Is There a Correlation Between Vaccination Rates and General Mortality/Economic Performance in a Country?

An Empirical Investigation

James P. Schwann

POLS 3704: Research Design - Data Analysis

December 16, 2022

INTRODUCTION:

As of December 2022, roughly six-hundred and fifty million confirmed Coronavirus cases have been recorded, and approximately six and a half million people have perished from the disease ("Coronavirus Death Toll"). The emergence of Covid-19, SARS-COV-2, and the subsequent devastation it has wrought around the world has fundamentally changed many aspects of how the modern world functions. Today, there is a greater emphasis on remote work/education, e-commerce, and health precautions such as masking, amongst other things. Whether or not these practices continue is a question that can only be answered as more time passes ("Covid 19: Facing the Pandemic of the 21st Century").

While the virus continues to spread, its effects have been nullified due to the miraculous vaccine rollout that took place in record time. By December 2020, the "... FDA approved the use of Pfizer's Comiranty vaccine - the first Covid-19 vaccine" ("COVID-19: Facing the Pandemic of the 21st Century"). By January 2021, vaccines manufactured by both Moderna and Pfizer were being given to healthcare workers. The Covid 19 pandemic highlighted the importance of vaccinations in the fight against prevalent viral diseases. Had no vaccine been made, the number of confirmed deaths from Covid 19 would be much larger; variants like the Delta and Omicron that emerged later would have had much more devastating effects on populations in the Western World and across the globe (COVID-19: Facing the Pandemic of the 21st Century").

Unfortunately, however, human irrationality "reared its head," as it always seems to do in moments of crisis. The lockdowns in many nations inspired specific sects of the population to question their government's and health authorities' decision-making. So much so, in fact, that conspiracy theories began propagating throughout the United States. A video titled "Plandemic" emerged that theorized that the super elites of the world - those in government, the ultra-rich, and others - had come together and planned the Covid 19 pandemic ("Plandemic Revisted: A Product

of Planned Disinformation…"). In response to this video, citizens began to resist vaccination against Covid 19 and other vaccinations. Over time, these ludicrous theories have proven to be demonstrably false, yet the vaccination statistics where vaccines are readily available are quite abysmal. Currently, only 68% of the eligible population in the United States is fully vaccinated against SARS-COV- 2. In addition, only 33% of the population has received a booster dose against the virus ("US Coronavirus Vaccine Tracker"). These statistics are disheartening when one realizes that the vaccine significantly reduces the risk of symptoms and hospitalization associated with SARS-COV-2 infection and that many developing nations across the globe have yet to be delivered a plentiful amount of vaccines ("COVID-19 Vaccine Effectiveness and Safety").

Given what has occurred over the last two years with the Coronavirus pandemic, I thought it would be enlightening to investigate if certain causal relationships and correlations exist between immunization rates in countries and macro variables of interest. In this report, I examine the immunization rates for Measles, Polio, DPT [Diphtheria-Tetanus-Pertussis], and BCG [Bacille Calmette-Guerin/Tuberculosis] over a period spanning a total of forty-one years: 1980-2021. I aim to determine the potential effects these vaccination rates have on variables such as the Crude Death Rate (per 1000 people), Death from Communicable Diseases, and economic indicators like GNI per Capita (Gross Net Income) and the Unemployment Rate in the countries of interest.

The literature on this topic has grown significantly over the last few decades as technology has enabled better data collection in the health/medical industry. Researchers have often examined the relationship between an economic measure like GDP Growth/GDP and the vaccination rates for a specific country/locale. In fact, in a paper titled "Vaccination and GDP

Growth Rates: Exploring the Links In a Conditional Convergence Framework," the authors conclude that:

"The link between better health and better economic performance is by now well established, and many vaccines have been shown to generate significant ROI in a classical cost-benefit framework" (Masia et al.).

Furthermore, in a paper published by the Center for International Development at Harvard University, authors Amar Hamoudi and Jeffrey Sachs assert the following:

"The correlation between health and economic performance is extremely robust across communities and over time" (1).

Thus, it is clear that some relationship exists between a country's economic outcomes and its people's health status. In addition, it is clear that in the 21st century, conscientious objectors to vaccination are causing the loss of herd immunity, leading pockets of the population to experience higher mortality rates for various infectious diseases like measles, amongst other diseases ("Becoming a vaccine champion…"). Overall, this report emphasizes the importance of vaccination measures by highlighting some of the consequences and benefits higher vaccination rates may garner for specific countries.

RESEARCH METHODOLOGY:

(A) Outcome of Interest & Explanatory Variables:

In this investigation, I wanted to analyze two sects of distinct response variables: Mortality/Death and Economic Performance Indicators. Thus, the response variables I analyzed for Mortality/Death were as follows: the Crude Death Rate within the country (per 1000 people) and the Cause of Death by Communicable Diseases (% of Total in the Country). In addition, the response variables I analyzed to determine Economic correlation/potential casualty were GNI per

Capita (Calculated via the Atlas Method) and the Total Unemployment Rate (% of total). These four variables are the response variables of interest in the regression analysis and other statistical measures calculated.

The Independent/Explanatory variables I examined to determine the degree of correlation/estimate potential causality in the relationships of interest were as follows: the BCG Vaccination Rate, Polio Vaccination Rate, Measles Vaccination Rate, DPT Vaccination Rate, Current Health Expenditure in the Country (% of GDP), and Domestic General Government Health Expenditure in the country (% of GDP). These variables were used as the primary regressors for the regression and descriptive statistical analysis, such as examining frequency distributions like histograms.

(B) Unit of Analysis:

The unit of analysis/entity of interest in this report is a country. The dataset consists of twenty countries I subsetted into two groups based on specific parameters. The first group is a set of countries labeled generally as "Developed Countries/Economies." This group consists of the following countries: **the United States**, **Canada**, **Australia**, **the United Kingdom**, **Ireland**, **Germany**, **France**, **Norway**, **Switzerland**, and **Mexico**. The second set of countries is labeled as "Developing Countries/Economies." This group consists of the following countries: **Turkmenistan, Belize, Bolivia**, **Philippines**, **Indonesia**, **Uzbekistan**, **Libya**, **Tonga**, **Paraguay**, and **Jordan**. I split the data into two groups of countries to conduct a better analysis: avoid confounding. I created these groups using the United Nations criteria known as the HDI: Human Development Index. The HDI is a robust measurement that considers a country's life expectancy, health, education, and quality of life to determine one's score. The best HDI score possible is 1.0, whereas a score of 0.8 and below indicates a label of "developing" ("Developing Countries.

2022"). The first group of ten countries is considered "Developed," whereas the second group is considered "Developing" based on their HDI scores.

Objectively, a country like the United States will have higher vaccination rates and better economic measures than a country like Turkmenistan. Furthermore, the possibility for confounders between these two countries is extremely high, given the noticeable cultural, geographical, and other differences that exist between the two. To strengthen the analysis, I examined each group's Explanatory/Response variables of interest separately. This separation ideally would lead to better causal estimates of the variables I examined.

(C) Description of Dataset:

I retrieved the data I used for this analysis from the World Bank Database, which contains data on "Health Nutrition and Population Statistics" in several countries across the globe. This data was free and is currently accessible on the World Bank website ("Health and Nutrition Population Statistics", *Databank*). In total, there are 845 observations/data points across fourteen variables of interest in the dataset: **Years** (1980-2021), **Country Name/Country Code** (name of country), **Measles Vaccination Rate**, **Polio Vaccination Rate**, **DPT Vaccination Rate**, **BCG Vaccination Rate**, **Crude Death Rate** (per 1000 people), **Death by Communicable Disease** (% Total), **GNI per Capita**, **Unemployment Rate** (% Total), **Domestic Health Expenditure** (% GDP), **Domestic Government Health Expenditure** (% GDP).

(D) Justification for Variables:

The response variables in the dataset capture the measures of interest. The Crude Death Rate and Death by Communicable Disease are relatively reasonable measures of general mortality in a country. Furthermore, economic indicators like the Total Unemployment Rate and

GNI per Capita are relatively decent measures of economic performance for a country. Thus, these variables represent the factors I wanted to analyze.

(E) Descriptive Statistics/Figures:

Descriptive Statistics Table for Response Variables Developed Countries/Economies

Descriptive Stats	DeathRate	DeathDisease	GNI	Unemployment Rate
Mean	8.433	6.602	33853	6.455
Median	8.600	5.799	29940	5.635
IQR	2.400	2.651	30752.5	3.87825
Standard Deviation	1.755	2.687	22230.82	2.790985

Descriptive Statistics Table for Response Variables Developing Countries/Economies

Descriptive Stats	DeathRate	DeathDisease	GNI	Unemployment Rate
Mean	6.537	18.968	2629.5	7.737
Median	6.213	18.774	1920.0	5.225
IQR	1.814	9.311	2647.5	7.266
Standard Deviation	1.989	7.150	2258.9	5.617

Histogram Distributions of Vaccination Rates:

Figure 1:

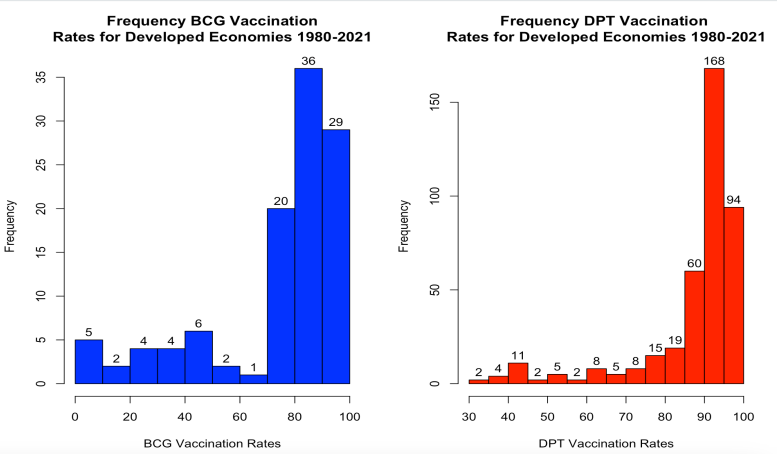


Figure 2:

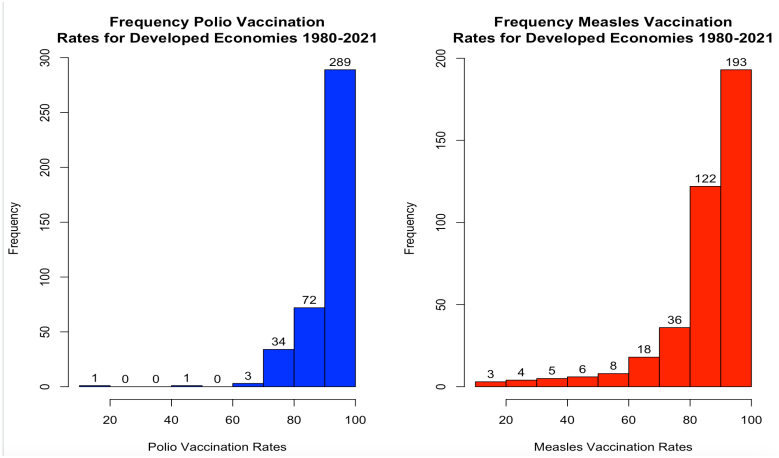


Figure 3:

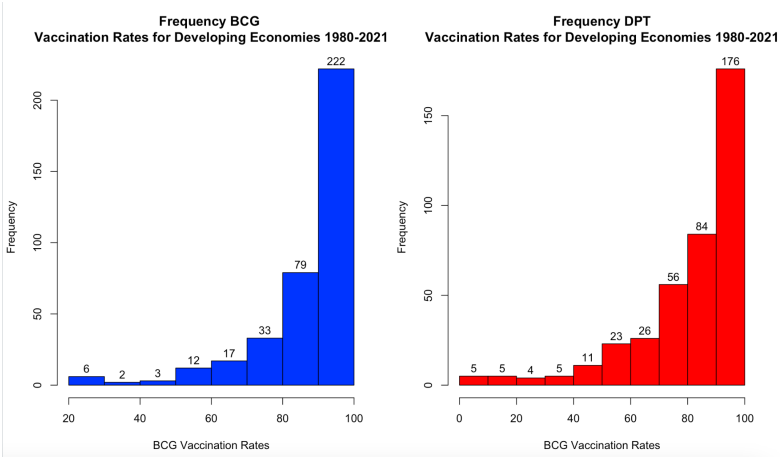
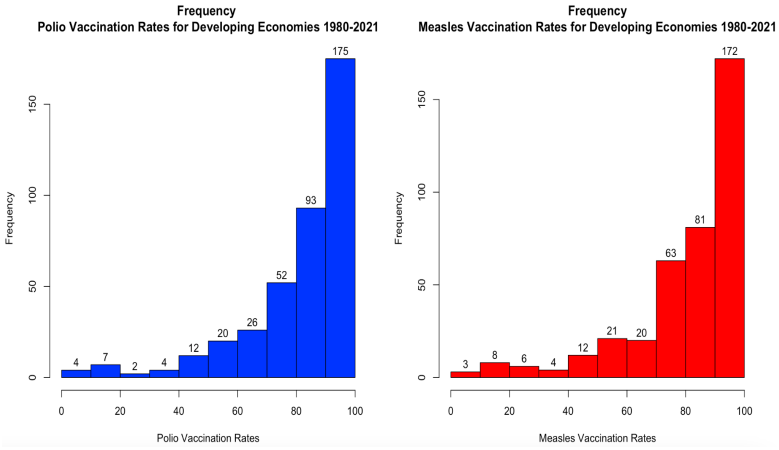


Figure 4:



RESEARCH DESIGN/SPECIFICATION:

In this investigation, I aimed to identify a potential causal link/correlation between vaccination rates for four prevalent diseases - Measles, Polio, DPT, & BCG, and the following response variables: Crude Death Rate, Death by Communicable Disease, GNI and Total Unemployment Rate. Understanding that an extensive analysis of both Developed and Developing Economies may have led to skewed results, I subsetted the data into two groups: Developed & Developing Economies/Countries. Following this, I performed the same operations on each group to ensure that my analysis was consistent and robust. The main instrument of analysis was regression. I created Least Squares Regression Lines and then conducted subsequent F-tests - H0: B1, B2, B3, etc., = 0 and Ha: At least one of the Bs is nonzero - to determine potential significance. In addition, I used both entity and time-fixed effects in the regressions to account for potential confounders. Entity/Country fixed effects account for the differences that could arise from factors specific to each country but remain stable over time: cultural makeup, demographic/ethnic makeup, and other factors. However, Time/Year fixed effects account for changes that vary over time: economic performance (booms & busts), political affiliations/allegiance, and other factors. Overall, this investigation was entirely an observational study. I analyzed data that was observed/collected over time, which requires specific methods/tools to avoid introducing a significant amount of artificial bias. Unfortunately, there will always be some degree of bias due to the fundamental problem of causal inference.

RESULTS:

Table 1: Regression Output Developed Economies Fixed Effects Mortality

	Dependent variable:	
	DeathRate	DeathDisease
	(1)	(2)
BCGVax	0.021***	0.090
	(0.002)	(0.043)
POL3Vax	−0.018*	−4.857
	(0.009)	(4.454)
MEASVax	−0.012	0.314
	(0.008)	(0.265)
DPTVax	−0.030***	4.819
	(0.006)	(4.717)
Constant	13.535***	−18.788
	(0.777)	(25.678)
Observations	104	9
R^2	0.975	0.935
Adjusted R^2	0.954	0.484
Residual Std. Error	0.367 (df = 57)	2.771 (df = 1)
F Statistic	47.535*** (df = 46; 57)	2.070 (df = 7; 1)
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 2: Regression Output Developed Economies Fixed Effects Economic Indicators

	Dependent variable:	
	GNI	UnemployRate
	(1)	(2)
MEASVax	−20.310	−0.034
	(211.497)	(0.051)
BCGVax	−150.571***	−0.003
	(36.260)	(0.009)
POL3Vax	1,523.406	0.284
	(1,587.711)	(0.385)
DPTVax	−613.992	−0.247
	(1,625.525)	(0.394)
HealthExpend	−4,621.436**	3.598***
	(1,967.873)	(0.477)
GovExpend	5,258.149*	−2.401***
	(2,855.695)	(0.693)
Constant	−41,639.610**	−8.481*
	(19,121.690)	(4.638)
Observations	48	48
R^2	0.990	0.985
Adjusted R^2	0.976	0.964
Residual Std. Error (df = 20)	2,766.114	0.671
F Statistic (df = 27; 20)	72.333***	47.110***
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 3: Regression Output Developing Economies Fixed Effects Mortality

	Dependent variable:	
	DeathRate	DeathDisease
	(1)	(2)
BCGVax	−0.035***	0.509**
	(0.006)	(0.245)
POL3Vax	−0.002	−0.031
	(0.008)	(0.484)
MEASVax	0.017**	0.117
	(0.008)	(0.231)
DPTVax	−0.047***	−0.896
	(0.010)	(0.543)
Constant	11.441***	48.560***
	(0.484)	(12.056)
Observations	358	39
R^2	0.880	0.637
Adjusted R^2	0.859	0.555
Residual Std. Error	0.750 (df = 304)	4.831 (df = 31)
F Statistic	41.946*** (df = 53; 304)	7.775*** (df = 7; 31)
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 4: Regression Output Developing Economies Fixed Effects Economic Indicators

	Dependent variable:	
	GNI	UnemployRate
	(1)	(2)
POL3Vax	0.351	0.008
	(32.615)	(0.047)
MEASVax	10.435	−0.080**
	(24.377)	(0.035)
DPTVax	5.310	0.017
	(40.610)	(0.059)
BCGVax	−15.269	0.030*
	(11.193)	(0.016)
HealthExpend	−194.578*	0.556***
	(117.107)	(0.167)
GovExpend	−352.839**	−0.341
	(176.140)	(0.251)
Constant	4,386.268**	11.314***
	(2,158.991)	(3.114)
Observations	188	190
R^2	0.866	0.947
Adjusted R^2	0.837	0.935
Residual Std. Error	848.727 (df = 153)	1.226 (df = 155)
F Statistic	29.140*** (df = 34; 153)	80.988*** (df = 34; 155)
Note:		*p<0.1; **p<0.05; ***p<0.01

The tables above illustrate the fixed effects regression results for the two groups within the data: Developed & Developing Economies/Countries. I ran a set of four regressions for each group. The first was a regression where I regressed the Crude Death Rate (Y_{it}) onto regressors BCG Vaccination Rate (BCGVax - X_1), Polio Vaccination Rate (POL3Vax - X_2), Measles Vaccination Rate (MEASVax - X_3), and DPT Vaccination Rate (DPT Vax - X_4). The second

regression was one in which I regressed Death by Communicable Disease (Y_{2it}) onto the same four regressors. For the second set of regressions, I regressed GNI (Y_{3it}) onto the same four vaccination rate regressors plus Domestic Health Expenditure (HealthExpend - X_5) and Domestic Government Health Expenditure (GovExpend - X_6). For the last regression, I repeated the same setup as the previous regression but changed the response variable from GNI to the Total Unemployment Rate (Y_{4it}).

DISCUSSION:

The results demonstrate that the regressions provide mixed signals about the relationships I aimed to identify/quantify. In the first table, the negative slope coefficients B_2 & B_4 are significant at the 0.10 (10%) and 0.01 (1%) alpha levels, respectively. The value of -0.018 demonstrates that the linear model predicts, on average, that a one percent increase in the Polio vaccination rate will lead to a roughly 0.018 decrease in the Crude Death Rate in a Developed Country. Similarly, the value of -0.030 demonstrates that the linear model predicts, on average, that a one percent increase in the DPT vaccination rate will lead to a roughly 0.030 decrease in the Crude Death Rate in a Developed Country. The regression results demonstrate a potential negative correlation between these regressors and the Crude Death Rate. That being said, the Measles regressor in this regression is not significant. Therefore, its directional effect on the response variable cannot be successfully determined[1]. Lastly, the BCG regressor has a positive slope coefficient which is the opposite result one would expect. Examining the second regression in Table One, only B_1 is significant at the 0.05 level, but this regressor has a positive slope coefficient which is illogical. Overall, this second regression, unfortunately, is not very

[1] Note - This regressor's confidence interval would contain the value of 0 making its effect unknowable

informative. It does not adequately elucidate the relationship between vaccination rates and Death by Communicable Disease.

The second table contains mixed results as well. Initially, the hypothesis was that vaccination rates would be positively correlated with GNI and negatively correlated with the Total Unemployment Rate. Thus, an increase in vaccination rates would - on average - lead to an increase in GNI, whereas an increase in vaccination rates would decrease the unemployment rate. Unfortunately, the regression results in Table Two are not definitive. In these regressions, the regressors that show statistical significance are: B_2, B_5, B_6. The B_2 estimate is significant at the 0.01 level and comes out to be roughly -150.5. This negative correlation, however, does not make sense. B_2 demonstrates that our linear model predicts that for every increase in the BCG vaccination rate, we should expect GNI to decrease by about 150 on average. This result seems illogical. In addition, the B_5 estimate also seems somewhat illogical for a similar reason. The only slope coefficient that makes sense in context is B_6. The slope coefficient estimate of roughly 5,258 illustrates that the linear model predicts that for every increase in government health expenditure, GNI is expected to increase, on average, by about 5,258. The B_6 estimate illustrates that increased government spending leads to greater economic performance: a principle of Modern Monetary Theory. For the second regression in Table Two, we see that slope coefficients B_5, B_6 are significant at the 0.01 level. In context, these estimates are logical. The positive slope coefficient on B_5 demonstrates that greater consumer spending on healthcare will increase unemployment. In other words, a LESS healthy population will lead to a reduction in the workforce. Conversely, the negative coefficient on the GovExpend variable demonstrates that the

linear model predicts a one percent increase in government expenditure (% of total GDP) will lead to a roughly 2.4 percent decrease in the unemployment rate.

Tables Three & Four contain the regression results for the Developing Economies/Countries within the dataset. In the first regression, slope coefficients B_1 , B_4 are statistically significant at the 0.01 level. Both slope coefficients are negative, demonstrating a possible negative correlation; if the BCG or DPT vaccination rate increases by one, the regression demonstrates that the Crude Death Rate, on average, should decrease by about 0.035 and 0.047, respectively. Once again, the positive coefficient on the Measles regressor is illogical in context, and the B_2 is not statistically significant. Like the Developed Economies, the second regression in Table Three did not produce the desired results.

Furthermore, the GNI regression in Table Four does not include significant slope coefficients. While most vaccination regressors have positive slope coefficients - logical in context - none are significant. Lastly, the slope coefficient for the Measles regressor in the Unemployment regression is significant at the 0.05 level.

CONCLUSION/CHALLENGES:

In conclusion, while some of our regressions indicate that the original hypothesis may hold some validity, they do not overwhelmingly confirm innate suspicions. The fixed effects regressions do not illustrate overwhelming evidence that Mortality/Death by Communicable diseases is negatively correlated with vaccination rates, that GNI is positively correlated with vaccination rates, and that the Unemployment Rate is negatively correlated with the vaccination rates. While some slope coefficients from our regression output are significant, not all are. Some slope coefficients are significant, but the sign of these coefficients is the opposite sign of the relationship initially expected. Thus, our final results demonstrate a mixed picture at best.

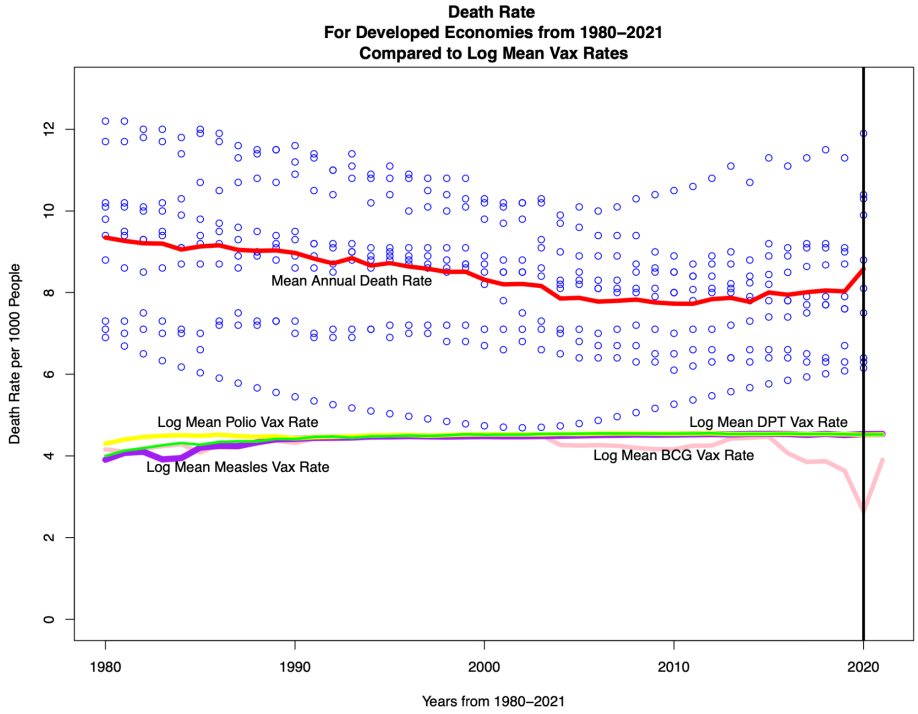
That being said, visible barriers/limitations exist that could influence the results to come out this way. For one, “missingness” needs to be addressed. The dataset contains 845 observations, yet most of the regressions only have data from a couple hundred data points, if that. The regression where “Death Disease” is the response variable only contains 9 observations for Developed and 39 observations for Developing Economies. NA data points most definitely made the regression results differ from the expected outcomes. If only so much data is available, one outlier or one “edge case” data point could influence the estimated least squares regression fit. In this circumstance, missingness - no doubt - contributes a degree of error to the final results. Unfortunately, health data is not readily available/collected compared to other data types. Initially, I wanted the dataset to span from 1960-2021. However, when I examined the data, most data points were missing for my fourteen variables throughout the '60s and '70s. This is why the "Years" variable in the dataset ranges from 1980-2021. Most likely, the regression results would have been better if missingness was not such an issue.

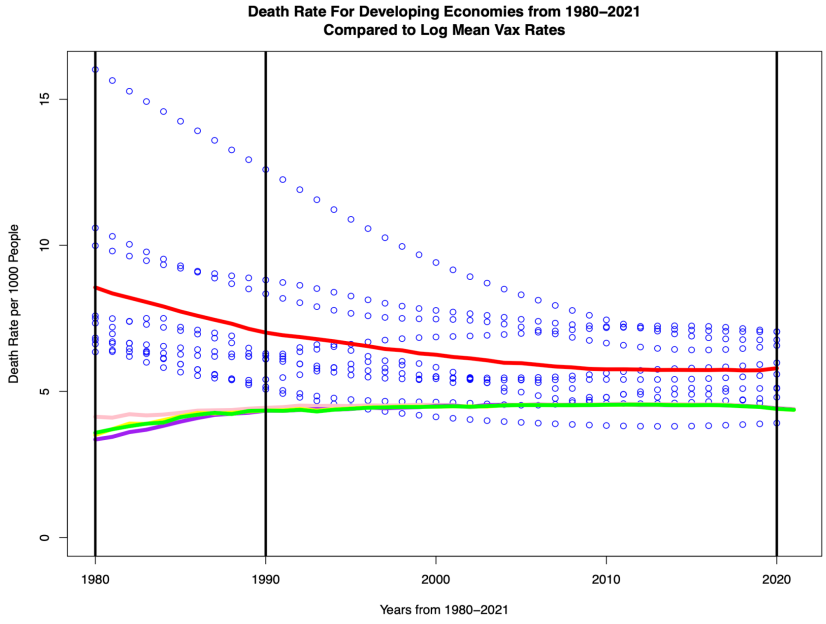
Furthermore, the regressions may also be experiencing an endogeneity problem: a circumstance in which explanatory variables change in response to a change in the response variable (“Endogeneity Problem”). One can use specific statistical tools to reduce this issue: implementing a wide array of control variables, making the Y more specific, or introducing a larger sample size. Unfortunately, none of these options were directly available. Therefore, it is reasonable to assert that the slope coefficients are inflated/biased because of these fundamental problems.

Despite the issues at hand, I think my initial hypothesis could be readily confirmed if I had a more capable dataset and I could better control for omitted variable bias - OVB - in addition to endogeneity problems. Despite the apparent issue of missingness, some of the slope

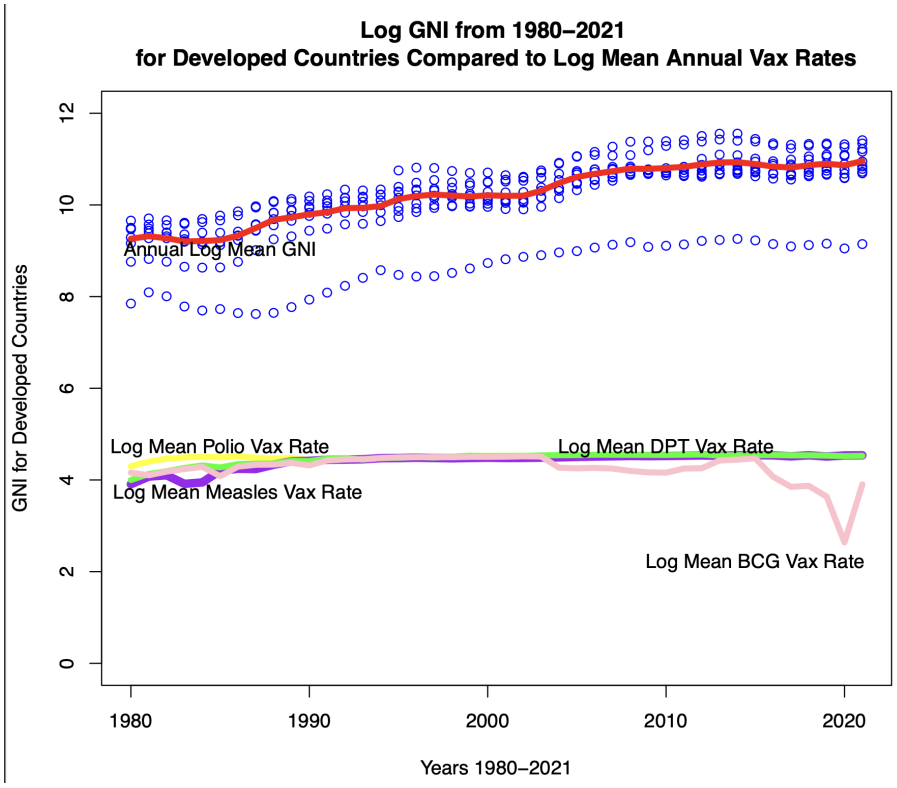
coefficients are significant in the direction I initially predicted. I am confident my hypotheses could be better confirmed if the analysis was more robust. This study/report demonstrates the current problem plaguing many health investigations: lack of concrete data. Hopefully, as time goes on, health data collection will become more efficient so that better conclusions can be made.

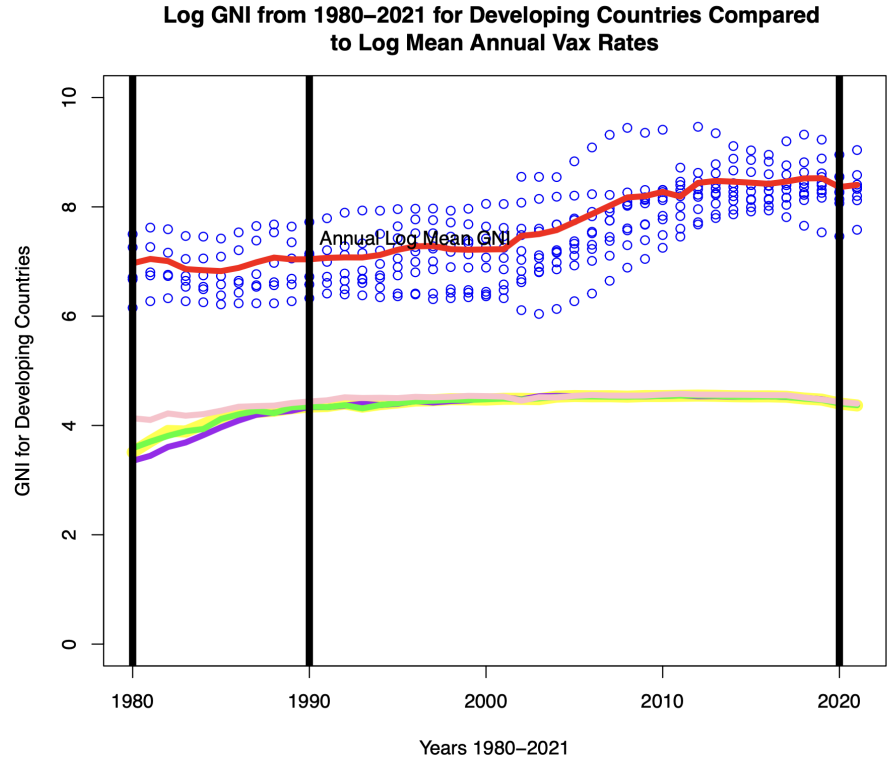
APPENDIX: [2]

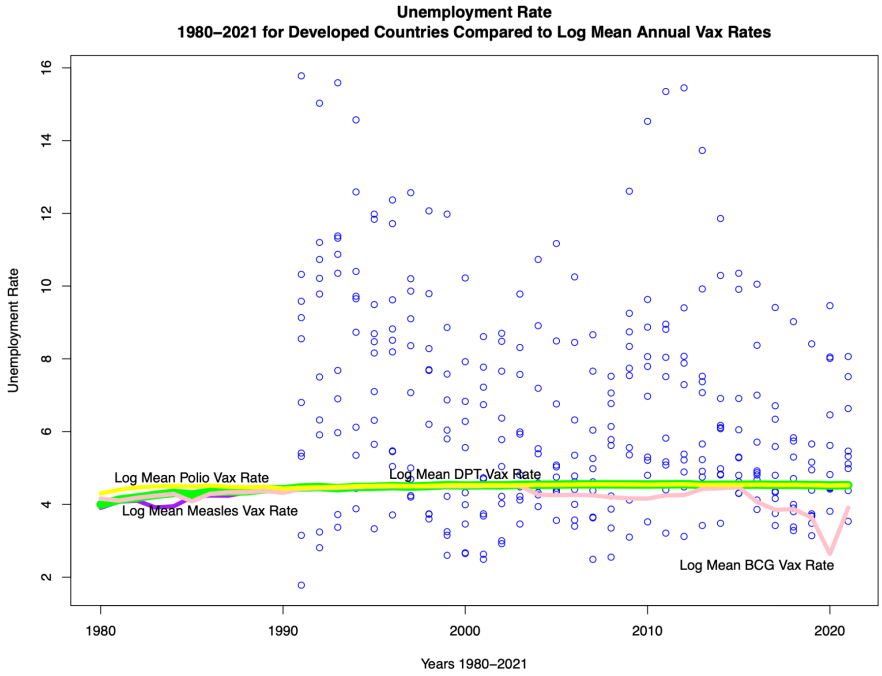


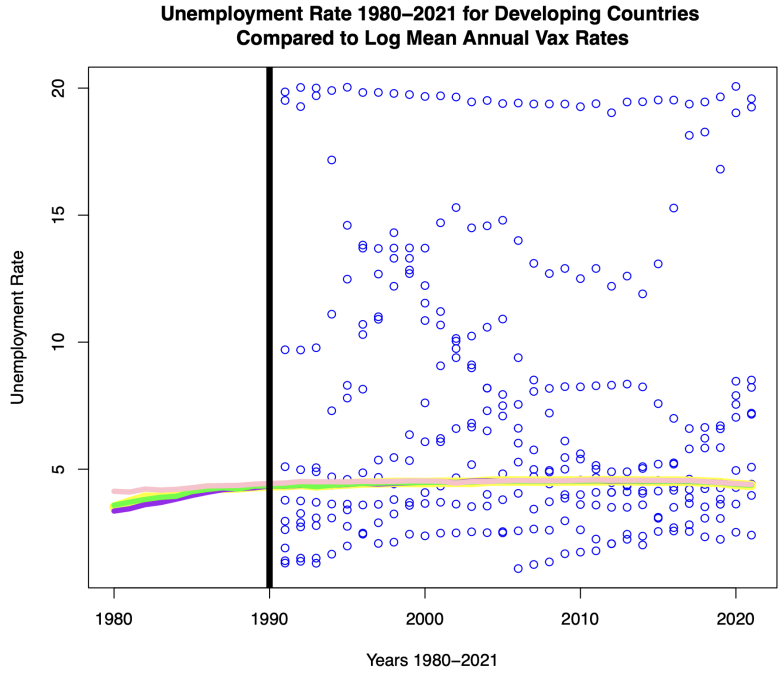


[2] Note: The color key for these images is as follows: **Red** - Mean Death Rate, **Pink** - Log Mean BCG Vax Rate, **Purple** - Log Mean Measles Vax Rate, **Yellow** - Log Mean Polio Vax Rate, **Green** - Log Mean DPT Vax Rate


Log GNI from 1980–2021
for Developed Countries Compared to Log Mean Annual Vax Rates
GNI for Developed Countries
0
2
4
6
8
10
12
1980
1990
2000
2010
2020
Years 1980–2021
Annual Log Mean GNI
Log Mean Polio Vax Rate
Log Mean DPT Vax Rate
Log Mean Measles Vax Rate
Log Mean BCG Vax Rate


Log GNI from 1980–2021 for Developing Countries Compared
to Log Mean Annual Vax Rates
GNI for Developing Countries
0
2
4
6
8
10
1980
1990
2000
2010
2020
Years 1980–2021
Annual Log Mean GNI


Unemployment Rate
1980–2021 for Developed Countries Compared to Log Mean Annual Vax Rates
Unemployment Rate
2
4
6
8
10
12
14
16
1980
1990
2000
2010
2020
Years 1980–2021
Log Mean Polio Vax Rate
Log Mean DPT Vax Rate
Log Mean Measles Vax Rate
Log Mean BCG Vax Rate


Unemployment Rate 1980–2021 for Developing Countries
Compared to Log Mean Annual Vax Rates
Unemployment Rate
5
10
15
20
1980
1990
2000
2010
2020
Years 1980–2021

Works Cited

Aaby P, Clements CJ. Measles immunization research: a review. Bull World Health Organ. 1989;67(4):443-8. PMID: 2680141; PMCID: PMC2491276.

"Coronavirus Death Toll." *worldometer*, www.worldometers.info/coronavirus/coronavirus-death-toll/. Accessed 5 Dec. 2022.

"COVID-19 Vaccine Effectiveness and Safety." *Centers for Disease Control and Prevention*, www.cdc.gov/mmwr/covid19_vaccine_safety.html.

"Developing Countries 2022." *World Population Review*, worldpopulationreview.com/country-rankings/developing-countries. Accessed 3 Dec. 2022.

"Endogeneity Problem." *ScienceDirect*, www.sciencedirect.com/topics/social-sciences/endogeneity-problem#:~:text=The%20basic%20problem%20of%20endogeneity,of%20selection%20bias%20discussed%20earlier.

Hamoudi, Amar A., and Jeffrey D. Sachs. "Economic Consequences of Health Status: A Review of the Evidence." CID Working Paper Series 1999.30, Harvard University, Cambridge, MA, December 1999

"Health Nutrition and Population Statistics." *The World Bank*, databank.worldbank.org/source/health-nutrition-and-population-statistics#. Accessed 3 Dec. 2022.

Masia, Neal A., et al. "Vaccination and GDP Growth Rates: Exploring the Links in a Conditional Convergence Framework." *World Development*, vol. 103, Mar. 2018, pp. 88-99. *ScienceDirect*, https://doi.org/10.1016/j.worlddev.2017.10.013. Accessed 3 Dec. 2022.

Nazar, Shahin, and Toine Pieters. "Plandemic Revisited: A Product of Planned Disinformation Amplifying the COVID-19 'infodemic.'" *Frontiers in Public Health*, vol. 9, 14 July 2021. *Frontiers in Public Health*, https://doi.org/10.3389/fpubh.2021.649930.

Temoka E. Becoming a vaccine champion: evidence-based interventions to address the challenges of vaccination. S D Med. 2013;Spec no:68-72. PMID: 23444594.

"USA's Most Similar Countries." *Objective Lists*, objectivelists.com/2022/07/21/which-countries-are-most-similar-to-the-united-states/. Accessed 3 Dec. 2022.

"US Coronavirus Vaccine Tracker." *USAFacts*, usafacts.org/visualizations/covid-vaccine-tracker-states. Accessed 5 Dec. 2022.