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

16/01/2025

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 central hypothesis is that while TB does indeed increase the risk of death, the magnitude of this risk is likely overestimated. Part of the observed increase in mortality is not directly attributable to TB. Instead, it reflects the broader impact of poor socioeconomic conditions.

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) Study designs

We will use two study designs to assess the risk associated with Tuberculosis. First, a cohort will be conducted, and the incidence of the outcomes of interest in pre-specified post-tuberculosis periods will be compared with unexposed individuals after adjustment for available confounding factors. Second, a self-controlled case series (SCCS) analysis will be undertaken in exposed individuals with the outcome event to assess the incidence in pre-specified post-tuberculosis risk periods compared to those outside these risk periods.[3] The SCCS analysis implicitly controls for any non-time varying factors not included in the cohort analysis that are associated with both tuberculosis and the outcome event.

4) Cohort analysis

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 or 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 on age in years at disease onset, sex, race or ethnicity, municipality of residence, household location, household's water supply type, material of the household, and household crowding, matching by these factors provided demonstrable control of bias in a previous study.[4]

Second cohort:

This cohort will be restricted to individuals who applied to the CadÚnico between Jan 1, 2004 and Dec 31, 2018 with a diagnosis of tuberculosis and non-TB cases from a household with a TB case (as the unexposed group). In this cohort, we will use the data from a previous publication.[5]

The restriction study sample of TB cases and their household contacts will provide a more homogeneous population. In this cohort, TB cases will be matched to non-TB cases using the same matching strategy of the first cohort. Due to the reduction in the pool of controls, if more than 20% of the TB cases did not find a match, we will conduct analysis using inverse probability treatment weighting (IPTW).

b. Outcome events

1) Natural death (all codes excluding external causes of death [Chapter XX] from International Classification of Diseases [ICD]-10)
2) External Deaths (Chapter XX ICD-10)
3) Natural death, excluding the deaths due to TB (A15-A19) and HIV (B20-B24)
4) Cause-specific: Cardiovascular (Chapter IX); Metabolic (Chapter IV); Respiratory System (Chapter X); Malignant neoplasms of respiratory and intrathoracic organs (C30-C39) [6]

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%.[5]

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 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, 360, 720, 1,080, and 1,440. Individuals with the same date of death and diagnosis of TB will be excluded from the analysis.

For the IPTW analysis, we will estimate the propensity score using logistic regression. The model will be adjusted for age in years at disease onset, sex, race or ethnicity, municipality of residence, household location, household's water supply type, material of the household, and household crowding. The weights derived from the logistic regression will be included in the Aalen-Johansen estimator.

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 15 years, the maximum time possible in the cohort. This will provide estimates for longer follow-up 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) using stratified models for those variables.

5) Self-controlled case series analysis (SCCS)

The SCCS compares the event rates in risk windows after exposure to the event rates of a baseline. This type of study relies on four assumptions: 1) Events arise independently within individuals; 2) Occurrence of an event does not influence the subsequent period of observation; 3) Occurrence of an event does not influence subsequent exposures; 4) Exposures do not influence the ascertainment of events. In the case of death as an event, there is a violation of assumptions 2 and 3. In this study, we will use the date of exposure as the beginning of the observation period and a pre-specified time as the end of observation, which solves the problems related to the violation of assumptions 2 and 3.

The difference in the pattern of the risk of death on external deaths versus natural deaths will be used as indicative of socioeconomic conditions directly affecting the risk of death after TB.

a. Study population

We will include all deaths from individuals with TB diagnosis between Jan 1, 2004, and Dec 31, 2018. We will exclude individuals with missing data on age or sex. The same outcomes of the cohort study will be evaluated using SCCS.

b. Statistical analysis

The self-controlled case series method uses conditional Poisson regression to compare intra-person mortality rates in different study periods relative to the date of disease onset. For each mortality outcome, we compared the mortality rate in the 0–5 years following the diagnosis of TB to the mortality rate in the period defined as the baseline (from 5 years after diagnosis until Dec 31, 2018). The nominal end of observation will be Dec 31, 2018. Therefore, we will include only post-disease person-time during the 14-year study period, which is necessary to meet assumptions of self-controlled case series when the outcome is death. For each mortality outcome, we will evaluate the period-specific risk of death during the following time windows (in days): 0, 1-30, 31-90, 91-180, 181-360, 361 - 720, 721 - 1080, 1081 - 1440; and the overall of 1 - 1440.

c. Subgroup analysis

We will use likelihood ratio tests to investigate heterogeneity between subgroups, sex and age (<18, 18-59, ≥60 years); HIV (yes/no); race/ethnicity (white/mixed-black/Indigenous/Asian) and diabetes (yes/no).

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 Petersen I, Douglas I, Whitaker H. Self controlled case series methods: an alternative to standard epidemiological study designs. *BMJ* 2016; **354**: i4515.

4 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.

5 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.

6 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.