Infectious Disease as a Mechanism Linking Health and Income Inequality

Joydeep Bhattacharya (joydeep@iastate.edu)

Jay Bhattacharya
Stanford University

Min Kim
Iowa State University

Article

Keywords: life expectancy, mental illness, obesity, infant mortality, teenage births, homicides, imprisonment, Tuberculosis (TB), HIV prevalence, public health expenditures, population density

Posted Date: October 8th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-951908/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Infectious Disease as a Mechanism Linking Health and Income Inequality

**Background** Within-country inequality has been rising worldwide rapidly since the 70s. An extensive literature has examined the effect of inequality on health, finding health outcomes to be worse in more unequal countries. Among the measures of health used are life expectancy, mental illness, obesity, infant mortality, teenage births, homicides, imprisonment, etc., or some weighted index of such measures. While these measures of health are informative, they are indirect. Secondly, many studies fail to establish an independent effect of inequality on health. Finally, as noted, if the individual-level relationship between health and income is non-linear, cross-section studies may detect a spurious association between inequality and health due to an aggregation problem.

This paper studies the relationship between the incidence of infectious disease, a direct and vital measure of health, and ambient income inequality. Our hypothesis is, the more income-unequal a society, the higher the chance a random mixing of people from different income strata brings the infected and uninfected closer, thereby raising disease spread. This implies two countries with similar per capita incomes but with varying levels of income inequality can exhibit very different trajectories of disease spread. We investigate this causal pathway by examining whether countries with elevated levels of income inequality have higher rates of Tuberculosis (TB) incidence per capita. The choice of TB is appropriate because it is an enduring, serious threat to global public health, the leading cause of death from infectious diseases worldwide. Moreover, it is well known that the four stages of TB pathogenesis (exposure to infection, progression to disease, late or inappropriate diagnosis and treatment, and treatment adherence) vary across rich and poor individuals.

**Methods** We used publicly available panel data for one hundred and thirty-three countries between 1997 and 2013. The data include TB incidence and prevalence per 100,000 people. It also contains data on income inequality (Gini coefficient) both across countries and across time within countries. Our multivariate regression model controlled, among other variables, for economic output per capita, HIV prevalence, public health expenditures, population density, and poverty, and also incorporated a country-level fixed effect and time-fixed effects. A novel correction for “aggregation bias” using data on diabetes (a non-communicable disease) is also applied.

**Findings** Overall, elevated levels of income inequality were positively associated and causally connected to tuberculosis prevalence. All else same, countries with income-Gini coefficients a mere 10% apart would likely see a 5% difference in tuberculosis prevalence.

**Interpretation** Like any air-borne infectious disease, TB is akin to a pollutant that spoils air quality and makes it unhealthy for all who breathe it. Our findings suggest a significant cause of this externality is ambient income inequality. In effect, TB is a negative externality whose reach amplified by income inequality. Around the world, the emergence of COVID-19 has renewed focus on the importance of reducing income differences. We join in that chorus by arguing that policy action aimed at reducing income inequities could directly contribute to a reduced TB burden by reducing the chance of infection spread via contact between the poor and the rich.
1. Introduction

If A is poor and B is rich, it is reasonable to expect B to be healthier than A. By contrast, the mechanism by which the gap in income between A and B could matter for their individual health is less well understood. Does this gap matter in its own right, apart from the effect of incomes on health?

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, are more likely to smoke, abuse drugs, be obese, and face chronic stressors like financial hardship. These factors, in concert, can reduce immunity, increase vulnerability to disease, and cause poor health.

The presence of a health-income inequality gradient, on the other hand, is more controversial. 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. 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. These explanations are usually indirect and rely on interpersonal, group-identity comparisons, or indirect “general equilibrium” considerations.

Our broad question is, does income inequality have a first-order direct effect on health? Specifically, do countries with similar levels of per capita income but very dissimilar levels of income inequality have vastly different health? We narrow the focus of this question substantially by focusing on health outcomes that are impacted by disease, specifically communicable or infectious disease. We explore the hypothesis that the income-inequality link is in part mediated by infectious disease: in locations with greater income inequality and residence-work integration, infectious disease spreads more easily.

The underlying logic is simple and relies on three links.

- First, assume A and B live in a high-income-inequality area (X), meaning X has few rich people like B and many poor like A, so the median income is low, yet the mean income is high. Counterparts of A and B also live in a comparable area, Y, with a similar mean albeit higher median income than X.
- Second, suppose the indigent like A, due to the above-discussed health-income gradient, is significantly more prone to catching and harboring latent infections from an infectious disease, such as tuberculosis (TB).
Finally, suppose people like A and B randomly mix in social settings (such as schools, food markets, malls, stores, stadia, restaurants, public transportation, and so on).

We posit it is more likely for infections to spread or linger in area X than in Y. The point is a random encounter between two people in a low-income-inequality area is very likely a meeting between people close in income (or health). Such contacts are unlikely to lead to new infections. The opposite is true in a high-income-inequality area.

Figure 1: Tuberculosis incidence

A first look at cross-country data suggests prima facie evidence in favor of our hypothesis. Figure 1a plots 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 same year. A quick look at the figures indicates a clear, positive association in Figure 1a and a negative association in Figure 1b. Figure 1b suggests a positive cross-country health-income gradient, and Figure 1b of a positive health-income-inequality gradient.

While this first look at the data is provocative, many extraneous factors apart from inequality might explain a correlation. In this paper, 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 incidence 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 adopt an analytic approach that treats the relationship between inequality and the prevalence of non-communicable diseases – specifically diabetes and cancer – as a test of our hypothesis. Since these diseases are not communicable, if the hypothesis is correct, we should observe no relationship between income inequality and the prevalence of these diseases.

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 133 countries during 1995–2013.

2.1 Disease incidence and prevalence

We take data on TB incidence and prevalence from the global TB database of the WHO. Incidence is the estimated number of new (and relapsed) TB cases in one year per 100,000 population.\(^1\) Tuberculosis prevalence is the number of cases of TB per 100,000 population at a given time. The WHO provides an annual estimate of TB prevalence using estimates of incidences combined with modeled assumptions about disease duration in each country. TB prevalence is a measure of the direct exposure to the infection.

Our data on diabetes prevalence is from NCD-RisT\(^5\), 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.

Cancer incidence is taken from Institute for Health Metrics and Evaluation\(^6\), which collected cancer registry data for the period 1980-2010. It contains incidence of breast and cervical cancer incidence among women aged 15-79. Since these cancers develop only for women, we normalize incidence per 100,000 female population aged 15-79. We use these data to test our hypothesis that income inequality affects infectious disease spread but not the incidence of non-communicable diseases.

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

---

\(^1\) 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.
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 Gini coefficient data provided by the World Income Inequality Database (WIID). We exclude countries from our analysis where the data are reported as low quality or where the data are not representative of the country's entire geographic area or population. We extrapolate data points for years where data are unavailable by imputing information from the closest preceding observation.

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 are years of schooling (from Penn World Table), GDP per capita, public health expenditure (% to GDP), poverty, the share of people living in urban areas (from World Bank). Data on HIV prevalence and BCG coverage are added to control for medical conditions related to TB risk.

3. Methods & Estimation Strategy

3.1 Dynamic Panel Data Analysis

We construct a panel of 133 countries over the period 1995-2013 for the variables discussed above. Our dependent variables are new incidence of disease, which includes TB, diabetes, and cancer in a year; 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 of 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. The controlled variables are discussed in the previous section. In modeling the infectious disease, we include lagged prevalence as a measure of direct exposure, itself a function of TB incidence. 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 Arellano-Bover\textsuperscript{17}/Blundell-Bond\textsuperscript{18} methods to correct this problem. Technical details are provided in the
Appendix. The same is not required while modeling non-communicable diseases since diabetes or cancer do not spread through interaction, so direct exposure to the disease does not cause a new incidence.

To see whether Gini has a differential effect on TB incidence depending on how rich a country is, we use the 2018 World Bank classifications to divide our sample into high, middle, and low-income countries.

Results

The main results are presented in Figure 2. It shows the estimated coefficient along with its 95% confidence interval. The coefficient of Gini ($\beta$) in the TB estimation is near 0.5 and statistically significant at the 1% level. This means if Gini increases by 10%, TB incidence is predicted to rise by 5%. 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, it is predicted that TB incidence in X is 5% higher than in Y. In sharp contrast, Gini does not significantly affect diabetes or cancer, meaning ambient income inequality does not affect the probability of developing diabetes or cancer.

Higher average income significantly lowers TB incidence; the estimated coefficient is -0.27 and statistically significant at the 1% level. This means a 10% increase in income leads to a 2.7% decrease in TB incidence. Diabetes is positively associated with GDP per capita: diabetes prevalence increases with income in low- and middle-income countries, presumably due to the nutritional transition (diet and physical activity) sweeping these countries. Cancer incidence is not statistically correlated with income.

Figure 3 presents the results of the TB estimation but for different country subgroups. Income inequality is positively associated with TB incidence in low- and high-income countries and is statistically significant at the 10% level. The effect of GDP per capita was not statistically significant when comparing across similar income subgroups.

Ruling out the statistical-artefact possibility In cross-country multivariate regressions involving inequality measures, there is always the danger the findings are simply statistical artefacts. This is because a positive correlation between a population health indicator and income inequality can arise in aggregate level data even when the two variables do not affect each other in individual-level data. Suppose there is a non-linear relationship between income and health. In that case, two countries with the same average income will differ in their average health if they have different income distributions. The difference in health arises
because of an aggregation effect positively related to the difference in income inequality between the countries: the larger the income variance in the population, the larger the statistical artefact. If this is the case, a finding of a negative relationship between health and income inequality no longer implies inequality is a health hazard in and of itself; the relationship could be driven by the aforementioned aggregation effect.

Our approach to overcome this is to run similar estimations using country-level data (and hence, subject to the aggregation effect) for diabetes and cancer and then compare results across infectious and non-infectious diseases. In this case, TB is significantly associated with income inequality, while the non-communicable disease is not, strengthening our argument that infectivity interplays with income inequality. Bakkeli\textsuperscript{21} also finds income inequality does not have a statistically significant influence on individuals’ probabilities of having health problems in China where health is measured by blood pressure or obesity.

![Figure 2: Estimated elasticities of disease with respect to income inequality and income.](image)

Notes: Coefficients of Gini and GDP per capita from each regression from TB, diabetes, and cancer. Model estimated as described in the Methods & Estimation Strategy section. Estimations adjusted for years of schooling, GDP per capita, public health expenditure, poverty, 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.
Figure 3 Estimated elasticities of disease with respect to income inequality and income, by income groups.

Notes: Coefficients of Gini and GDP per capita from TB regression by income subgroups. Model estimated as described in the Methods & Estimation Strategy section. Estimations adjusted for years of schooling, GDP per capita, public health expenditure, poverty, the share of people living in urban areas, HIV prevalence, and BCG coverage. Error bars are 90%, 95%, and 99% CIs based on robust standard errors.

Discussion

Principal findings

This paper studies a connection between health and income inequality and finds the two related in the case of health affected by infectious disease. Longitudinal panel data from a large set of countries reveal that even after controlling for average income, income inequality within a country increases the incidence of an infectious disease, TB, but not the incidence of a non-communicable disease, such as diabetes and cancer. The direct effect of inequality is strong: All else same, countries with income-Gini coefficients 10% apart show a 5% difference in tuberculosis prevalence.
Comparisons with other studies  Nearly every study that explores the relation between health and income inequality uses a measure of mortality, such as life expectancy, as a proxy for health. Most studies report age-standardized or age-adjusted mortality rates; others report period life expectancy, combining age-specific mortality rates to create a period survival curve from which life expectancy can be estimated. The majority of such studies find consistent and strong effects: “in a relatively inegalitarian country life expectancy may be between five and ten years lower than in a more egalitarian country.”22 Our approach, instead, is to use incidences of diseases instead of a broad measure such as life expectancy. Our choice is partly guided by our research question, the connection between disease and inequality. It is also motivated by the fact that disease is a large, and more importantly, direct contributor to health, morbidity and mortality, unlike life expectancy, a more indirect measure that is impacted by wars, famines, and so on.

Another important point of departure for us is that, unlike existing work, we go beyond simply reporting a correlation and present a causal connection between health and income inequality. Our statistical approach (Arellano and Bover/ Blundell and Bond) offers a work-around the standard endogeneity concerns and, unlike extant work, permits us to claim causality.

Caveats and limitations  Our analysis hints at but is unable to test a causal mechanism connecting disease spread and income inequities. Such testing would require micro data on mixing patterns among individuals with different incomes. Our argument relies on the mixing being mostly random. In reality, it may not be entirely so with rich people mixing mostly with the other rich and deliberately staying away from places where mixing with the poor is unavoidable. Future research using contact tracing data may help shed more light on this matter.

The nature of the disease, its pathogenesis, is 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 communicable diseases from spread even under random mixing but to the extent the poor are less likely to receive the vaccinations, our message would still be active.
1. 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

2. Khullar D, Chokshi DA. *Health, Income, & Poverty: Where We Are & What Could Help.* 2018. doi:10.1377/hpb20180817.901935

3. 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

4. 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

5. 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

6. 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

7. 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:10.1007/978-978-981-10-2911-0_4

8. Do N. *The Effects of Poverty and Allostatic Load on the Development of Chronic Disease.* 2017.

9. Wagstaff A, van Doorslaer E. Income Inequality and Health: What Does the Literature Tell Us? *Annu Rev Public Health.* Published online 2000. doi:10.1146/annurev.publhealth.21.1.543

10. 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

11. Marmot MG. Understanding social inequalities in health. *Perspect Biol Med.* 2003;46(3 SUPPL.). doi:10.1353/pbm.2003.0056

12. Deaton A. Health, Inequality, and Economic Development. *J Econ Lit.* 2003;41(1):113-158. doi:10.1257/002205103321544710

13. Sapolsky RM. *Why Zebras Don’t Get Ulcers.* Owl Book/Henry Holt and Co.; 2004.

14. Kawachi I, Kennedy BP, Lochner K, Prothrow-stith D. Social Capital, Income Inequality, and Mortality. *Am J Public Health.* 1996;87(9).

15. 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
16. Forouzanfar MH, Foreman KJ, Delossantos AM, et al. Breast and cervical cancer in 187 countries between 1980 and 2010: a systematic analysis. *Lancet*. 2011;378(9801):1461-1484. doi:https://doi.org/10.1016/S0140-6736(11)61351-2

17. 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

18. 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

19. Popkin BM. Nutrition Transition and the Global Diabetes Epidemic. *Curr Diab Rep*. 2015;15(9). doi:10.1007/s11892-015-0631-4

20. Gravelle H. How much of the relation between population mortality and unequal distribution of income is a statistical artefact? *BMJ*. 1998;316(7128):382-385. doi:10.1136/bmj.316.7128.382

21. Bakkeli NZ. Income inequality and health in China: A panel data analysis. *Soc Sci Med*. 2016;157:39-47. doi:https://doi.org/10.1016/j.socscimed.2016.03.041

22. Rodgers GB. Income and inequality as determinants of mortality: An international cross-section analysis. *Popul Stud (NY)*. 1979;33(2):343-351. doi:10.1080/00324728.1979.10410449

23. Halter D, Oechslin M, Zweimüller J. Inequality and growth: The neglected time dimension. *J Econ Growth*. 2014;19(1):81-104. doi:10.1007/s10887-013-9099-8

24. Deininger K, Squire L. A new data set measuring income inequality. *World Bank Econ Rev*. 1996;10(3):565-591.
### Tables

| Variable         | Description                                                                 | Source                      | Mean    | Std. Dev. |
|------------------|-----------------------------------------------------------------------------|-----------------------------|---------|-----------|
| TB               | Estimated tuberculosis incidence rate (per 100,000 people)                   | WHO TB database             | 193.22  | 239.63    |
| **Prev**: TB     | Estimated number of cases of TB at a given point in time (per 100,000 people) | WHO TB database             | 241.01  | 245.30    |
| Gini             | Gini index                                                                  | UNU-Wider                   | 43.92   | 10.63     |
| GDP per capita   | Real GDP per capita (in 2010 USD)                                           | World Bank                  | 5374.10 | 10014.01  |
| Years of schooling | Average years of secondary schooling in the population aged over 25         | Penn World Table 9.0        | 6.52    | 3.21      |
| Health expenditure | Public health expenditure (% of GDP)                                      | World Bank                  | 3.03    | 1.62      |
| Poverty          | Poverty headcount ratio at $1.90 a day in 2011 PPP                           | World Bank                  | 24.31   | 25.00     |
| Urban            | People living in urban areas (% of total)                                   | World Bank                  | 48.85   | 20.31     |
| HIV prevalence   | Percentage of people ages 15-49 who are infected with HIV                   | World Bank                  | 2.53    | 5.11      |
| BCG              | BCG coverage (%)                                                            | WHO                         | 90.02   | 12.60     |
| Diabetes Prevalence | Proportion of adults over 18 who have diabetes.                        | NCD_RisC Institute for Health Metrics and Evaluation | 0.121   | 0.044     |
| Cancer incidence (Women) | Cancer incidence of breast and cervical cancer per 100,000 female population 15-79 | Institute for Health Metrics and Evaluation | 90.54   | 53.35     |

**Table 1: Description of variables & Summary statistics**
Appendix

Data

Income Inequality

World Income Inequality Database (WIID) collects Gini coefficient from various surveys, so it is possible that it has several estimates per country and year. It contains details of each survey and the quality of the estimates to enable the researcher to choose the appropriate estimate. Based on this, the inequality data is constructed as follows; keep only high and average quality data, and only include observations that cover the entire area and entire population of the country. If there is more than one observation that satisfied the conditions, income-based Gini is preferred over expenditure or consumption. If there were more than one Gini after the refinement, the following criteria were considered in the following order: income definition, income sharing unit, and unit of analysis. If the best possible Gini is consumption based, it is systematically lower than income inequality, so we follow the suggestion from Deininger and Squire (1996) to add 6.6 on expenditure-based Gini. If the coefficient is not available for one of the years, the observation is taken from the closest preceding observation, so it forms a step function.

Estimation Strategy

Our model estimating TB is as follows:

\[
TB_{it} = \gamma \text{Prev}_{i,t-1} + \beta \text{Gini}_{i,t} + x_{it} \delta + \zeta_i + \alpha_i + \nu_{it},
\]

where \(i\) denotes country and \(t\) represents time. The dependent variable is log of new TB incidence per 100,000 people, and the independent variable includes the prevalence of TB of the previous year, log of Gini index. In addition, we have a column vector \(x_{it}\) consisted of log of real GDP per capita, log of years of schooling, log of health expenditure, HIV prevalence, poverty headcount, share of people living in urban area, and BCG coverage. All variables are defined in Table 1. \(\gamma\) captures the effect of time-lagged TB prevalence on current TB incidence. Our parameter of interest is \(\beta\), the elasticity of new TB infections with respect to changes in the income Gini: If Gini increases by 1 percent, then new TB infections are predicted to increase by \(\beta\) percent. \(\alpha_i\) represent country fixed-
effects and $\zeta_t$ denotes time fixed-effects. 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. In our setting, 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 for the inequality changes.

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\(^{17}\) and Blundell and Bond\(^{18}\). GMM is a statistical method combining observed economic data with the information in population moment conditions to produce estimates of the unknown parameters. System GMM estimates equation (1) instrumenting endogenous variables using their first-differences and estimates the first-differenced equation of (1) by instrumenting the endogenous variables by their lagged values. The error term is assumed to be serially uncorrelated: the Arellano and Bond autocorrelation test rejected 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 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 effect. Third, the use of lagged dependent variables as (“internal”) instruments is an advantage, as valid external instruments are normally difficult to construct. When we estimate equation (1) by income subgroups, we use 5-year panel data to prevent the problem of number of instruments exceeds the number of countries and restricted to lag 2. High-income countries consist of the high and upper-middle income countries, middle-income countries are upper and lower middle-income countries, and low-income countries are countries classified as lower middle- and low-income countries by the World Bank in 2018.

Since non-communicable diseases do not spread through interaction, we do not need to control the prevalence of the disease estimating the association of non-communicable diseases and income inequality.
\[(2) \quad \ln(\text{incidence})_{it} = \beta_0 + \beta_1 \text{Gini}_{it} + x_{it} \delta + \zeta_{it} + \alpha_t + \nu_{it} \]

We estimate equation (2) using fixed effect estimation. Dependent variable uses two different diseases. The first one is diabetes, new diabetes incidence proxied by the difference of prevalence. The second one is the incidence of cancer; cervical and breast cancer per 100,000 female population.

**Aggregation Effect**

In cross-country multivariate regressions using measures of inequality, there is always the danger that the findings are simply statistical artefacts. To see this, suppose the health-income gradient is convex shaped. In that case, two countries with the same average income will differ in their average health if they have different income distributions. The difference in health arises because of an aggregation effect positively related to the difference in income inequality between the countries: the larger the income variance in the population, the larger the statistical artefact. If this is the case, a finding of a negative relationship between health and income inequality no longer implies inequality is health hazard in and of itself; the relationship could be driven by the aforementioned aggregation effect.

To ensure we are correctly identifying the extent to which income inequality directly raises TB infection risk, we correct for the aggregation effect by running a regression similar to (1) using diabetes (and, separately cancer) as the dependent variable. Our argument is that diabetes is also an aggregate measure so it also suffers from aggregation effect, but it is a non-communicable disease, and hence, does not has externalities. Thus, TB and diabetes both suffers from aggregation effect but the unique mechanism of infectious disease that infectivity interplays with income inequality is only present in TB. The difference on the association of the disease and income inequality will be because of infectivity.
### Appendix Tables

| 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        | https://www.wider.unu.edu/project/world-income-inequality-database-wiid |
| GDP per capita              | World Bank       | https://data.worldbank.org/                                           |
| Years of schooling          | Penn World Table 9.0 | 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/                                             |
| Diabetes Prevalence         | NCD_RisC         | https://ncdrisc.org/data-downloads-diabetes.html                      |
| Cancer incidence (Women)    | Institute for Health Metrics and Evaluation | http://ghdx.healthdata.org/record/ihme-data/breast-and-cervical-cancer-incidence-and-mortality-estimates-country-1980-2010 |

*Table A 1 Data sources and URL links*
| Variable                           | Coefficient | Std. Error | T-Value | Significance |
|-----------------------------------|-------------|------------|---------|--------------|
| L.in(prevalence)                  | 0.735***    | (0.103)    |         |              |
| ln(Gini)                          | 0.502***    | (0.182)    |         |              |
| ln(GDP per capita)                | -0.267***   | (0.088)    |         |              |
| ln(years of schooling)            | 0.162       | (0.204)    |         |              |
| ln(health expenditure)            | -0.060      | (0.103)    |         |              |
| HIV prevalence                    | 0.056***    | (0.010)    |         |              |
| poverty                           | -0.002      | (0.004)    |         |              |
| urban                             | 0.002       | (0.007)    |         |              |
| BCG                               | 0.008*      | (0.005)    |         |              |

Observations: 1,521  
Number of N: 95  
Instruments: 45  
AR(2): 0.710  
Hansen: 0.418  
Diff-in-Hansen: 0.253

Robust standard errors in parentheses  
*** p<0.01, ** p<0.05, * p<0.1

Table A 2 TB, Arellano-Bover/ Blundell-Bond estimation
|                           | High income (1) | Middle income (2) | Low income (3) |
|---------------------------|-----------------|-------------------|----------------|
| L.\ln(prevalence)        | 0.816***        | 0.773***          | 0.751***       |
|                           | (0.171)         | (0.107)           | (0.122)        |
| lnGini                    | 0.596*          | 0.248             | 0.762*         |
|                           | (0.340)         | (0.460)           | (0.418)        |
| ln(GDP per capita)        | -0.255          | -0.057            | -0.324         |
|                           | (0.186)         | (0.179)           | (0.276)        |
| ln(years of schooling)    | 1.016*          | 0.713***          | 0.176          |
|                           | (0.564)         | (0.207)           | (0.184)        |
| ln(health expenditure)    | -0.054          | -0.544**          | -0.457         |
|                           | (0.342)         | (0.275)           | (0.282)        |
| HIV prevalence            | 0.096**         | 0.080***          | 0.072***       |
|                           | (0.049)         | (0.019)           | (0.017)        |
| poverty                   | -0.028          | -0.002            | -0.002         |
|                           | (0.018)         | (0.006)           | (0.006)        |
| urban                     | 0.006           | 0.001             | -0.000         |
|                           | (0.011)         | (0.008)           | (0.011)        |
| BCG                       | -0.020*         | -0.008            | 0.003          |
|                           | (0.012)         | (0.009)           | (0.013)        |
| Observations              | 124             | 195               | 164            |
| Number of N               | 40              | 59                | 54             |
| Year FE                   | Yes             | Yes               | Yes            |
| Instruments               | 21              | 21                | 21             |
| AR(2)                     | 0.840           | 0.394             | 0.902          |
| Hansen                    | 0.358           | 0.424             | 0.588          |

Robust standard errors in parentheses

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

Table A 3 TB estimation by income group
|                                | Crude diabetes (1) | Crude diabetes (2) | Crude diabetes (3) | Crude diabetes (4) | Age-standardized diabetes (3) | Age-standardized diabetes (4) |
|--------------------------------|---------------------|--------------------|--------------------|--------------------|-----------------------------|-----------------------------|
| lnGini                         | -0.268              | -0.348*            | -0.318*            | -0.224             | 0.703***                    | 1.030***                    |
|                                | (0.215)             | (0.187)            | (0.191)            | (0.259)            |                             |                             |
| ln(GDP per capita)             | 0.866***            | 0.985***           | 0.703***           | 1.030***           |                             |                             |
|                                | (0.227)             | (0.194)            | (0.185)            | (0.267)            |                             |                             |
| ln(years of schooling)         | 0.640*              | 0.788***           | 0.196              | 0.336              |                             |                             |
|                                | (0.385)             | (0.295)            | (0.295)            | (0.273)            |                             |                             |
| ln(health expenditure)         | -0.163**            | -0.153**           | -0.151**           | -0.133*            |                             |                             |
|                                | (0.074)             | (0.072)            | (0.069)            | (0.074)            |                             |                             |
| HIV prevalence                 | -0.032**            |                   | -0.017             |                   |                             |                             |
|                                | (0.016)             |                   | (0.017)            |                   |                             |                             |
| poverty                        | 0.004               | 0.004              | 0.004              | 0.006*             |                             |                             |
|                                | (0.003)             | (0.003)            | (0.003)            | (0.003)            |                             |                             |
| urban                          | -0.027**            | -0.022**           | -0.045***          | -0.042***          |                             |                             |
|                                | (0.013)             | (0.011)            | (0.010)            | (0.011)            |                             |                             |
| Constant                       | -0.245              | -1.610             | 2.883              | -0.595             |                             |                             |
|                                | (2.180)             | (1.876)            | (1.842)            | (2.627)            |                             |                             |
| Observations                   | 1,755               | 2,112              | 1,684              | 1,990              |                             |                             |
| R-squared                      | 0.249               | 0.217              | 0.189              | 0.134              |                             |                             |
| Number of N                    | 104                 | 126                | 104                | 124                |                             |                             |
| Year FE                        | Yes                 | Yes                | Yes                | Yes                |                             |                             |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A 4 Diabetes, Fixed effect estimation
|                     | ln(cancer) | ln(cervical) | ln(breast) |
|---------------------|------------|--------------|------------|
|                     | (1)        | (2)          | (3)        |
| lnGini              | -0.020     | 0.004        | -0.070     |
|                     | (0.077)    | (0.082)      | (0.068)    |
| ln(GDP per capita)  | 0.062      | 0.080        | 0.040      |
|                     | (0.049)    | (0.057)      | (0.063)    |
| ln(years of schooling) | -0.108   | -0.144       | 0.170      |
|                     | (0.101)    | (0.117)      | (0.121)    |
| ln(health expenditure) | -0.013   | -0.002       | 0.011      |
|                     | (0.026)    | (0.022)      | (0.025)    |
| urban               | 0.006      | 0.002        | 0.010**    |
|                     | (0.004)    | (0.005)      | (0.005)    |
| Constant            | 3.844***   | 2.685***     | 2.879***   |
|                     | (0.637)    | (0.670)      | (0.782)    |
| Observations        | 1,829      | 1,829        | 1,829      |
| R-squared           | 0.256      | 0.256        | 0.621      |
| Number of N         | 132        | 132          | 132        |
| Year FE             | Yes        | Yes          | Yes        |

Robust standard errors in parentheses

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

Cancer incidence in column (1) is the sum of cervical and breast cancer.

Table A 5 Cancer, Fixed effect estimation