Assessing the Surveillance of Drug-Resistant Tuberculosis Linkage-to-Care in Free State from 2016-2022: A Surveillance System Evaluation of EDRWeb

Evaluation performed in accordance with the CDC guidelines, and completed as part of the requirments for the South African Field EPidemiology Training Programme.

Brian Brummer1,2, Boitumelo Fanampe4, Leole Setlhare4, Motshabi Modise1,7, Ramasedi Mokoena5, Joy Ebonwu6, and Hetani Mdose1

2023-04-24

Table of Contents

[1 Executive Summary 3](#_Toc133246215)

[2 Background 4](#_Toc133246216)

[3 Research Question 6](#_Toc133246217)

[4 Aims 6](#_Toc133246218)

[5 Objectives 6](#_Toc133246219)

[6 Methodology 7](#_Toc133246220)

[7 Outline 7](#_Toc133246221)

[7.1 De-duplication and Data-Linkage Method 7](#_Toc133246222)

[7.1.1 Exact String Matching 7](#_Toc133246223)

[7.1.2 Fuzzy Matching 7](#_Toc133246224)

[7.1.3 Deduplcaition and linking process 7](#_Toc133246225)

[7.1.4 Deduplication and Linkage by numbers 8](#_Toc133246226)

[7.2 Design 8](#_Toc133246227)

[7.3 Setting and population 8](#_Toc133246228)

[8 Results 9](#_Toc133246229)

[8.1 Epidemiology of DRTB in Free State 9](#_Toc133246230)

[8.2 Data Quality 9](#_Toc133246231)

[8.3 Sensitivity and Positive Predictive Value 10](#_Toc133246232)

[8.3.1 Sensitivty 10](#_Toc133246233)

[8.3.2 Positive Preditive Value (PPV) 10](#_Toc133246234)

[8.3.3 Discussion 10](#_Toc133246235)

[8.4 Representativeness 11](#_Toc133246236)

[9 Timeliness 12](#_Toc133246237)

[9.1 Reporting Delay of EDRWeb 12](#_Toc133246238)

[9.2 Time to Drug Initiation 12](#_Toc133246239)

[9.3 Questionnaire Results Narrative 13](#_Toc133246240)

[10 Modelling 13](#_Toc133246241)

[10.1 Effect of Surveillance Timeliness on Mortality 13](#_Toc133246242)

[10.2 Effect of initiation time on Mortality 13](#_Toc133246243)

[11 Discussion 13](#_Toc133246244)

[11.1 Limitations 13](#_Toc133246245)

[11.2 Potential Benefits 13](#_Toc133246246)

[12 Cnclusion 14](#_Toc133246247)

[13 Tables and Figures 14](#_Toc133246248)

[14 Appendices 19](#_Toc133246249)

[References 19](#_Toc133246250)

1 South African Field Epidemiology Training Programme, NICD, Johannesburg, South Africa  
2 School of Public Health, University of the Witwatersrand, Johannesburg, South Africa  
3 Free State Department of Health, Bophelo House, Bloemfontein, South Africa  
4 TB Programme, Free StateDepartment of Health, Bophelo House, Bloemfontein, South Africa  
5 Community Health, Faculty of Health Sciences, University of Free State, Bloemfontein, South Africa  
6 Africe Centres for Disease Control and Prevention, African Union, Addis Ababa, Ethiopia  
7 Division of Public Health, Surveillance and Response, National Institute for Communicable Diseases,

# 1 Executive Summary

**Background:** Tuberculosis (TB) is a critical public health issue globally, and the World Health Organization (WHO) aims to treat more than 90% of eligible TB patients using appopriate regimens by 2025 as part of its “end TB strategy” and fulfill the 2030 Sustainable Development Goals (SDGs) for tuberculosis (TB) established the United Nations. Public health surveillance programs supply a significant amount of data for action, and the evaluation of these systems is vital in ensuring they fulfill their desired outcomes and contribute to meeting their objectives. Recent efforts in the Free State province of South Africa to improve the delivery of TB and HIV services and improve linkage-to-care merit the evaluation of the TB surveillance with a focus on linkage-to-care.

**Objectives:** Conducting a surveillance evaluation aims to assess the effectiveness, efficiency, and quality of a surveillance system by reporting and discussing the systems attributes. The evaluation will provide recommendations for improving DR-TB surveillance when used to monitor linkage-to-care in Free State province, South Africa.

**Methods:** The study will use the attributes of a surveillance System defined in the CDC “Morbidity and Mortality Weekly Report” Surveillance Evaluation guide. Qualitative and quantitative methodologies will be appropriately utilised based on each objective and attribute. Broadly, qualitative methods will involve questionnaires collected via RedCap from personnel involved in TB surveillance, while quantitative methods will gathered data from EDRWeb and TB testing data and compared. Data will be analyzed using Rstudio. The author’s laptop will be protected by an alphanumeric password.

**Ethical considerations:** Ethical clearance will be sought from the Health Sciences Research Ethics Committee of University of Free-State and relevant permissions from the Free State Department of Health will be obtained. There is no increased risk in data linkage compared to programmatic conditions.

**Expected Benefits:** The surveillance system evaluation provides valuable insight for the Free State province to identify weaknesses in their TB program, introduce measures for improvement thereby contributing to better healthcare outcomes and the realisation of “end TB strategy” and SDGs.

# 2 Background

Tuberculosis (TB) is a significant and urgent public health concern worldwide, and achieving the World Health Organization (WHO)‘s goal of treating over 90% of eligible TB patients with appropriate regimens by 2025 is an important part of the “end TB strategy” and meeting the United Nations’ Sustainable Development Goals (SDGs) for tuberculosis.1 South Africa, in particular, faces a heavy burden from TB, with the disease being the leading cause of death among all age groups in 2018.2

Drug-resistant tuberculosis (DR-TB) is a form of tuberculosis caused by Mycobacterium tuberculosis that does not respond to standard first-line anti-TB drugs such as isoniazid and rifampicin. DR-TB is typically treated with second-line drugs, which are less effective, more toxic, and disproportionately expensive compared to Drug Sensitive TB.3 An analysis of the South African TB care cascade published in 2017 found that only 47% of DR-TB incident cases are initiated on treatment (also referred to as “linkage-to-care”) and concluded that significant effort would be required to meet the “end TB strategy” goals of having at least 90% of patients on an appropriate regimen4.

Surveillance systems provide “information for action” and are essential for effective planning, resource management, response and evaluation of health programs5. Surveillance systems should be evaluated to ensure they are meeting their objectives and uses6. In the case of DR-TB linkage-to-care in South Africa, evaluating the surveillance system is important to help reach SDG targets and ultimately save lives.

Delayed linkage-to-care (initiation) is a critical issue in TB control. Delayed linkage-to-care has an acute impact on TB-related deaths and increases TB incidence, as patients remain contagious for longer7. Therefore, early treatment is essential to reduce morbidity and mortality from DR-TB7. South Africa recommends that DR-TB patients should initiate treatment within 5 days of diagnosis. However, a nested trial in South Africa found that the mean time-to-initiation of DR-TB patients was 11 days8.

Without near real-time surveillance, delayed initiation may only be realized once the window for recommended initiation within the 5 days has passed -the opportunity for action is therefore lost. A study in the Cape Town metropole demonstrated that patients with initial loss to follow up were 2.4 times more likely to die compared to those initiated timeously9 This highlights the critical importance of early linkage-to-care to patients’ survival.

While recent efforts are being made to SMS TB results to patients, it has been demonstrated that it does not increase initiation rates to 95%, but does decrease the time to initiation to an acceptable level of 3-5 days.10 Hence, leveraging technology such as electronic systems and mobile health applications can aid in near real-time surveillance, monitoring, and response to improve TB patients’ timely initiation, adherence, and treatment outcomes.

As it stands, linkage-to-care largely relies on passive strategies - patients return to clinics to find out their result and are then initiated at the facility; Active case-finding and linkage-to-care is implemented in high-risk and vulnerable groups by Community Health Workers, however, these strategies are often hindered by resource-limitations11 making the majority of case-finding and initiation active conditional on a patient presenting to a health facility.

TB and HIV related treatment surveillance systems have been through digital permutations since the early 2000s. In their current forms, Drug Sensitive TB and HIV are recorded in TIER.net (previously ETR) and Drug Resistant TB in EDRWeb. Evaluations of systems related to TB have been published in the following areas: ETR in Western Cape found digital data to be less complete than paper-based records12 TIER.net in Northern parts of SA found good completeness except in treatment outcomes13, ETR in Mpumalanga found over reporting of successful treatment outcomes14 and in Gauteng and KZN EDRWeb was found to under-report treatment outcomes and HIV status15. Furthermore, a study in a paediatric setting in Cape Town found that only 64% of clinically diagnosed DR-TB paediatric cases appeared on EDRWeb16, this finding was reproduced with a nationally representative facility sample recorded on ETR which demonstrated that one-third of incident TB cases were not captured on ETR17.

Another system used for surveillance worth mentioning is the District Health Information System (DHIS) which was established with the vision of providing “comprehensive, timeous, reliable and good quality evidence”.18  Despite this laudable aim, DHIS data has been reported to be highly variable and fragmented19 and lack highly skilled personnel to analyse and manage the system.20 An optimal number of a skilled professional such as a field epidemiologist is 1 per 200 000 population21. The limited capacity for analysis due to the low-skill capacity, variable data completeness, fragmentation, low interoperability and crucially the system’s focus on aggregated facility-level data instead of individual patient data is a significant constraint on producing robust evidence.

These limitations render the system somewhat outdated in this current technological era of heightened interoperability among patient-level platforms. For instance the OpenSafely electronic health records platform in the United Kingdom has been used in twenty COVID-19 related studies as of July 202222 demonstrating that it offers data used to generate robust evidence compared to DHIS.

While we acknowledge the timely study by Benade and colleagues23 who used DHIS data to measure TB initiation rates during the first two waves of COVID-19 in South Africa, the limitations of the paper speak for themselves: months of missing data from facilities without explanation rendered only 29% of the original DHIS dataset fit for analysis, significantly limiting the generalisability and inferential capabilities of the study.

The Free State is considered a province with a high burden of HIV and TB24,25,26 and it has been demonstrated that successful and unsuccessful treatment outcomes25 and factors associated with defaulting26 are affected by HIV coinfection highlighting the need for integrating currently fragmented TB and HIV care and surveillance structures.27 Governance structures have rightly been improved,28 yet their role in surveillance systems seems to be lacking especially where integration is concerned.

DR-TB surveillance in Free-State draws on a combination of sources. TB Testing data is sent as a linelist to the Free State TB programme from the centre for TB and positive cases can be exported from the Notifiable Medical Conditions database; DR-TB treatment is recorded on EDRWeb which is an online treatment register. Data from EDRWeb can be exported by identified stakeholders within the TB programme. Aggregated headcount data on key perfomrance indicators is also reported on District Health Information System (DHIS). This information gathered by TB coordinators and used for the surveillance of DR-TB in Free State which is reported to the TB programme coordinator of the Free-State every two weeks as a powerpoint over teleconference.

Harnessing the power of programmatic data can help achieve a better understanding of TB epidemiology and lead to evidence-based practices and better health outcomes.29 To ensure that programmatic data is useful, it is important to ensure surveillance systems capture, store, and process data accurately and efficiently through an evaluation. Therefore, conducting a surveillance evaluation of DR-TB surveillance in Free State is crucial for improving the efficiency and effectiveness of the TB program. Evaluating the attributes, strengths, and limitations of the current system provides an opportunity for meaningful data analysis and refinement of the system, contributing to improvements in DR-TB surveillance and overall TB control in Free State with evidence-based practices.

# 3 Research Question

How can the DR-TB surveillance in the Free State be improved to enhance linkage-to-care for patients with DR-TB and improve patient outcomes?

# 4 Aims

The aim of this study is to describe the attributes of DR-TB surveillance and identify attributes that may be affecting DR-TB patient outcomes and recommend specific improvements in DR-TB surveillance in Free State. Additionally, the study aims to provide digital surveillance tools to enhance surveillance efforts.

# 5 Objectives

1. To characterize the epidemiology of drug-resistant TB (DR-TB) in the Free State province of South Africa, including trends over time, demographic and clinical characteristics of patients, and treatment outcomes.
2. To compare three common data linkage methods for integrating information from TB line lists and EDRWeb surveillance data in the Free State province.
3. To describe the attributes of DR-TB surveillance as defined by the CDC guidelines. These can be further conceptualised in Table 1 under methodology.
4. To determine whether the time to initiation of DR-TB treatment has an effect on mortality.
5. To determine whether the timeliness of EDRWeb data has an effect on mortality.

# 6 Methodology

# 7 Outline

Due to the nature of evaluations, different methods are employed depending on the objective. Broadly speaking there are qualitative and quantitative methodologies. Methodology, per objective are summarised and tabulated in Table 13.1. Quantitative methods largely involved a de-duplicaton and data linkage process. Extensive technical methodology regarding de-duplication and data-linkage are provided as a sub-heading below.

## 7.1 De-duplication and Data-Linkage Method

Data is standardised. Standardisation of data includes eliminating leading or trailing spaces, ensuring standardised cases, and separating names and dates into indivdual components. Deduplcaition and Linkage was implemented with *Tidyverse* verbs and the *RecordLinkage* package in RStudio IDE. Once a dataset is de-duplicated or linked, we refer to the de-duplicated obsrvations or linked observations as entities.

### 7.1.1 Exact String Matching

We performed exact string matching to identify duplicates in each dataset using the firstname and surname or an available unique identifier. The number of exact string matches is described.

### 7.1.2 Fuzzy Matching

After exact string matching we used fuzzy matching to further identify duplicates and form entities. We identified duplicates using average Jarowinkler scores on available identification data To maximise the number of links we used a relatively low cut off to consider a link between observations with an average Jaro-Winkler score of 0.75.

### 7.1.3 Deduplcaition and linking process

We initially started with the data “as is” (this was used to report data quality and representativeness), de-duplicated it with exact string matching then with fuzzy matching using Jaro-winkler scoring and then linked the two datasets using weighted Jaro-Winkler Scores.

When de-duplicating two observations where a person had two different test dates, we used a “first contact” scenario and kept the earliest testing date reported. This mean that both patients who never had a re-test or loss-to-follow-up and those who did, would both be classified as “first contact” scenarios.

### 7.1.4 Deduplication and Linkage by numbers

For laboratory Data between Inf and -Inf , there were 300 tests. after deduplicating on exact string matches of name and surname, there were 288 “unique” observations. After considering matched observations using variables and an average Jaro-Winkler score of >0.77, we were left with 245 test entities.

To account for the expected delay from test date to date of first regstration on EDRWeb we used the TB test date reported on EDRWeb to select the minimum and maximum dates for the EDRWeb dataset. There were initially 192 registrations which we deduplcaited using the EDRWeb system assigned ‘personid’ variable and were left with 173 observations. We performed fuzzy matching however no definite or probable links were estabished using a an average Jarowinkler score of >0.78. as being considered a match.

Prior to linkage, we further standardised common variables so that they were uniform between the two datasets, such as location, district, sub-district variables and removed missing variables so that no JaroWinkler score would be performed on missing variables and skewing the Jaro-Winkler scores.

Using a Weighted cumulative Jaro-Winkler score for the variables(weighting) . Which resulted in 152 linked observations.

## 7.2 Design

We used open-ended and close-ended questionnaires directed at personell involved in TB surveillance within the Free State TB programme such as programme co-ordinators, managers and epidemiologists. Results are presented as a narrative. Quantitative methods used cross-sectional type designs from data collected on EDRWeb and from DR-TB testing data. Details on the approach for each objective will be described in more detail.

## 7.3 Setting and population

For qualitative methods, we will aim to gather insights from key stakeholders of the TB programme in Free State. These stakeholders will include programme directors and TB coordinators within the Free Statedepartment of Health. To ensure meaningful representation, we aim to obtain a minimum of four responses from stakeholders operating at the Provincial Level, as well as at least two responses from stakeholders situated in each of the five districts of the Free State. This approach will enable us to collect a minimum of 14 responses.

For quantitative methods, we will utilise routinely collected data of DR-TB patients from health facilities that utilize NHLS laboratories and record patient treatment on EDRWeb. As such, the study will include all patients who have accessed diagnosis and/or treatment from the Free Statedepartment of health.

# 8 Results

## 8.1 Epidemiology of DRTB in Free State

## 8.2 Data Quality

The missingness for common sociodemographic variables, and variables to be used in matching was assessed as the total number of observations in each dataset and the number of non-missing observations. These were presented as proportions for each dataset.

Table 8.1: Table showing the missingness of common Sociodemographic variables in betwwen Laboratory Data and EDRWeb

| **Characteristic** | **EDRWeb**, N = 1921 | **TB Laboratory**, N = 3001 |
| --- | --- | --- |
| **firstname** |  |  |
| Present | 192 (100%) | 300 (100%) |
| Missing | 0 (0%) | 0 (0%) |
| **secondname** |  |  |
| Present | 76 (40%) | 32 (11%) |
| Missing | 116 (60%) | 268 (89%) |
| **surname** |  |  |
| Present | 192 (100%) | 300 (100%) |
| Missing | 0 (0%) | 0 (0%) |
| **dateofbirth** |  |  |
| Present | 192 (100%) | 300 (100%) |
| Missing | 0 (0%) | 0 (0%) |
| **gender** |  |  |
| Present | 192 (100%) | 297 (100%) |
| Missing | 0 (0%) | 0 (0%) |
| Unknown | 0 | 3 |
| **district** |  |  |
| Present | 192 (100%) | 300 (100%) |
| Missing | 0 (0%) | 0 (0%) |
| **subdistrict** |  |  |
| Present | 187 (97%) | 276 (92%) |
| Missing | 5 (2.6%) | 24 (8.0%) |
| **location** |  |  |
| Present | 192 (100%) | 225 (75%) |
| Missing | 0 (0%) | 75 (25%) |
| **facility** |  |  |
| Present | 97 (51%) | 300 (100%) |
| Missing | 95 (49%) | 0 (0%) |
| 1n (%) | | |

## 8.3 Sensitivity and Positive Predictive Value

### 8.3.1 Sensitivty

Sensitivity is the probability of a positive test in a truly diseased person. In terms of surveillance evaluation, the sensitive simply represents the number of cases from EDRWeb we were able to link to a laboratory test. A synonym for the Sensitivity, in this case, is the initiation coverage. We calculated the Sensitivity as the number of entities appearing on EDRWeb that were linked to entities on the TB laboratory list out of the total entities on the TB laboratory list over the same period.

### 8.3.2 Positive Preditive Value (PPV)

PPV is the probability that if a patient tests positive, that they are a true positive. In the case of the evaluation, the PPV represents the proportion of patients who were on treatment who were tested within the same time period. We calclated the PPV as the number of entities who were linked on EDRWeb over the total number of entities on EDRWeb over the same time period.

### 8.3.3 Discussion

Both of these measures are influeced by robust Recordlinkage techniues and the timeliness of the system, a delayed evaluation system will decrease the sensitivty and PPV in this instance. We would have expected the PPV would be close to 100% however due to the timeliness this is likeley to have caused the drop. Since sensitivty is, in this case, the initation coverage, we expected it to be lower. It ounfortunately confirms the poor initiation coverage of DRTB.

Table 8.2: Table of the Sensitivity and PPV of EDRWeb and The TB laboratroy testing list using deudplicated and Linked datasets.

| System | On EDR | Not on EDR | Total |
| --- | --- | --- | --- |
| On Lab | 152 | 93 | 245 |
| Not on Lab | 25 | 0 | 25 |
| Total | 177 | 93 | 270 |

## 8.4 Representativeness

Common Sociodemographic variables were compared, There were view variables that were common between datasets. We foudn that there was no statitsically significant difference between the age and gener reported by the two systems however the geographical distribution was was statistically different between the two sources at the 5% level.

The time from test results according to the laboratory linelist and the time the patient is first registered on EDRWeb are presented in Table 9.1. The median difference is zero days with an IQR or -1 indicating that the two systems largely represent the same date of testing.

Table 8.3: Comparing the Sociodemographic Factors of DRTB patients from Non-Deduplciated EDRWeb and TB laboratory Datasets in Free State

| **Characteristic** | **EDRWeb**, N = 1921 | **TB Laboratory**, N = 3001 | **p-value**2 |
| --- | --- | --- | --- |
| **age** | 40 (30, 51) | 39 (31, 51) | 0.9 |
| **gender** |  |  | 0.3 |
| Female | 61 (32%) | 104 (35%) |  |
| Male | 131 (68%) | 193 (64%) |  |
| Unknown | 0 (0%) | 3 (1.0%) |  |
| **district** |  |  | 0.028 |
| Fezile Dabi | 30 (16%) | 55 (18%) |  |
| Lejweleputswa | 34 (18%) | 73 (24%) |  |
| Mangaung Metro | 98 (51%) | 109 (36%) |  |
| Thabo Mofutsanyana | 18 (9.4%) | 41 (14%) |  |
| Xhariep | 12 (6.2%) | 22 (7.3%) |  |
| 1Median (IQR); n (%) | | | |
| 2Wilcoxon rank sum test; Fisher's exact test; Pearson's Chi-squared test | | | |

# 9 Timeliness

## 9.1 Reporting Delay of EDRWeb

We calcualted the timeliness of EDRWeb as the difference between the test date on the laborary testing dataset and the date of first registration on EDRWeb. THe Median (IQR) difference was 17 (10, 26). DRTB patients in SA should be initiated within 5 days. We demonstrate that EDRWeb is too slow to provide information within this time period. It is therfore evident that TB coordinators do not have reliable information on who is initiated on treatment. This lag reduces the opportunity for action.

If EDRWeb’s timeliness were drastically improved and was reported in near-Real-time, it would allow TB coordinators to react to information on patients who have not started TB treatment within 5 days and prioritise their initiation.

## 9.2 Time to Drug Initiation

The time from from test result to treatment initiation is calculated using the result date on the TB laboratory dataset until the reported treatment start date on EDRWeb. This would be the “real” time to initiation and is presented in Table 9.1.

Table 9.1: Differences in Days between Key dates between Laboratory Test Data and EDRWeb to look at aspects of Timeliness with regards to Surveillance and Time to Initiation

| **Characteristic** | **Overall**, N = 1431 | **Alive**, N = 1191 | **Died**, N = 241 | **p-value**2 |
| --- | --- | --- | --- | --- |
| Laboratory Date of Test to EDRWeb Registration | 17 (10, 26) | 17 (11, 26) | 18 (8, 28) | >0.9 |
| EDRWeb reported Test Date to EDRWeb Reported Date of RX Initiation (Time to Initiation from Single Dataset) | 9 (6, 14) | 9 (6, 14) | 8 (7, 13) | 0.4 |
| Laboratory Reported Test Date to EDRWeb Reported Date of RX Initiation (Time to Initation from Linked Dataset | 8 (5, 13) | 8 (6, 12) | 6 (4, 13) | 0.15 |
| Difference between Laboratory reported Test date | 0 (-1, 0) | 0 (-1, 0) | -1 (-1, 0) | 0.2 |
| 1Median (IQR) | | | | |
| 2Wilcoxon rank sum test | | | | |

## 9.3 Questionnaire Results Narrative

# 10 Modelling

## 10.1 Effect of Surveillance Timeliness on Mortality

Table 10.1: Differences in Days between Key dates between Laboratory Test Data and EDRWeb to look at aspects of Timeliness with regards to Surveillance and Time to Initiation

| **Characteristic** | **Beta** | **95% CI**1 | **p-value** |
| --- | --- | --- | --- |
| lab2edr | 0.00 | 0.00, 0.00 | 0.7 |
| edr2initiation | 0.00 | -0.01, 0.00 | 0.3 |
| 1CI = Confidence Interval | | | |

## 10.2 Effect of initiation time on Mortality

Table 10.2: Differences in Days between Key dates between Laboratory Test Data and EDRWeb to look at aspects of Timeliness with regards to Surveillance and Time to Initiation

| **Characteristic** | **Beta** | **95% CI**1 | **p-value** |
| --- | --- | --- | --- |
| lab2edr | 0.00 | 0.00, 0.00 | 0.7 |
| edr2initiation | 0.00 | -0.01, 0.00 | 0.3 |
| 1CI = Confidence Interval | | | |

# 11 Discussion

## 11.1 Limitations

One noteworthy constraint in comparing various surveillance systems is that EDRWeb is incapable of capturing patient outcomes prior to the initiation of treatment. Therefore, patients who pass away before receiving treatment will only be recorded on the TB linelist data, and not on EDRWeb. This constraint is present in both the research study and the surveillance system being evaluated, and may be a tentative early recommendation. Additionally, the private, military, and mining sectors may not be accounted for in the laboratory linelist and EDRWeb, which would require significant ethical permissions, data harmonization, and cleaning efforts - an impractical undertaking for the purposes of this report.

## 11.2 Potential Benefits

This assessment has the potential to provide valuable insights into the effectiveness of the connection between EDRWeb and patient testing. It has the capability to identify high-risk patients who may delay or fail to initiate necessary treatments. Additionally, this evaluation can help determine whether timely reporting of initiation rates has an impact on patient outcomes. The findings of this evaluation could inform strategies to enhance EDRWeb’s performance. The analytical methods presented in this report can be integrated into a dashboard for the TB programme in the Free State province, enabling the creation of easy-to-use and reproducible digital surveillance tools.

# 12 Cnclusion

# 13 Tables and Figures

Table 13.1: Research Objectives for the DR-TB surveillance Surveillance Evaluation Methods

| Objectives | Methods |
| --- | --- |
| To characterize the epidemiology of drug-resistant TB (DR-TB) in the Free State province of South Africa, including trends over time, demographic and clinical characteristics of patients, and treatment outcomes. | Analyse routinely collected DR-TB data from the Free StateDepartment of Health involving visualising incidence per District from 2016 to 2022, reporting the sociodemographic and clinical characteristics of DR-TB patients recorded on EDRWeb and their outcomes using descriptive statistics. Treatment outcomes of interest are cured, treatment completed, treatment failed, died, lost to follow-up, or not evaluated |
| To compare three common data linkage methods for integrating information from TB line lists and EDRWeb surveillance data in the Free State province. | Present the sensitivity of exact string matching, weighted string matching and fuzzy matching utilising the Leventsheim distance and disucss the strengths and weaknesses of each as they pertain to linking DR-TB testing data to EDRWeb data |
| To describe the attributes of DR-TB surveillance as defined by the CDC guidelines. These can be further conceptualised in Table 1 under methodology. | Review the CDC guidelines and collect inforamtion on each attributes as they pertains to the DR-TB surveillance in the Free State context. More detailed infromation on each attribute can be found in the table on Attributes. |
| To determine whether the time to initiation of DR-TB treatment has an effect on mortality. | The association between the time to initiation of DR-TB patinets and treatment outcomes will be analyzed using appropriate statistical methods, such as the chi-squared test and logistic regression analysis. Adjustments will be made for potential confounding variables, such as sociodemographics and clinical factors. |
| To determine whether the timeliness of EDRWeb data has an effect on mortality. | The association between the timeliness of EDRweb and DR-TB treatment outcomes will be analyzed using appropriate statistical methods, such as the chi-squared test and logistic regression analysis. Adjustments will be made for potential confounding variables, such as sociodemographics and clinical factors. Timeliness can be defined as the length of time between the submission of the patient's sample for drug susceptibility testing and the reporting of the results via EDRweb |
| To develop user-friendly and simple dashboards that enable effective surveillance of DR-TB surveillance by leveraging the analyses and measures generated during the evaluation process. | Develop the dashboards using RMarkdown, Shinydashboard, and flexdashboard packages which can be made available via a private internet link to the TB programme.The analyses for timeliness, sensitivity, time to initiation, incidence and patient outcomes will be automatically reproduced when data is updated. |

Table 13.2: Attributes of a Surveilllance System adapted from CDC MMWR guidelines

| Attributes | Definition | Methods |
| --- | --- | --- |
| Usefulness | The extent to which surveillance data provide useful information that can be used for public health decision making | Conduct surveys with stakeholders to assess usefulness of surveillance for instance is the data able to be easily used and disseminated to infrom public health response. |
| Simplicity | The extent to which surveillance data are easy to understand, analyze, interpret and communicate | Conduct surveys with stakeholders to assess usefulness of surveillance data and identify areas for improvement with speicfic focus on skills needed and training provided |
| Flexibility | The degree to which surveillance data systems can be modified or enhanced to meet evolving public health data needs | Conduct surveys with stakeholders to assess whether changes to the Surveillance system are easy to action and what their experience may have been with actioning such a change. |
| Acceptability | The extent to which the surveillance system is acceptable to stakeholders and the population under surveillance | Conduct surveys with stakeholders to assess whether they find the system appropriate for their level of training and scope of practice within the TB programme. |
| Timeliness | The extent to which surveillance data are available in a timely manner and can be used for timely public health action | Timeliness can be defined as the length of time between the submission of the patient's sample for drug susceptibility testing and the patients first beign recorded on EDRWeb |
| Data quality | The extent to which the data collected by the system is accurate, complete and valid | The completeness of observations related to sociodemographic and clinical information will be assessed. Observations will be categorized as either complete or incomplete based on, but not limited to HIV status, age, and sex. |
| Sensitivity | The ability of the system to accurately detect changes in health status, behavior or risk factors over time | Use statistical methods to calculate sensitivity of EDRWeb, more detail can be found under the contingency table |
| Positive Predictive Value | The proportion of reported cases or events that are truly positive | Use statistical methods to calculate the positive predictive value of EDRWeb, more detail can be found under the contingency table |
| Representativeness | The extent to which the surveillance system provides an accurate representation of the population under surveillance | To compare common sociodemographic variables available on both the linelist and EDRWeb and use Chi-Square and t-tests will be used to test for differences of common variables |

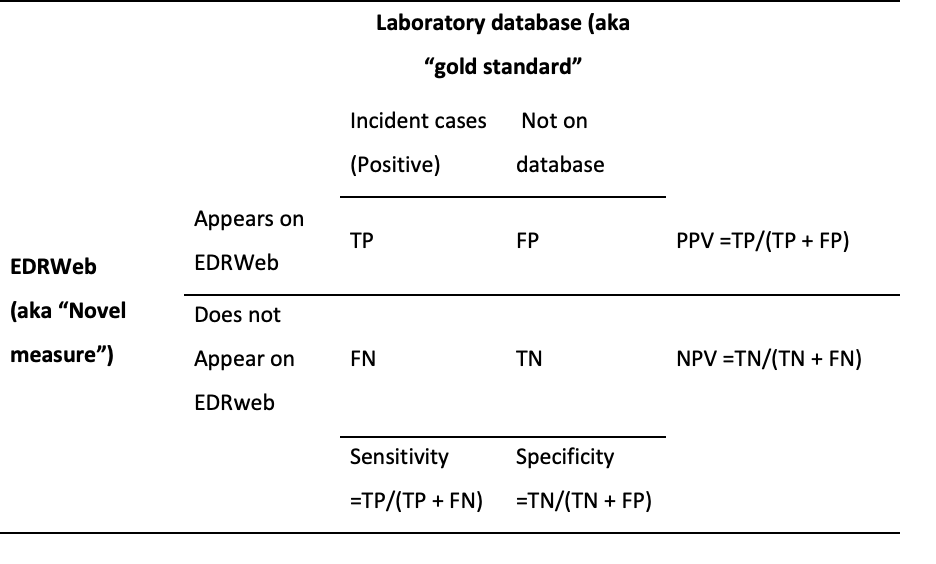


Figure 13.1: Contingency table showing how to calculate Sensitivity and Positive Predictive Value. The True Positive (TP) will be based on the data linkage method chosen in Objective 2

# 14 Appendices

# References

1. (WHO) TWHO. *The End TB strategy*. WHO, 2015.

2. *Mortality and causes of death in South Africa: Findings from death notification 2018*. STATISTICAL RELEASE P0309.3, Department of Statistics South Africa, June 2021.

3. Pooran A, Pieterson E, Davids M, et al. [What is the Cost of Diagnosis and Management of Drug Resistant Tuberculosis in South Africa?](https://doi.org/10.1371/journal.pone.0054587) *Plos One* 2013; 8: e54587.

4. Naidoo P, Theron G, Rangaka MX, et al. [The South African Tuberculosis Care Cascade: Estimated Losses and Methodological Challenges](https://doi.org/10.1093/infdis/jix335). *The Journal of Infectious Diseases* 2017; 216: S702–S713.

5. Alexander Langmuir. [The Surveillance of Communicable Disease of National Importance](https://doi.org/10.1056/NEJM196301242680405). *New Engl J Med* 1963; 268: 182–192.

6. German R, Horan J, Milstein B, et al. Updated guidelines for evaluating public health surveillance systems; recommendations from the Guidelines Working Group. *Centre for Disease COntrol and Prevention(US)*; 50.

7. Organization WH. *Global Tuberculosis Report 2022*. Geneva: World Health Organization, 2022.

8. van de Water BJ, Prvu Bettger J, Silva S, et al. [Time to Drug-Resistant Tuberculosis Treatment in a Prospective South African Cohort](https://doi.org/10.1177/2333794X17744140). *Global Pediatric Health* 2017; 4: 2333794X1774414.

9. Osman M, Meehan S-A, von Delft A, et al. [Early mortality in tuberculosis patients initially lost to follow up following diagnosis in provincial hospitals and primary health care facilities in Western Cape, South Africa](https://doi.org/10.1371/journal.pone.0252084). *Plos One* 2021; 16: e0252084.

10. Mwansa-Kambafwile JRM, Chasela C, Levin J, et al. [Treatment initiation among tuberculosis patients: The role of short message service (SMS) technology and Ward-based outreach teams (WBOTs)](https://doi.org/10.1186/s12889-022-12736-6). *Bmc Public Health* 2022; 22: 318.

11. Ajudua FI, Mash RJ. [Implementing active surveillance for TB views of managers in a resource limited setting, South Africa](https://doi.org/10.1371/journal.pone.0239430). *Plos One* 2020; 15: e0239430.

12. Mlotshwa M, Smit S, Williams S, et al. [Evaluating the electronic tuberculosis register surveillance system in Eden District, Western Cape, South Africa, 2015](https://doi.org/10.1080/16549716.2017.1360560). *Global Health Action* 2017; 10: 1360560.

13. Murphy JP, Kgowedi S, Coetzee L, et al. [Assessment of facility-based tuberculosis data quality in an integrated HIV/TB database in three South African districts](https://doi.org/10.1371/journal.pgph.0000312). *PLOS Global Public Health* 2022; 2: e0000312.

14. Dreyer AW, Mbambo D, Machaba M, et al. [Tuberculosis cure rates and the ETR.Net: Investigating the quality of reporting treatment outcomes from primary healthcare facilities in Mpumalanga province, South Africa](https://doi.org/10.1186/s12913-017-2128-0). *Bmc Health Serv Res* 2017; 17: 190.

15. Jamieson L, Evans D, Berhanu R, et al. [Data quality of drug-resistant tuberculosis and antiretroviral therapy electronic registers in South Africa](https://doi.org/10.1186/s12889-019-7965-9). *Bmc Public Health* 2019; 19: 1638.

16. Rose PC, Schaaf HS, du Preez K, et al. [Completeness and accuracy of electronic recording of paediatric drug-resistant tuberculosis in Cape Town, South Africa](https://doi.org/10.5588/pha.13.0041). *Public Health Action* 2013; 3: 214–219.

17. Podewils LJ, Bantubani N, Bristow C, et al. [Completeness and Reliability of the Republic of South Africa National Tuberculosis (TB) Surveillance System](https://doi.org/10.1186/s12889-015-2117-3). *Bmc Public Health* 2015; 15: 765.

18. DISTRICT HEALTH MANAGEMENT INFORMATION SYSTEM (DHMIS) POLICY.

19. Garrib A, Stoops N, McKenzie A, et al. An evaluation of the District Health Information System in rural South Africa. 98.

20. English R, Masilela T, Barron P, et al. Health Information Systems in South Africa.

21. Williams SG, Fontaine RE, Turcios Ruiz RM, et al. [One Field Epidemiologist per 200,000 Population: Lessons Learned from Implementing a Global Public Health Workforce Target](https://doi.org/10.1089/hs.2019.0119). *Health Security* 2020; 18: S-113-S-118.

22. Andrews C, Schultze A, Curtis H, et al. [OpenSAFELY: Representativeness of electronic health record platform OpenSAFELY-TPP data compared to the population of England](https://doi.org/10.12688/wellcomeopenres.18010.1). *Wellcome Open Research* 2022; 7: 191.

23. Benade M, Long L, Meyer-Rath G, et al. [Reduction in initiations of drug-sensitive tuberculosis treatment in South Africa during the COVID-19 pandemic: Analysis of retrospective, facility-level data](https://doi.org/10.1371/journal.pgph.0000559). *PLOS Global Public Health* 2022; 2: e0000559.

24. Heunis C, Wouters E, Kigozi G, et al. Accuracy of Tuberculosis Routine Data and Nurses’ Views of the TB-HIV Information System in the Free State, South Africa. 7.

25. Engelbrecht MC, Kigozi NG, Chikobvu P, et al. Unsuccessful TB treatment outcomes with a focus on HIV co-infected cases: A cross- sectional retrospective record review in a high-burdened province of South Africa. 2017; 11.

26. Kigozi G, Heunis C, Chikobvu P, et al. [Factors influencing treatment default among tuberculosis patients in a high burden province of South Africa](http://dx.doi.org/doi:10.1016/j.ijid.2016.11.407). *Internation Journal of Infectious Diseases* 2017; 54: 95–102.

27. Malakoane B, Heunis JC, Chikobvu P, et al. Public health system challenges in the Free State, South Africa: A situation appraisal to inform health system strengthening. 2020; 15.

28. Malakoane B, Heunis JC, Chikobvu P, et al. Improving public health sector service delivery in the Free State, South Africa: Development of a provincial intervention model. 2022; 17.

29. Taaffe J, Croda J, Moultrie H, et al. [Advancing TB research using digitized programmatic data](https://doi.org/10.5588/ijtld.21.0325). *The International Journal of Tuberculosis and Lung Disease* 2021; 25: 890–895.