Tuberculosis and labour market participation: Evidence from South Africa

Received: 23 March 2023 | Revised: 24 April 2024 | Accepted: 21 June 2024

DOI: 10.1111/saje.12385

ORIGINAL ARTICLE

Tuberculosis and labour market participation: Evidence from South Africa

Helena Ting[1] | **Martina Bozzola**[2] | **Chiara Ravetti**[3]

[1]IHEID Graduate Institute of International and Development Studies, Geneva, Switzerland

[2]School of Biological Sciences, Institute for Global Food Security (IGFS), Queen's University Belfast, Belfast, UK

[3]DIGEP, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129, Turin, Italy

Correspondence
Martina Bozzola, Queen's University Belfast, 19 Chlorine Gardens, Belfast BT9 5DL, UK.
Email: m.bozzola@qub.ac.uk

Funding information
Ministero dell'Università e della Ricerca, Grant/Award Number: DM1062; MAVA Foundation, Grant/Award Number: Human Niche Project

Abstract

Tuberculosis (TB) is one of the leading causes of death from an infectious disease, but its micro-economic impacts on labour market participation remain poorly understood. We analyse the relationship between TB and employment outcomes in South Africa, one of the countries with the highest TB disease burden worldwide, using individual-level panel data from 2008 to 2017. Applying a coarsened exact matching methodology, we find that contracting TB entails a 5% lower probability of entering the formal labour market. Moreover, TB and its associated employment changes go hand in hand with corresponding reductions in individual income but not in household income and expenditure.

KEYWORDS

health, infectious disease, labour market, respiratory illnesses, tuberculosis

JEL CLASSIFICATION

I15, J01, J71

1 | INTRODUCTION

Tuberculosis (TB) is one of the deadliest diseases in the world. In 2022, it was the second leading cause of death due to a single infectious agent worldwide, ranking only after the coronavirus (COVID-19) and above HIV/AIDS. (WHO, 2022, 2023). The COVID-19 pandemic had detrimental effects on access to TB diagnosis and treatment. The number of people newly diagnosed with TB dropped substantially during the pandemic, suggesting an increase in the number of undiagnosed and untreated cases globally (WHO, 2022). The pandemic also fuelled growing concerns that developing countries with high levels of TB would have a higher vulnerability since both diseases affect the respiratory system (Houben & Dodd, 2016; Schluter et al., 2021).

Even though TB in its active form is a highly debilitating illness with a high mortality rate, patients can generally resume their normal activities in society once they complete appropriate treatment, since the infectiousness of TB and the likelihood of transmission decreases rapidly after treatment initiation

(Fitzwater et al., 2010; Petersen et al., 2016; Turner et al., 2017; WHO, 2010). While effective medical treatments for TB have existed since the 1940s, uptake of such treatments and the control of the disease are hindered by challenges in access, stigmatisation in the community and lack of information, motivation or support (Yoeli et al., 2019). Moreover, TB is often associated with socio-economic indicators of low income, such as crowded living conditions and malnutrition of the urban poor (Lonnroth et al., 2009). Despite these well-known obstacles to care and the vulnerability of the affected communities, the effect of TB infection on the working populations' employment prospects, household income and expenditure has not been sufficiently studied.

Within the African continent, South Africa is one of the countries with the highest burden of infectious diseases: Its population has the world's largest number of people living with HIV, as well as one of the highest incidence rates of TB, and this context creates unique socio-epidemiological challenges for the country (Mojola et al., 2022). For South African workers, TB and HIV/AIDS represent a significant additional challenge in a labour market already characterised by low labour force participation and high unemployment rates (Ingle & Mlatsheni, 2017; Statistics South Africa, 2023). Less than 60% of the working-age population is active in the labour market, and the unemployment rate has been consistently above 20% in the past 15 years and has climbed to over 30% since 2020 (Statistics South Africa, 2023). However, the literature on the impact of contracting TB on workers in South Africa is still relatively scarce. Most empirical studies on the effects of infectious diseases on labour market outcomes focus primarily on HIV/AIDS (*e.g.* French et al., 2019), despite the fact that TB is more infectious and lethal if not well controlled. Few micro-econometric studies of developing countries look at individual labour outcomes deriving from health conditions and include TB as one of the key health indicators (Nwosu, 2018)[1].

Our study analyses the labour market challenges faced by working-age individuals with a recent TB infection in South Africa. We use individual-level panel data from the National Income Dynamic Study (NIDS, 2018b), waves 1 to 5, which covers the period 2008 to 2017. Since the occurrence of TB is not random, we implement a coarsened exact matching (CEM) procedure (Iacus et al., 2011, 2012) to balance observations in TB-affected and TB-free individuals on a broad set of covariates that affects the selection into TB and the likelihood of employment. These covariates include individual and household demographic characteristics, skills, health and financial and household conditions. Our results are applicable to the relatively small proportion of the population whose demographic profile matches those who acquire an active TB infection. Thus, our findings should be placed in the context of this higher-risk group for TB. Depending on the nature of the outcome variable (employment status, income and expenditure), we implement a multinomial discrete choice model or OLS regressions after matching. As a robustness check, we also compare our results with a propensity score regression adjustment estimator.

We find that an individual who contracted a TB infection is 5% less likely to enter the formal employment market. This result is robust across a range of alternative specifications. In contrast, the relationship between contracting TB and exiting the formal employment market is not statistically significant. We also investigate the effect of TB on earnings and expenditure and find that household income is not significantly linked to TB and the reduced access to formal employment. This may be explained by the availability of alternative income sources from the informal sector (through casual or self employment, agriculture and household work), which we find to be unaffected by TB. The fact that we do not observe an effect of a family member contracting TB on household income could also be due to the support of other family members or the availability of public social welfare programmes for those in need. Finally, household expenditure is also not significantly linked to disease-related employment changes. These results suggest that policies addressing the economic vulnerability of TB patients should provide support especially for those workers that are searching for employment and do not have substantial access to family networks that can compensate for the labour market disadvantage associated with TB infections.

[1]Several studies present summary statistics on small surveyed populations indicating that attachment to the job market may be hindered by TB but do not model the relationship with controls for individual characteristics or balancing "treated" and "untreated" populations. See, for example, Meghji et al. (2021) and Rupani et al. (2020).

The article is organised as follows. In Section 2, we present the context and relevant literature on TB. Section 3 describes our data. Section 4 presents our empirical models and Section 5 the main results for individual employment, individual and household income and expenditure. Section 6 concludes with a discussion of policy implications and avenues for further research.

2 | BACKGROUND AND RELATED LITERATURE

First, we contextualise the issue of TB in South Africa, especially under the labour market conditions faced by the working population. Next, we review the literature on the economic impacts of health shocks in South Africa and elsewhere. Lastly, we focus on the few studies closer to our analysis, which use evidence from African countries to study labour outcomes of HIV and TB.

2.1 | TB in South Africa

TB is caused by infection with *Mycobacterium tuberculosis*. It spreads when droplets carrying the active disease are inhaled by susceptible individuals. The probability of transmission is determined by the severity of pulmonary TB, the rate of cough aerosol production and closer proximity or longer duration of contact between a source and susceptible individuals (Mathema et al., 2017; Yates et al., 2016). There are significant and ongoing efforts to control TB in South Africa, and programs are in place for earlier diagnosis and free treatment at local clinics. The incidence rate of TB in South Africa has decreased steadily from 1270 in 2008 to 513 per 100,000 in 2021 (WHO, 2022). However, roughly 50% of individuals with TB do not successfully complete the full course of treatment over six months (Botha et al., 2008; Classens et al., 2013; Naidoo et al., 2017). A qualitative study found that the leading reasons for lost-to-follow-up include not feeling ill or feeling better after some treatment, conflicts with daily activities (*e.g.* difficulty accessing facilities or conflicts with work), feeling too ill and social stigma—particularly for those who are also HIV positive (Skinner & Classens, 2016).

For better adherence to medication, clinic programs promote 'directly observed therapy (DOT)', whereby a health care worker watches the TB patient swallow each dose of the prescribed drugs. All patients are counselled to start DOT at the time of diagnosis; however, only a small fraction receives the complete long term treatment. A cross-sectional study in the Free State province shows that only about 25% of TB patients are undergoing DOT at a given time either at home or at the clinic, about 28% report not receiving any support for DOT, while the remaining receive support from family, friends and employers (Howell & Heunis, 2018). Alternative approaches that are less disruptive for people's schedules to increase the chances that a patient successfully completes the full course of the TB treatment include methods based on providing comprehensive and personalised support (Alipanah et al., 2018) and digital approaches such as interactive two-way mobile phone text message reminders and video assisted DOT (Ngwatu et al., 2018). Both of these approaches have been associated with good results.

The disease burden of TB is closely associated with industrialisation and urbanisation. TB rates are often high in urban settings, where transmission increases due to higher population density and crowded living conditions, while poor nutritional status and other risk factors increase the risk of transitioning to the active state in vulnerable groups, such as the urban poor (Lonnroth et al., 2009). TB may undermine households' economic conditions and increase poverty through several channels. Direct and indirect costs associated with TB diagnosis and treatment can amount to more than 10% of patients' or households' annual income (Barter et al., 2012; Russell, 2004). Disease-related costs lead to reduced food spending, reduced spending on education or other health care and borrowing or selling assets (Chimbindi et al., 2012).

Foster et al. (2015) carried out a study at 10 clinical centres across four provinces in South Africa and found that patients incur most of the costs (about 40%) before starting treatment, with income loss as the leading contributor to the overall cost of disease. Direct costs such as nutritional supplements, transportation and the cost of someone accompanying the patient to the clinic account for 12% of annual pre-symptom income. There is, however, a lack of evidence on the specific mechanisms driving the impoverishment caused by TB and particularly on the consequences for workers with respect to their occupations and their household consumption.

2.2 | Employment in South Africa

The labour market in South Africa is characterised by persistently high unemployment and rapid transitions in and out of the market. Youth unemployment is a particularly harsh problem for the country and many youths suffer prolonged unemployment. This may lead to discouragement and depression (Mlatsheni, 2012). A study from the Limpopo province shows that skills training and work experience are associated with higher odds of being employed (Dagume & Gyekye, 2016). Using data from the first four waves of the National Income Dynamic Study, Ingle and Mlatsheni (2017) find that employment of youth in South Africa is unstable, with layoffs, rather than voluntary quits, as the main driver of this instability. The authors suggest that the availability of a large number of youths for a small selection of jobs has given employers the flexibility to be selective in their hiring decisions. Espi-Sanchis et al. (2022) studied labour market outcomes during COVID-19 and found that although the employment-to-population ratio was nearly identical before and after lockdown, there had been extensive churning as 23% of those employed has lost jobs, and 30% of those unemployed has found jobs. Overall, we suspect that health shocks can be particularly detrimental for the individuals experiencing them in this context of high labour market volatility and persistent unemployment, as discussed in the next section.

2.3 | Health shocks and economic outcomes

Early studies on the economic consequences of health shocks in developing countries focus on how families facing health issues use informal mechanisms to cope with such episodes. This literature has shown that typically households are able to smooth consumption when the shocks are moderate but not when they are severe (Gertler & Gruber, 2002; Kochar, 1995; Townsend, 1994). To address the concerns about possible endogeneity regarding the relationship between health shocks and economic outcomes, studies often rely on objective, unpredictable and relatively exogenous measures of health changes, such as urgent hospitalisations (Garcia-Gomez et al., 2013); sudden onset of major diseases such as cancer, stroke or myocardial infarction (Jones et al., 2020), bus accidents (Mohanan, 2013); or a combination of events such as sudden death in the family, a sudden long spell of hospitalisation and a large drop in body mass index (Wagstaff, 2007). Several of these studies implement matching algorithms to further eliminate potential biases introduced by selection into healthy or unhealthy groups (Garcia-Gomez et al., 2013; Jones et al., 2020; Mohanan, 2013).

A few recent studies evaluate the relationship between health and labour market outcomes in South Africa. On general health, the study of Alfers and Rogan (2015) uses the first wave of data from NIDS to show that the formality of employment is significantly correlated with better health and argue that this protective effect derives largely from the higher level of income associated with more formal types of employment. Nwosu and Woolard (2017) study the effect of self-assessed health on labour force participation using four waves of NIDS and find a positive and significant effect of self-reported good health on labour force participation. Lawana et al. (2020) use a structural estimation model with multivariate probit to identify a negative effect of noncommunicable diseases on labour force participation using data from NIDS waves 1 to 4.

The studies that come closest to our analysis look at labour market outcomes in Sub-Saharan Africa following HIV and, in one case, specifically TB. A first study evaluating the effect of HIV on South Africa's labour market, Levinsohn et al. (2013), shows that being HIV-positive is associated with a 6% to 7% increase in the likelihood of being unemployed, with a stronger effect for those with less education. However, other studies show that proper treatment can counteract these effects. Thirumurthy and Zivin (2012) find that antiretroviral therapy greatly increases the individual labour supply of treated patients in rural Kenya, starting already six months after the initiation of treatment. They use a panel with individual fixed effects following patients for 36 months after the start of treatment and observe almost doubling in the number of hours worked from the time of treatment initiation. Similarly, Bor et al. (2012) evaluate the employment effects of a public-sector HIV treatment program. Using a 10-year panel on more than 100,000 people in the province of northern KwaZulu Natal (South Africa), they find that 4 years after the initiation of antiretroviral therapy, employment among HIV patients recovered to about 90% of rates at three to five years prior to treatment. Likewise, French et al. (2019) examine the effect of HIV on labour outcomes in a mining company in South Africa and find that when treatment is initiated at an earlier stage, the risk of separating from the company is 37% lower.

Given the importance of TB as an infectious disease, its widespread diffusion in South Africa and its high comorbidity with HIV, this literature suggests that similar effects could be expected for TB and its treatment. Using data from NIDS waves 1 to 4, Nwosu (2018) examines the effect of physical and mental health on labour market participation and employment, including TB as one of the physical health indicators. The author uses pooled ordinary least squares (OLS), random effects (RE) and fixed effects (FE) models to discern the effects and finds that while there is a 12% to 13% lower likelihood of employment compared to being noneconomically active for individuals with TB using pooled OLS and RE models, this effect disappears when a fixed effect model is used. Furthermore, the study did not find any significant effects on the likelihood of employment for those who are active in the labour market, comparing the population of employed and unemployed but searching workers. The study draws the conclusion that health issues, such as TB or other chronic conditions, can negatively affect the likelihood of being employed compared to being economically inactive, and the effect extends over multiple periods.

Our study contributes to this literature by exploiting the panel structure of the NIDS dataset over the time period of 2008 to 2017 (waves 1 to 5) to evaluate the effect of a recent TB infection on the likelihood of transition in the formal employment market. We complement existing studies by exploring the difference between formal and informal employment markets, across demographic subgroups, and in lower-income households. Moreover, we present these labour market consequences in relation to income and expenditure changes for individuals and households. Our study takes a different approach from Nwosu (2018), as we focus exclusively on labour market transitions in and out of employment linked to TB, due to its specific policy dimensions. Because of its infectiousness and the seriousness of an infection, TB patients must be carefully treated and monitored, and the success of South Africa's campaign against TB in recent years points to a robust system to control the spread of the disease. Therefore, people with TB infections can, under the right circumstances, be active in the labour market. However, policymakers need to be aware of the challenges posed by TB for the labour market participation of TB patients and design and adjust policy instruments accordingly. Our study intends to contribute specifically to this policy debate.

3 | DATA

We use individual-level panel data from South Africa's National Income Dynamic Study, waves 1 to 5, which cover the time period from 2008 to 2017. The survey collects data from over 7000 households in wave 1 with additional households added in subsequent waves (NIDS, 2018b, 2018c, 2018d, 2018e, 2018f). It provides representative household information in two-year increments on a broad range of characteristics, including individual demographic characteristics (age, gender, education, parents' education, employment, financial status), household characteristics (household size, dwelling type, energy source, water source and amenities) and health characteristics (self-rated health and life satisfaction, diseases such as hypertension, diabetes, heart disease, asthma and cancer, and behaviours such as smoking and alcohol consumption). Employment status is reported for wage employment (full-time or part-time) in the formal sector and casual income, self-employed income, agriculture work and household work are recorded. Total household income and expenditure for the previous 30 days is also captured by the survey. We adjust income and expenditure for inflation using the consumer price index (CPI) of South Africa over the same time period.

These data have been used to study a variety of household-level outcomes, such as the determinants of household savings in South Africa (Zwane et al., 2016), the prevalence of co-morbidities and their association with socio-economic disadvantage (Weimann et al., 2016), the impact of health on labour force participation (Nwosu & Woolard, 2017) and the relationship between employment and physical/mental health (Nwosu, 2018). Leibbrandt et al. (2009) provide a detailed description of the sampling and data collection methodology. We restrict the sample to adults aged 15 to 60, which captures the working-age population of South African adults. After the age of 60, South Africans are eligible to apply for an old age pension from the government.

There are three questions in the survey related to an individual's TB status. The first question asks whether the person 'has ever been told by a doctor, nurse or health care professional that they have TB.' This question captures the population who has had TB but has recovered from it as well as those who are currently living with TB. Since the TB episode could have occurred in the past and we do not have good information about the time that has elapsed since the occurrence of TB for this population, we do not use this information to represent the population who currently or recently had TB. The second question asks whether the individual is currently taking medication for TB. We consider that if the individual is currently under active treatment, then it is almost certain that the individual currently carries the bacteria. Thus, any individual who answers positively to this question is considered in our data as having TB. The third question asks 'Do you still have TB?'. This question directly asks if the person is carrying the bacteria. If the individual answers 'yes' to either of the last two questions, then we consider this person as carrying the bacteria. In the sample, all individuals who report that they are taking medication also report that they still have TB. Out of 1379 individuals with TB, a large majority (88%) reports that they are taking medication. However, since there is only one "yes" or "no" question on the medication, we do not have full information on the extent of medication adherence of the individual.

Our data are a lower-bound estimate of people who have an infection of TB. Since close to 90% of individuals with TB included in our sample are taking medications, we assume that they have some symptoms or risk factors which has led them to seek treatment. Therefore, in this analysis, we capture individuals with a confirmed diagnosis of TB which has led to a prescription of some treatment. If individuals are not diagnosed but have TB, or if they know about their conditions but choose to not disclose it, then they would report that they do not have TB presumably in both periods. This could lead to an under-reporting of TB in the treated category, and our result may be underestimated. Note that there is also a population of individuals who responded that they have TB across multiple periods, which could be an indication that they have persistent, potentially treatment-resistant TB in a more severe form, but in our analysis, we excluded this population.

Table 1 outlines the disease burden from TB in our sample. As expected, the disease burden in the sample is higher than the incidence reported by the WHO (WHO, 2022) since this measure of disease burden includes new cases, relapsed cases and potentially longer term carriers of the disease. In Table 2, we also report the transition frequency of TB from one wave to another. This transition frequency shows four categories: first, for individuals who do not have TB in a given period, it presents (i) the share of those who remain TB-free in the next wave and (ii) those who contract TB. Second, for individuals already with TB, it presents (iii) the share of those who become TB-free and (iv) those who remain TB-infected in the following period. About 1% of individuals with no TB contract and start treating the disease in the following period and about 78% of those with medicated TB in one wave are no longer sick and taking medication in the following period. For our analysis, we look at the population that either remained TB-free in both periods ($n = 44{,}904$) or transitioned from being TB-free in period 0 to having TB in period 1 ($n = 492$). For our analysis, we define a change in TB status as a recent episode of TB infection, compared to the condition of a healthy individual that experienced no change in health status.

The sample includes data on employment (part-time and full-time) in the formal sector and activities in the informal sector (casual employment, self-employment, agriculture work and household work). From

TABLE 1 Disease burden of TB.

	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5
Sample statistics					
TB observations	320	221	272	349	217
No TB observations	15,163	17,273	18,373	22,027	23,304
Disease burden	2067	1263	1459	1560	923
WHO statistics					
Incidence	1270	1200	1160	988	738
Case detection rate	54%	58%	52%	52%	58%
Treatment success rate	73%	77%	77%	81%	77%

Note: Disease burden is estimated as the number of individuals with TB over the total sample, scaled to per 100,000 population (TB observations / (TB observations + no TB observations) * 100,000). This includes new, recurrent and potentially longer-term TB cases. Incidence data are reported on the scale of per 100,000 population (WHO, 2022). Observations in which respondents did not provide an answer for all three questions regarding TB are excluded. These observations represent 0.9% of all observations with data on both employment status and TB status.
Abbreviation: TB, tuberculosis.

TABLE 2 Transition of TB status between any two periods.

Number of cases		Period 1: No TB	Period 1: TB	Total
Period 0	No TB	44,904 (98.9%)	492 (1.1%)	45,396
	TB	415 (78.6%)	113 (21.4%)	528
	Total	45,319	605	

Note: This table shows the transition matrix of TB infections across two periods. In this sample, close to 99% of the sample do not have TB in both periods. About 1.1% of the sample is infected with TB, while 78.6% of the sample recovers from having TB to not having TB. For our analysis, we focus on individuals who remain healthy and those who contract a TB infection.
Abbreviation: TB, tuberculosis.

this data, we derive our definition of unemployment, which excludes those individuals who have no desire to work but, to avoid any measurement error in the surveys, does not further distinguish between those who are searching from those who are not actively searching for a job. Including formal and informal employment, the labour force participation rate in those years is 46%–55%, and this measure is lower than the official statistics of 56%–60% in the same period[2]. Using a strict definition of unemployment, which conforms to the ILO definition and is used by Statistics South Africa, the unemployment rate in this period ranges from 24% to 30%[3] (ILO, 2016; NIDS, 2018a; Statistics South Africa, 2023).

The main outcome variables for our analysis are changes in employment status, income and expenditure, while the explanatory variable is a change in TB status, namely, contracting TB. The summary statistics are outlined in Table 4, and the description of all variables used in the regression models is outlined in Table A1 in the appendix.

[2]This could be due to missing data on whether an individual desires to seek employment, which is labelled as "Other unemployed" in Table 3. Some proportion of these individuals could be in the labour force, and some could be considered as economically inactive. However, since we do not have data on their desire to seek employment, we cannot make this determination. Since these individuals are included in the total population, but not in the labour force, they depress the statistic for labour force participation. Excluding this group from the total population will yield labour force participation rates closer to the official statistics.

[3]The strict definition characterises unemployment as those who are not working, available, and are actively searching for jobs. The official statistics from Statistics South Africa uses this strict definition for its unemployment rate calculations; however, it also presents analyses for a broader definition more aligned with ours, which does not further distinguish between those who are actively searching or not searching a job (ILO, 2016; Statistics South Africa, 2023).

TABLE 3 Employment statistics.

	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5
Employed (formal)	3717	3924	4586	6206	6708
Employed (informal)	2083	1178	1508	2181	2043
Unemployed - desire to work	2757	2440	3026	2992	2978
No desire to work	4956	7755	7195	8234	8660
Other unemployed	1988	1185	1471	2119	1948
Total	15,501	16,482	17,786	21,732	22,337
Rates from the sample (%)					
Labour force participation	55.2	45.8	51.3	52.4	52.5
Unemployment rate					
Strict definition	24.5	24.4	30.3	24.5	23.6
Broad definition	32.2	32.4	33.2	26.3	25.4
Rates from Statistics South Africa (%)					
Labour force participation	59.2	55.5	56.8	58.6	59.9
Unemployment rate	22.8	24.8	25.2	26.4	27.7

Note: Labour force participation is calculated as the ratio of those who are employed (in both formal and informal sectors including agriculture and household work) and those who are unemployed over the total sample. Unemployment is calculated by the ratio of those who are unemployed over the total labour force. The broad definition of unemployment refers to those who are not working and available to work, and includes both those who are searching and not actively searching. The strict definition of unemployment includes only those who are actively searching in the unemployed population. This is calculated using the derived variable $wx_{e}mpl_{s}tat$. The strict definition is based on the ILO definition of unemployment. (ILO, 2016) Those who are unemployed but does not indicate whether they desire to seek work are labelled as "Other unemployed" and included in the total sample size but not in the labour force. Statistics South Africa data is from QLFS trends 2008-2023Q1 and is based on the strict definition of unemployment (Statistics South Africa, 2023).

4 | EMPIRICAL APPROACH

4.1 | CEM

Since the population who is likely to be infected with TB has different characteristics than the general population, we apply a matching technique to reduce the model dependence by creating balance in covariates across treated and control individuals prior to the parametric estimation. Successful matching brings the data closer to the conditional independence assumption, leading to less dependence on model specification of any subsequent parametric regression modelling (Ho et al., 2007; Iacus et al., 2011, 2012; Jones et al., 2020). The matching technique seeks to ensure that adequate balance is achieved on a set of covariates (confounders) deemed to be the most relevant, *a priori* to the analysis. These confounders include variables related to selection into the treatment (in our case, TB) and to the outcome of interest (in our case, employment transitions) (Stuart, 2010). There is, however, still no full consensus on which ones to select, and our approach is to identify the list of confounders that affect treatment selection and the outcome through a set of logistic regressions. We then evaluate the resulting balance after matching on these coarsened confounders. We iterate this process until a satisfactory level of balance is achieved across all covariates. The selection of this set of confounders is based on their effect on the treatment and on the outcome (Jones et al., 2020).

Our identification approach relies on the assumption that, conditional on the set of balanced covariates, selection into TB could be thought of as being close to exogenous, and no other factors should be relevant (ignorability assumption). This is a strong assumption as there could be unobserved characteristics and sample selection bias that could affect the joint distribution of the covariates on the outcome. Therefore, while our method attempts to improve identification through the matching technique, yet we cannot claim that all characteristics (both observed and unobserved) are balanced through our approach,

TABLE 4 Summary statistics of variables.

	Obs.	Mean	S.D.	Min	Max
Outcome variables					
Employment	84,837	0.30	0.46	0	1
Household income	82,084	4433	9249	0	440,529
Household expenditure	78,762	2163	4996	0	475,436
Individual income	83,961	1102	3340	0	199,377
Health shock					
TB	84,777	0.01	0.12	0	1
Demographic control variables					
Age	84,934	32.7	13.0	15	60
Gender	84,934	1.58	0.49	1	2
Race	84,921	1.22	0.50	1	3
Education	84,194	2.33	0.88	0	3
Mother's education	59,216	1.11	1.19	0	3
Father's education	62,243	1.02	1.20	0	3
Computer literacy	82,780	0.35	0.48	0	1
Driver license	82,748	0.16	0.36	0	1
Health status	84,839	2.13	1.05	1	5
Hypertension	82,448	0.10	0.30	0	1
Diabetes	84,107	0.03	0.16	0	1
Stroke	84,639	0.01	0.08	0	1
Asthma	84,079	0.03	0.16	0	1
Heart disease	84,354	0.02	0.12	0	1
Cancer	84,616	0.01	0.08	0	1
Smoke	84,160	0.22	0.41	0	1
Alcohol	63,896	1.97	0.16	1	2
Medical aid	84,100	0.10	0.30	0	1
Life satisfaction	82,684	2.88	1.19	1	5
Pension	83,906	0.03	0.18	0	1
Bank account	83,834	0.44	0.50	0	1
Household size	84,934	5.44	3.43	1	41
Dwelling type	84,597	1.55	0.81	1	3
Urban/Rural	84,761	1.67	0.61	1	3
Owned house	84,778	0.76	0.43	0	1
Land grant	83,646	0.03	0.18	0	1
Subsidy	83,555	0.19	0.39	0	1
Water source	84,791	1.21	0.59	1	4
Toilet	84,765	1.51	0.58	1	3
Electricity	84,368	0.83	0.38	0	1
Cooking energy	84,729	1.23	0.42	1	2
Heating energy	84,621	1.37	0.48	1	2
Lighting energy	84,733	1.14	0.35	1	2
HH has mobile	84,787	0.86	0.34	0	1

(Continues)

TABLE 4 (Continued)

	Obs.	Mean	S.D.	Min	Max
HH has refuse collection	84,704	0.48	0.50	0	1
HH has street light	84,792	0.39	0.49	0	1

and thus, we do not claim a precise causal link but only a first attempt to come closer to conditional independence of the parametric model.

We take advantage of the rich set of background data available through NIDS to construct a set of confounders that affect treatment selection or the outcome independent of the treatment selection. Information about the individual's demographics, health status and behaviour, background financial situation and various household settings are used to construct a set of confounders to support the matching approach used in CEM. We also take a first-difference approach across two consecutive time periods to eliminate any time-indifferent individual heterogeneities. However, we are well aware that there could be other unobserved characteristics that affect the changes in treatment (contracting TB) and outcome (finding or losing employment). Since we can only select on observed characteristics, our method could suffer from potential sample selection bias, as we cannot rule out some predisposition both at the physical level and linked to behavioural choices that could also influence the chances of success in the formal labour market, for example, frequent visits to facilities where TB transmissions occur, preferences for seeking healthcare and individual's risk aptitude. An important missing variable in this case is the HIV status of the individual, which is not available through our data. HIV positive individuals are more likely to be immune-compromised, and co-infection between HIV and TB is a significant element of the health risk that we cannot account for, unfortunately, due to the lack of data on HIV status. All these factors could increase the chances of contracting TB and simultaneously of losing a job or not finding one.

CEM is a stratification technique to prune the observational database based on exact matching of a set of coarsened core variables thought to affect both the treatment assignment and the outcome, controlling for the treatment. CEM generates treatment and control samples of observations with balanced covariates, in an attempt to improve overlap in covariate distribution and approach conditional independence (Iacus et al., 2011, 2012). For any matching technique, the most ideal case would be to find an exact match from the control group for each observation in the treatment group since this ensures perfect balance of observed covariates. This is, however, oftentimes unattainable due to data limitations. In the case of a data set where there is a large number of control group observations and a set of confounders to be balanced at baseline, the coarsened exact match procedure approximates exact matching within groups created by meaningfully coarsening select variables. By eliminating groups where there is no match between control and treatment groups, the procedure brings balance to the baseline profile of the observations based on a set of observed characteristics.

The drawback of this procedure is, however, that a selected number of confounders, which are variables that may determine selection into the treatment group or the outcome, must be determined *a priori* which would allow such variables to be used to generate coarsened stratifications. A study using insurance claims datasets and simulations suggests that CEM implemented with default auto-coarsening algorithms may introduce bias and low precision due to the significant pruning of control group observations that do not match treatment group observations (Ripollone et al., 2020). For complex health services datasets with many covariates and complicated relationships among covariates, a small set of meaningful confounders may not be determined prior to the analyses. This, however, does not seem to be the case in our analysis, where we identify a set of confounders that leads to improved balance among covariates and robust results. However, CEM, like other matching techniques, takes effect on a set of observed covariates and therefore, by design, does not adjust for any potentially unobserved covariates that may affect the outcome.

We take a two-stage approach to stratification, similar to the approach taken by Jones et al. (2020). By adding a coarsened propensity score as an input to stratification in the second step, we further finetune local imbalances and improve balance among treatment and control groups within the coarsened stratum resulting from the first stage (Iacus et al., 2012)[4].

The first step of the CEM pre-processing is to ensure that adequate balance is achieved on a set of covariates (confounders) that are deemed to be most relevant, *a priori*, to the analysis (Ho et al., 2007; Iacus et al., 2011, 2012; Stuart, 2010). We determine covariates that may be related to being infected with TB (choice into the treatment group) and to employment, controlling for TB (outcome, controlling for the treatment), through a set of logistic regressions following the approach of Jones et al. (2020). Table A2 in the appendix shows the list of covariates for the coarsened matching procedure, including demographic characteristics such as age; education and skills; health indicators such as self-reported health status, health comorbidities (hypertension, diabetes, stroke, heart disease, smoking etc.)[5]; financial indicators; household conditions and energy source. We then stratify our data based on the combinations of coarsened categories and match treatment and control samples within each coarsened stratum. We prune control observations that are unmatched for the treatment within each stratum. In the second step, we add a coarsened propensity score. The propensity score is generated from a logistic regression of the treatment variable on the list of covariates. We weigh the propensity scores by the inverse of the estimated propensity of treatment (Imbens, 2000; Stuart, 2010). We repeat the procedure of eliminating strata where there does not exist a match for the treatment group. Further, we generate weights to be included in the regression that incorporate the disproportionate number of observations that remain in each stratum between the treated and control groups (Iacus et al., 2012). The weight for each treated observation is 1, while the weight for each matched control observation is $\frac{m_C}{m_T}\frac{m^s_T}{m^s_C}$, where m_C is the total number of matched control observations, m_T is the total number of matched treatment observations, m^s_T is the number of treatment observations in the stratum and m^s_C is the number of control observations in the stratum. Table 5 highlights the pre-matching and post-matching balance of key covariates, and a full list of balanced covariates can be found in Table A3 in the appendix[6].

Our matching is based on observations which are person-time pairs. Therefore, it is possible for one individual to represent multiple observations in the treatment or control groups. Prior to matching, multiple observations from the same individual represent 79% of the observations in the control group, while only 4% in the treated group. After matching, 25% of the observations in the control group represent duplicate individuals, while 4% remains in the treatment group. The duplicates with TB could represent either treatment failure, or reinfection, or a person not treating the infection. It could also be related to measurement error in the recording of the disease data in the survey. There are two ways in which measurement error could occur in this case. If an individual has TB but reports that they do not currently have TB either because they do not know or because of an intentional misstatement, then our TB observations are underestimated indicating a potential bias of the result towards the null. This could be one source of measurement error. Although unlikely, it could also be the case that an individual reports having TB although they no longer carry the bacteria. In this case, our TB observations could be overestimated. We believe that this case is less likely than an underestimation of the TB observations.

4.2 | Empirical model

After preprocessing with CEM, we estimate the effect of TB as a difference in means of the outcome variable, our measures of employment, income and consumption on a balanced set of covariates, assuming that the preprocessing step has brought us closer to conditional independence (Ho et al., 2007).

[4]Of course, we are not controlling for unobservable factors uncorrelated to our matching variables, and these could still play a role in determining who contracts TB and what kind of labour market outcomes they experience.

[5]Since the use of alcohol is not recorded in Wave 5, we remove it from the propensity score estimation in the second step of the CEM stratification.

[6]We also assessed pre- and post-match balance of other variables not included in CEM. Two variables, household mobile phone possession and receiving land grants, are unbalanced after the matching procedure. As a robustness check, we tested our model including these two additional variables. Our results remain robust (data not shown and available upon request). This does not rule out, however, that there could be other confounders in the matching process, and these other imbalances could be driving the result.

TABLE 5 Prematching and postmatching balance of select covariates.

	Prematching			Postmatching		
	Mean TB	**Diff.**	***p***	**Mean TB**	**Diff.**	***p***
Age	37.4	−5.18	0.00	37.4	−0.30	0.66
Gender	1.56	0.04	0.10	1.56	0.01	0.74
Parents education	0.59	0.50	0.00	0.60	−0.03	0.52
Education	1.81	0.50	0.00	1.88	0.00	0.96
Computer literacy	0.17	0.15	0.00	0.16	0.02	0.32
Driver's license	0.07	0.06	0.00	0.07	0.01	0.55
Self-rated health	2.58	−0.48	0.00	2.56	0.03	0.54
Hypertension	0.13	−0.03	0.05	0.14	−0.00	0.88
Diabetes	0.03	−0.002	0.83	0.02	0.00	0.86
Stroke	0.01	−0.005	0.21	0.01	−0.00	0.70
Heart disease	0.02	−0.001	0.90	0.02	0.00	0.98
Smoke	0.32	−0.12	0.00	0.33	−0.04	0.11
Alcohol[a]	1.96	0.02	0.07	1.96	0.00	0.87
Receives medical aid	0.03	0.07	0.00	0.03	0.00	0.69
Self-rated life satisf.	2.62	0.23	0.00	2.61	0.02	0.67
Pension	0.01	0.02	0.03	0.01	0.00	0.71
Bank account	0.32	0.09	0.00	0.34	0.03	0.31
Household size	5.41	0.07	0.66	5.20	−0.02	0.93

Note: The description of each variable is outlined in Appendix Table A1. Two sample *t*-tests are reported. The prematching sample size for TB is 492 and for those without TB is 44,904. The postmatching sample size for TB is 394 and for those without TB is 5278. The difference between TB sample sizes is due to missing data which are not used in propensity score stratification and in group assignment. Abbreviation: TB, tuberculosis.
[a]Alcohol refers to data from waves 1 to 4 only since this question was not asked in Wave 5.

Our outcome variable, employment, has four potential values that indicate a change from one time point to another. These four categories are (1) those who are unemployed at baseline and continue to be unemployed, (2) those who are employed at baseline and continue to be employed, (3) those who find a job, meaning that their employment status has changed from 0 in period zero to 1 in period one, and (4) those who had a loss in employment, meaning that their employment status has changed from 1 in period zero to 0 in period one to assess the effect of our treatment variable on employment; the main model we adopt is a multinomial discrete choice model (Imbens & Wooldridge, 2007; McFadden, 1973), where the predicted probabilities are represented by

$$\begin{aligned} \pi_j(EMP_j) &\equiv Pr(EMP_i = j | TB_i) \\ &= \frac{exp(TB_i^T \beta_j)}{\sum_{k=1}^{J} exp(TB_i^T \beta_k)} \\ &= \frac{exp(TB_i^T \beta_j)}{1 + \sum_{k=1}^{J-1} exp(TB_i^T \beta_k)} \end{aligned} \tag{1}$$

where EMP_i is the change in employment status for individual i and TB_i is the change in whether the individual has TB from one period to the next. $\pi_j(EMP_j)$ is the marginal predicted probability, which is the difference between the predicted probability of a change in employment between the treatment (those who contracted TB) and control groups (those who remained healthy). To interpret the results on changes in employment relative to baseline employment, we set the reference groups of the regressions in relation to their respective changes. The effect of a gain in employment is interpreted relative to the group that is unemployed in both periods, and the effect of a loss in employment is interpreted relative to the group that is employed in both periods.

We estimate the effect first with a specification of contracting TB on change in employment status, including only fixed effects on districts and year. Since we estimate the first-differenced effect across two periods, we also remove the effect of all time-invariant covariates on the analysis. In the second specification, to test the robustness of our estimates, we further include a list of covariates. If the CEM procedure is successful, there should be sufficient balance between the treatment and control groups. Thus, including the covariates should not change the estimate compared to the first specification.

To evaluate the effect of the infection of TB on employment across subgroups (*e.g.* age, gender, education and skills), we interact a dummy variable indicating the index of the subgroup with the disease state, to estimate the differential effect of the disease on employment in particular subgroups. We test the interaction effect of TB in various subgroups in the following specification.

$$\begin{aligned}\pi_j(EMP_j) &\equiv Pr(EMP_i = j | TB_i, STB_i)\\ &= \frac{exp(TB_i^T\beta_j + STB_i^T\beta_j)}{\sum_{k=1}^{J} exp(TB_i^T\beta_k + STB_i^T\beta_k)}\\ &= \frac{exp(TB_i^T\beta_j + STB_i^T\beta_j)}{1+\sum_{k=1}^{J-1} exp(TB_i^T\beta_k + STB_i^T\beta_k)}\end{aligned} \tag{2}$$

where S is a discrete variable that represents the subgroup category. We also assess the effect of symptoms, such as persistent cough or bloody cough, on the relationship between TB and employment. Similar to the specification for subgroups, we include a dummy variable that indicates whether the individual reports symptoms of persistent or bloody cough and interacts that variable with the treatment variable.

Our second set of outcome variables for analysis is changes in income and expenditure, which are continuous variables. We apply linear OLS regressions where the coefficient, β, represents the expected mean on the effect of contracting TB on income or expenditure. The income captures total individual income, which includes income from formal employment, informal employment, aids and grants. Since changes in income are related to changes in employment, we regress the income and expenditure outcomes with the interaction of TB and employment, according to the following specification

$$\begin{aligned}(Y_1 - Y_0) &= \beta_1 * (TB_1 - TB_0) + \beta_2 * (EMP_1 - EMP_0)\\ &+\beta_3 * [(TB_1 - TB_0) * (EMP_1 - EMP_0)]\\ &+(X_1 - X_0)'\gamma + (\epsilon_1 - \epsilon_0)\end{aligned} \tag{3}$$

where Y represents variables of interest (changes in income or expenditure). TB represents changes in TB illness status, EMP represents changes in employment status and X represents a set of covariates. For the interaction term, we consider specific combinations of changes in personal conditions, such as having TB and losing employment, finding employment or remaining with the same job status separately. We also include district and year fixed effects, which control, respectively, for district-level exogenous variables, such as local regulations or other characteristics that we do not measure, and for known technologies assumed to be changing over time.

4.3 | Propensity score regression adjustment estimator

We compare the results of our analyses with a propensity score regression adjustment estimator. The propensity score is denoted as $e(x) = Pr(z = 1|x)$, the score at which the conditional distribution of the control and treatment groups is balanced. In a nonrandomised setting, it is estimated from observed data using a logit model. It is the coarsest among a set of such balancing scores. At any value of the propensity score, the difference between means of treatment and control groups is an unbiased estimate of the average treatment effect if treatment assignment is ignorable (Rosenbaum & Rubin, 1983). The distribution of the propensity scores for the treatment and control groups is presented in Figure 1. We estimate the propensity score with the entire list of covariates used in the CEM procedure. The regression adjusted with propensity score is specified as follows:

$$\begin{aligned}(Y_1 - Y_0) = \; & \beta_1 * (TB_1 - TB_0) + \beta_2 * (EMP_1 - EMP_0) \\ & + \beta_3 * [(TB_1 - TB_0) * (EMP_1 - EMP_0)] \\ & + (X_1 - X_0)'\gamma + PS'\delta + (\epsilon_1 - \epsilon_0)\end{aligned} \tag{4}$$

where PS is the scalar propensity score.

The main difference between the CEM approach and propensity score adjustment is the number of observations that enter the regressions. With CEM preprocessing, we prune a large number of observations where there is no match between the treated and control groups. We retain about 6000 observations, while the propensity score regressions have a sample size of over 35,000 observations. We show the mean and standard deviation of income and expenditure for those that are included in the CEM and those that are eliminated in Figure 2. As can be seen from the figure, CEM preprocessing eliminated observations at the higher and lower ends of the distribution, leading to a narrower range as compared with the overall data-set.

5 | RESULTS

In this section, we present the results of our analysis, starting from the relationship between TB and employment outcomes for workers across different specifications and robustness checks, and then looking at financial outcomes in terms of total income and expenditure, both at the individual and at the household levels.

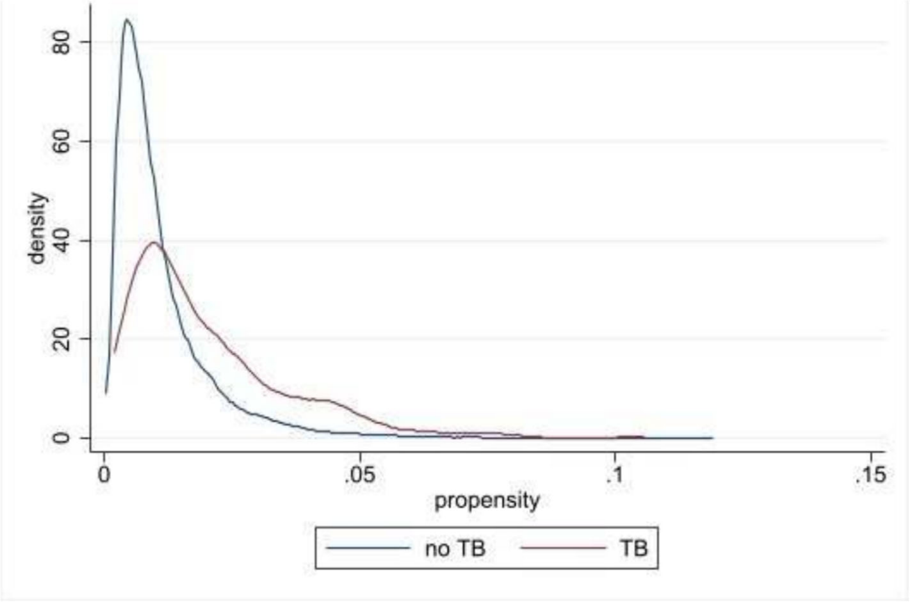

FIGURE 1 Density plot of propensity scores by treatment group. [Color figure can be viewed at wileyonlinelibrary.com]

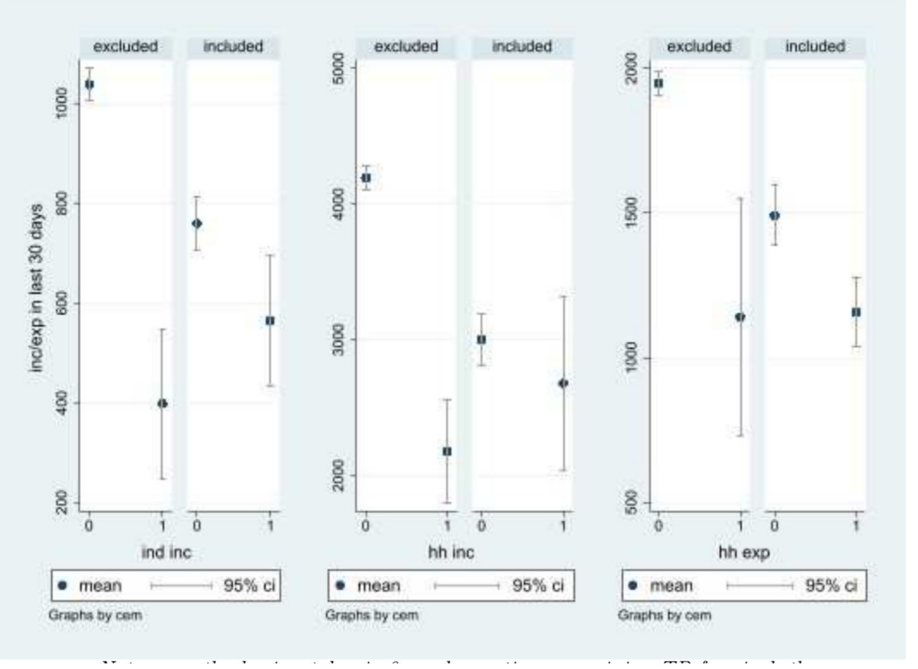

FIGURE 2 Income and expenditure for sample included or excluded in the coarsened exact matching (CEM) dataset. [Color figure can be viewed at wileyonlinelibrary.com]

5.1 | Employment

We first look at the number of individuals who enter and exit the formal labour force, depending on their TB status. The unconditional proportion of individuals who desire to work in the 4 weeks prior to the survey is 34% for those with TB and 30% for those without TB, and the mean difference between the groups is not statistically significant[7].

The results of the multinomial logit of a change in employment, following Equation (1), are outlined in Tables 6 (for exiting the formal labour market if the individual is already employed) and 7 (for entering the formal labour market if the individual is unemployed). In the first column, we present the predicted probability at the margin for contracting TB on a change in employment, controlling for district and year fixed effects. We also include weights that take into consideration the differences in the number of treatment and control observations within each stratum from CEM. In the second column, we include baseline covariates. If the CEM procedure is successful, then the estimates in columns 1 and 2 should not be different. The results indicate that TB does not lead to a statistically significant change in the probability of exiting formal employment if the individual is currently employed, but it leads to a 5% reduced probability of entering formal employment if one is not employed formally. An infection of TB seems to create a penalty in the formal job market, despite the fact that almost 90% of the individuals in our sample with TB declare to be treating the disease by taking medications for it. The disadvantage is specifically for entering the formal employment market, compared to those who already have a formal employment contract.

[7]Table A4 shows that when the individual is TB-free in both periods, 28% of the observations exit the formal employment market while 16% of the observations enter the formal employment market. When the individual has an infection of TB, 42% of the observations exit the formal employment market, while 10% of the observations enter the formal employment market.

TABLE 6 TB and exits from the formal employment market.

			Subgroups					
Employment	**All(1)**	**All(2)**	**Age**	**Gender**	**Education**	**Skills**	**Urban/rural**	**Symptoms**
All	0.02	0.02						
	(0.01)	(0.01)						
> = 35 years old			0.01					
			(0.02)					
< 35 years old			0.05					
			(0.03)					
Male				0.02				
				(0.02)				
Female				0.02				
				(0.02)				
> = Secondary					0.02			
					(0.02)			
< = Primary					0.03			
					(0.03)			
Skills						0.06		
						(0.04)		
No skills						0.01		
						(0.02)		
Urban							0.03	
							(0.02)	
Rural							0.02	
							(0.02)	
Symptoms								0.05
								(0.03)
No symptoms								0.01
								(0.02)
Covariate controls								
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
CEM weights	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
No. of obs.	5661	5661	5661	5661	5661	5661	5661	5661

Note: Marginal predicted probability of exiting formal employment is reported. Standard error is reported in the parenthesis. Skills refer to having a driver license or the ability to use a computer. Symptoms refer to the self-reported symptoms of persistent cough or bloody cough.
Abbreviation: TB, tuberculosis
*$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$.

5.1.1 | Informal employment

The effects identified above are all relative to the formal labour market, but it is important to consider also informal sector employment, including casual employment, self employment, agriculture work and other household work. We find that, for informal jobs, TB plays no significant role. This indicates that informal

TABLE 7 TB and entries in the formal employment market.

Employment	All(1)	All(2)	Subgroups Age	Gender	Education	Skills	Urban/rural	Symptoms
All	−0.05*	−0.05*						
	(0.02)	(0.02)						
> = 35 years old			−0.03					
			(0.02)					
< 35 years old			−0.06*					
			(0.03)					
Male				−0.05**				
				(0.02)				
Female				−0.03				
				(0.02)				
> = Secondary					−0.03			
					(0.02)			
< = Primary					−0.08***			
					(0.02)			
Skills						−0.03		
						(0.03)		
No skills						−0.05**		
						(0.02)		
Urban							−0.07**	
							(0.02)	
Rural							−0.02	
							(0.02)	
Symptoms								−0.07**
								(0.03)
No symptoms								−0.04*
								(0.02)
Covariate controls								
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
CEM weights	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
No. of obs.	5661	5661	5661	5661	5661	5661	5661	5661

Note: Marginal predicted probability of entering the formal employment market is reported. Standard error is reported in the parenthesis. Skills refer to having a driver license or the ability to use a computer. Symptoms refer to the self-reported symptoms of persistent cough or bloody cough.
Abbreviations: CEM, coarsened exact matching; TB, tuberculosis.
* $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

economic activities do not change with an infection of TB. When formal and informal sector employment are considered together, the predicted probability on the margin of being employed still remains negative, but becomes statistically insignificant. Table 8 compares the results on formal, informal and total employment. The significant negative coefficient on the likelihood of entering employment disappears when the outcome variable is informal employment only or total employment. Thus, overall, having a TB infection

TABLE 8 TB and formal, informal and total employment.

	Probability of exiting			Probability of entering		
Employment markets	**(1)**	**(2)**	**(3)**	**(1)**	**(2)**	**(3)**
Formal	0.02			−0.05*		
	(0.01)			(0.02)		
Informal		−0.01			0.00	
		(0.02)			(0.01)	
Both			0.03			−0.02
			(0.02)			(0.02)
Controls						
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	Yes	Yes	Yes	Yes	Yes	Yes
CEM weights	Yes	Yes	Yes	Yes	Yes	Yes
No. of obs.	5661	5625	5632	5661	5625	5632

Note: Specification (1) refers to the formal employment outcome. Specification (2) refers to the informal employment outcome. Specification (3) refers to the total employment outcome. Informal employment refers to casual employment, self-employment, agriculture work and other household help. For total employment, an individual is considered employed if he/she holds a formal employment, an informal employment, or both. The marginal predicted probability of the formal employment market is the same as those reported in Tables 6 and 7. Standard error is reported in the parenthesis.
Abbreviations: CEM, coarsened exact matching; TB, tuberculosis.
*$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$.

does not seem to have a significant link with starting or stopping any informal forms of employment[8]. Although a TB infection makes an individual less likely to enter the formal employment market, the individual continues their informal activities and since informal jobs are a significant share of the labour market in South Africa, when we look at total employment altogether, we do not see any significant effect of TB.

5.1.2 | Lower income households

Since TB disproportionately affects marginalised and vulnerable populations, we estimate the effect for those who are most likely to be affected by TB by limiting the initial sample to the lower half of the household income distribution and implementing the CEM procedure on this population only. Thus, we evaluate the effect of TB on employment specifically in poorer households. The median household income that separates the top half and bottom half of the population is ZAR 2100, and the average household income for the lower half is ZAR 1,139. Table 9 outlines the results. The employment market effect for poorer households is stronger than the total population. Those with TB infections are less likely to enter the formal employment market and the size of the effect is stronger compared to the overall population. Moreover, they are significantly more likely to exit the formal employment market.

Thus, having a recent TB infection disproportionately affects the poorer households in the formal job market. Workers from low income households are more vulnerable to illness and to adverse employment outcomes, possibly because on average they are less educated and can only access unskilled and less

[8]To unpack one possible mechanism why households are able to maintain stable expenditure when individual loss of employment occurs, we also evaluate whether any changes in the intensity of work (*i.e.* hours worked) compensates for some of the income loss due to employment loss. When a job loss in formal employment occurs, we find that there is a moderate shift to informal work, with around 4 extra hours per week for those without TB and 6 extra hours per week for those with TB (see Appendix Table A5 in the appendix). For both groups, this increase in informal work is significantly different from zero. However, it compensates only to a small extent for the loss of formal employment, whose average intensity is of 34–36 h per week).

TABLE 9 TB and formal employment in poorer households.

	Exiting employment		**Entering employment**	
Poorer households	**(1)**	**(2)**	**(1)**	**(2)**
	0.03	0.04**		
	(0.01)	(0.01)		
			−0.09**	−0.09**
			(0.03)	(0.03)
Controls				
Year FE	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes
Covariates	No	Yes	No	Yes
CEM weights	Yes	Yes	Yes	Yes
No. of obs.	2768	2768	2768	2768

Note: Poorer households are defined as households in the lower half of the household income distribution. In this analysis, we implement the CEM preprocessing and estimate the effects on the poorer households only. Specification (1) does not include baseline covariates in the regression. Specification (2) further includes baseline covariates in the regression. Standard error is reported in the parenthesis.
Abbreviation: CEM, coarsened exact matching; TB, tuberculosis.
*$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$.

"protected" formal jobs, and may therefore have less secure employment status with fewer contractual protections for sick leave. It is also possible that, among formal jobs, those held by poorer workers have shorter duration and therefore may not be renewed by the employers in the case of a severe illness.

5.1.3 | Subgroup analysis on employment

To evaluate the effect of TB on employment across different groups of individuals, we estimate the marginal predicted probability of employment changes with indicators to represent groups based on age, gender, education, skills, and urban/rural residence, interacted with the treatment variable (TB). Adding the interaction with these individual characteristics, the probability of exiting the formal employment market is still not significant in any subset (Table 6). The probability of entering the formal employment market appears significantly different by subgroups (Table 7). While all ages are significantly affected by the disease, the young (those < 35 years old) are more negatively affected than the older cohort. Males and those with less than secondary education levels and no skills are significantly negatively affected, as well as those living in urban areas. Some of the insignificant interaction effects could be due to sample size limitations, but we are reassured that for many categories, a TB infection is confirmed to be negatively associated with the chances of entering the formal employment market.

5.1.4 | Symptoms as a proxy for TB visibility and severity

To get some further insights into possible mechanisms behind our results, we test a visible signal of disease, which is the appearance of symptoms like persistent and bloody coughs. These symptoms can have the effect of making the TB condition more visible to third parties, such as employers, and identifying individuals with particularly severe illness, suffering physical impairments that are more substantial than those with no coughs. A higher proportion of individuals who contract TB report cough symptoms than those who do not have TB in both periods. An individual is more likely to enter the formal job market if s/he does not have TB (12% vs 8%). If the person contracts TB, then she/he is more likely to exit

TABLE 10 TB and income and expenditure.

	TB	**TB+Δ*emp* = 0**	**TB+Δ*emp* = −1**	**TB+Δ*emp* = 1**
Individual income	−194**	−135*	−1646***	1097***
	(65)	(53)	(143)	(112)
Household income	−361	69	−332	1221*
	(291)	(244)	(471)	(367)
Household expenditure	−149	−35	−6	519
	(105)	(145)	(260)	(406)
Covariate controls				
Year FE	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes
Covariates	Yes	Yes	Yes	Yes
CEM weights	Yes	Yes	Yes	Yes

Note: In the first column, TB represents the pooled effect of TB on income and expenditure. TBTB+Δ*emp* = 0 represents the effect of TB and no employment change on income and expenditure. Separating those who experience no change by remaining unemployed and by remaining in employment yields the same expenditure effects (results available upon request). TBTB+Δ*emp* = −1 represents the effect of TB and exit from employment and TBTB+Δ*emp* = 1 represent the effect of TB and gain of employment. The sample size of individual income is 5373, for household income is 5164 and for household expenditure is 4712.

Abbreviations: CEM, coarsened exact matching; TB, tuberculosis.

*$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$.

the formal job market (8% vs 13%)[9]. The final column of Tables 6 and 7 presents the estimated marginal predicted probability of cough symptoms interacting with disease state. We find that having symptoms does not lead to a significantly higher probability of individuals to exit the formal employment market. However, it leads to a significant negative effect on the probability of entering the formal employment market. The negative effect for TB individuals is also present for those without symptoms, although the probability estimate is smaller. So, for workers who have TB but do not display external symptoms or who experience a less severe form of the illness, the chances of entering the formal employment market are also significantly different from zero. It is not possible for us to disentangle whether these symptoms signal to employers specifically the TB status of the patient, or overall bad health[10]. Moreover, we cannot distinguish whether coughs give rise to explicit discrimination from employers in the hiring process or simply capture those individuals that suffer from more severe forms of TB that impair their productivity and thus suitability for the formal employment market. All these effects are probably at play for individuals with TB, so we should be cautious not to over-interpret these results as indicators of a definitive mechanism.

5.2 | Income and expenditure: Individuals and households

In Table 10, we present results on the analysis of the effect of TB on income and expenditure following Equation (3). Individual income represents total individual income over the last 30 days, which includes income from formal employment and any income reported from casual employment, self employment, agriculture and household work. We include the usual set of controls and fixed effects. In the first

[9]Summary statistics are available upon request.

[10]Since persistent coughs and bloody coughs are symptoms that are strongly associated with a disease like TB, we can use other information from the survey about other symptoms to give us an alternative proxy of poorer overall health. To try to disentangle the role of poor health from TB, we can broaden our definition of symptoms to include fever, chest pain, swelling ankles, and sudden weight loss. When we examine the interaction between TB and these broad symptoms, we do not find a significant marginal effect of having bad health symptoms, and we confirm a robust marginal effect of having TB over the average effect of symptoms similar to our other results (results available upon request).

specification (labelled 'TB(1)'), we do not specify the employment change that is associated with the disease. In the second, third, and fourth specifications, we interact TB with changes in employment status. A loss in employment in theory should represent a sudden drop in labour market income. In fact, we see in the third column of Table 10 that when TB is associated with an employment exit, there is a sizeable and significant individual income decline. We further test the effects of TB on household income and household expenditure. This captures the extent to which this individual loss translates into a loss for the entire household. Notably, other adjustments might take place among family members of working age, also due to possible aids and grants or other public benefits. These results are also presented in Table 10.

Household income is represented by the total income that the household received in the previous 30 days. This includes income from jobs, as well as any informal sector income, pensions and grants. When there is employment exit, the sign of the estimate indicates that there is also a decline in household income. However, this decline is not statistically significant due to the large standard error. When TB is associated with a positive change in employment status, we find also a significant positive effect in household income. A possible explanation of this result is that households rely on wages from multiple family members that have some flexibility in the amount of work they each perform, possibly also through informal jobs (as discussed before) or if they have access to alternative sources of income that are different from employment income, such as grants or other social benefits.

Household expenditure captures household-level spending in the past 4 weeks. Our results show that household expenditure is not significantly affected by TB disease state or disease-induced employment changes. This result indicates that households are able to smooth their consumption when they are faced with an infection of TB within the household, regardless of whether this disease state is associated with an employment change. Since household income is not significantly negatively affected by the infection of TB, households are able to continue with their normal consumption patterns even when the disease strikes a member of the household. Interestingly, when there is a positive change in employment (in other words, a member of the household enters the formal employment market), despite a significant increase in household income, the positive effect on household expenditure is not significant due to a large standard error.

5.3 | Propensity score regression adjustment estimator

To compare our results, we implement a propensity score regression adjustment estimator on the overall dataset. Note that with PS, we are implementing a much less restrictive definition of pairing of treated

TABLE 11 TB and formal employment with propensity score regression adjustment.

Formal employment	(1) Exit	(2) Entrance
	0.02	
	(0.01)	
		−0.05*
		(0.02)
Covariate controls		
Year FE	Yes	Yes
District FE	Yes	Yes
Covariates	Yes	Yes
No. of obs.	35,320	35,320

Note: Exit refers to the marginal probability of exiting the formal employment market if the individual is employed at baseline. Entrance refers to the marginal probability of entering the formal employment market if the individual is not formally employed at baseline.
Abbreviation: TB, tuberculosis.
*$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$.

TABLE 12 TB and income and expenditure using propensity score regression adjustment.

	TB	TB+$\Delta emp = 0$	TB+$\Delta emp = -1$	TB+$\Delta emp = 1$
Individual income	−159**	−83	−2164***	1773***
	(61)	(50)	(98)	(87)
Household income	−442	−96	−1181**	1072***
	(296)	(207)	(353)	(257)
Household expenditure	−189*	−146	−240	−9
	(82)	(95)	(164)	(136)
Covariate controls				
Year FE	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes
Covariates	Yes	Yes	Yes	Yes

Note: In the first column, TB represents the pooled effect of TB on income and expenditure. TBTB+$\Delta emp = 0$ represents the effect of TB and no employment change on income and expenditure. TBTB+$\Delta emp = -1$ represents the effect of TB and exit from employment, and TBTB+$\Delta emp = 1$ represents the effect of TB and gain of employment. The sample size of individual income is 34,826, for household income is 33,546 and for household expenditure is 31,259.

Abbreviation: TB, tuberculosis.

*$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$.

and control individuals; thus, we have a significantly larger sample of around 35,000 observations. Table 11 presents the estimation results for employment outcomes (see Equation (4)), and Table 12 presents results on income and expenditure. The overall results show that individuals with TB are not more likely to exit the formal employment market. This is similar to the finding from CEM preprocessed data. Individuals are 5% less likely to enter the formal employment market, which is also in line with the finding from CEM preprocessed data.

Similar to the result with CEM preprocessing, TB-associated employment changes in the formal market are associated with respective losses and gains in individual income. Household income is also significantly affected by employment losses or gains while household expenditure is not significantly affected. The point estimates of the effect are different in magnitude to the estimates after CEM preprocessing, sometimes to the order of three times.

6 | CONCLUSION

In this study, we evaluate the microeconomic role of TB infections for employment, income and expenditure at the individual and household level using representative individual-level panel data from South Africa's National Income Dynamic Study, collected from 2008 to 2017. Depending on the model, we find that contracting TB is systematically associated with around 5% decreased probability of entering the formal labour market. Our results complement those of the study by Nwosu (2018), as we confirm that the probability of overall employment, including both formal and informal jobs, is not significantly linked to the illness. Informal employment (including casual and self-employment, agriculture and other household work) is also not significantly changed, suggesting that although individuals have a lower probability of entering the formal employment market, they can continue with their informal activities throughout the health episode, leading to a null overall effect on employment. Exits from the formal employment market, also, are not significantly modified when contracting TB. Thus, we conclude that most of the impact of TB on labour market outcomes derives from the barrier to entering formal employment, and contracting the disease may be less harmful for those who already have a formal occupation.

Beyond changes in employment status, we find that individual income is significantly reduced in individuals with TB through the employment channel. Household expenditure, however, is not significantly affected, indicating some further adjustment capacity at the household level through informal jobs, kinship-based support and other means not impacted by individual income. Policy-makers designing support programmes for the population affected by TB could make use of these support networks to better reach TB patients but should also be particularly attentive to those who do not have employment and lack a strong family structure that could mitigate the adverse financial and labour impacts of the illness.

Several possible channels can lead to our results regarding the link between employment and TB. First, due to the productivity impacts of the disease itself, individuals may be weakened even if they are taking medications and thus might be less able to pursue and perform in the formal employment market. Moreover, the long duration of anti-TB treatment (typically at least 6 months) poses significant logistic challenges to TB patients undergoing DOT. While some studies and meta-analyses have found that DOT improves treatment outcomes (Zimmer et al., 2021), this approach has also been shown to cause issues and inconveniences in terms of healthcare access, stigmatisation, confidentiality, impact on work and financial hardship (Karumbi & Garner, 2015; Salehitali et al., 2019; Zimmer et al., 2021). Due to the rigidity of working time in the formal sector, the intention to comply with DOT may in fact lead to conflicts with workplace schedules and work separation. Furthermore, the additional demands of DOT may prevent individuals from effectively seeking employment in the formal sector. Even for patients that discontinue their participation in DOT, just a few weeks of adherence can have disruptive effects on their participation in formal employment, which could explain our observed link between TB and worse labour market outcomes.

Lack of proper and credible information on the infectiousness of the disease is another possible mechanism, as the disease is well-known to be infectious, while nuances about the nature of infectiousness are not straightforward. This could lead to suspicion towards those who have any signs that could suggest an infection of TB. Our analysis of the 'visibility' and 'severity' of the disease provides some results that indicate that these channels might be at play. Therefore, public information campaigns raising awareness on the modes of transmission of TB, and the effectiveness of drugs in reducing contagion could have the additional benefit of reducing stigma against workers displaying some symptoms.

In our analysis, we do not have any evidence of explicit discrimination occurring against sick workers in this context, but other studies have shown how segmented the South African labour market tends to be (*e.g.* Ravetti et al., 2019), so it is possible that health issues could introduce a further bias in hiring processes. Furthermore, although we control for individual effects through a first differencing between two periods, there may be other individual effects that we do not observe in the data. For example, if the individual has had a string of short-term jobs in the past, this may be not observable, but it could be a factor that influences the likelihood of the individual finding a job in the current period. Nonetheless, we cannot claim that the lack of access to formal jobs for TB infected workers is necessarily the outcome of discrimination: For a precautionary principle, employers may prefer avoiding giving a job to a (suspect) TB patient, given their inability to perfectly monitor adherence to treatment and thus risk of contagion for other employees. Further research is needed to understand better the interplay between the disease and the cultural and socio-economic mechanisms affecting the working population.

TB is a deadly infectious disease, and any program related to making TB patients better integrated in society must include precautions to protect the public from contracting this disease, including in the context of the workplace. The profile of reported TB infections based on our data suggests that awareness of the need for TB treatment is high, leading to close to 90% of individuals reporting that they are taking some medication. However, we cannot rule out some degree of misreporting, namely individuals having TB but reporting otherwise due to barriers to healthcare access and lack of diagnosis, stigma, or other causes. Moreover, we do not have sufficient data to fully understand whether the individual is fully or only partially compliant with medication. Our data suggest that workers with TB and taking medication are active in the economy, through some form of informal or formal employment activities. While the incidence of TB has declined over the course of the data time frame according to WHO statistics, our analysis raises new questions about the need to further understand and improve the measurement of TB, the coverage of diagnoses and compliance with medication and what levers are needed to support economic activities after contracting the disease.

A final caveat should be noted regarding our results on the relation between TB on labour market outcomes. Our analysis focuses on a short-medium term horizon but does not consider indirect effects on the labour market participation of individuals through modifications in human capital investments. For the HIV/AIDS pandemic, there is already substantial evidence that, lacking treatment, the negative correlation with schooling, especially for women, dramatically reduces human capital accumulation in sub-Saharan Africa (Chicoine et al., 2020). Further research is needed to understand the additional long-term effects to labour force participation from TB and its effects in informal labour markets.

The economic impacts of TB still remain heavily understudied, despite the dramatic effects of this disease in the world. With the COVID-19 pandemic, in high-burden settings like South Africa, deaths due to TB could increase up to 20% over the next 5 years, according to recent estimates (Hogan et al., 2020). Moreover, TB incidence is expected to worsen due to COVID-19's effects on GDP and undernutrition. The results of our analysis indicate that the pre-existing economic disruptions caused by TB could be amplified. Protecting workers' livelihood and ensuring appropriate diagnoses and treatments for all (especially for those searching for employment and without safety networks) should be seen not only as a public health priority but also as an economic one.

ACKNOWLEDGEMENTS

We thank the participants attending the Research Seminars at the Centre for International Environmental Studies (CIES) and the Graduate Institute Department of Economics, Carlos Riumallo Herl and participants at EuHEA Conference 2020, Martina Viarengo and Chiara Toniolo for helpful comments. We are also grateful to the data centre at National Income Dynamic Study and the European Centre for Medium Range Forecasting, who provided valuable assistance in data access. Helena Ting acknowledges funding from the MAVA Foundation - "The Human Niche Project". Chiara Ravetti acknowledges funding from Ministero dell'Universit'a e della Ricerca, Ministerial Decree no. 1062/2021 and the FSE REACT-EU - PON Ricerca e Innovazione 2014-2020. This manuscript reflects only the authors' views and opinions, neither the European Union nor the European Commission can be considered responsible for them.

ORCID

Chiara Ravetti https://orcid.org/0000-0002-8034-5265

REFERENCES

Alfers, L. & Rogan, M. (2015) Health risks and informal employment in South Africa: does formality protect health? *International Journal of Occupational and Environmental Health*, 21(3), 207–215.

Alipanah, N., Jarlsberg, L., Miller, C., Linh, N., Falzon, D., Jaramillo, E. & Nahid, P. (2018) Adherence interventions and outcomes of tuberculosis treatment: a systematic review and meta-analysis of trials and observational studies. *PLoS Medicine*, 15(7), e1002595.

Barter, D., Agboola, S., Murray, M. & Barnighausen, T. (2012) Tuberculosis and poverty: the contribution of patient costs in sub-Saharan Africa: a systematic review. *BMC Public Health*, 12(980). Available from: https://doi.org/10.1186/1471-2458-12-980

Bor, J., Tanser, F., Newell, M. & Barnighausen, T. (2012) In a study of a population cohort in South Africa, HIV patients on antiretrovirals had nearly full recovery of employment. *Health Affairs*, 31(7), 1459–1469.

Botha, E., DenBoon, S., Verver, S., Dunbar, R., Lawrence, K. & Bosman, M. (2008) Initial default from tuberculosis treatment: how often does it happen and what are the reasons? *International Journal of Tuberculosis and Lung Disease*, 12(7), 820–823.

Chicoine, L., Lyons, E. & Sahue, A. (2020) The impact of HIV/AIDS on human capital investment in sub-Saharan Africa: new evidence. (*DP No. 13609*): IZA.

Chimbindi, N., Barnighausen, T. & Newell, M. (2012) Time and money: the costs of utilizing HIV and TB treatment in rural KwaZulu-Natal. AIDS 2012, Washington, DC.

Classens, M., duToit, E., Dunbar, R., Lombard, C., Enarson, D. & Beyers, N. (2013) Tuberculosis patients in primary care do not start treatment. What role do health system delays play? *International Journal of Tuberculosis and Lung Disease*, 17(5), 603–607.

Dagume, M. & Gyekye, A. (2016) Determinants of youth unemployment in South Africa: evidence from the Vhembe district of Limpopo province. *Environmental Economics*, 7(4), 59–67.

Espi-Sanchis, G., Leibbrandt, M., & Ranchhod, V. (2022) Age, employment and labour force participation outcomes in COVID-era South Africa. *Development South Africa*, 39(5), 664–688.

Fitzwater, S., Caviedes, L., Gilman, R., Coronel, J., Lachira, D., Salazar, C., et al. (2010) Prolonged infectiousness of tuberculosis patients in a directly observed therapy short-course program with standardized therapy. *Clinical Infectious Diseases*, 51(4), 371–378.

Foster, N., Vassall, A., Cleary, S., Cunnama, L., Churchyard, G. & Sinanovic, E. (2015) The economic burden of TB diagnosis and treatment in South Africa. *Social Science and Medicine*, 130, 42–50.

French, D., Brink, J. & Barnighausen, T. (2019) Early HIV treatment and labour outcomes: a case study of mining workers in South Africa. *Health Economics*, 28(2), 204–218.

Garcia-Gomez, P., van Kippersluis, H., O'Donnell, O., & van Doorslaer, E. (2013) Long-term and spillover effects of health shocks on employment and income. *The Journal of Human Resources*, 48(4), 873–909.

Gertler, P. & Gruber, J. (2002) Insuring consumption against illness. *American Economic Review*, 92(1), 51–70.

Ho, D., Imai, K., King, G. & Stuart, E. (2007) Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis*, 15(3), 199–236.

Hogan, A., Jewell, B., Sherrard-Smith, E., Vesga, J., Watson, O., Whittaker, C., et al. (2020) Potential impact of the COVID-19 pandemic on HIV, tuberculosis and malaria in low-income and middle-income countries: a modelling study. *The Lancet Global Health*, 8(9), e1132–e1141.

Houben, R. & Dodd, P. (2016) The global burden of latent tuberculosis infection: a re-estimation using mathematical modelling. *PLOS Medicine*, 13(10).

Howell, E. & Heunis, J. (2018) Community-based directly observed treatment for TB patients to improve HIV services: a cross-sectional study in a South African province. *BMC Health Services Research*, 18(1), 1–10.

ILO (2016) *Key indicators of the labour market* (9th ed.).: International Labour Organization.

Iacus, S., King, G. & Porro, G. (2011) Multivariate matching methods that are monotonic imbalance bounding. *Journal of the American Statistical Association*, 106(493), 345–361.

Iacus, S., King, G. & Porro, G. (2012) Casual inference without balance checking: coarsened exact matching. *Political analysis*, 20(1), 1–24.

Imbens, G. (2000) The role of the propensity score in estimating dose-dependent functions. *Biometrika*, 87, 706–710.

Imbens, G. & Wooldridge, J. (2007) Lecture notes 11: discrete choice models. *NBER: What's new in econometrics*, 1-24.

Ingle, K. & Mlatsheni, C. (2017) The extent of churn in the South African youth labor market: evidence from NIDS 2008 - 2015. (*Working Paper No 201*): SALDRU.

Jones, A., Rice, N. & Zantomio, F. (2020) Acute health shocks and labour market outcomes: evidence from the post crash era. *Economics and Human Biology*, 36, 100811.

Karumbi, J. & Garner, P. (2015) Directly observed therapy for treating tuberculosis. *Cochrane Database of Systematic Reviews*, 2015(5): CD003343.

Kochar, A. (1995) Explaining household vulnerability to idiosyncratic income shocks. *American Economic Review*, 85(2), 159–164.

Lawana, N., Booysen, F., Tsegaye, A., Mingiri, F. & Hongoro, C. (2020) Lifestyle risk factors, non-communicable diseases and labour force participation in South Africa. *Development South Africa*, 37(3), 446–461.

Leibbrandt, M., Woolard, I. & de Villiers, L. (2009) Methodology: report on NIDS wave 1. (*Technical Paper No. 1*): NIDS.

Levinsohn, J., McLaren, Z., Shisana, O. & Zuma, K. (2013) HIV status and labour market participation in South Africa. *Review of Economics and Statistics*, 95(1), 98–108.

Lonnroth, K., Jaramillo, E., Williams, B. G., Dye, C. & Raviglione, M. (2009) Drivers of tuberculosis epidemics: the role of risk factors and social determinants. *Social Sciences and Medicine*, 68, 2240–2246.

Mathema, B., Andrews, J., Cohen, T., Borgdorff, M., Behr, M., Glynn, J., et al. (2017) Drivers of tuberculosis transmission. *Journal of Infectious Diseases*, 216(Suppl6), S644–S653.

McFadden, D. (1973) Conditional logit analysis of qualitative choice behavior. In: *Frontiers in Econometrics*, Edited by Zarembka, P., Academic Press,105–135.

Meghji, J., Gregorius, S., Madan, J., Chitimbe, F., Thomson, R., Rylance, J., et al. (2021) The long term effect of pulmonary tuberculosis on income and employment in a low income, urban setting. *Thorax*, 76, 387–395.

Mlatsheni, C. (2012) *The challenges unemployment imposes on youth*. In: Shaping the Future of South Africa's Youth: Rethinking Post-School Education and Skills Training, Edited by. Perold, H., Cloete, N. and Papier, J. African Mind, 31-41.

Mohanan, M. (2013) Causal effects of health shocks on consumption and debt: quasi-experimental evidence from bus accident injuries. *Review of Economics and Statistics*, 95(2), 673–681.

Mojola, S. A., Ice, E., Schatz, E., Angotti, N., Houle, B. & Gómez-Olivé, F. X. (2022) The meaning of health in rural south africa: gender, the life course, and the socioepidemiological context. *Population and Development Review*, 48(4), 1061–1095. https://onlinelibrary.wiley.com/doi/pdf/10.1111/padr.12494

NIDS (2018a). National income dynamic study (NIDS) panel user manual version 1.

NIDS (2018b). *National income dynamic study (NIDS) wave 1, 2008 [dataset]. Version 7.0.0*: Southern Africa Labour and Development Research Unit. Pretoria: SA Presidency [funding agency]. Cape Town: Southern Africa Labour and Development Research Unit [implementer]. Cape Town: DataFirst [distributor].

NIDS (2018c). *National income dynamic study (NIDS) wave 2, 2010-2011 [dataset]. Version 4.0.0*: Southern Africa Labour and Development Research Unit. Pretoria: SA Presidency [funding agency]. Cape Town: Southern Africa Labour and Development Research Unit [implementer]. Cape Town: DataFirst [distributor].

NIDS (2018d). *National income dynamic study (NIDS) wave 3, 2012 [dataset]. Version 3.0.0*: Southern Africa Labour and Development Research Unit. Pretoria: SA Presidency [funding agency]. Cape Town: Southern Africa Labour and Development Research Unit [implementer]. Cape Town: DataFirst [distributor].

NIDS (2018e). *National income dynamic study (NIDS) wave 4, 2014-2015 [dataset]. Version 2.0.0*: Southern Africa Labour and Development Research Unit. Pretoria: SA Presidency [funding agency]. Cape Town: Southern Africa Labour and Development Research Unit [implementer]. Cape Town: DataFirst [distributor].

NIDS (2018f). *National income dynamic study (NIDS) wave 5, 2017 [dataset]. Version 1.0.0*: Southern Africa Labour and Development Research Unit. Pretoria: SA Presidency [funding agency]. Cape Town: Southern Africa Labour and Development Research Unit [implementer]. Cape Town: DataFirst [distributor].

Naidoo, P., Theron, G., Rangaka, M., Chihota, V., Vaughan, L., Brey, Z. & Pillay, Y. (2017) The South Africa tuberculosis care cascade: estimated losses and methodological challenges. *Journal of Infectious Disease*, 216(S7), S702–S713.

Ngwatu, B., Nsengiyumva, N., Oxlade, O., Mappin-Kasirer, B., Nguyen, N., Jaramillo, E., et al. (2018) The impact of digital health technologies on tuberculosis treatment: a systematic review. *European Respiratory Journal*, 51(1): 1701596.

Nwosu, C. (2018) The relationship between employment and mental and physical health in South Africa. *Development South Africa*, 35(2), 145–162.

Nwosu, C. & Woolard, I. (2017) The impact of health on labour force participation in South Africa. *South African Journal of Economics*, 85(4), 481–490.

Petersen, E., Khamis, F., Migliori, G., Bay, J., Marais, B., Wejse, C. & Zumla, A. (2016) De-isolation of patients with pulmonary tuberculosis after start of treatment - clear, unequivocal guidelines are missing. *International Journal of Infectious Diseases*, 56, 34–38.

Ravetti, C., Sarr, M., Munene, D. & Swanson, T. (2019) Discrimination and favouritism among South African workers: ethics, identity and union membership. *World Development*, (Vol. 123, No. 104604).

Ripollone, J., Huybrechts, K., Rothman, K., Ferguson, R. & Franklin, J. (2020) Evaluating the utility of coarsened exact matching for pharmacoepidemiology using real and simulated claims data. *American Journal of Epidemiology*, 189(6), 613–622.

Rosenbaum, P. & Rubin, D. (1983) The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41–55.

Rupani, M., Cattamanchi, A., Shete, P., Vollmer, W., Basu, S. & Dave, J. (2020) Costs incurred by patients with drug-susceptible pulmonary tuberculosis in semi-urban and rural settings of western India. *Infectious Diseases of Poverty*, 9(1):144.

Russell, S. (2004) The economic burden of illness for households in developing countries: a review of studies focusing on malaria, tuberculosis and human immunodeficiency virus/acquired immunodeficiency syndrome. *American Journal of Tropical Medicine and Hygiene*, 71(S2), 147–155.

Salehitali, S., Noorian, K., Hafizi, M. & Dehkordi, A. H. (2019) Quality of life and its effective factors in tuberculosis patients receiving directly observed treatment short-course (dots). *Journal of Clinical Tuberculosis and Other Mycobacterial Diseases*, 15, 100093. https://www.sciencedirect.com/science/article/pii/S2405579418300810

Schluter, J., Sorensen, L., Bossert, A., Kersting, M., Staab, W. & Wacker, B. (2021) Anticipating the impact of Covid-19 and comorbidities on the South African healthcare system by agent-based simulations. *Scientific Reports*, 11(7901).

Skinner, D. & Classens, M. (2016) It's complicated: why do tuberculosis patients not initiate or stay adherent to treatment? A qualitative study from South Africa. *BMC Infectious Diseases*, 16, 712.

Statistics South Africa (2023) QLFS Trends 2008 - 2023Q1. *Statistics South Africa*, 1-25.

Stuart, E. (2010) Matching methods for causal inference: a review and a look forward. *Statistical Science*, 25(1), 21.

Thirumurthy, H. & Zivin, J. (2012) Health and labor supply in the context of HIV/AIDS: the long-run economic impacts on antiretroviral therapy. *Economic Development and Cultural Change*, 61(1), 73–96.

Townsend, R. M. (1994) Risk and insurance in village India. *Econometrica*, 62(3), 539–591.

Turner, R., Chiu, C., Churchyard, G., Esmail, H., Lewinsohn, D., Gandhi, N. & Fennelly, K. (2017) Tuberculosis infectiousness and host susceptibility. *The Journal of Infectious Diseases*, 216(S6), S636–S643.

WHO (2010) Guidelines for treatment of tuberculosis. *WHO, Fourth Edition*, 1-160.

WHO (2022) Globals Tuberculosis Report 2022. *World Health Organization*, 1-68.

WHO (2023) Global Tuberculosis Report 2023. *World Health Organization*; 1-75.

Wagstaff, A. (2007) The economic consequences of health shocks: evidence from Vietnam. *Journal of Health Economics*, 26(1), 82–100.

Weimann, A., Dai, D. & Oni, T. (2016) A cross-sectional and spatial analysis of the prevalence of multimorbidity and its association with socioeconomic disadvantage in South Africa: a comparison between 2008 and 2012. *Social Science and Medicine*, 163, 144–156.

Yates, T., Khan, P., Knight, G., Taylor, J., McHugh, T., Lipman, M., et al. (2016) The transition of mycobacterium tuberculosis in high burden settings. *The Lancet Infectious Diseases*, 16(2), 227–238.

Yoeli, E., Rathauser, J., Bhanot, S., Kimenye, M., Masini, E., Owiti, P. & Rand, D. (2019) Digital health support in treatment for tuberculosis. *New England Journal of Medicine*, 381, 986–987.

Zimmer, A. J., Heitkamp, P., Malar, J., Dantas, C., O'Brien, K., Pandita, A., & Waite, R. C. (2021) Facility-based directly observed therapy (dot) for tuberculosis during Covid-19: a community perspective. *Journal of Clinical Tuberculosis and Other Mycobacterial Diseases*, 24, 100248. https://www.sciencedirect.com/science/article/pii/S2405579421000371

Zwane, T., Greyling, L. & Maleka, M. (2016) The determinants of household savings in South Africa: a panel data approach. *International Business and Economics Research Journal*, 15(4), 209–218.

How to cite this article: Ting, H., Bozzola, M. & Ravetti, C. (2024) Tuberculosis and labour market participation: Evidence from South Africa. *South African Journal of Economics*, 92(4), 444–474. Available from: https://doi.org/10.1111/saje.12385

APPENDIX: Variables and descriptive data

TABLE A1 Definition of variables.

Variable	Definition
Outcome variables	
Employment	Employment status of the individual. 0 = unemployed; 1 = employed
Household income	Household income of the individual for the last 30 days.
Household expenditure	Household expenditure of the individual for the last 30 days
Individual income	Individual total income for the last 30 days
Health shock	
TB	TB status of the individual. 0 = no TB; 1 = currently sick and/or taking medication
Demographic control variables	
Age	Age of the individual in years
Gender	Gender of the individual. 1 = male; 2 = female
Race	Race of the individual. 1 = African; 2 = colored; 3 = others
Education	Education level of the individual. 0 = no schooling; 1 = primary; 2 = lower secondary; 3 = upper secondary and above
Mother's education	Education level of mother, and if not available, father. 0 = no schooling; 1 = primary; 2 = lower secondary; 3 = upper secondary and above
Father's education	Education level of father. 0 = no schooling; 1 = primary; 2 = lower secondary; 3 = upper secondary and above
Computer literacy	Computer literacy of the individual. 0 = no; 1 = yes
Driver license	Person has a driver license. 0 = no; 1 = yes
Health status	Self-reported health status. 1 = excellent; 5 = poor
Hypertension	Whether the individual has hypertension. 0 = no; 1 = yes
Diabetes	Whether the individual has diabetes. 0 = no; 1 = yes
Stroke	Whether the individual has had a stroke. 0 = no; 1 = yes
Heart disease	Whether the individual has heart disease. 0 = no; 1 = yes
Asthma	Whether the individual has asthma. 0 = no; 1 = yes
Cancer	Whether the individual has cancer. 0 = no; 1 = yes
Smoke	Whether the individual ever smoked cigarettes. 1 = yes; 2 = no
Alcohol	Alcohol consumption level of the individual. 1 = more than 3 per week; 2 = less than 3 per week
Medical aid	Whether the individual is covered by medical aid. 0 = no; 1 = yes
Pension	Whether the individual receives a pension. 0 = no; 1 = yes
Bank account	Whether the individual has a bank account. 0 = no; 1 = yes
Life satisfaction	Self-rated life satifaction. 1 = very dissatisfied; 5 = very satisfied
Household size	Size of the household.
Urban/rural	Whether the household lives in urban or rural area. 1 = urban; 2 = traditional; 3 = farms
Dwelling type	Type of dwelling of the household. 1 = house; 2 = huts; 3 = other
Owned dwelling	Whether the household owns the dwelling. 0 = no; 1 = yes
Land grant	Whether the household received land grant. 0 = no; 1 = yes
Subsidy	Whether the household received subsidy. 0 = no; 1 = yes
Water source	Household water source. 1 = tap; 2 = tank; 3 = flow; 4 = other
Toilet	Type of toilet for the household. 1 = toilet; 2 = pit; 3 = other
Electricity	Whether the household has electricity. 0 = no; 1 = yes

TABLE A1 (Continued)

Variable	Definition
Cooking	Energy source for cooking. 1 = electricity; 2 = others
Heating	Energy source for heating. 1 = electricity; 2 = others
Lighting	Energy source for lighting. 1 = electricity; 2 = others
Mobile	Whether the household has a mobile. 0 = no; 1 = yes
Refuse	Whether the household has refuse collection. 0 = no; 1 = yes
Street light	Whether the household has street lights. 0= no; 1= yes

Selection of confounders for CEM procedure

TABLE A2 Logistic regression of employment and TB on potential confounders.

	TB		*Employment*	
	b	**Std. err.**	***b***	**Std. err.**
Parents education	−0.201***	0.044	−0.166***	0.011
Education	−0.079	0.046	0.170***	0.017
Computer literacy	−0.235*	0.117	0.180***	0.027
Driver's License	−0.571***	0.159	0.322***	0.032
Self-rated health	0.731***	0.035	−0.122***	0.012
Hypertension	−0.691***	0.117	−0.222***	0.038
Diabetes	−0.417*	0.213	−0.555***	0.071
Stroke	−0.263	0.323	−0.633***	0.158
Asthma	0.322*	0.152	−0.163*	0.067
Heart disease	0.119	0.200	−0.193*	0.089
Cancer	0.215	0.374	−0.624***	0.148
Smoke	0.349***	0.095	0.181***	0.029
Receives medical aid	−0.058	0.197	0.974***	0.040
Self-rated life satisf.	−0.098**	0.033	0.036***	0.010
Pension	−0.768	0.398	1.572***	0.075
Bank account	−0.034	0.084	1.582***	0.025
Age	0.019***	0.004	0.037***	0.001
Gender	−0.257**	0.087	−0.342***	0.025
Race	−0.087	0.230	−0.785***	0.051
Household size	−0.005	0.012	−0.042***	0.004
Dwelling type	0.091	0.048	0.088***	0.014
Urban/Rural	−0.057	0.083	0.658***	0.025
Owned living	0.074	0.099	−0.564***	0.026
Land grant	0.278	0.181	−0.018	0.061
Govt subsidy	−0.071	0.102	0.015	0.030
Water source	0.026	0.065	−0.073**	0.025
Toilet type	0.003	0.084	0.176***	0.028
Electricity	0.003	0.175	0.057	0.053
Cooking energy	−0.027	0.135	−0.312***	0.046

(Continues)

TABLE A2 (Continued)

	TB		Employment	
	b	Std. err.	b	Std. err.
Heating energy	−0.161	0.101	0.078*	0.031
Lighting energy	0.040	0.203	0.178**	0.067
HH has mobile	−0.301**	0.090	0.008	0.033
Refuse collection	0.142	0.119	0.015	0.034
Street light	0.011	0.101	−0.003	0.028
Year FE	Yes		Yes	
District FE	Yes		Yes	
Languages FE	Yes		Yes	

Note: Confounders are considered as variables that potentially affect both labour market behaviour and the risk of having a recent TB infection vs not having an infection (Stuart, 2010; Jones et al., 2020). In this table, we show the coefficients and standard errors of the logistic regressions of TB and employment (controlling for TB) on a list of potential confounders. We include the bold-faced variables in the CEM procedure which includes confounders that affects TB and employment. Some variables that have a strong correlation with another variable already included in the CEM (*e.g.* self-declared health status with health conditions like hypertension and heart diabetes) are excluded to avoid redundancy.
Abbreviations: CEM, coarsened exact matching; TB, tuberculosis.
$^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$.

Prematching and postmatching balance of all covariates

TABLE A3 Prematching and postmatching balance of covariates.

	Prematching			Postmatching		
	Mean TB	Diff	p	Mean	Diff	p
Gender	1.56	0.04	0.10	1.56	0.01	0.74
Race	1.16	0.03	0.12	1.16	0.01	0.74
Parents education	0.59	0.50	0.00***	0.60	−0.03	0.52
Education	1.81	0.50	0.00***	1.88	0.00	0.96
Computer literacy	0.17	0.15	0.00***	0.16	0.02	0.32
Driver's license	0.07	0.06	0.00***	0.07	0.01	0.55
Self-rated health	2.10	2.58	0.00***	2.56	0.03	0.54
Hypertension	0.13	−0.03	0.05	0.14	−0.00	0.88
Diabetes	0.03	−0.00	0.83	0.02	0.00	0.86
Stroke	0.01	−0.01	0.21	0.01	−0.00	0.70
Heart disease	0.02	−0.00	0.90	0.02	0.00	0.98
Cancer	0.00	0.00	0.40	0.00	−0.00	0.92
Smoke	0.32	−0.12	0.00***	0.33	−0.04	0.11
Alcohol[a]	1.96	0.02	0.07	1.96	0.00	0.87
Receives medical aid	0.03	0.07	0.00***	0.03	0.00	0.69
Self-rated life satisf.	2.62	0.23	0.00***	2.61	0.02	0.67
Pension	0.01	0.02	0.03*	0.01	0.00	0.71
Bank account	0.32	0.09	0.00***	0.34	0.03	0.31
Age	37.4	−5.18	0.00***	37.4	−0.29	0.66
Household size	5.41	0.07	0.66	5.20	−0.02	0.93
Dwelling type	1.63	−0.09	0.01**	1.59	0.05	0.27

TABLE A3 (Continued)

	Prematching			Postmatching		
	Mean TB	**Diff**	***p***	**Mean**	**Diff**	***p***
Urban/rural	1.69	−0.04	0.21	1.50	0.01	0.69
Owned living	0.78	0.01	0.66	0.78	−0.01	0.56
Water source	1.28	−0.06	0.02*	1.35	0.02	0.69
Toilet type	1.58	−0.06	0.03*	1.55	−0.02	0.57
Cooking energy	1.29	−0.04	0.05	1.28	−0.01	0.85
Heating energy	1.45	−0.07	0.00**	1.44	0.01	0.82
Lighting energy	1.22	−0.05	0.00**	1.20	0.01	0.68

Note: Results from two sample t-tests are presented prior to the CEM procedure and after the CEM procedure.
Abbreviations: CEM, coarsened exact matching; TB, tuberculosis.
[a]Alcohol refers to data from waves 1 to 4 only since this question was not asked in wave 5.
$^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$.

Descriptive statistics

TABLE A4 Number of individuals who enters or exits the formal employment market.

Employment at baseline	Exits market	No change	Enters market
TB-free in both periods ($n = 5268$)			
Employed	430	1,116	
	(28%)	(72%)	
Not employed		3,141	581
		(84%)	(16%)
TB infection in period 1 ($n = 393$)			
Employed	42	59	
	(42%)	(58%)	
Not employed		263	29
		(90%)	(10%)

TABLE A5 Changes in working hours after exiting the formal employment market.

Weekly working hours	No TB	Has TB
Formal hours		
t_0	36.36	33.93
$\Delta_{t_1-t_0}$	−36.36	−33.93
Informal hours		
t_0	1.20	1.05
$\Delta_{t_1-t_0}$	4.01	6.44
Total hours		
t_0	37.58	34.98
$\Delta_{t_1-t_0}$	−32.36	−27.49

Note: t_0 refers to the number of hours worked at baseline. $\Delta_{t_1-t_0}$ refers to the change in the number of hours worked after the loss of formal employment. The increases in hours worked in informal employment (means of 4.01 for the *No TB* group and 6.44 for the *Has TB* group) are significantly different from zero (*p<0.001* in both cases). However, the difference between TB groups (4.01 h vs 6.44 h) is not statistically significant.