REGRESSION DISCONTINUITY AND HERD IMMUNITY

The Empiricist's Approach to Eliminating Tuberculosis

Team 11: Oscar Chang, Ben Goodman, Tim Marsh, and Tia Lim
Citadel DataOpen Executive Report

EXECUTIVE SUMMARY

process, conclusions, and content

THE EMPIRICIST'S APPROACH TO ELIMINATING TUBERCULOSIS

The World Health Organization (WHO) launched a campaign in 2015 to "End Tuberculosis (TB)." The aim is to reduce TB deaths by 95% from 2015 to 2035, and new TB cases by 90% over the same time range. Unfortunately, we live in a world where finite financial resources constrain the ability of even large organizations like WHO from achieving their goals. The impacts of TB are still tremendous; in 2016, over 1.7 million people died from TB. In order for such an ambitious campaign to be realized, the intelligent leveraging of financial and medical resources will be needed to maximize the reduction in TB incidences.

For this report, we look at one specific aspect of TB prevention in the form of the BCG vaccination. Within this sub-field of vaccination, we test the specific hypothesis that crossing a certain level of "herd immunity" leads to a discontinues "jump-down" of TB cases and deaths.

Given data from WHO, we first assess visually how TB immunization, TB incidences, and mortality data have changed over time for individual countries.

Then, using available data on BCG vaccine efficacy, we employ a quasi-experimental, regression discontinuity design (RDD) and find strong statistical evidence that our hypothesis is true. From both statistical and machine learning perspectives, we find significant evidence of **herd immunity effects** existing at **86% immunization**. For developing countries, we ultimately find that crossing that herd immunity threshold results in **7.3 fewer deaths per 100,000** people and around **95 fewer TB cases per 100,000** people.

Following these results, we propose a new policy initiative to allocate TB vaccination funding in a more efficient method that prioritizes the multiplicative effects of pushing countries over the herd immunity threshold to maximize funding impact. Hopefully, following similar analysis and resource allocation, the WHO can do more, with less money, and win the battle to End Tuberculosis.

SUMMARY OF CONTENTS

PROBLEM BACKGROUND

medical logic driving the analysis

TUBERCULOSIS & BCG

Despite being mostly eradicated in the developed world, tuberculosis (TB) is still one of the top 10 causes of death globally. In 2016, approximately 10.4 million people developed TB and 1.7 million people died from it, which amounts to 5,000 people a day[1]. The effects of TB mainly plague the developing world, which accounts for >95% of the cases and deaths[2]. As a result, routine immunization against TB is still common in high risk countries but less widespread across the developed world.

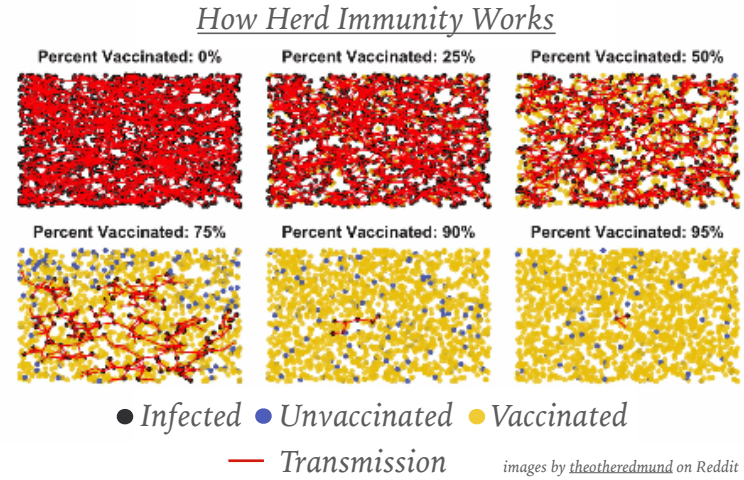
The world has committed to ending the TB epidemic by 2030. However, the World Health Organization (WHO) Director-General notes that existing actions and investments have not been effective in achieving this goal. This failure has compounding effects, since TB carries significant social and economic implications beyond just bad health, including poverty, stigma, and discrimination[3].

As of today, the only commercially available vaccination against TB is Bacillus Calmette-Guérin (BCG)[1]. A vaccine that was first used on humans in 1921 to protect against tuberculosis, BCG has been reported to have high efficacy against TB for children but modest and/or variable efficacy for older individuals. As a result, there has been ongoing development of other forms of vaccination against TB, as well as booster vaccines that aim to complement BCG. However, BCG is still considered the most reliable form of protection against tuberculosis today, and is the sole vaccination against TB that is recognized by the WHO.

BCG is usually administered via intra-dermal injection to newborns, due to a marked increase in efficacy when administered during the neonatal period and diminishing effectiveness on older children and adults.

In this report we propose an improved method for allocating TB vaccination funding that leverages an understanding and application of herd immunity.

HERD IMMUNITY



For infectious diseases, herd immunity describes the phenomena whereby disease immunity within a given population subset can diminish overall incidences of the disease within the entire population by preventing the spread of the disease. Essentially, when a high percentage of the population is vaccinated, diseases are 'roadblocked' since there are not many people who can get infected, inhibiting transmission of the disease to other susceptible pockets of the population. In the above graphic, this is seen in the large drop in transmission when vaccination levels jumps from 75 to 90%.

Herd immunity only takes effect when a large majority (>80%) of the population is immune, with the herd immunity threshold level for a specific disease varying by specific population characteristics and how contagious the particular disease is.

Some groups of people who cannot be safely vaccinated depend on herd immunity for protection. These groups include people with damaged immune systems, people on chemotherapy, people with HIV, newborns, and the elderly.

The **multiplicative effects of reaching herd immunity thresholds** is the core line of reasoning behind our research exploration, model development, and ultimately our policy recommendations. By finding a herd immunity threshold for TB, we are able to recommend more efficient vaccine allocations when funding is limited.

ANALYTICAL APPROACH

empirical approach

HYPOTHESIS

Herd immunity significantly lowers TB incidence and mortality rates. We test this hypothesis with the B_1 parameter in the regression discontinuity model outlined below. If our hypothesis is valid, we expect to see a statistically significant B_1 coefficient. This hypothesis will be tested from both a statistical and machine learning perspective.

QUESTION & APPROACH

Policy makers need tangible actions to take when confronting public health issues regarding infectious diseases. As a result, our empirical analysis should suggest causal relationships.

In the literature, the herd immunity effect is described as an all-or-nothing cutoff. A population either has herd immunity, or it does not. We assume that for any given country, regardless of the time period, and other characteristics of the country, the country desires to have a population immunization level that is as high as possible. When given a collection of countries, after controlling for individual characteristics, we would expect countries close to herd immunity to randomly distribute on either side of the herd immunity boundary. That is, a country may randomly fall short (or not) of herd immunity one year due to random factors like natural disasters.

If a herd immunity effect exists, then we would expect a country that attains herd immunity to see a sharp decrease in tuberculosis death and incidence rates.

This framework naturally leads us to pursue a regression discontinuity analysis. While we cannot randomly assign whether or not a country has achieved herd immunity, we can use regression discontinuity to parse out a herd immunity effect; both in terms of whether or not it exists, and what the magnitude in the data appears to be.

Based on thresholds determined by regression discontinuity, policy makers can leverage the herd immunity effect to efficiently pursue cost-effective immunization programs that maximize funding utilization.

REGRESSION DISCONTINUITY MODEL

Regression discontinuity applies to situations where treatment is assigned based on a cutoff point along the range of a continuous variable. It treats observations falling to the left of a cutoff value as a quasi-control group and those falling to the right as a treatment group. It then determines if there is a non-linear change in the outcome variable when moving from the control group to the treatment group. Our hypothesis is that when crossing the herd threshold from the control group (which is below the herd effect level), the additional population immunity will result in a herd effect and a non-linear drop in the outcome (TB incidence rate). As a result, regression discontinuity is a rigorous, quasi-experimental way to provide evidence for herd immunity having a causal effect on disease prevalence and mortality. With i denoting some observation, our model takes the form:

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 I_i + \vec{\beta_3} X_i + D_i \vec{\beta_4} X_i + \beta_5 year_i + \epsilon_i$$

Y_i is the dependent disease variable we are fitting, be it mortality due to the disease per capita, or disease incidence per capita. D_i is a dummy variable which takes on the value of 1 if the observation has crossed the herd immunity threshold, and 0 otherwise. We control for immunity percentage with I_i terms, and for other control variables with the feature vectors X_i. We allow coefficient estimates to change for both I_i and X_i if $D_i = 1$, disambiguating potential confounding affects with our coefficient of interest, B_1, once the herd immunity threshold has been crossed. We also include country-level fixed effects in our model to control for different baselines in a disease's treatment and incidence within a population due to intangible factors outside the scope of the control variables included in our model.

DATA PRE-PROCESSING

about the data

DESCRIPTION OF DATA

Our analysis depended on six raw datasets: The immunization, mortality, health_indicators, and country_codes datasets provided in the Competitor's Packet, as well as external datasets on world GDP per capita (PPP) by country across time from the World Bank, and on tuberculosis incidences by country across time from the WHO.

Table 1: Raw Datasets before Pre-Processing

Dataset name	#Datapoints	#Features	Source
immunization.csv	849	40	Competitor's Packet (WHO)
mortality.csv	1006027	37	Competitor's Packet (WHO)
health_indicators.csv	12936	246	Competitor's Packet (WHO)
country_codes.csv	284	2	Competitor's Packet (WHO)
tuberculosis_incidence.csv	3848	12	WHO
gdp_Capita_PPP.csv	264	64	World Bank

Table 2: Datasets obtained after Pre-Processing

Dataset name	#Datapoints	#Features
mortality_measles.csv	189	270
mortality_neonatal.csv	159	270
mortality_pneunomia.csv	391	270
mortality_polio.csv	483	270
mortality_rota.csv	79	270
mortality_tuberculosis.csv	567	292
TB_burden.csv	3,850	47

CLEANING

We processed the diseases listed in the mortality data and found 6 diseases—Tuberculosis, Measles, Neonatal Tetanus, Polio, Rotavirus and Streptococcus Pneumococcus Infections—that had a corresponding vaccination in the immunization dataset. We filtered and split the mortality and immunization datasets into 6 smaller datasets.

The filtering step was done by examining the *cause code* column of *mortality.csv* and manually determining if the cause code was related to the disease. This was tricky, labor intensive and necessary, because an automatic fuzzy search (as opposed to human search) will count deaths not preventable by the given vaccine, for example pneumonia caused by non-Streptococcus bacteria.

For each disease-vaccine dataset pair, we performed a left-join on *country* and *year* with columns from the *health_indicators*, *country_codes*, and *gdp_Capita_PPP* datasets. The *mortality* dataset includes years from 1988 to 2017, while the *immunization* dataset includes years from 1980 to 2017, so the pre-processed dataset has *NA*s in some of the relevant rows. The same goes for the *country* information, where the *mortality* dataset includes 135 countries and the *immunization* dataset includes 192 countries.

Finally, we removed rows corresponding to years where there were less than 5 years of data found in both the *mortality* and *immunization* datasets.

Focusing on the TB data

We would have liked to have analyzed all diseases for which we have vaccine data. However, lack of data prevents a rigorous analysis of measles and rota. Tetanus and neonatal vaccinations protect against non-communicable diseases, and this medical fact makes herd immunity irrelevant for these diseases. Lastly, Polio seems to have been primarily eradicated, with very low variability, which is not ideal for our analysis.

Hence, we decided to perform our analysis primarily on TB using both mortality and incidence data.

DATA PRE-PROCESSING

engineering the data

CONTROLLING VARIABLES

We analyze the effect of immunization rates on TB incidence rates (per 100,000 people) and mortality rates. That being said, an observation's TB incidence depends on more than just immunization rates. We chose control variables for the following reason:

We use four core controls: 1. **Year**, to capture how overall medical quality has improved; 2. **GDP per capita (PPP)**, to proxy overall societal development and sanitation (see next page); 3. **Health care expenditure per GDP** to proxy the strength and sophistication of a country's healthcare system; 4. **Death rate per 1,000 people,** if people die a lot in a country, it may be do to war or other factors which drive up TB contagion and death; 5. Lastly, **country fixed-effects**, which incorporate the intangible country-level traits which may affect deaths (geography, urban vs. rural distribution, government, etc.).

FEATURE ENGINEERING

We engineered extra features to understand how the data interacts and has changed over time.

Metric for Relative Performance

As part of immunization policy guidance, it is important to understand how vaccination rates are developing in different countries relative to their peers, so we created a relative performance metric to call out the lowest performing countries.

This metric was created to analyze the vaccination rates in individual countries over time from 2000 to 2017. We first calculated the percentage growth (or decline) of vaccination by country. These results were then calibrated against the country with the strongest rates to understand relative lagging growth. We then made a further magnification to call out the countries that are lagging the most.

Though this diminished the interpretability of the actual metric, it created a dataset that allowed for easy geographic visualization, as presented on page 7. By transforming our data to tailor it to our analysis, we were able to clearly display our findings in a digestible format.

Factor Variable for Herd Immunity

We also needed to track whether a population reached herd immunity. We did this by creating a new factor column in our relevant datasets that encoded "1" if the country's vaccination rate was above X% in the given year for herd threshold X, and "0" if not.

Adjusting to Per Capita

To account for differences in population, we adjusted GDP, health expenditure, and mortality rate data to be per capita to allow for comparison across countries.

Adjusting to Different Age Groups

The datasets sourced from the Competitor's Packet contained different age formats, so we aggregated them at the coarsest level possible to use all the data, and adjusted mortality rates based on different age groups.

Interaction Variables

As part of the regression discontinuity design, interaction variables were incorporated to ensure that the herd effect is isolated from potential changes in control variable behavior post-herd.

DATA LIMITATIONS

We might be over-counting mortality, if the BCG vaccine has variable effect on different variants of the TB disease, and undercounting immunization rate, if some people are vaccinated when they get older than 1 year old.

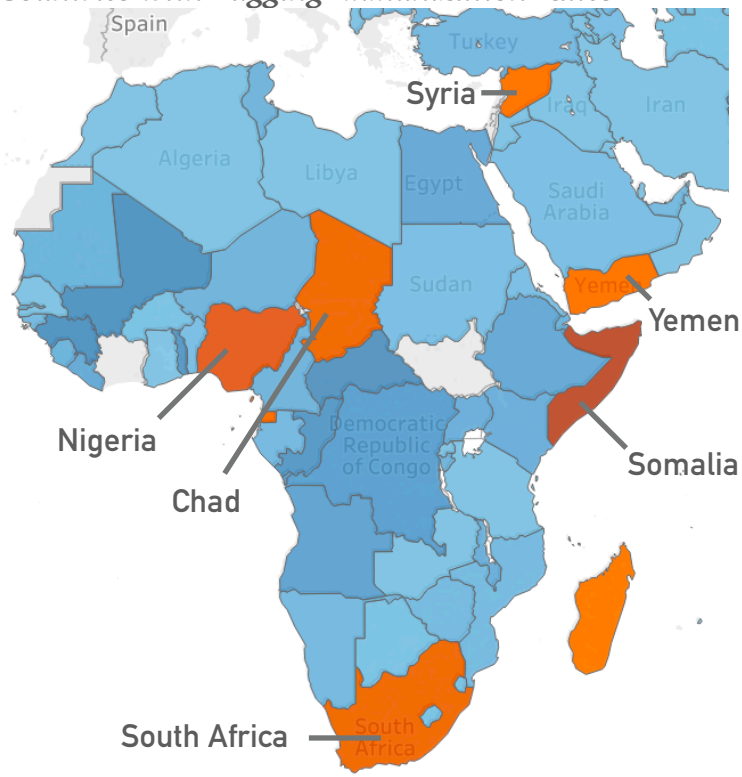
There may be some time discrepancy between immunization year and mortality in that year. For example, some people might have gotten the vaccine after contracting the disease or the herd immunity effect may go into some time after the appropriate immunization level has been reached.

It is difficult to account for these potential limitations found in the data. Fortunately, the effects go in both directions, which means these are more random, but less systematic errors.

LAGGING COUNTRIES

As part of an informed policy recommendation, we include an analysis of TB immunization progress over time to identify specific countries that are falling behind their peers on immunization rate improvement. Details of the metric used for this analysis are discussed under "Metric for Relative Performance" on page 6. The goal is to identify specific countries that need to be recognized as having slipping immunization rates.

Countries with Lagging Immunization Rates



The above graph shows that Somalia (-46%), Nigeria (-40%), Chad (-40%), South Africa (-40%), Syria (-33%), and Yemen (-33%) are some of the worst performing countries that have shown decline in TB immunization rates from 2000-2017. The current policies and funding should therefore be explored in these countries to understand how progress can be made, to ensure they don't continue falling behind.

Data Considerations

Though South Africa has low vaccination rates, they also display low TB incidences, which could reflect the country's transition from a developing country in 2000 towards a more developed country that no longer faces TB challenges in 2017.

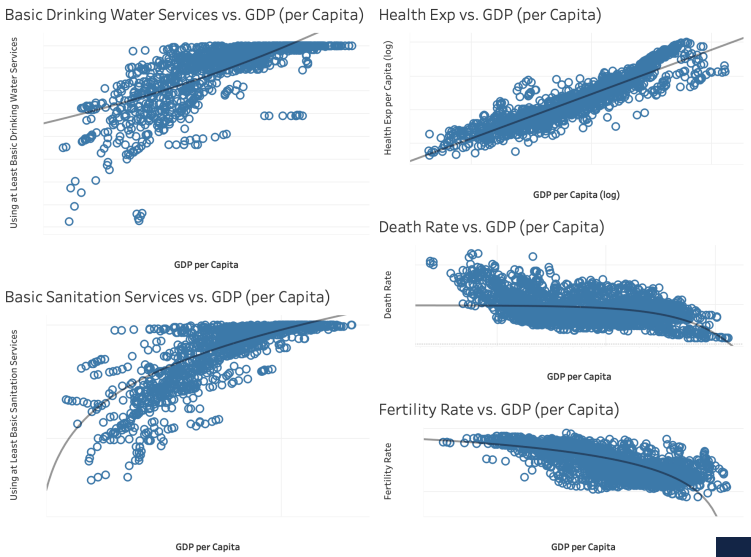
In addition, Somali has been trapped in an ongoing civil war since 2009. Other low performing countries like Yemen and Syria have also been war ridden during the analysis horizon. This situation could have two possible effects on the data and this graph: 1. TB vaccination rates severely dropped as a result of the war so this graph is an accurate representation of where attention is needed. 2. WHO data recording of TB vaccination rates are underreporting the actual rates since data gathering was blocked by war.

GDP AS A CONTROL

GDP per capita (PPP) (GDPc) is one of our core control variables. This EDA shows that GDPc is an accurate conglomerate of other health data that may interact with TB incidence, mortality, and vaccination rates. Two core health statistics (basic drinking water and sanitation services) both show relatively strong positive correlations with GDPc, implying GDPc helps account for their variation. Similar results can be drawn for health expenditures per capita and fertility rates. Using GDPc as a proxy is particularly useful because we have many more observations for it than the health data.

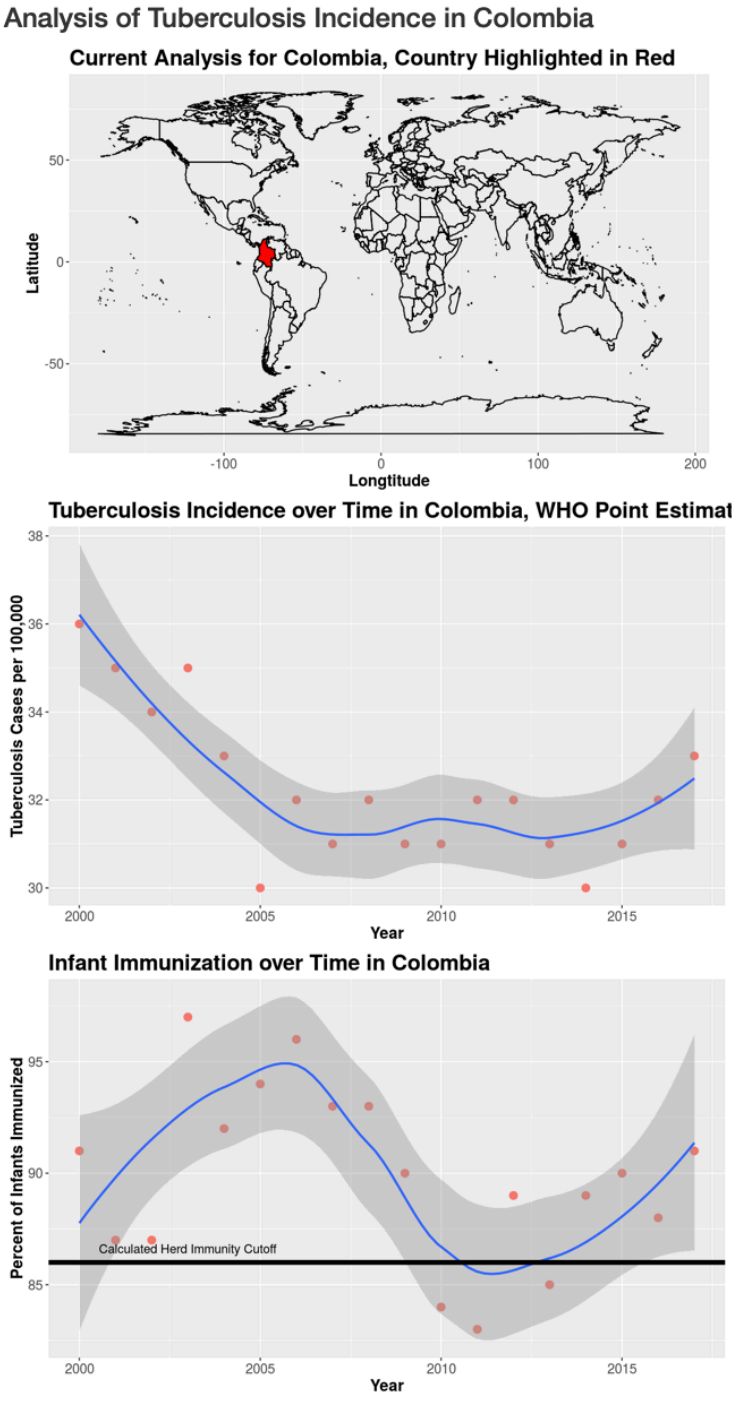
Death rate does not show any strong correlation; as a result, it is also used as an additional control in our model.

EDA on GDP as a Control Variable



HERD IMMUNITY GUI

Before diving into a statistical analysis and other modeling to establish the herd immunity effect, we first needed to casually explore the data to see if this is a trend worth pursuing. Also, our GUI was updated after our modeling to include the herd immunity cutoff of 86% in the existing version as a reference point. The current GUI can be explored by clicking the example GUI screenshot of Colombia below, linking automatically to your web browser:



GUI Explained

This GUI allows a user to toggle between countries to observe whether herd immunity effects are prevalent at the country level. As in the example with Colombia, the country is highlighted on the world map and users observe the country's TB incidences and infant immunization rates over time. The plots include both point estimates and their associated trend lines over time, with error band estimates.

In addition, our TB dataset from WHO included point estimates along with upper and lower bound values. The user can also toggle between these three options to observe any differences between these base dataset values.

In this example, we see that Colombia dips below herd immunity around 2012. After this incident, TB incidences seem to become concave and begin trending upward. Similar observations were seen for other countries.

Interpreting the Results

Based on our analysis comparing data from a variety of countries, while trends are not overwhelmingly supportive, we felt comfortable moving forward with more statistical herd immunity modeling.

Interesting Finding in Developed Countries

EDA on the GUI unveiled Sweden and Ireland as having rather peculiar trends. Although both countries maintain low vaccination rates, well below herd immunity, they also maintain incredibly low incidences of TB.

Further research revealed that developed countries actually do not maintain high vaccination rates because TB is essentially extinguished from these populations already. Therefore, they can only immunize vulnerable populations without risk of widespread TB infection. This finding complements our previous EDA.

From the previous section "Lagging Countries," this GUI explains South Africa's trend in-line with these developed countries, making their lagging immunization rates less worrisome. In addition, it explains why most developed countries, like the USA, were excluded from our original datasets.

THRESHOLD SELECTION

We employed two methods to select the threshold for which the herd immunity effect is most prominent, i.e. the cutoff in the Regression Discontinuity Model. We consider integer thresholds from 0 to 99, where a threshold of 0 corresponds to the basic Linear Regression model.

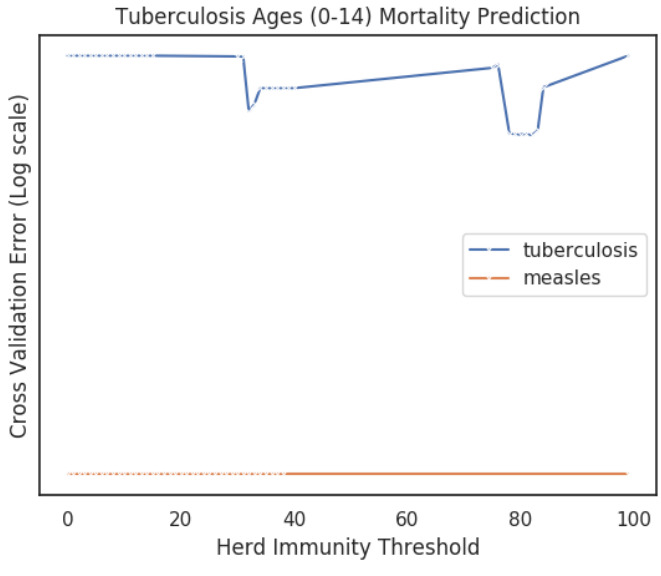
Method 1: Subsampling Cross-Validation

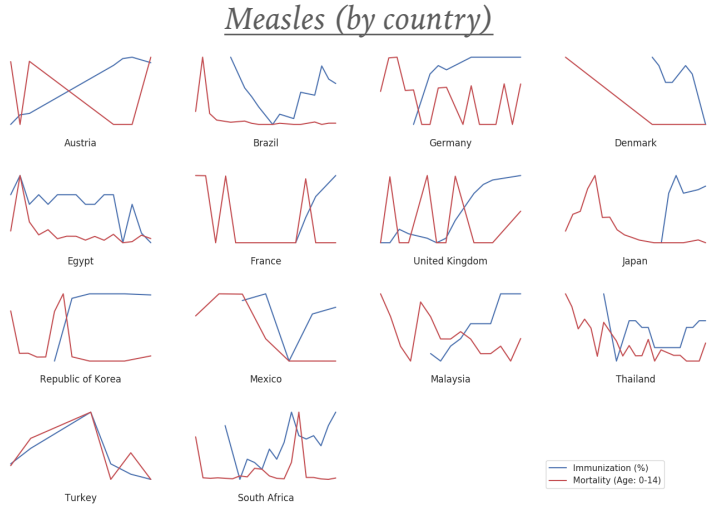
We repeatedly split all the data into training and testing datasets. For each of 100 splits, we use the testing dataset to evaluate the models fitted with the training dataset with a mean squared error loss function.

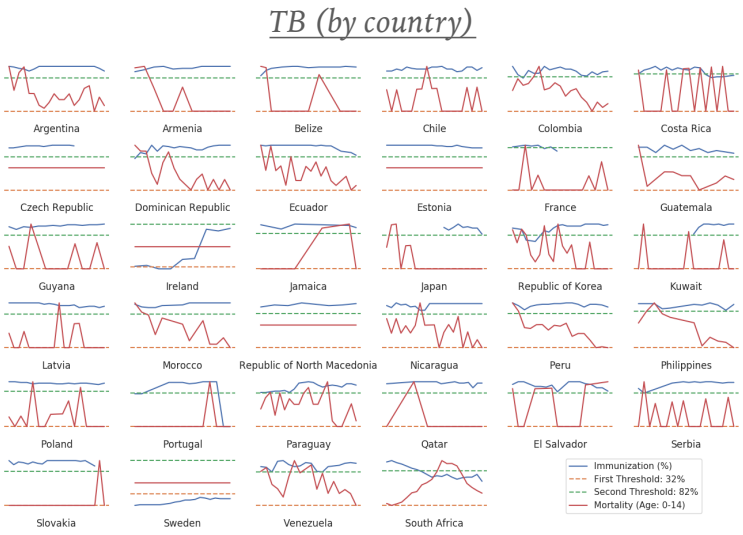
The average error of all the splits forms the statistic for our cross validation. If we plot out the cross validation error across our considered thresholds, we find two points at which there is a local minimum. We visualize the prominence of this effect in TB by contrasting it with measles, where the effect is clearly absent. This also reaffirms our decision to focus on TB.

The two thresholds for TB mortality chosen for the model across age groups 0-4, 0-14, and all ages correspond to (33, 80), (32, 82), and (32, 80) respectively (see figure below). If we combine all the age groups into a single target variable, we get a model which minimizes cross validation error at (33, 80).

The two thresholds that minimize TB incidence







The uniformity of the recommended thresholds justify the use of the Regression Discontinuity model over the basic Linear Regression Model. We can interpret the higher threshold as the *strong* herd immunity effect and the lower one as the *weak* herd immunity effect.

Figures (Above):

In the figures above, we observe that countries which stay above the strong herd immunity threshold (the green line) generally see a decreasing trend in mortality. The weak herd immunity effect (the orange line) displays a similar effect, but it is most prominent in two countries: Ireland and Sweden, where the general state of healthcare is extremely high, and thus even lower immunization rates are effective.

The existence of these two separate effects suggest that we should prescribe different solutions for countries with different levels of wealth and healthcare to maximize the efficiency of their healthcare spending on immunization.

Method 2: Hypothesis Testing

To further confirm that our findings from before are significant, we tested our Regression Discontinuity model against the null hypothesis that the herd immunity effect does not decrease incidence and mortality rates of TB when the immunization rate is already accounted for. We found that the most statistically significant herd immunity threshold on only TB mortality is 86%, which roughly agrees with the threshold found using cross-validation, giving us additional confidence in this selection.

The Final Model:

At the 86% threshold, our results are consistent whether we are analyzing the herd immunity effect on TB Mortality or the range of WHO tuberculosis incidence estimates. Consistent with the WHO data, all outcomes are results per 100,000 people. **We find that the herd immunity threshold is both robust and statistically significant. Given our model design, this provides strong evidence that attaining herd immunity causes a drop in TB incidences and mortalities.**

Moving to coefficient interpretation, we can interpret the Mortality coefficient as follows: **crossing the herd threshold results in 7.309 fewer deaths per 100,000 people**, all else being equal. Additionally, we can interpret the TB incident point-estimate (P.E.) as follows: **crossing the herd immunity threshold results in around 95 fewer TB cases per 100,000 people**, all else being equal. The other model's interpretations naturally follow this structure (U.B. = Upper Bound), (L.B. = Lower Bound).

In terms of the control variables, the positive coefficients on the Percent Immunized variables should be discussed. One would think that the more immunizations, all else being equal, the less disease incidences. However, almost all of the Percent Immunized coefficients are positive. This may be due to a directionality issue; that is, countries with more TB could be more likely to have more robust immunization programs to try to counter TB prevalence.

Otherwise, the other control variables have coefficients moving in the direction we would expect.

Table 1: Primary Regression Discontinuity Results for Tuberculosis (TB), Herd Immunity Cutoff at 86%

	TB Mortality	TB Incident P.E.	TB Incident U.B.	TB Incident L.B.
	(1)	(2)	(3)	(4)
Percent Immunized	−0.014	0.420**	0.723***	0.182
	(0.014)	(0.192)	(0.269)	(0.139)
Percent Immunized ∗Herd	0.133***	0.530	0.543	0.517*
	(0.025)	(0.402)	(0.561)	(0.290)
Herd	−7.309***	−95.529**	−111.838**	−80.729***
	(2.335)	(38.034)	(53.102)	(27.506)
Death Rate (per 1,000)	0.709***	14.368***	19.049***	10.279***
	(0.118)	(1.068)	(1.491)	(0.772)
Death Rate (per 1,000) ∗Herd	−0.729***	6.008***	7.620***	4.491***
	(0.082)	(0.910)	(1.270)	(0.658)
GDP per Capita (PPP)	0.00002	−0.00000	0.00003	0.00000
	(0.00002)	(0.001)	(0.001)	(0.0004)
GDP per Capita (PPP) ∗Herd	−0.00005**	−0.0001	0.00000	−0.0001
	(0.00002)	(0.001)	(0.001)	(0.0004)
Health Expend. per Capita (PPP)	−0.0003	0.005	0.009	0.002
	(0.0003)	(0.007)	(0.009)	(0.005)
Health Expend. per Capita (PPP) ∗Herd	0.0004	−0.014**	−0.019*	−0.010**
	(0.0003)	(0.007)	(0.010)	(0.005)
Year	−0.054***	−0.899***	−1.696***	−0.320
	(0.017)	(0.302)	(0.422)	(0.218)
Constant	104.778***	1,840.890***	3,462.352***	662.279
	(34.902)	(611.297)	(853.473)	(442.086)
Country FE	Yes	Yes	Yes	Yes
Observations	781	2,270	2,270	2,270
R^2	0.879	0.965	0.965	0.959
Adjusted R^2	0.864	0.962	0.963	0.956
Degrees of Freedom	1.101 (df = 695)	38.962 (df = 2109)	54.398 (df = 2109)	28.177 (df = 2109)
Note:				*p<0.1; **p<0.05; ***p<0.01

SMART FUNDING MODEL

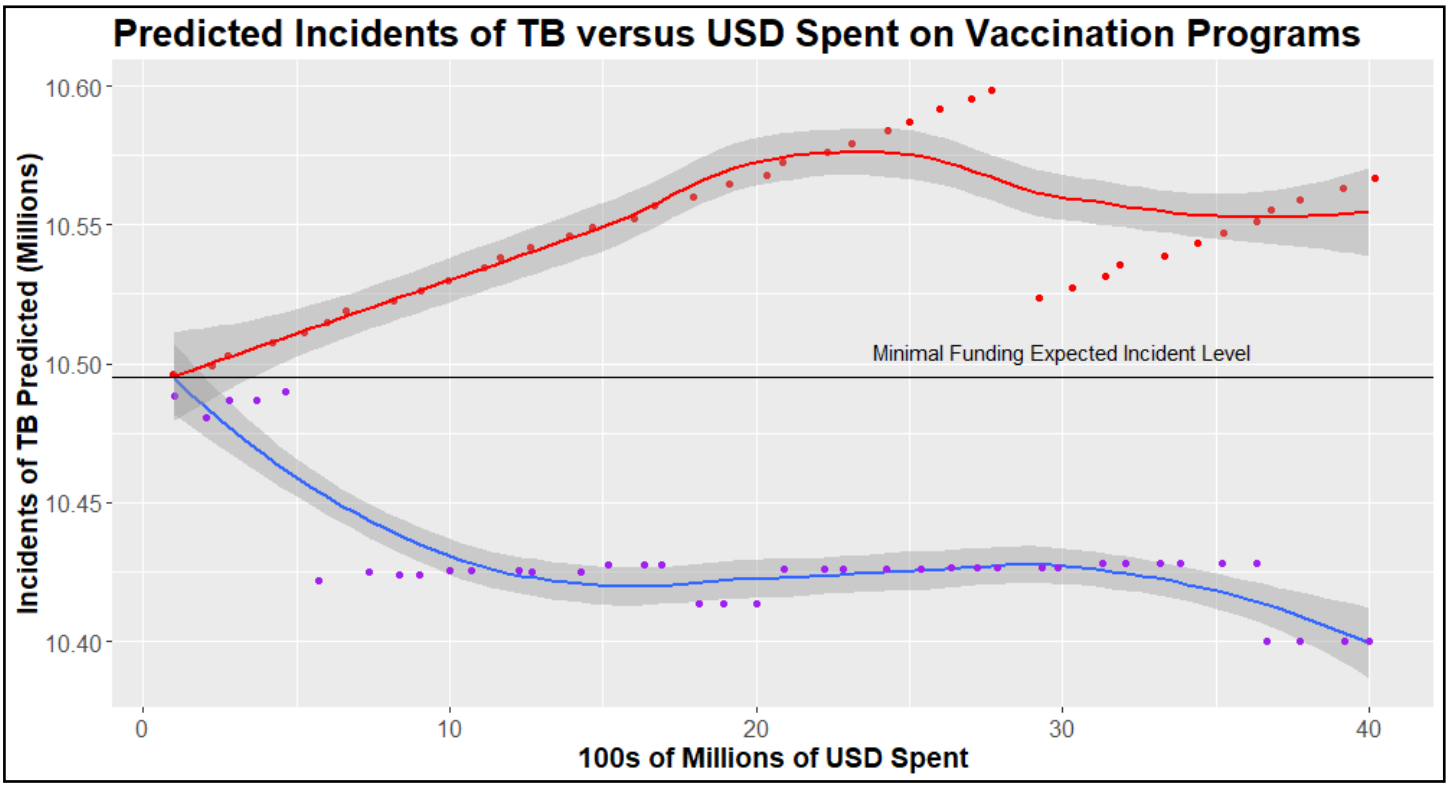
As a result, based on the model presented and shown to be robust over a variety of different conditions, our analysis implies that immunization funding would see **improved returns if allocated to push countries over the herd immunity threshold**. These returns can be measured with both TB incidence rates and mortality rates.

The fundamental idea is that by pushing populations across the TB herd immunity threshold of 86%, the TB incidence and mortality rates show a significant decline relative to the progress leading up to the threshold. In addition, after this threshold is sufficiently crossed, any additional vaccinations see decreasing returns to scale, since this segment is already benefiting from herd immunity effects.

We recommend the **prioritization of vaccination funding based on how close countries are to hitting herd immunity by "number of immunizations to go."** Therefore, assuming equal populations across countries, some country A with 84% vaccination rates should be prioritized over country B with 50% vaccination rates because the funding needed to get A to herd immunity will reduce tuberculosis incidence much more than if that same funding was applied to country B by leveraging the Herd Immunity effect.

To create the below graph, we established two allocation schemes. (1) The red line and dots correspond to results predicted by our model if funding was distributed evenly per capita amongst the bottom 40% of countries by income. (2) The blue line and dots correspond to results predicted by our model if funding is targeted towards countries for whom it is easiest to pass the herd immunity threshold (ie, less new vaccinations needed to cross the threshold). A single immunization was assumed to cost $50 in both allocation schemes.

While by no means should the point forecasts be taken as gospel (for example, it is unclear how increasing vaccinations leads to more tuberculosis — this may be a directionality issue within the model), **the qualitative implications of the forecasts are clear: Those who are serious about reducing Tuberculosis incidents could operate more efficiently and effectively by leveraging funding to push countries past the herd immunity threshold.**



FUTURE WORK

We were unable to capture all relevant detail that would have ideally produced a more accurate model due to time constraints and shortcomings in the available data. Below, we discuss these limitations, and suggest possible extensions for a more comprehensive study of this problem in the future.

Currently, there is no consensus on the lifespan of the BCG vaccine's efficacy. Different studies on BCG from various countries report protection duration from as low as 15 years to as high as 60 years. That being said, WHO suggests that young-to-middle-aged adults are most susceptible to tuberculosis[3], which gave us sufficient confidence that BCG would be sufficient to protect adults from the disease during their most vulnerable period. To improve the rigor of our model, though, we would, with sufficient data, ideally be able to perform a new regression for each country to be able to control for internal country climates that affect the efficacy of the BCG vaccine. In our proposed analysis we attempted to control for these internal country traits using country-level fixed effects.

In fact, further accuracies might be obtained if we could performing our analyses on cities rather than countries. This is because the herd immunity threshold varies according to population density, as we expect dense cities to have a higher threshold value as compared to sparser communities. Identifying a separate threshold for each community would allow us to identify the specific communities that would gain the most from increased immunization rates. However, we realize that this solution is highly unlikely due to the lack of immunization data by community.

Last in terms of data, incidences of tuberculosis are closely linked to that of HIV, where likelihood of contracting tuberculosis disease increases up to 30-fold for individuals with HIV[2]. Ideally, our model would have been able to control for HIV, as tuberculosis mortality trends might closely mimic that of HIV. We became aware of this too late in the competition to adjust.

In our funding model, we assumed a constant cost per dose of BCG vaccine, as we were unable to find data on the exact costs of the vaccine by country. We acknowledge reality is not quite so simplistic, and other factors such as transportation overheads and even corruption need to be taken into account before making informed, cost-effective policy decisions.

Also, in the funding model, we used *Incidents of Predicted TB* as our dependent variable in which we want to suppress. We might obtain further insight by using *Cost per Life Saved from TB* as the dependent variable, as this would enable policy makers to compare exact tradeoffs in immunization spending. Intuitively, the cost per life saved should also be at its highest when a country's immunization rate falls just short of the herd immunity threshold.

BIBLIOGRAPHY

[1] "BCG vaccines: WHO position paper – February 2018". Releve Epidemiologique Hebdomadaire. 93 (8): 73–96. 23 February 2018. PMID 29474026.
[2]Tuberculosis (TB). (2018, September 18). Retrieved March 31, 2019.
[3]World Health Organization Strategic and Technical Advisory Group for TB. Use of high burden country lists for TB by WHO in the post-2015 era (discussion paper). Geneva: WHO; 2015 (https://www.who.int/tb/publications/global_report/high_tb_burdencountrylists2016-2020.pdf, accessed 31 March 2019).
[5]Dockrell, H. M., & Smith, S. G. (2017). What Have We Learnt about BCG Vaccination in the Last 20 Years?. *Frontiers in immunology*, *8*, 1134. doi: 10.3389/fimmu.2017.01134
[6]Dowdy, D. W., & Chaisson, R. E. (april 2009). The persistence of tuberculosis in the age of DOTS: Reassessing the effect of case detection. *Bulletin of the World Health Organization, 87*(4), 296-304. Retrieved March 31, 2019.