Factors Associated with Catastrophic Costs due to Tuberculosis. COMMIT Tuberculosis Data Analysis.

Vernetta Huang

SPH3101

GitHub Repository

Introduction

This analysis explores the impact of tuberculosis (TB) on catastrophic costs, exploring data from the Community Mobilization Initiatives to End Tuberculosis (COMMIT) cohort. TB presents a significant burden in many low-income countries, such as Cambodia, not only in terms of health but also financially, making it a critical area for intervention. Catastrophic costs can hinder access to necessary care for those with TB, leading to worse health outcomes and pushing households further into poverty.

Given the implications for health equity and severe financial consequences associated with TB, understanding the economic toll of the disease has become a key area of focus in public health research. This study aims to identify the primary drivers for catastrophic costs related to TB and to analyze their relationships, providing a foundation for policy recommendations and support strategies for those affected by TB.

This analysis is based on data from the COMMIT cohort, which includes two rounds of survey data collected at baseline in 2022 and follow-up in 2023. The baseline survey included 837 individuals over the age of 18, all newly diagnosed with TB, across 10 operational districts in Cambodia. Information was collected on sociodemographic characteristics, medical history, TB exposure risk, and other relevant personal and situational factors. The follow-up survey was conducted a year later with 620 participants to assess changes post-treatment, using similar questions to gather comparable data.

Methods

Data preparation included cleaning incomplete records and calculating standardized cost factors. All statistical analyses were conducted in R version 4.3.2[1], utilizing chi-squared and t-tests, as well as logistic regression to assess factors such as sex, known medical conditions, and relapse, in relation to catastrophic costs.

Calculating Costs

To standardize and breakdown cost factors, select data from each survey was grouped into three cost categories: direct, non-medical, and indirect costs. For both baseline and follow-up, direct costs considered expenses for hospital visits, private clinic fees, self-medication, traditional medicine, and other healthcare-related fees. Non-medical costs for both baseline and follow-up included travel, food during health facility visits or hospital stay, and accommodation costs. The baseline survey data did not directly provide indirect costs, and the value was estimated using gamma distribution sampling as employed by Fukushi Morishita[2]. Follow-up indirect costs were calculated with the survey data, considering income lost during treatment, borrowed money, and assets sold to cover expenses.

The total baseline costs were the sum of these categories, while follow-up costs combined both baseline and follow-up expenses. A 20% threshold, as defined by the World Health Organization (WHO)[3], was used to identify individuals experiencing catastrophic costs based on their annual income.

Data Analysis

For data analysis, descriptive statistics were first used to explore TB-related costs and factors potentially associated with catastrophic costs. This initial data exploration involved examining the distribution of direct, non-medical, and indirect costs, as well as their relationships with catastrophic costs and various thresholds. Additionally, factors such as age, sex, employment status, TB type, and known medical conditions were explored to identify general trends related to catastrophic costs.

Following the exploratory analysis, hypothesis testing and logistic regression were done to investigate associations between selected factors and catastrophic costs, first using the baseline dataset and then the follow-up dataset.

For hypothesis testing, chi-squared tests were conducted to determine statistically significant associations between selected factors and catastrophic costs. The variables tested included sex, employment, TB type, previous known medical conditions, and relapse occurance. For continuous variables, t-tests were performed to similarly assess the association between selected factors and catastrophic costs, specifically looking at average age and time required to reach health facilities.

After the hypothesis testing, logistic regression analysis was conducted to refine the findings and identify the most significant predictors of catastrophic costs among those tested.

Results

Cost Composition

Cost breakdown was explored using boxplots (Figure 1). It was observed that baseline costs were fairly evenly distributed, with direct costs contributing slightly more than non-medical and indirect costs. Analyzing the quartiles and whiskers also revealed that baseline costs had fewer outliers compared to follow-up costs. Additionally, indirect costs comprise a significant portion of the follow-up costs.

Given that indirect costs constitute a large portion of follow-up costs, a closer analysis was done to explore the components of these costs. It was found that 72.1% of indirect costs at follow-up were due to income loss, 21.9% from borrowed money, and 6.0% from property sold (refer to A1, A2).

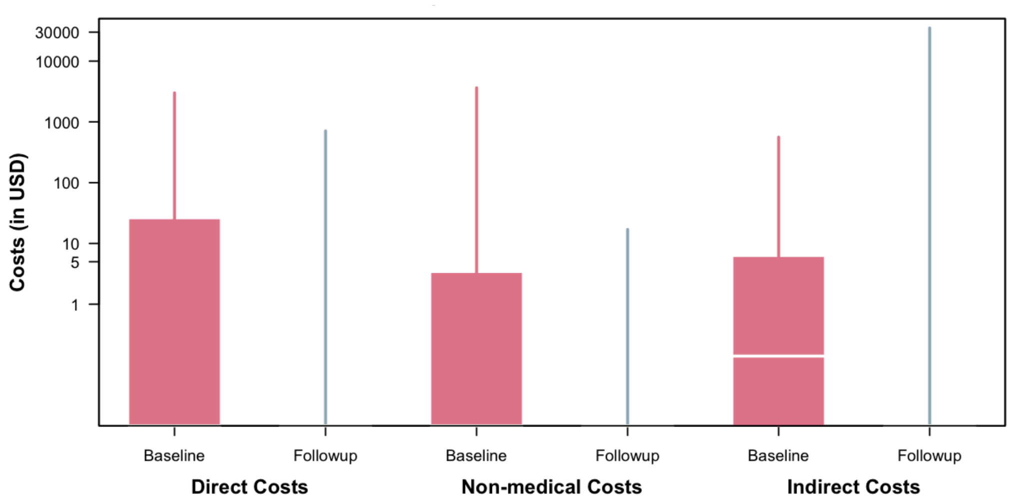


Fig 1. Comparison of cost distribution between baseline and follow-up.

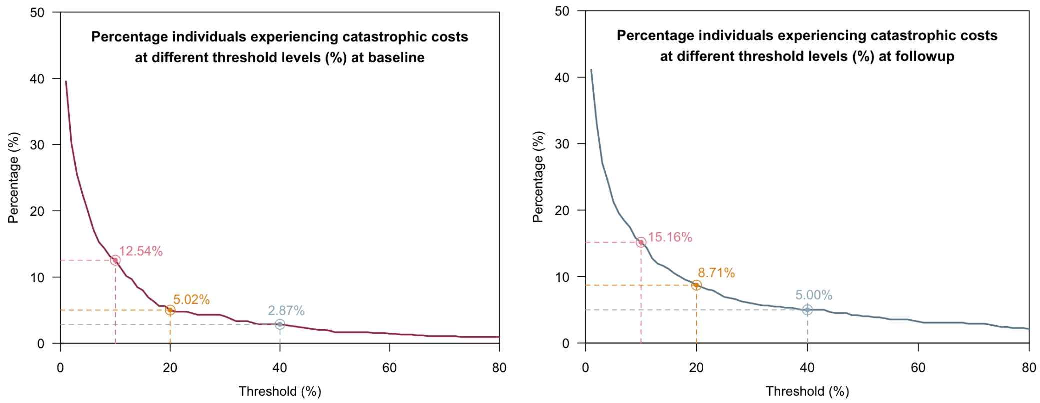


Fig 2. Percentage of individuals experiencing catastrophic costs at different threshold levels at baseline and follow-up.

Catastrophic Costs

Analysis of the calculated costs and the percentage of individuals experiencing catastrophic costs revealed that from baseline to follow-up, 217 individuals dropped out, 8 went from catastrophic to non-catastrophic, 30 became catastrophic, and 24 remained catastrophic throughout (refer to A3).

When examining the threshold rates and sensitivity, expanding beyond WHO's 20%, the proportion of individuals who spent more than 10% of their annual household income on total costs was lower at baseline (12.54%) than at follow-up (15.16%). This trend is consistent across both the 20% and 40% thresholds, with proportions 5.02% vs 8.71% and 2.87% vs 5.00%, respectively (Figure 2).

To support the exploratory analysis of catastrophic costs, we examined how the cost breakdown might differ between individuals experiencing and not experiencing catastrophic costs. Previously, we observed that the baseline cost breakdown was fairly evenly distributed and that indirect costs were the greatest contributor to follow-up costs (Figure 1). However, when breaking down the costs into non-catastrophic and catastrophic individuals, a different trend was found.

Indirect costs were the largest contributor to expenses for individuals not experiencing catastrophic costs at both baseline (45.6%) and follow-up (47.5%). In contrast, direct costs accounted for the greatest proportion of costs for those experiencing catastrophic costs, representing 65.9% at baseline and 57.5% at follow-up. For individuals experiencing catastrophic costs, indirect costs were the second largest contributor to TB-related expenses, accounting for 24.0% at baseline and increasing to 36.1% at follow-up (Table 1).

Table 1. Baseline and Follow-Up Catastrophic Cost Breakdown.

	Direct		Non-Medical		Indirect	
	Baseline	**Follow-Up**	**Baseline**	**Follow-Up**	**Baseline**	**Follow-Up**
Non Catastrophic	31.1%	29.5%	23.3%	23.0%	45.6%	47.5%
Catastrophic	65.9%	57.5%	10.1%	6.4%	24.0%	36.1%

Baseline

For the baseline data analysis, factors selected for testing associations with catastrophic costs included sex, employment, TB type, known medical conditions, HIV status, previous TB, age, and time to reach the nearest health facility. During preliminary data exploration, HIV status and previous TB were excluded due to insufficient data (refer to A4).

Chi-squared and t-tests conducted on the remaining factors identified sex and known medical conditions as having statistically significant associations with catastrophic costs, with p-values of 0.05 and 0.02, respectively, both meeting the threshold of $p \leq 0.05$ (refer to A5, A6).

Logistic regression was then applied to model the association between catastrophic costs and all considered factors, sequentially eliminating the least correlated variable at each step until reaching a final model where all remaining factors were significant (refer to A7). The initial logistic regression model for the baseline dataset had an AIC of 334.39, with RR TB and known medical conditions as significant factors. RR TB had a p-value of 0.02 and odds ratio of 4.17 at a 95% confidence interval (1.20, 12.95). Known medical conditions had a p-value of 0.03 and OR of 0.47 at a 95% CI (0.24, 0.92).

Successive models were run, resulting in a final model with an AIC of 331.04, where known medical conditions was the only factor with a statistically significant association with catastrophic costs (refer to A7). TB type was ultimately excluded from the final model despite its small p-value, as it was not found significant in the prior chi-squared tests (refer to A5).

Therefore, we concluded at baseline that known medical conditions was key factor for those with TB experiencing catastrophic costs, the final logistic regression model showing a p-value of 0.01 and OR of 0.45 at 95% CI (0.24, 0.85) (refer to A7).

Follow-Up

The same analyses were conducted on the follow-up dataset. Factors selected for testing associations with catastrophic costs included sex, employment, TB type, and relapse (refer to A8).

Chi-squared tests indicated that relapse was the only significant factor, with a p-value of 0.02 (refer to A9). The initial logistic regression model for the follow-up dataset, which included all considered factors, had an AIC of 365.23, with relapse and TB Bac- (TB type) emerging as significant factors. Relapse had an OR (95% CI) of 7.38 (2.13, 46.61) and TB Bac- had an OR (95% CI) of 0.53 (0.28, 0.99). Successive models were developed, resulting in a final model with an AIC of 362.20, where relapse was the only factor significantly associated with catastrophic costs (refer to A10). TB type was again excluded from the final model despite a small p-value, as it was not found significant in the previous chi-squared tests (refer to A9).

Therefore, we concluded at follow-up that relapse was a key factor for those with TB experiencing catastrophic costs, the final logistic regression model showing a p-value of 0.02 and OR of 5.24 at 95% CI (1.59, 32.39) (refer to A10).

Key Findings

The hypothesis tests and logistic regression models provide statistical evidence that known medical conditions were significantly associated with catastrophic costs at baseline, while relapse was a significant contributor to catastrophic costs at follow-up.

Discussion

The findings from this analysis highlight key factors that impact catastrophic costs among people with TB. Relapse of TB was strongly correlated with catastrophic costs, as relapses lead to a significant increase in medical expenses, which in turn increases financial strain. This is often worsened by the stigma associated with TB, potential job loss, and other consequences that can lead to experiencing catastrophic costs. For patients with known medical conditions, pre-existing health conditions can lead to higher costs since they require ongoing medical attention, increasing the likelihood of catastrophic expenses. These findings highlight the need for targeted interventions, such as subsidized healthcare and income support, to alleviate the financial toll of TB on vulnerable populations.

Limitations

While this analysis provides critical insight into factors contributing to catastrophic cost among individuals with TB, there are limitations to this study. The cohort data is restricted to one specific geographic area, limiting generalizability of results to other regions. Moreso, some factors such as previous TB status and HIV comorbidities were excluded from the analysis due to insufficient data, preventing a comprehensive analysis of potentially significant factors with catastrophic costs as well as relapse[4]. These limitations may reduce the accuracy of our findings or their applicability to broader populations.

Additionally, some cost estimates may be over or underestimated due how indirect costs were calculated, thus making this analysis not comparable to others. From baseline to follow-up, the decrease in sample size could also prevent us from observing significant statistical relationships.

Conclusion

Overall, the implications of this study emphasize the urgent need for financial protection mechanisms to support individuals with TB in low-income settings. The findings reveal the significant financial burden TB places on households, particularly for those with pre-existing medical conditions and those who experience a relapse during treatment. These insights are critical for designing interventions that reduce out-of-pocket expenses and improve access to care.

In practical terms, this analysis suggests that policymakers should prioritize targeted interventions for high-risk groups. Programs designed to reduce out-of-pocket expenses, provide transportation or accommodation assistance, and support patients with recurrent TB could significantly reduce the economic toll on households and improve treatment adherence.

Through this study, we found a strong relationship between catastrophic costs and two key factors: pre-existing medical conditions at baseline and TB relapse at follow-up. Patients with pre-existing conditions face higher medical expenses due to ongoing healthcare needs, while relapse leads to additional treatment costs, both of which can push households into financial hardship. These relationships suggest that targeted financial interventions for these high-risk groups could help mitigate the financial toll of TB.

This study provides critical insights into the economic challenges faced by individuals with TBs, results highlighting the importance of reducing financial barriers to TB treatment to improve health outcomes and prevent further financial hardships. This study's findings offer a base for future research and policy development to help mitigate the financial consequences of TB, ensuring that individuals can access the care they need without facing catastrophic costs.

References

1. R Core Team (2023). _R: A Language and Environment for Statistical Computing_. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

2. Morishita F, Yadav RP, Eang MT, Saint S, Nishikiori N. Mitigating Financial Burden of Tuberculosis through Active Case Finding Targeting Household and Neighbourhood Contacts in Cambodia. Lubell Y, editor. PLOS ONE. 2016 Sep 9;11(9):e0162796.

3. Wingfield T, Boccia D, Tovar M, Gavino A, Zevallos K, Montoya R, et al. Defining Catastrophic Costs and Comparing Their Importance for Adverse Tuberculosis Outcome with Multi-Drug Resistance: A Prospective Cohort Study, Peru. Ruger JP, editor. PLoS Medicine. 2014 Jul 15;11(7):e1001675.

4. Vega V, Cabrera-Sanchez J, Rodríguez S, Verdonck K, Seas C, Otero L, et al. Risk factors for pulmonary tuberculosis recurrence, relapse and reinfection: a systematic review and meta-analysis. BMJ open respiratory research [Internet]. 2024 Mar 13 [cited 2024 Jun 11];11(1):e002281. Available from: https://pubmed.ncbi.nlm.nih.gov/38479821/

Appendices

A1. Table of Follow-Up Indirect Cost Breakdown

A2. Histograms of Follow-Up Indirect Cost Breakdown

A3. Table of Number of People with Catastrophic Costs

A4. Data Exploration for Selected Variables at Baseline

A5. Baseline Chi-Squared Test Results

A6. Baseline T-Test Results

A7. Baseline Logistic Regression Models

A8. Data Exploration for Selected Variables at Follow-Up

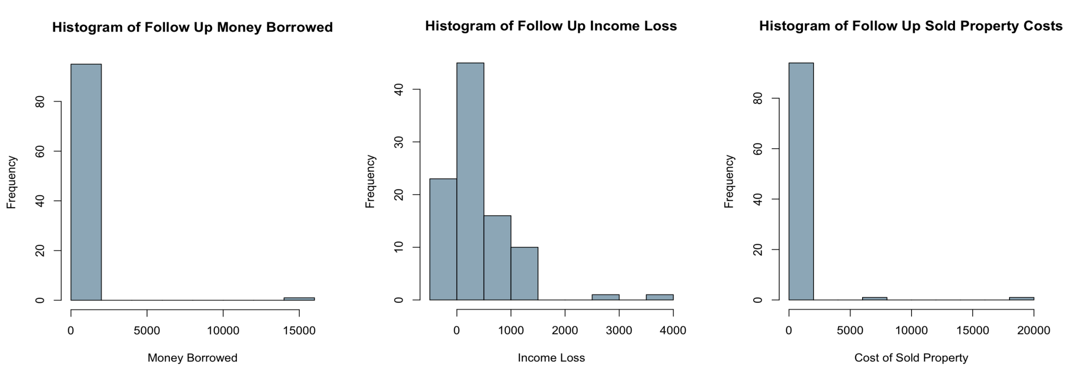
A9. Follow-Up Chi-Squared Test Results

A10. Follow-Up Logistic Regression Models

A1. Table of Follow-Up Indirect Cost Breakdown

	Median	IQR	Mean	Range	Proportion of Follow-Up Indirect Costs
Income Loss	$300	$60 - $600	$409.40	-$270 - $3600	72.1%
Money Borrowed	$0	$0 - $42.50	$285.80	$0 - $15000	21.9%
Cost of Property Sold	$0	$0 - $0	$341.10	$0 - $20000	6.0%

A2. Histograms of Follow-Up Indirect Cost Breakdown



A3. Table of Number of People with Catastrophic Costs

Time	True	False	Total
Baseline	42	795	837
Follow-Up	54	566	620

A4. Data Exploration for Selected Variables at Baseline

Catastrophic Costs vs Sex

	Male	Female
Non Catastrophic	417	378
Catastrophic	15	27

Catastrophic Costs vs TB Type

	TB Bac+	TB Bac-	RR TB
Non Catastrophic	312	445	38
Catastrophic	12	25	5

Catastrophic Costs vs Employment

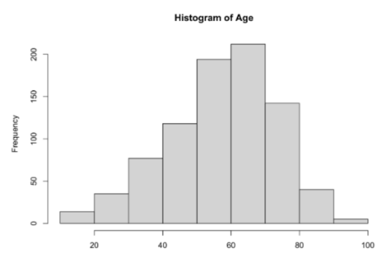
	Employed	Unemployed
Non Catastrophic	385	410
Catastrophic	15	27

Catastrophic Costs vs Known Medical Conditions

	Known Medical	No Known Medical
Non Catastrophic	318	447
Catastrophic	25	17

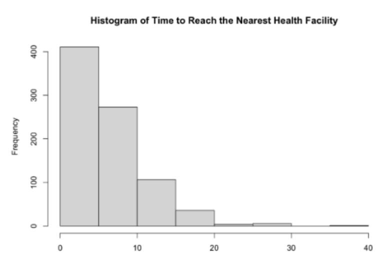
Catastrophic Costs vs Previous TB

	TB Bac+	TB Bac-	EPTB	Others	Don't Know
Non Catastrophic	19	22	3	2	25
Catastrophic	0	0	0	0	4



Catastrophic Costs vs HIV Status

	HIV Pos	HIV Neg	Unsure
Non Catastrophic	7	723	65
Catastrophic	0	34	8



Bac+: bacteriologically-confirmed TB
Bac-: clinically-diagnosed TB
RR: rifampicin-resistant
EPTB: extrapulmonary TB

A5. Baseline Chi-Squared Test Results

Variable	P-Value
Sex	0.05
Employment	0.15
TB Type	0.07
Known Medical Conditions	0.02

A6. Baseline T-Test Results

Variable	P-Value
Age	0.54
Time to Reach Health Facility	0.68

A7. Baseline Logistic Regression Models

Initial Baseline Logistic Regression			
Parameter	**OR (95% CI)**	**P-Value**	**AIC**
Female	1.51 (0.76, 3.10)	0.24	334.39
Unemployed	1.38 (0.68, 2.8)	0.34	
TB Bac-	1.21 (0.58, 2.62)	0.62	
RR TB	4.17 (1.20, 12.95)	0.02	
Age	1.00 (0.98, 1.03)	0.78	
Time to Reach Health Facility	1.01 (0.95, 1.07)	0.70	
Known Medical Conditions	0.47 (0.24, 0.92)	0.03	
Final Baseline Logistic Regression			
Known Medical Conditions	0.45 (0.24, 0.85)	0.01	331.04

TB Bac-: clinically-diagnosed TB
RR: rifampicin-resistant

A8. Data Exploration for Selected Variables at Follow-Up

Catastrophic Costs vs Sex

	Male	Female
Non Catastrophic	284	282
Catastrophic	30	24

Catastrophic Costs vs TB Type

	TB Bac+	TB Bac-	RR TB
Non Catastrophic	196	347	23
Catastrophic	22	29	3

Catastrophic Costs vs Employment

	Employed	Unemployed
Non Catastrophic	314	252
Catastrophic	32	22

Catastrophic Costs vs Relapse

	Relapse	No Relapse
Non Catastrophic	95	471
Catastrophic	2	52

Bac+: bacteriologically-confirmed TB
Bac-: clinically-diagnosed TB
RR: rifampicin-resistant

A9. Follow-Up Chi-Squared Test Results

Variable	P-Value
Sex	0.54
Employment	0.70
TB Type	0.53
Relapse	0.02

A10. Follow-Up Logistic Regression Models

Initial Baseline Logistic Regression			
Parameter	**OR (95% CI)**	**P-Value**	**AIC**
Female	0.88 (0.47, 1.62)	0.67	365.23
Unemployed	0.97 (0.53, 1.77)	0.93	
TB Bac-	0.53 (0.28, 0.99)	0.05	
EPTB	0.78 (0.17, 2.53)	0.71	
Relapse	7.38 (2.13, 46.61)	0.01	
Final Baseline Logistic Regression			
Relapse	5.24 (1.59, 32.39)	0.02	362.20

Bac-: clinically-diagnosed TB
EPTB: extrapulmonary TB