in selected clinics and multilevel TB programme personnel. Data collection will be conducted over six months from July 2024 to December 2024. A process flow framework will be used to document the current TB data recording systems in THD. Validated data quality assessment tools (DAQs) will be used for data auditing purposes. Furthermore, a self-administered questionnaire, and one-to-one semi-structured interviews will be used for qualitative data collection. Triangulation will be achieved by integrating both research method findings. The study will be initiated after obtaining approvals from relevant authorities.

Conclusion. This research aims to gain insight and understanding of behavioural, systemic, organisational and technical factors that influence health data quality using TB programme data as a lens. This will enable quantifying of data quality elements and contribute to the body of research with regards to data quality in health.

Keywords. Data quality, Tshwane Health District, Tuberculosis, programme decision-making.

S Afr J Public Health 2025;8(1):e2698. <https://doi.org/10.7196/SHS.2025.v8i1.2698>

An estimated two billion people - about one-quarter of the world's population - are infected with *mycobacterium tuberculosis* (TB).^[1] Mortality and morbidity from TB remain alarmingly high, with ~10 million people developing TB disease and 1.6 million people dying from the infection each year.^[2] TB is still regarded as the world's leading cause of death from an infectious agent.^[3,4] The goal of ending TB by 2035, set by the World Health Organization (WHO) has been adopted by the global community, however, developing countries continue to battle with rising TB cases, largely due to poor social determinants of health and systemic healthcare challenges.^[5] Current global TB targets are described in the United Nations Sustainable Development Goals (SDGs) under

target 3.3 as a 90% reduction in the number of TB deaths and an 80% reduction in the TB incidence rate by 2030 compared with levels in 2015.

South Africa (SA) has made commendable progress in reducing the incidence and prevalence of TB within its national population, despite multiple socio-economic challenges and the impact of the COVID-19 pandemic. SA was one of six countries to achieve the End TB Strategy milestones for 2020, including a 20% reduction in new TB cases (incidence rate) between 2015 and 2020. However, SA remains burdened by TB-related morbidity and mortality, with some performance indicators - such as reductions in TB mortality - still falling short of global targets.^[6] Despite the decline in new cases

of TB, SA continues to have one of the highest TB incidence rates globally, at 615 cases per 100 000 population in 2019. This burden is further exacerbated by a high HIV co-infection rate among TB cases, estimated at 28%.^[7]

TB performance is influenced by various healthcare system factors, including the quality of data collected and reported at all levels of the SA healthcare system. Data quality directly determines the accuracy of TB performance estimates at national and provincial levels, which in turn depend on data collected at facility and district levels. TB tracer indicator data can be tracked from its source to provincial and national levels, forming the basis of TB performance monitoring and evaluation.^[8] Effective strategic planning and goal setting for national TB policies require a continuous feedback loop of reliable information. Identifying gaps in TB performance depends on accurate data, which is only possible with high quality data collection and reporting.^[9] The WHO has set standards for measuring TB data quality through multiple *Data Quality Assessment* tools. These tools evaluate key dimensions of data quality, including completeness, reliability, accuracy, timeliness, precision, and integrity.^[10]

Various health system factors have been identified as determinants of data quality in routine health information systems.^[11] Data quality has been linked to multiple health system components, including human resources, training and development of healthcare personnel, health financing, infrastructure, information systems and technology, as well as leadership and governance structures.^[11] Although a substantial body of literature exists on the determinants of TB data quality, a deeper understanding is needed to explore the underlying explanations for the relationship between health system factors and TB data quality.

Given the inconsistencies and discrepancies between TB data reporting sources – which pose a continuous threat to data quality and may result in a misleading representation of TB performance in health districts – this study aims to assess TB data quality in THD. It seeks to address challenges related to data verification and discrepancies between source documents and reporting systems used for decision-making.

Objectives

This study aims to identify factors influencing TB data quality for decision-making within the TB programme in THD. To achieve this, the study will focus on the following objectives: (i) assessing and describing current sources and level of TB data quality; (ii) identifying factors contributing to TB data quality; and (iii) exploring the perspectives of TB program personnel on factors affecting TB data quality.

Methods

This study will employ a cross-sectional analytical design incorporating both quantitative and qualitative components. Data collection will take place over six months (July–December 2024). The study will be conducted in the Tshwane Health District across its seven sub-districts. Different sampling methods will be used for each of the three objectives.

For Objective 1, a stratified random sampling will be used to select clinics from each of the seven subdistricts. Monthly reports of TB data from these clinics will be audited against District Health Information System (DHIS) reports. Based on the Tshwane District Health Plan and TB data quality assessment conducted in other SA Health Districts,^[1] 30% of the 73 fixed PHC clinics in THD will be selected. This equates to 21 clinics, or three clinics per subdistrict.

The variables to be collected will be based on the facility data quality audit, including data transmission, data processing and analysis obtained from monthly facility TB reports compared with DHIS reports, the percentage of completeness of data elements between monthly reports and DHIS, timeliness of report submissions, and agreement and accuracy between monthly reports and DHIS. Audit findings will be reported as a percentage of completeness, timeliness and accuracy according to WHO data quality acceptance standards ($\geq 75\%$ Completeness, $\geq 80\%$ Timeliness, $\geq 90\%$ Accuracy (less than 10% discrepancy)

For Objective 2, all TB programme personnel employed at the selected health facilities, along with TB programme coordinators and managers at subdistrict and district-levels will be invited to participate in the study. Purposive sampling will be used to select the TB programme personnel. To ensure a good representation of the study, we have determined a minimum sample size of 181 participants (namely 173 PHC clinic TB programme staff members, seven subdistrict TB programme coordinators and one district level TB programme manager) using G-Power statistical software, with an effect size of 0.3, 80% power, at the 0.05 (two-tailed) significance level.

Variables to be collected will be based on TB Programme personnel survey including socio-demographic profiles of staff, behavioural factors, knowledge and technical factors and organisational factors, and their association with data quality score. A multiple logistic regression will be performed to assess factors associated with TB data quality among the clinics grouped as 'good' or 'poor'. The variables with a significance level of 0.2 will be selected for this analysis.

For Objective 3, participants for one-to-one interviews will be selected from each level of the THD, and the sample size will be determined when saturation is reached, when no new themes emerge from the interviews. Literature estimates data saturation between 10 to 17 interviews. A process flow framework will be used to document the current TB data recording systems in THD. Data will be collected using semi-structured and probing questions. Bracketing will be performed before interviews to increase the researcher's self-awareness and minimise bias throughout the data collection process.

A deductive thematic analysis will be used following the 'framework method', to analyse qualitative data.^[12]

Summary

The study will employ various data collection methods, requiring thoroughness in each aspect to meet the set objectives. The mixed methods approach will allow for quantifying TB data quality and provide an assessment at the district level, which can be applied to other health districts and healthcare system settings. Gaining

a comprehensive understanding of the thoughts and insights of TB programme personnel will also provide valuable insights into the factors influencing TB data quality and healthcare data as a whole.

Conclusion

This research aims to gain insight and understanding of the behavioural, systemic, organizational and technical factors that influence health data quality, using TB data as a focal point. It will quantify data quality elements and contribute to the body of research with regards to data quality in healthcare.

Ethics approval and consent to participate. The study has been approved by the University of Pretoria MMed Committee, the University of Pretoria Ethics committee, the Tshwane District Research committee and Tshwane Local Municipality.

Conflict of interest. None.

1. Murphy JP, Girdwood SJ, Wilkinson L, et al. Assessment of facility-based tuberculosis data quality in an integrated HIV/TB database in three South African districts. *PLOS Glob Public Health* 2022;2(9):e0000312. <https://doi.org/10.1371/journal.pgph.0000312>.
2. Alsayed SSR, Gunosewoyo H. Tuberculosis: Pathogenesis, current treatment regimens and new drug targets. *Int J Mol Sci* 2023;24(6):3481. <https://doi.org/10.3390/ijms24065202>

3. MacNeil A, Glaziou P, Sismanidis C, Date A, Maloney S, Floyd K. Global epidemiology of tuberculosis and progress toward meeting global targets – worldwide, 2018. *MMWR Morb Mortal Wkly Rep* 2020;69(11):281-285. <https://doi.org/10.15585/mmwr.mm6911a2>.
4. Brunero S, Dunn S, Lamont S. Development and effectiveness of tabletop exercises in preparing health practitioners in violence prevention management: A sequential explanatory mixed methods study. *Nurse Educ Today* 2021;103:104976. <https://doi.org/10.1016/j.nedt.2021.104976>.
5. Trajman A. The social drivers of tuberculosis, reconfirmed. *Lancet Infect Dis*. 2024;24(1):5-6. [https://doi.org/10.1016/S1473-3099\(23\)00390-0](https://doi.org/10.1016/S1473-3099(23)00390-0).
6. World Health Organization. Global tuberculosis report 2020. Geneva: WHO; 2020.
7. Conan N, Ramsay A, Du Cros P. Prevalence of TB and health-seeking behaviour. *Int J Tuberc Lung Dis* 2022;26(5):463-465. <https://doi.org/10.5588/ijtld.22.0001>
8. World Health Organization. Data quality review: module 1: framework and metrics. Geneva: WHO; 2017.
9. World Health Organization. Definitions and reporting framework for tuberculosis – 2013 revision: updated December 2014 and January 2020. Geneva: WHO; 2013.
10. Brown W. Data Quality Assurance Tool for Program-Level Indicators. Measure Evaluation, Chapel Hill, NC. 2007. Report MS-07-19.
11. Chekol A, Kassie DG, Fentahun N, Dagnaw A, Abera M. Data quality and associated factors of routine health information system among health centers of West Gojjam Zone, northwest Ethiopia, 2021. *Front Health Serv* 2023;3:1059611. <https://doi.org/10.1186/s12911-022-01898-3>
12. Gale NK, Heath G, Cameron E, Rashid S, Redwood S. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. *BMC Med Res Methodol* 2013;13(1):117. <https://doi.org/10.1186/1471-2288-13-117>

Submitted 8 October 2024. Accepted 20 March 2025.