Protocol: Analysis of risk of death after Tuberculosis diagnosis in Brazil between 2004 and 2018.

Version History

Version	Date	author	comments
1	20/02/2025	TCS, JMP	First Draft
2	03/03/2025	TCS	Update methodology

20/02/2025

1) Background

Tuberculosis (TB) is closely linked to poverty and other poor socioeconomic conditions. These factors not only increase the risk of TB infection but may also increase the impact on health, even mortality. Numerous studies have confirmed the association between TB and an elevated risk of death. However, much of this evidence lacks a detailed exploration of the specific causes of death, potentially ignoring the context in which these deaths occur.[1]

One critical oversight in previous research is the failure to account for deaths due to external causes, such as accidents and violence, which usually reflect poor socioeconomic factors that disproportionately affect disadvantaged populations. Individuals with TB often belong to the most vulnerable groups in society, facing challenges such as inadequate access to healthcare, unstable housing, food insecurity, and systemic inequities. These same factors may contribute to higher mortality rates from causes unrelated to the disease itself.[1,2]

Our object is to quantify the additional risk of death in individuals with TB, accounting for the different effects in multiple causes of death (cardiovascular, metabolic, etc).

2) Data sources

SINAN - National Notifiable Disease Information System

Suspected and/or confirmed cases of a list of infectious diseases must be reported to the Epidemiological Surveillance service on a specific enumerated notification form, which is available in any local health facility. This form can be filled out by any health professional who suspects disease. It is disease-specific. In this study, we will use the Tuberculosis database. We

will extract the type of notification (new case or reinfection), date of diagnosis, clinical classification (pulmonary or extrapulmonary tuberculosis); HIV coinfection, and comorbidities (diabetes/etc). People diagnosed with pulmonary plus extrapulmonary tuberculosis were classified as pulmonary tuberculosis.

SIM - Mortality Information System

This system uses the death certificate, a legal document only a physician can complete, and records information about all deaths in Brazil. We will extract the date and basic cause of death.

CadÚnico -Unified Registry for Social Programs

We will extract all the socioeconomic and demographic variables at the individual and household level for the index patient with tuberculosis and the cohabitants (i.e., age, sex, education, self-identified race or ethnicity, Brazilian region and area of residence [rural or urban], household density, housing materials, water supply, sewage, lighting, and garbage disposal), as well as information on which individuals were living in the same household at the time of registration (i.e., identified through a family code)

3) Cohort

a. Study population

We will conduct two cohorts.

First cohort:

Inclusion:
All individuals who applied to the CadÚnico between Jan 1, 2004 and Dec 31, 2018.

Exclusion:
We will exclude individuals:

- Aged 100 years or older at CadÚnico registration
- Diagnosed with tuberculosis before the CadÚnico registration.
- Missing data: age, race/ethnicity, municipality of residence, household location, household's water supply type, material of the household, and household crowding.
- Data inconsistencies: date of death before diagnosis date, date of death before date of CadÚnico registration
- Homelessness due to the impossibility of assessing variables related to the household

In this cohort, we will conduct exact matching, without replacement, on age in years at disease onset (5-year bin), sex, race or ethnicity, municipality of residence, household location, household's water supply type, material of the household, year of registration in the CadÚnico

(3-year bin) and household crowding, matching by these factors provided demonstrable control of bias in a previous study.[3] Controls matched on a given day who acquired TB on a subsequent date became a case and could be matched to a new control.

Second cohort:

The inclusion and exclusion criteria from the first cohort will be used, but exposed individuals will be those diagnosed with tuberculosis and confirmation of cure. In the criterium of data inconsistencies, individuals with a date of cure equal to the date of death will be excluded, as it probably reflects an error in data entry. The time zero in this cohort will be the date of confirmation of cure.
The restriction study sample of cured TB cases will provide an estimate of residual risk.

Third cohort:

This cohort will be conducted by matching TB contacts (individuals residing in the same household of a TB case) to individuals free of TB (no tuberculosis, no TB contact). TB contacts will be eligible if they are alive at the moment of the first TB case diagnosis in the household, i.e., persons born after the first TB case won't be included. A TB contact who becomes a TB case afterwards will be censored on the day of TB diagnosis (pair censoring)
The time zero in this cohort will be the date of the first TB case in the household.
This cohort aims to estimate the risk related to social vulnerability and not directly related to TB.

b. Outcome events

The chosen outcomes aim to capture both the direct impact of TB on mortality and the broader social factors influencing TB development. Cause 3—natural death, excluding deaths due to TB (A15-A19) and HIV (B20-B24)—was selected as the primary outcome to assess the additional mortality risks associated with TB. Given the well-established link between HIV and TB, excluding deaths directly attributed to TB and HIV allows for a clearer assessment of TB's broader adverse effects.

Cause 4 aims to capture aspects of social vulnerability. Cause 5 was chosen as it represents the leading cause of death worldwide.[4] Additionally, Causes 6, 7, and 8 were included due to their established association with TB.[5,6]

1) All-cause mortality
2) Natural death (all codes excluding external causes of death [Chapter XX] from International Classification of Diseases [ICD]-10)
3) Natural death, excluding the deaths due to TB (A15-A19) and HIV (B20-B24).
4) External Deaths (Chapter XX ICD-10)
    a) Group V01-X59
    b) Group X85-Y09

5) Cardiovascular (Chapter IX)
    a) Ischaemic heart diseases (I20-I25)
    b) Cerebrovascular diseases (I60-I69)
6) Metabolic (Chapter IV);
7) Respiratory System (Chapter X);
8) Cancer (Chapter II)
    a) Malignant neoplasms of respiratory and intrathoracic organs (C30-C39)
    b) Malignant neoplasms of digestive organs (C15-C26)

c. Data linkage

Linkage was performed using Centro de Integração de Dados e Conhecimentos para Saúde Record Linkage (CIDACS-RL), which uses a two-step strategy. The first step is a fully deterministic linkage based on five identifying variables (i.e., name, mother's name, sex, date of birth, and the municipality of residence). The second step is non-deterministic and uses the same five variables to produce a similarity score; matched registries are based on scores of optimal sensitivity and specificity thresholds. Linkage accuracy between tuberculosis registries and the 100MCohort was measured in terms of sensitivity (94·6%) and specificity (93·6%) calculated based on false or true links between the two databases. Linkage accuracy between mortality registries and the 100MCohort was similarly calculated by year: sensitivity ranged between 97·8% and 100·0% and specificity between 96·6% and 99·9%.[7]

d. Statistical analysis

Exposed individuals will be exactly matched without replacement to an unexposed individual on the day of diagnosis of TB. Controls matched on a given day who acquired TB on a subsequent date became a case and could be matched to a new control.

Each matched pair will be followed up from the matching date until the earliest of the following events: diagnosis of TB, death, 1,825 days (5 years) of follow-up, or Dec 31, 2018 (final data collection date). The start of the timescale of the study will be the time since diagnosis or the time since matching for controls.

We estimated the cumulative incidence function of each outcome using the Aalen-Johansen estimator, considering the competing risk of death for other causes. We will estimate period-specific risks, risk differences, and risk ratios (RRs), comparing the exposed group against the unexposed group for each outcome. The period-specific intervals were demarcated on days 30, 90, 180, 365, 730, 1,095, and 1,460. Individuals with the same date of death and diagnosis of TB will be excluded from the analysis.

We will use a non-parametric bootstrapping (resampling only matched pairs when conducted in matched analysis) with 500 iterations to calculate percentile-based 95% CIs for risk differences and RRs.

e. Sensitivity analysis/subgroup analysis

As a sensitivity analysis, we will extend the follow-up until the end of follow-up demarcated by years (365, 730, 1,095, …5110). This will provide estimates for longer follow-ups but with decreased precision.

Subgroup analysis will be conducted by sex, age at diagnosis of TB (<18, 18-59 and ≥60 years) and race (white/mixed-black/Indigenous/Asian), stratifying the models for those variables. We will also conduct subgroup analysis by clinical characteristics: type of TB (pulmonary, extrapulmonary and extrapulmonary+pulmonary); diagnosis of HIV/AIDS; diagnosis of diabetes.

4) **Limitations**

The clinical variables, such as HIV/AIDS and diabetes, are only available for individuals diagnosed with TB. The expected effect is the combined impact of HIV/AIDS + TB or diabetes + TB compared to healthy individuals. However, because it is not possible to assess the status of HIV/AIDS or diabetes in the unexposed group, this can result in a downward bias, as some individuals with HIV/AIDS or diabetes may be included in the unexposed group.

References

1 Ferreira MRL, Bonfim RO, Bossonario PA, *et al.* Social protection as a right of people affected by tuberculosis: a scoping review and conceptual framework. *Infect Dis Poverty* 2023; **12**: 103.

2 Liyew AM, Clements ACA, Akalu TY, Gilmour B, Alene KA. Ecological-level factors associated with tuberculosis incidence and mortality: A systematic review and meta-analysis. *PLOS Glob Public Health* 2024; **4**: e0003425.

3 Cerqueira-Silva T, Pescarini JM, Cardim LL, *et al.* Risk of death following chikungunya virus disease in the 100 Million Brazilian Cohort, 2015–18: a matched cohort study and self-controlled case series. *Lancet Infect Dis* 2024; **24**: 504–13.

4 Naghavi M, Ong KL, Aali A, *et al.* Global burden of 288 causes of death and life expectancy decomposition in 204 countries and territories and 811 subnational locations, 1990–2021: a

systematic analysis for the Global Burden of Disease Study 2021. *The Lancet* 2024; **403**: 2100–32.

5 Nordholm AC, Andersen AB, Wejse C, *et al.* Mortality, risk factors, and causes of death among people with tuberculosis in Denmark, 1990-2018. *Int J Infect Dis* 2023; **130**: 76–82.

6 Choi EH, Coyle WJ. Gastrointestinal Tuberculosis. *Microbiol Spectr* 2016; **4**: 10.1128/microbiolspec.tnmi7-0014–2016.

7 Pinto PFPS, Teixeira CSS, Ichihara MY, *et al.* Incidence and risk factors of tuberculosis among 420 854 household contacts of patients with tuberculosis in the 100 Million Brazilian Cohort (2004–18): a cohort study. *Lancet Infect Dis* 2024; **24**: 46–56.