NBER WORKING PAPER SERIES

IS INCOME INEQUALITY LINKED TO INFECTIOUS DISEASE PREVALENCE? A HYPOTHESIS-GENERATING ECOLOGICAL STUDY USING TUBERCULOSIS

Jay Bhattacharya
Joydeep Bhattacharya
Minkyong Kim

Working Paper 31053
http://www.nber.org/papers/w31053

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2023

We have no funding sources or relevant financial conflicts of interest to report. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Is Income Inequality Linked to Infectious Disease Prevalence? A Hypothesis-Generating Ecological Study Using Tuberculosis

Jay Bhattacharya, Joydeep Bhattacharya, and Minkyong Kim

NBER Working Paper No. 31053
March 2023
JEL No. I1,I14,O57

ABSTRACT

We study the association between infectious disease prevalence and income inequality. We hypothesize that random social mixing in an income-unequal society brings into contact a) susceptible and infected poor and b) the infected-poor and the susceptible-rich, raising infectious disease prevalence. We investigate this association by examining whether countries with elevated levels of income inequality have higher rates of pulmonary Tuberculosis (TB) incidence per capita. We analyzed publicly available country-level panel data for a large cross-section of countries between 1995 and 2013. A "negative control" using anemia (a non-communicable disease, and hence impervious to the hypothesized mechanism) is also applied. We find that elevated levels of income inequality were positively associated with tuberculosis prevalence. All else equal, countries with income-Gini coefficients 10% apart are a statistically significant 4% different in tuberculosis incidence. Income inequality had a null effect on anemia, the negative control. Our cross-country regression results suggest that income inequality may create conditions where TB spreads more easily.

Jay Bhattacharya
117 Encina Commons
CHP/PCOR
Stanford University
Stanford, CA 94305-6019
and NBER
jay@stanford.edu

Joydeep Bhattacharya
Department of Economics
Iowa State University
Ames, IA 50011-1070
joydeep@iastate.edu

Minkyong Kim
Trinity University
mkkim0802@gmail.com

1. Introduction

Consider two areas, X and Y, both poor (with similarly low per capita incomes), but with income inequality substantially higher in X than in Y. Suppose individual A is poor and B is rich, and both reside in X. It is reasonable to expect B to be healthier than A because B is richer.[iv] Specifically, suppose people like A are significantly more prone to catching, harboring, and transmitting airborne infectious diseases, such as tuberculosis (TB). Finally, suppose people like A and B randomly mix and linger in social settings of area X (such as schools, workplaces, public transportation, etc.). We posit infections are more likely to spread or linger in area X than in Y. We label this the "mixing hypothesis." [v]

Two mixing models underly the hypothesis are at play and challenging to separate.

- First, in a poor community (low per capita income) with *low*-income inequality, most are impoverished and likely to be infected and infectious. In such settings, income-*assortative* mixing among the poor means the susceptible poor are very likely to come in touch with other (infected) poor and get infected. Since the per capita incomes in X and Y are similar, poor residents of X (such as A) are as likely to get infected as their counterparts in Y.
- Second, in a poor community with *high*-income inequality, the affluent are more likely to come in contact with the (infected and infectious) poor under income-*disassortative* mixing. Consequently, richer individuals in such communities would be at a higher risk of infection than similar individuals elsewhere: richer people like B are more likely to be infected in X than in Y.

This simple reasoning implies that the rich are an *added contributor* to overall disease prevalence in X than in Y. This suggests that overall infection levels will be higher in high-income-inequality areas compared to comparable low-income-inequality areas with *similar* per capita incomes. This paper documents cross-country

[iv] There is a well-studied and documented direct connection between income and health -- high-income people and countries are likely to be healthier. The positive slope is intuitive: after all, low-income people often face significant barriers to medical care access[47,48], are more likely to smoke[49,50], abuse drugs[51,52], be obese[53], and face chronic stressors[53,54] like financial hardship. These factors, in concert, can reduce immunity, increase vulnerability to disease, and cause poor health.

[v] A related literature explores the presence of a health-income *inequality* gradient[55–58], a more controversial hypothesis[59]. One set of explanations assert that A may have poorer health if they feel economically disadvantaged relative to B in a reference group, precipitating stressful social comparisons[60]. Ambient income inequality may also influence the health of both A and B if affluent people like B support disinvestment in public health or education, or social capital[61,62]. These explanations are usually indirect and rely on interpersonal, group-identity comparisons, or indirect "general equilibrium" considerations.

evidence in support of this hypothesis. It finds a robust positive association between TB prevalence and income inequality across countries after controlling for per capita income.

This empirical association in cross-country data is consistent with the mixing hypothesis and supports further analyses using individual-level data to test it. We discuss several threats to our interpretation and partially defend against them using a "negative control" check. Consider one such threat. Suppose in place of social mixing, another distinct channel (such as low political support for public health from the affluent) drives the association between TB and inequality. It stands to reason that such a channel would also lead to an association between anemia, a *non*-communicable disease of poverty, and inequality. We find little evidence of that, which makes the political support hypothesis less likely. Of course, there may be other phenomena we cannot check using cross-country data. As discussed below, we do not claim to have shown that income inequality is an *independent* contributor to TB prevalence.

Our choice of infectious disease, tuberculosis, is motivated by previous work on the worldwide prevalence of the disease, its pathology, modes of transmission, and related factors. Pulmonary Tuberculosis (TB) is an infectious disease caused by *Mycobacterium tuberculosis*, with an estimated 10 million new cases and 1.5 million deaths annually. Its incidence varies widely and is over 500 per 100,000 in many low- and middle-income countries. Tuberculosis is distinct among infectious diseases because it is transmitted almost exclusively through the air.[vi] Moreover, the transmission does not require close contact and may occur by 'sharing air' in insufficiently ventilated indoor spaces [1,2]. At the population level, Janssens and Rieder (2008) document that doubling GDP per capita is associated with a 38.5% decrease in TB incidence across countries[3]. Oxlade and Murray (2012) and Andrews et al. (2015) present evidence from India that the lowest wealth quintiles have a five-fold higher incidence of TB than the wealthiest quintile[4,5].

Our mixing hypothesis relies on two links. First, low-income people are more likely to catch, harbor, and transmit TB onward. Second, there is considerable social mixing and extended contact between low-and high-income people in a high-income-inequality area.

Regarding the first link: At the micro level, it is undisputed that the vast majority of the less affluent live and work in poorly ventilated and overcrowded conditions ideal for spreading TB bacteria. Many suffer from malnutrition and disease – particularly HIV – which reduces resistance to TB. And they have limited access

[vi] A necessary step in TB transmission is "the expulsion of bacilli from the lungs of infected individuals and their release into the surrounding air in a form that can remain viable during aerial transit." Coughing is the primary transmission channel, but breathing and talking are also important [63].

to healthcare. Given how infectious the disease is, just one person with untreated TB can pass the illness on to 10-15 other people annually. There is a strong association between income/poverty and TB, as documented by numerous WHO reports; country-specific studies include Brazil [6], Portugal [7], and Uganda [8].

Incomplete adherence to tuberculosis (TB) treatment increases the risk of continued transmission in the community[9]. At the neighborhood level, there is substantial evidence that low socioeconomic status areas in developing countries such as India and Brazil have higher TB rates. Low socioeconomic individuals have a higher risk of treatment abandonment[10,11]. Some studies find a similar pattern in the US[12]. Non-compliance with treatment regimens, once detected with active TB, is also related to household incomes. Researchers in various country settings report that TB patients with higher household incomes were less likely to report non-compliance with standard treatment regimens. Also, patients who felt that the economic burden related to TB treatment was high were more likely to report non-compliance[13].

Regarding the second link: for our mixing hypothesis to have credence, a substantial fraction of transmission must happen in community settings where people of different income strata repeatedly congregate. While routes of transmission differ in high vs. low-incidence countries, the literature suggests that about 8-19% of transmission happens within households, while the rest occurs in the community, such as in workplaces and schools.[2,14] Community transmission of tuberculosis depends on the burden of prevalent tuberculosis; how individuals live, work, and interact; and the quality and capacity of public health systems[15].

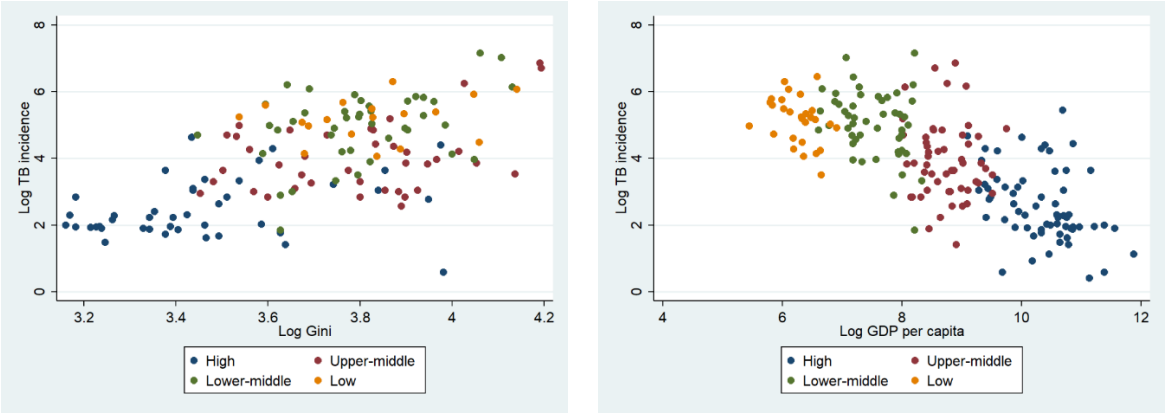
Social-mixing patterns depend on age and demographic structure, cultural behaviors, population density, and migration patterns[16, vii]. There is evidence of widespread sex-assortative social mixing and high rates of workplace contact for men[17]. Closer proximity and longer duration of contact (such as in workplaces, churches, bars, in-home help, crowded cafeterias, and cinema halls) between infected and susceptible individuals increase the risk of onward transmission[15]. A study by Classen and colleagues find that in high TB incidence areas, contact outside the household (during recreation, drinking in groups) may be more critical for the transmission of TB than household contacts[18].

A key source of community transmission is the presence of urban slums in many of the populous cities in the developing world. For example, the prevalence of TB is significantly higher in the slums than in the non-slum

[vii] A study in Hong Kong found participants with more years of education and higher income levels were more likely to report *more* social contacts, "showing a tendency for individuals with more years of education and higher income levels to spend more time on social and leisure activities, while the individuals with less education and lower income levels spent more time on household commitments."[64]

areas of Mumbai, Chennai, and Kolkata.[19,20,viii] Many residents of urban slums work in non-slum areas as maids, carters, babysitters, masons, hawkers, drivers, cleaners, secretaries, or clerks in supermarkets[21], where repeated and prolonged contact across income strata is commonplace.[ix] For our current purposes, the upshot is that income-disassortative social mixing is an essential determinant of TB transmission. In many mixing events or spaces listed above, the infected and susceptible individuals are not segregated by income, even if they are segregated by housing and neighborhoods[22,x].

Figure 1a is a plot of the logarithm of tuberculosis (TB) incidence (new cases per 100,000 people in a year) versus the logarithm of the Gini coefficient of income for ninety-seven countries in 2010. Figure 1b plots the logarithm of tuberculosis (TB) incidence versus the logarithm of the per capita income across the same countries and year.



(a) Log TB incidence versus Log Gini **(b) Log TB incidence versus Log GDP per capita**

Figure 1: Tuberculosis incidence

A quick look at the figures indicates a clear, positive association in Figure 1a and a negative association in Figure 1b.

[viii] Not just TB: In the case of Covid, Sahasranaman and Jensen studied slum and non-slum areas in six major cities[65]. They report that neighborhoods with slums are found to contain the highest density of cases across all cities under consideration, revealing that slums "constitute the most at-risk urban locations in this epidemic."

[ix] Rains and colleagues provide satellite imagery evidence to document that slum and non-slum areas are often found right next to each other in Indian cities[66]. Indeed, the affluent often go to specific slums looking for workers. In Delhi and Mumbai, the biggest slums actively participate in the handicraft industry, food-processing business, and the leather works business.

[x] The evidence cited above is suggestive but not conclusive: there does not appear to be any direct, micro-level, income-salient contact data on such mixing, a point we will return to later[5].

While this first look at the data is provocative, many extraneous factors apart from inequality might impact the correlation. We address this problem in three ways. First, we adopt a longitudinal approach in our data analysis, focusing on the relationship between changes in inequality and disease prevalence *within* countries rather than cross-country comparisons. Second, we adopt a statistical method that adjusts for measurable variables that may explain our correlation. Finally, we use a negative control (discussed below) using anemia and diabetes, both non-communicable diseases and immune to our mixing hypothesis.

We find that Gini is positively associated with TB incidence but not with anemia or diabetes. A 10% increase in the Gini coefficient is associated with a 4% increase in TB incidence. In contrast, such a relationship was either not seen with anemia or diabetes or was only weakly related.

With Covid raging across the planet, many ecological studies have documented an association between Covid-related outcomes (such as morbidity) and income inequality. For example, Oronce *et al*. show that US states with higher income inequality experienced higher deaths due to Covid[23]. Similarly, Elgar *et al*. found income inequality was associated with 30-day mortality rates in 84 countries[24]. These findings suggest that income inequality may explain why some parts of the world are hit harder by the Covid pandemic than others.

Our findings for TB are in line with these studies. They support the general notion, pioneered by Holtgrave and Crosby, that income inequality is associated with TB prevalence[25]. Our research design and methodology move the literature further: using system-GMM and negative controls, we provide a state-of-the-art analysis that shows a robust correlation for TB and a weak one for non-infectious diseases such as anemia.

These findings offer *indirect* credence to our proposed causal mechanism involving social mixing. But like all ecological studies, ours fail to establish a causal connection; specifically, we do not (nor can we) show that income inequality *independent*ly contributes to TB prevalence. In this context, we view our results as hypothesis-generating instead of hypothesis-confirming.[xi]

[xi] As Ranstam makes clear, "Many null hypotheses are tested in order to generate study hypotheses for further research, others to confirm an already established study hypothesis." [62]

2. Data

We draw on data from the World Health Organization (WHO), the World Bank, and the World Income Inequality Database (WIID) to construct a panel for a large cross-section of countries during 1995–2013.

2.1 TB incidence and prevalence

TB data are reported to WHO annually by its member states and compiled in the global TB database. They are reviewed and validated in collaboration with reporting entities. For countries where case notifications did not capture all incident cases that occurred within a year (based on a standardized checklist), special studies, including TB prevalence surveys or inventory studies, contribute to incidence estimates.

Disease incidence is a flow while prevalence is a stock, computed using past incidence and correcting for both the duration of the disease and recovery. Incidence is the estimated number of new (and relapsed) TB cases in a year per 100,000 population.[xii] It is calculated by dividing the number of subjects developing the disease over a certain period by the total number of subjects followed over that period. Intuitively, this is the probability a subject within a population will develop TB over a specified period. By contrast, TB prevalence is the number of cases of TB per 100,000 population at a given time, the proportion of people in a population having TB. The prevalence of TB tells us how much transmission is occurring. There are important reasons to use both as independent variables. The prevalence measure is of value because it reflects the burden of a disease in a population, which can help governments decide where to target healthcare investments.

There are several well-recognized problems with country-level TB data that may threaten the robustness of our findings. First, there is often a wide gap between the number of people who develop TB and those recorded and identified by public or private health centers. These "missing people" were not diagnosed and, therefore, not treated. This problem is particularly common among children under the age of fifteen. Second, almost all ecological work in this area uses data from national prevalence surveys that are subsequently adjusted: "notification data adjusted by a standard factor to account for under-reporting, over- and under-diagnosis; inventory studies that measure under-reporting; and expert opinion on TB detection gaps.[26]" Incidence estimates, which rely on these prevalence surveys, may therefore underestimate true prevalence. Additionally, poorer countries that lack high-quality surveillance systems and rely primarily on notification

[xii] Estimates were made based on case notifications from each country and the reference year with the most accurate incidence, the year in which a survey was carried out, or the year in which incidence was first estimated.

data. However, notification policy changes over time may be driven by changes in notification rates, which introduces multiple biases in the data. Nevertheless, the data we use is the best aggregate-level data available and is the same used by Chakaya *et al*.[26] in computing the global burden of TB.

As discussed below, we use anemia and diabetes as negative controls to check the mixing hypothesis. Our data on anemia are taken from Global Health Observatory Data from World Health Organization. We analyze anemia data from 2000 and 2013 (the longest continuous publicly available series). The data measures the prevalence of anemia among non-pregnant women with hemoglobin levels less than 12 g/dL. The data on hemoglobin in women of reproductive age was obtained from household surveys that collected blood hemoglobin concentration; these data are adjusted for elevation, smoking, and time trends using a Bayesian hierarchical mixture model[27]. Our data on diabetes prevalence is from NCD-Risc[28], which collects data from 751 studies that are population-based and had collected biomarkers and then estimated the trend using a Bayesian hierarchical model. Since diabetes data is available only in prevalence form, a cumulative variable of new incidence, and since diabetes is an incurable disease, we take the first difference to measure new incidence.

2.2 Income inequality

The Gini coefficient is the most frequently used measure of the extent of income inequality for a given community or society. It is defined as half of the arithmetic average of the absolute differences between all pairs of incomes in a population, the total then being normalized by mean income. If income in a population is distributed completely equally, the Gini value is 0; if a single person has all the income (maximum inequality), the Gini is 1.0. We used the Gini coefficient data from the World Income Inequality Database (WIID)[29]. We exclude countries from our analysis where the data are reported as low quality or are not representative of the country's entire geographic area or population [xiii] . We extrapolate data points for years where data are unavailable by imputing information from the closest preceding observation.

The Gini coefficient measures the gap between each person's income and that of others in their community. For our purposes, lacking direct evidence of social mixing across income strata, we use the income Gini as a proxy for community heterogeneity. Such heterogeneity, as argued before, may bring about increased

[xiii] If the best possible Gini is expenditure-based and is systematically lower than income inequality, we follow Deininger and Squire[58] and add 6.6 to the expenditure-based Gini.

disassortativity of social mixing by increasing the proportion of income-heterogenous people and lead to increased infections if disease prevalence differs by income groups[30] [xiv].

The Gini is undoubtedly the most popular measure of income inequality. It is a popular choice in ecological studies such as ours mainly because of its wide availability and ease of interpretation. There is some debate about its usefulness because it can have the same value for two different income distributions. This concern is not unique to the Gini: any inequality measure which compresses the Lorenz curve into a single parameter faces this problem. Blesch *et al.*[31] advocate for multi-parameter models, while Mirza *et al.*[32] suggest using night-light data to fill gaps in inequality measurement. We follow much of the epidemiology literature and use the Gini, as do Tan *et al.*[33] and Sepulveda and Brooker.[34]

2.3 Control variables

We select control variables based on what is most often used in the literature. These include economic variables that affect new TB infections, such as years of schooling (from the Penn World Table), GDP per capita, public health expenditure (% to GDP), poverty, and the share of people living in urban areas (from World Bank). We include data on HIV prevalence (from World Bank) and BCG coverage (from BCG Atlas) to control for medical conditions related to TB risk. We also include data on age composition, including population ages 65 and above and males between the ages of 15 and 40. Because nutritional intake heavily influences anemia prevalence, we include the population whose food consumption is insufficient to give the dietary energy levels necessary to maintain a normal, healthy life.

One difficulty we face is a bi-directional correlation: while inequality may be linked to TB prevalence, the latter may also influence income inequality. It is challenging to isolate exogenous variation in income inequality within a country, much less across a broad cross-section of countries. Nevertheless, the economics literature studying the *shifters* of income inequality has found that changes in foreign direct investment (FDI) and trade are robust shifters of within-country inequality but are plausibly exogenous to disease transmission[35]. We include them in our instrumental variable regression analysis. The World Bank provides data on trade (the sum of exports and imports as a share of GDP). The IMF provides data on foreign direct

[xiv] In mathematical modeling studies of infection spread, it is common to showcase the Lorenz curves (and associated Ginis), which plot the cumulative proportion of incident infections (y-axis) against that of the population (x-axis), after ranking population subunits (in our case, income) in order of their incidence rates. As Chen *et al.* argue, the Gini is interpretable as a summary measure for the degree to which infections are asymmetrically distributed between subpopulations[67].

investments (net inflows of investment to acquire a lasting management interest in an enterprise operating in an economy).

Negative controls

Our regression model tries to quantify the relationship between TB and income inequality after controlling for various variables. Even if we find a significant relationship, there may be several explanations beyond the mixing hypothesis we are testing that are consistent with the finding.

a. By their very nature, high-income-unequal countries have an overwhelming presence of poor people and few rich people. If TB is a disease of poverty, then the poor with comprised nutrition are more likely to catch, harbor, and transmit it. That, in and of itself, can explain why two countries with similar per capita incomes but different income inequality can show different levels of prevalence of TB. In other words, irrespective of the validity of the mixing hypothesis, an association between TB and income inequality may appear at the cross-country level.
b. Classic aggregation bias arises whenever analysts test hypotheses about individual-level relationships with data that averages individual-level data. Specifically, the issue may arise from a) a non-linear relationship between TB infectivity and income at the individual level[36] and b) aggregating individual input data, each with idiosyncratic measurement errors, across spatial, temporal, and activity scales (as would be the case for our variables of interest). Analyses unaware of the aggregation bias may report associations between population-level TB prevalence and income inequality, even though income inequality has no causal effect on the chance of any individual contracting the disease.
c. Another channel, distinct channel social mixing such as low political support for public health from the affluent[37], drives the association between TB and inequality.

We contend that our testing of negative control conditions – anemia (or, to a lesser extent, diabetes) may serve a useful purpose. First, anemia, like TB, is also a disease of poverty. It should be subject to arguments (a) and (c) above. Second, anemia is negatively associated with individual income mainly because nutrition improves with income. (Diabetes, on the other hand, is positively associated with income, but that picture is changing[xv].) Both diseases likely exhibit a non-linear relationship between individual income and incidence and may be subject to aggregation bias. Third, measurement errors arising from aggregating

[xv] Four out of five people in the world with diabetes now live in low- and middle-income countries.

individual input data across spatial, temporal, and activity scales would likely also affect our negative control analyses on anemia and diabetes data. Therefore, by the logic of (b) or (c), anemia or diabetes prevalence should also be associated with country-level income inequality. Fourth, if our result for TB is driven purely by the overwhelming presence of the poor in income-unequal countries, anemia should also show a strong correlation with income inequality. As shown below, it does *not*: the correlation of anemia or diabetes prevalence with income inequality is low and, notably, much lower than the same for TB prevalence. Of course, one critical difference between TB and anemia and diabetes is that the former is communicable while the latter two are not.

3. Methods & Estimation Strategy

3.1 Dynamic Panel Data Analysis

Our dependent variables are new incidences of TB in a year and anemia prevalence; the explanatory variable of interest is the Gini. We log-transform the variables to achieve log-linearity, which has the added benefit that the regression coefficient of Gini measures its elasticity (how responsive it is to changes in the variable). We estimate the relationship between changes in disease incidence and income inequality rather than the cross-sectional correlation, thereby reducing omitted-variable bias by controlling unobservable country characteristics, such as socio-cultural and geographical variations, that are time-invariant.

We employ a multivariate panel data model by controlling potential variables that might alter disease incidence and inequality. Our model is as follows:

$$(1)\quad TB_{it} = \gamma Prev_{i,t-1} + \beta Gini_{it} + x_{it}\delta + \zeta_t + \alpha_i + \nu_{it}$$

γ captures the effect of the lagged TB prevalence on its current TB incidence, x_{it} is a vector of control variables and β is the elasticity of Gini on new TB incidence. If the Gini increases by 1 percent, TB incidence increases by β percent. We also include country and time-fixed effects in the specification.[xvi]

[xvi] Time fixed-effects allow us to reduce bias from unobservables that change over time but are constant over countries, and country fixed-effects control for factors that differ across countries but are constant over time. For us, this means differences in income inequality across countries are not because they are at different stages of development or have different climates or different political structures – these factors do not change quickly over short lengths of time - but because there are time-varying factors within a country that are responsible.

We include lagged prevalence ($Prev_{i,t-1}$) as a measure of direct exposure, itself a function of TB incidence. Thus, statistically, it is equivalent to having a lagged dependent variable as an explanatory variable. Such an inclusion introduces potential statistical bias and inconsistency when using a standard panel fixed-effect model because of the correlation between the error and the lagged regressor. We applied the system-GMM estimation developed by Arellano-Bover/Blundell-Bond[38,39] to correct this problem [xvii]. The primary estimation method of system-GMM uses internal instruments, which are lagged values of the covariates. Using lagged instruments allows us to rule out reverse causality, but we cannot be sure that an unobservable factor affects both income inequality and disease spread. So we also include trade and foreign direct investment as external instruments well-known in the economics literature as external shifters of income disparity.

Dynamic panel-data analysis is well suited to the data we have. We use the System-Generalized Method of Moments (GMM) estimator developed for dynamic models by Arellano and Bover[38] and Blundell and Bond[39]. GMM is a statistical method combining observed economic data with information in population moment conditions to estimate unknown parameters. System GMM estimates equation (1) by instrumenting endogenous variables using their first differences and estimates the first-differenced equation of (1) by instrumenting the endogenous variables by their lagged values. We assume the error term to be serially uncorrelated because an Arellano and Bond autocorrelation test rejects autocorrelation on the first-differenced errors at the second order. We restricted the set of instruments to lag 3 and collapsed the instruments into one column to prevent them from overfitting the endogenous variable.

Using the system-GMM method has several advantages. First, it permits us to account for the effect of past realizations of the dependent variable (TB) on the current independent variable (*Prev*). Second, it controls the unobserved ecological conditions such as geography and weather by including country-fixed effects. Third,

[xvii] Ordinary least square (OLS) estimation and fixed effect estimation are biased due to the presence of the lagged dependent variable[68]. The difference GMM estimator proposed by Arellano and Bond[69] has been widely used in the literature on dynamic panel data. The first difference GMM uses higher-order lags as an instrument for the first-differenced variables assuming errors are serially uncorrelated and independent across countries. Maintaining the assumption that TB incidence and income inequality (and other regressors) are uncorrelated contemporaneously, the Arellano and Bond[69] tests show that lagged values of the predetermined $Gini_{it}$ and x_{it} are valid instruments for the first-differenced equation. However, solely applying first-difference GMM in this study could be problematic because variables such as income distribution and education tend to be persistent causing a weak instrument problem. Blundell and Bond[39] suggested a system GMM that adds a set of level equations instrumented by their own lagged differences to solve this challenge when the initial deviation of explanatory variables, TB incidence, income inequality, and GDP, are uncorrelated with the country fixed effect. In addition, the validity depends on the AR(2) test statistics, which test the serial correlation of the original error term. Our test statistics shows that serial correlation does not appear to be an important concern.

using lagged dependent variables as ("internal") instruments is an advantage, as valid external instruments are typically challenging to construct.

Although non-communicable diseases (such as anemia or diabetes) do not spread through interaction, their prevalence tends to be highly persistent. Given that the only accessible data on non-communicable disease is the prevalence data, we include lagged prevalence to account for the persistent nature of the disease and use the same estimation method as TB.

4. Results

Figure 2 displays the estimated correlation between the logged GDP, the logged Gini coefficient, and the 95% confidence interval. TB1 shows the estimated coefficient from the estimation results using only lagged instruments. TB2 shows the result when employing additional external instruments -- FDI and openness of the economy. Anemia and diabetes in Figure 2 shows the coefficients estimates for logged GDP, the logged Gini coefficient but with anemia and diabetes prevalence as dependent variables. The coefficient of Gini (β) in the TB estimation is near 0.4 and statistically significant at the 5% level. If Gini increases by 10%, TB incidence is predicted to rise by 4%. More precisely, compare two countries, X and Y, with income inequality in X being 10% higher in X than in Y. Even correcting for differences in mean incomes across the two, our estimates predict that TB incidence in X is 4% higher than in Y. The coefficient of Gini in the anemia estimation is near -0.07, meaning a 10% increase in Gini is associated with a 0.7% reduction in anemia prevalence. Gini was positively associated with diabetes prevalence by 0.025. In both cases of the non-communicable disease, the elasticity with Gini is substantially *smaller* than that of TB.

Higher average income significantly lowers TB incidence; the estimated coefficient is -0.24 and statistically significant at the 5% level. These results imply that a 10% increase in income is associated with a 2.4% decrease in TB incidence. The association between GDP and anemia and diabetes prevalence is not statistically significant.

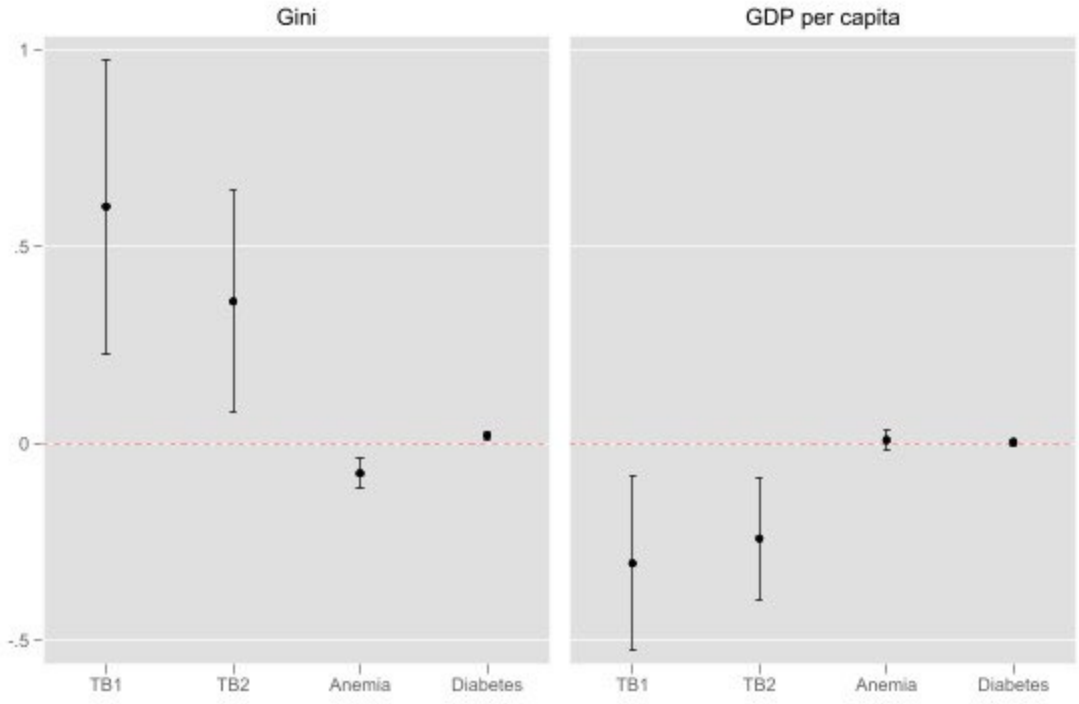


Figure 2: Estimated elasticities of disease concerning income inequality and income.

Notes: The coefficients of Gini and GDP per capita from each regression are for TB, anemia, and diabetes. TB1 uses only lagged instruments, while TB2 employs additional external instruments -- FDI and openness of the economy. The model is estimated as described in the Methods & Estimation Strategy section. This estimation uses external Gini shifters, foreign direct investment, and openness of the economy. Estimations adjusted for years of schooling, GDP per capita, public health expenditure, poverty, and the share of people living in urban areas. In addition, TB estimations control for HIV prevalence and BCG coverage. Error bars are 90%, 95%, and 99% CIs based on robust standard errors. The complete set of result tables is presented in the Appendix.

5. Discussion and caveats

Proper testing of our mixing hypothesis would require micro data on mixing patterns among individuals with different incomes. To our knowledge, while such data exists for mixing by age and gender, it does not yet exist for income[40–42]. Without such data, we do not know if social mixing is predominantly random and ensuring. In reality, it may be that rich people congregate primarily with the other rich and deliberately stay away from places where mixing with the poor is unavoidable. More importantly, our ecological study cannot definitively rule out other possible causal mechanisms. Future research analyzing contact tracing data may help shed more light on this matter[41].

That we rely on the system GMM estimator may be a limitation. While system GMM estimation is an improvement over standard panel data estimation, it is known to be sensitive to specification assumptions[43,44]. According to Bellemare *et al.*, identification using lag instruments is only reliable under particular conditions[43], though we do employ trade and foreign direct investment as external instruments as well.

The nature of the disease, and its pathogenesis, are also crucial. Our results rely on TB data and may not extend to other infectious diseases such as HIV or malaria. Mass vaccinations may prevent some contagious diseases from spreading even under random mixing. However, to the extent the poor are less likely to receive the vaccinations, our message would still be active.

To close, our ecological analysis, with all its caveats, hints that TB, like other infectious diseases, may be causally linked to income inequality. If our mixing hypothesis is confirmed by future work using income-social-contact data, it will assert that income redistribution could be an essential policy lever to improve population health.[45]

Bibliography

1. Raffalli J, Sepkowitz KA, Armstrong D. Community-based outbreaks of tuberculosis. *Arch Intern Med*. 1996;156(10). doi:10.1001/archinte.156.10.1053
2. McCreesh N, White RG. An explanation for the low proportion of tuberculosis that results from transmission between household and known social contacts. *Sci Rep*. 2018;8(1). doi:10.1038/s41598-018-23797-2
3. Janssens JP, Rieder HL. An ecological analysis of incidence of tuberculosis and per capita gross domestic product. *Eur Respir J*. 2008;32(5):1415-1416. doi:10.1183/09031936.00078708
4. Oxlade O, Murray M. Tuberculosis and Poverty: Why Are the Poor at Greater Risk in India? *PLoS One*. 2012;7(11). doi:10.1371/journal.pone.0047533
5. Andrews JR, Basu S, Dowdy DW, Murray MB. The epidemiological advantage of preferential targeting of tuberculosis control at the poor. *Int J Tuberc Lung Dis*. 2015;19(4):375-380. doi:10.5588/ijtld.14.0423
6. Pelissari DM, Diaz-Quijano FA. Household crowding as a potential mediator of socioeconomic determinants of tuberculosis incidence in Brazil. *PLoS One*. 2017;12(4). doi:10.1371/journal.pone.0176116
7. Apolinário D, Ribeiro AI, Krainski E, Sousa P, Abranches M, Duarte R. Tuberculosis inequalities & socio-economic deprivation in Portugal. *Int J Tuberc Lung Dis*. 2017;21(7). doi:10.5588/ijtld.16.0907
8. Nidoi J, Muttamba W, Walusimbi S, et al. Impact of socio-economic factors on Tuberculosis treatment outcomes in north-eastern Uganda: a mixed methods study. *BMC Public Health*. 2021;21(1). doi:10.1186/s12889-021-12056-1
9. Alipanah N, Jarlsberg L, Miller C, et al. Adherence interventions and outcomes of tuberculosis treatment: A systematic review and meta-analysis of trials and observational studies. *PLoS Med*. 2018;15(7). doi:10.1371/journal.pmed.1002595
10. Harling G, Lima Neto AS, Sousa GS, Machado MMT, Castro MC. Determinants of tuberculosis transmission and treatment abandonment in Fortaleza, Brazil. *BMC Public Health*. 2017;17(1). doi:10.1186/s12889-017-4435-0
11. Muniyandi M, Thomas BE, Karikalan N, et al. Association of Tuberculosis with Household Catastrophic Expenditure in South India. *JAMA Netw Open*. 2020;3(2). doi:10.1001/jamanetworkopen.2019.20973
12. Oren E, Koepsell T, Leroux BG, Mayer J. Area-based socio-economic disadvantage and tuberculosis incidence. *Int J Tuberc Lung Dis*. 2012;16(7). doi:10.5588/ijtld.11.0700
13. Munro SA, Lewin SA, Smith HJ, Engel ME, Fretheim A, Volmink J. Patient adherence to tuberculosis treatment: A systematic review of qualitative research. *PLoS Med*. 2007;4(7). doi:10.1371/journal.pmed.0040238
14. Martinez L, Shen Y, Mupere E, Kizza A, Hill PC, Whalen CC. Transmission of Mycobacterium Tuberculosis in Households and the Community: A Systematic Review and Meta-Analysis. In: *American Journal of Epidemiology*. Vol 185. ; 2017. doi:10.1093/aje/kwx025
15. Mathema B, Andrews JR, Cohen T, et al. Drivers of Tuberculosis Transmission. *J Infect Dis*. 2017;216. doi:10.1093/infdis/jix354
16. Johnstone-Robertson SP, Mark D, Morrow C, et al. Social mixing patterns within a South African township community: Implications for respiratory disease transmission and control. *Am J Epidemiol*. 2011;174(11). doi:10.1093/aje/kwr251
17. Horton KC, Hoey AL, Béraud G, Corbett EL, White RG. Systematic review and meta-analysis of sex differences in social contact patterns and implications for tuberculosis transmission and control. *Emerg Infect Dis*. 2020;26(5). doi:10.3201/eid2605.190574
18. Classen CN, Warren R, Richardson M, et al. Impact of social interactions in the community on the transmission of tuberculosis in a high incidence area. *Thorax*. 1999;54(2). doi:10.1136/thx.54.2.136
19. Marimuthu P. Tuberculosis prevalence and socio-economic differentials in the slums of four metropolitan cities of India. *Indian J Tuberc*. 2016;63(3). doi:10.1016/j.ijtb.2016.08.007

20. Banu S, Rahman MT, Uddin MKM, et al. Epidemiology of Tuberculosis in an Urban Slum of Dhaka City, Bangladesh. *PLoS One*. 2013;8(10). doi:10.1371/journal.pone.0077721
21. Cavalcanti ARC. Work, Slums and Informal Settlement Traditions : Architecture of the Favela Do Telegrafo. *Tradit Dwellings Settlements Rev*. 2017;XXVIII(II).
22. Dowdy DW, Golub JE, Chaisson RE, Saraceni V. Heterogeneity in tuberculosis transmission and the role of geographic hotspots in propagating epidemics. *Proc Natl Acad Sci U S A*. 2012;109(24). doi:10.1073/pnas.1203517109
23. Oronce CIA, Scannell CA, Kawachi I, Tsugawa Y. Association Between State-Level Income Inequality and COVID-19 Cases and Mortality in the USA. *J Gen Intern Med*. 2020;35(9). doi:10.1007/s11606-020-05971-3
24. Elgar FJ, Stefaniak A, Wohl MJA. The trouble with trust: Time-series analysis of social capital, income inequality, and COVID-19 deaths in 84 countries. *Soc Sci Med*. 2020;263(September):113365. doi:10.1016/j.socscimed.2020.113365
25. Holtgrave DR, Crosby RA. Social determinants of tuberculosis case rates in the United States. *Am J Prev Med*. 2004;26(2):159-162. doi:10.1016/J.AMEPRE.2003.10.014
26. Chakaya J, Khan M, Ntoumi F, et al. Global Tuberculosis Report 2020 – Reflections on the Global TB burden, treatment and prevention efforts. *Int J Infect Dis*. 2021;113. doi:10.1016/j.ijid.2021.02.107
27. Stevens GA, Paciorek CJ, Flores-Urrutia MC, et al. National, regional, and global estimates of anaemia by severity in women and children for 2000–19: a pooled analysis of population-representative data. *Lancet Glob Heal*. 2022;10(5):e627-e639. doi:10.1016/S2214-109X(22)00084-5
28. Zhou B, Lu Y, Hajifathalian K, et al. Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4.4 million participants. *Lancet*. 2016;387(10027):1513-1530. doi:10.1016/S0140-6736(16)00618-8
29. UNU-WIDER : World Income Inequality Database - WIID. Accessed February 2, 2023. https://www.wider.unu.edu/database/world-income-inequality-database-wiid
30. Harling G, Subramanian S V., Bärnighausen T, Kawachi I. Income inequality and sexually transmitted in the United States: Who bears the burden? *Soc Sci Med*. 2014;102. doi:10.1016/j.socscimed.2013.11.025
31. Blesch K, Hauser OP, Jachimowicz JM. Measuring inequality beyond the Gini coefficient may clarify conflicting findings. *Nat Hum Behav*. Published online November 1, 2022. doi:10.1038/s41562-022-01430-7
32. Mirza MU, Xu C, van Bavel B, van Nes EH, Scheffer M. Global inequality remotely sensed. *Proc Natl Acad Sci U S A*. 2021;118(18). doi:10.1073/pnas.1919913118
33. Tan AX, Hinman JA, Abdel Magid HS, Nelson LM, Odden MC. Association between Income Inequality and County-Level COVID-19 Cases and Deaths in the US. *JAMA Netw Open*. Published online 2021. doi:10.1001/jamanetworkopen.2021.8799
34. Sepulveda ER, Brooker AS. Income inequality and COVID-19 mortality: Age-stratified analysis of 22 OECD countries. *SSM - Popul Heal*. 2021;16. doi:10.1016/j.ssmph.2021.100904
35. Furceri D, Ostry JD. Robust determinants of income inequality. *Oxford Rev Econ Policy*. 2019;35(3). doi:10.1093/oxrep/grz014
36. Gravelle H, Wildman J, Sutton M. Income, income inequality and health: What can we learn from aggregate data? *Soc Sci Med*. 2002;54(4):577-589. doi:10.1016/S0277-9536(01)00053-3
37. Dickman SL, Himmelstein DU, Woolhandler S. Inequality and the health-care system in the USA. *Lancet*. 2017;389(10077):1431-1441. doi:10.1016/S0140-6736(17)30398-7
38. Arellano M, Bover O. Another look at the instrumental variable estimation of error-components models. *J Econom*. 1995;68(1):29-51. doi:10.1016/0304-4076(94)01642-D
39. Blundell R, Bond S. Initial conditions and moment restrictions in dynamic panel data models. *J Econom*. 1998;87(1):115-143. doi:10.1016/S0304-4076(98)00009-8
40. McCreesh N, Mohlamonyane M, Edwards A, et al. Improving Estimates of Social Contact Patterns for Airborne Transmission of Respiratory Pathogens. *Emerg Infect Dis*. 2022;28(10):2016-2026. doi:10.3201/eid2810.212567
41. Moro E, Calacci D, Dong X, Pentland A. Mobility patterns are associated with experienced income segregation in large US cities. *Nat Commun 2021 121*. 2021;12(1):1-10. doi:10.1038/s41467-021-24899-8

42. Hoang T, Coletti P, Melegaro A, et al. A Systematic Review of Social Contact Surveys to Inform Transmission Models of Close-contact Infections. *Epidemiology*. 2019;30(5):723-736. doi:10.1097/EDE.0000000000001047
43. Bellemare MF, Masaki T, Pepinsky TB. Lagged Explanatory Variables and the Estimation of Causal Effect. *Source J Polit*. 2017;79(3):949-963. doi:10.2307/26551051
44. Bun MJG, Windmeijer F. The weak instrument problem of the system GMM estimator in dynamic panel data models. *Econom J*. 2010;13(1):95-126. doi:10.1111/J.1368-423X.2009.00299.X
45. Bor J, Cohen GH, Galea S. Population health in an era of rising income inequality: USA, 1980–2015. *Lancet*. 2017;389(10077):1475-1490. doi:10.1016/S0140-6736(17)30571-8
46. Feenstra RC, Inklaar R, Timmer MP. The Next Generation of the Penn World Table. *Am Econ Rev*. 2015;105(10):3150-3182. doi:10.1257/aer.20130954
47. Miller S, Wherry LR. Health and Access to Care during the First 2 Years of the ACA Medicaid Expansions. *N Engl J Med*. 2017;376(10):947-956. doi:10.1056/nejmsa1612890
48. Khullar D, Chokshi DA. *Health, Income, & Poverty: Where We Are & What Could Help*.; 2018. doi:10.1377/hpb20180817.901935
49. Cambron C, Kosterman R, Hawkins JD. Neighborhood poverty increases risk for cigarette smoking from age 30 to 39. *Ann Behav Med*. 2019;53(9):858-864. doi:10.1093/abm/kay089
50. Green MJ, Stritzel H, Smith C, Popham F, Crosnoe R. Timing of poverty in childhood and adolescent health: Evidence from the US and UK. *Soc Sci Med*. 2018;197:136-143. doi:10.1016/J.SOCSCIMED.2017.12.004
51. Gibbs A, Jewkes R, Willan S, Washington L. Associations between poverty, mental health and substance use, gender power, and intimate partner violence amongst young (18-30) women and men in urban informal settlements in South Africa: A cross-sectional study and structural equation model. *PLoS One*. 2018;13(10):1-19. doi:10.1371/journal.pone.0204956
52. Walker ER, Druss BG. Cumulative burden of comorbid mental disorders, substance use disorders, chronic medical conditions, and poverty on health among adults in the U.S.A. *Psychol Health Med*. 2017;22(6):727-735. doi:10.1080/13548506.2016.1227855
53. Alavi Hojjat T, Hojjat R. Socioeconomic Factors: Poverty and Obesity. In: *The Economics of Obesity: Poverty, Income Inequality and Health*. Springer Singapore; 2017:19-25. doi:10.1007/978-981-10-2911-0_4
54. Do N. *The Effects of Poverty and Allostatic Load on the Development of Chronic Disease*.; 2017.
55. Wagstaff A, van Doorslaer E. Income Inequality and Health: What Does the Literature Tell Us? *Annu Rev Public Health*. 2000;21(1):543-567. doi:10.1146/annurev.publhealth.21.1.543
56. Pickett KE, Wilkinson RG. Income inequality and health: A causal review. *Soc Sci Med*. 2015;128:316-326. doi:https://doi.org/10.1016/j.socscimed.2014.12.031
57. Marmot MG. Understanding social inequalities in health. *Perspect Biol Med*. 2003;46(3 SUPPL.). doi:10.1353/pbm.2003.0056
58. Deininger K, Squire L. A new data set measuring income inequality. *World Bank Econ Rev*. 1996;10(3):565-591.
59. Deaton A. Health, Inequality, and Economic Development. *J Econ Lit*. 2003;41(1):113-158. doi:10.1257/002205103321544710
60. Sapolsky RM. *Why Zebras Don't Get Ulcers*. Owl Book/Henry Holt and Co.; 2004.
61. Kawachi I, Kennedy BP, Lochner K, Prothrow-stith D. Social Capital, Income Inequality, and Mortality. *Am J Public Health*. 1996;87(9).
62. Ranstam J. Hypothesis-generating and confirmatory studies, Bonferroni correction, and pre-specification of trial endpoints. *New Pub Med Journals Sweden*. 2019;90(4):297. doi:10.1080/17453674.2019.1612624
63. Turner RD, Bothamley GH. Cough and the transmission of tuberculosis. *J Infect Dis*. 2015;211(9). doi:10.1093/infdis/jiu625
64. Leung K, Jit M, Lau EHY, Wu JT. Social contact patterns relevant to the spread of respiratory infectious diseases in Hong Kong. *Sci Rep*. 2017;7. doi:10.1038/s41598-017-08241-1

65. Sahasranaman A, Jensen HJ. Spread of COVID-19 in urban neighbourhoods and slums of the developing world: Spread of COVID-19 in urban neighbourhoods and slums of the developing world. *J R Soc Interface*. 2021;18(174). doi:10.1098/rsif.2020.0599
66. Rains E, Krishna A, Wibbels E. Combining satellite and survey data to study Indian slums: evidence on the range of conditions and implications for urban policy. *Environ Urban*. 2019;31(1). doi:10.1177/0956247818798744
67. Chen MI, Ghani AC, Edmunds WJ. A metapopulation modelling framework for gonorrhoea and other sexually transmitted infections in heterosexual populations. *J R Soc Interface*. 2009;6(38). doi:10.1098/rsif.2008.0394
68. Nickell S. Biases in Dynamic Models with Fixed Effects. *Econometrica*. 1981;49(6):1417. doi:10.2307/1911408
69. Arellano M, Bond S. Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *Rev Econ Stud*. 1991;58(2):277. doi:10.2307/2297968

Appendix Tables

Variable	Description	Mean	Std. Dev.	Min	Max
TB	Estimated tuberculosis incidence rate (per 100,000 people)	193.220	239.630	5.000	1660.000
Prev: TB prevalence	Estimated number of cases of TB at a given point in time (per 100,000 people)	241.010	245.300	6.400	1686.000
Gini	Gini index	43.920	10.630	23.500	72.360
GDP per capita	Real GDP per capita (in 2010 USD)	5528.461	10207.990	182.709	101380.800
School	Average years of schooling in the population aged over 25	6.632	3.184	0.623	13.267
Health expenditure	Public health expenditure (% of GDP)	3.081	1.643	0.552	9.087
Poverty	Poverty headcount ratio at $1.90 a day in 2011, PPP	23.592	24.293	0.000	85.722
Urban	People living in urban areas (% of total)	49.012	19.948	7.211	94.843
HIV prevalence	Percentage of people ages 15-49 who are infected with HIV	2.578	5.225	0.100	28.400
BCG	BCG coverage (%)	90.020	12.600	21.000	99.000
Population 65	Population ages 65 and above (% of total population)	6.990	4.740	1.871	24.626
Male 15-40	Male population ages 15-40 (% of total population)	20.060	3.160	14.564	25.178
Anemia	Prevalence of anemia among non-pregnant women (% of women ages 15-49)	26.390	14.660	7.100	66.500
Undernourish	Prevalence of undernourishment (% of the population)	10.940	10.320	2.500	53.700
Diabetes Prevalence	The proportion of adults over 18 who have diabetes (age standardized).	0.069	0.019	0.022	0.174

Table A1: Description of variables & Summary statistics

Variable	Source	URL
TB	WHO TB database	https://unstats.un.org/unsd/mdg/Metadata.aspx
TB prevalence	WHO TB database	https://unstats.un.org/unsd/mdg/Metadata.aspx
Gini	UNU-Wider[29]	https://www.wider.unu.edu/project/world-income-inequality-database-wiid
GDP per capita	World Bank	https://data.worldbank.org/
School	Penn World Table 9.0[46]	https://www.rug.nl/ggdc/productivity/pwt/pwt-releases/pwt9.0?lang=en
Health expenditure	World Bank	https://data.worldbank.org/
Poverty	World Bank	http://iresearch.worldbank.org/PovcalNet/povOnDemand.aspx
Urban	World Bank	https://data.worldbank.org/
HIV prevalence	World Bank	https://data.worldbank.org/
BCG	BCG Atlas	http://www.bcgatlas.org/
Population 65	World Bank	https://data.worldbank.org/
Male 15-40	World Bank	https://data.worldbank.org/
Anemia	WHO	https://www.who.int/data/gh
Undernourish	World Bank	https://data.worldbank.org/
Diabetes Prevalence	NCD_RisC	https://ncdrisc.org/data-downloads-diabetes.html

Table A2: Data sources and URL links

VARIABLES	(1) ln(TB incidence)	(2) ln(TB incidence)	(3) ln(TB incidence)	(4) ln(TB incidence)
L.ln(prevalence)	0.756***	0.667***	0.861***	0.819***
	(0.066)	(0.097)	(0.065)	(0.069)
ln(Gini)	0.602***	0.658**	0.361**	0.473**
	(0.191)	(0.278)	(0.144)	(0.189)
ln(GDP per capita)	-0.304***	-0.116	-0.241***	-0.092
	(0.113)	(0.101)	(0.079)	(0.064)
ln(school)	0.027	0.001	0.020	0.020
	(0.170)	(0.204)	(0.137)	(0.142)
ln(health expenditure)	-0.045	-0.142	0.069	-0.052
	(0.092)	(0.132)	(0.074)	(0.096)
HIV prevalence	0.058***	0.058***	0.051***	0.047***
	(0.009)	(0.014)	(0.008)	(0.011)
poverty	-0.006*	-0.007	-0.005*	-0.006**
	(0.004)	(0.005)	(0.003)	(0.003)
urban	0.004	-0.008	0.005	-0.003
	(0.005)	(0.007)	(0.004)	(0.004)
Population 65	0.019	-0.005	0.012	-0.006
	(0.016)	(0.024)	(0.013)	(0.014)
BCG	0.007**	0.010**	0.005	0.006*
	(0.003)	(0.004)	(0.003)	(0.003)
Male 15-40		-0.032		-0.042
		(0.040)		(0.026)
Observations	1,521	1,521	1,371	1,371
Number of N	95	95	92	92
Year FE	No	No	No	No
Instruments	58	62	60	64
AR(2)	0.507	0.370	0.763	0.423
Hansen	0.242	0.946	0.418	0.542

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Column (1)-(2) uses only lagged instruments, while Column (3)-(4) uses FDI and trade as external instruments.

Table A3: TB, Arellano-Bover/ Blundell-Bond estimation

VARIABLES	(1) ln(anemia)	(2) ln(anemia)	(3) ln(anemia)	(4) ln(anemia)	(5) ln(anemia)	(6) ln(anemia)
L.ln(anemia)			1.106***	1.068***	1.065***	1.060***
			(0.043)	(0.029)	(0.040)	(0.035)
ln(Gini)	-2.474	-0.887	-0.069**	-0.076***	-0.068***	-0.075***
	(2.873)	(0.799)	(0.032)	(0.020)	(0.025)	(0.020)
ln(GDP per capita)	0.015	-0.039	0.030*	0.024**	0.009	0.009
	(0.191)	(0.084)	(0.016)	(0.011)	(0.016)	(0.013)
poverty	0.002	0.002	0.000	-0.000	0.001**	0.000
	(0.003)	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)
ln(school)	-0.224	-0.059	0.028	-0.009	0.035	0.007
	(0.291)	(0.149)	(0.026)	(0.018)	(0.026)	(0.014)
ln(health expenditure)	-0.022	-0.024	0.023	0.002	0.020	0.004
	(0.049)	(0.026)	(0.016)	(0.017)	(0.017)	(0.015)
urban	0.009	0.005	-0.001	-0.000	0.000	0.000
	(0.013)	(0.006)	(0.001)	(0.001)	(0.001)	(0.001)
ln(undernourish)	0.127	0.061	0.014	0.003	-0.006	-0.003
	(0.116)	(0.041)	(0.012)	(0.010)	(0.011)	(0.011)
HIV prevalence		0.001		0.003		0.002
		(0.013)		(0.003)		(0.002)
Observations	1,306	1,077	1,443	1,203	1,306	1,077
Number of N	111	92	116	97	111	92
Year FE			No	No	No	No
Instruments			45	45	47	47
AR(2)			0.181	0.147	0.470	0.368
Hansen			0.158	0.0214	0.248	0.0366

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Column (1)-(2) uses panel fixed effect estimation, Column (3)-(4) uses system GMM with only lagged instruments, and Column (5)-(6) uses system GMM with additional external instruments.

Table A4: Anemia

VARIABLES	(1) ln(diabetes case)	(2) ln(diabetes prevalence)	(3) ln(diabetes prevalence)
L. ln(diabetes prevalence)		1.018***	1.018***
		(0.006)	(0.007)
ln(Gini)	-0.977	0.023***	0.021***
	(3.858)	(0.005)	(0.004)
ln(GDP per capita)	0.948***	0.002	0.004
	(0.316)	(0.003)	(0.004)
poverty	0.007	-0.000	-0.000
	(0.007)	(0.000)	(0.000)
ln(school)	0.241	0.002	0.002
	(0.360)	(0.004)	(0.004)
HIV prevalence	-0.008	-0.000	-0.000
	(0.023)	(0.000)	(0.000)
ln(health expenditure)	-0.184**	-0.005*	-0.002
	(0.090)	(0.003)	(0.003)
urban	-0.048	-0.001**	-0.001***
	(0.029)	(0.000)	(0.000)
Population 65	-0.009	-0.000	-0.001
	(0.059)	(0.000)	(0.000)
Observations	1,491	1,769	1,574
Number of N	99	104	100
Year FE		No	No
Instruments		54	56
AR(2)		0.0712	0.0233
Hansen		0.00483	0.00749

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Column (1) uses panel fixed effect estimation the dependent variable is the first difference of diabetes prevalence, Column (2) uses system GMM with only lagged instruments, and Column (3) uses system GMM with additional external instruments. The dependent variable of the system GMM estimation is age standardized diabetes prevalence.

Table A5: Diabetes