Article

Health in the speeches of different income groups, predicted by Economic factors

Marius Ursu [1]

[1] marius.ursu@student.tue.nl 10916865

Abstract

Access to healthcare has greatly improved over the last decades. This paper looks at whether health-related topics in the speeches at the UN General Debate differs between different income group countries. The speeches have great variability in the rate of sentences that talk about health, both within the same country, and within one income group. The second part aims to investigate whether three economic indicators are good predictors for childhood mortality in data covering 3 decades. These predictors are: GDP per capita, Public investment in healthcare and Electricity access. The results are robust and generalize well, despite good data only being available for high-income countries. More dense data covering lower income groups is required for a followup panel that can uncover cause-effect links or prescribe policy.

1. Introduction

Over the last decades, economic growth has led to more people being lifted out of poverty - at a faster rate than ever. This paper explores the topic of Health and Wellbeing (SDG3) and how it relates to the economic strength of a country.

An exploratory text analysis of the UN Debate Corpus aims to relate speech topics to the development level of countries (4 groups as defined by the World Bank, with data starting in 1987).

In part 2, 3 regression models and an ensemble method based tree is trained on a range of years, aiming to see if we can relate economic strength, access to electricity, and public investment in healthcare to a lower child mortality rate.

Child mortality under 5 can be used to represent the quality of a country's healthcare system. Max Roser at Our World in Data puts it clearly:

> *"Child mortality played a substantial role in increasing overall life expectancy, historically. But in recent decades, as child mortality rates have been at much lower levels, further declines have made much smaller contributions to gains in life expectancy."*

This metric is chosen as an accurate indicator of health quality because it is *"less noisy than other health measures (e.g., adult life expectancy can be influenced by lifestyle diseases, accidents, wars)."* [1].

[1] https://ourworldindata.org/health-meta

Received:
Revised:
Accepted:
Published: 29 September 2025

Citation: . Title. *Journal Not Specified* **2025**, *1*, 0. https://doi.org/

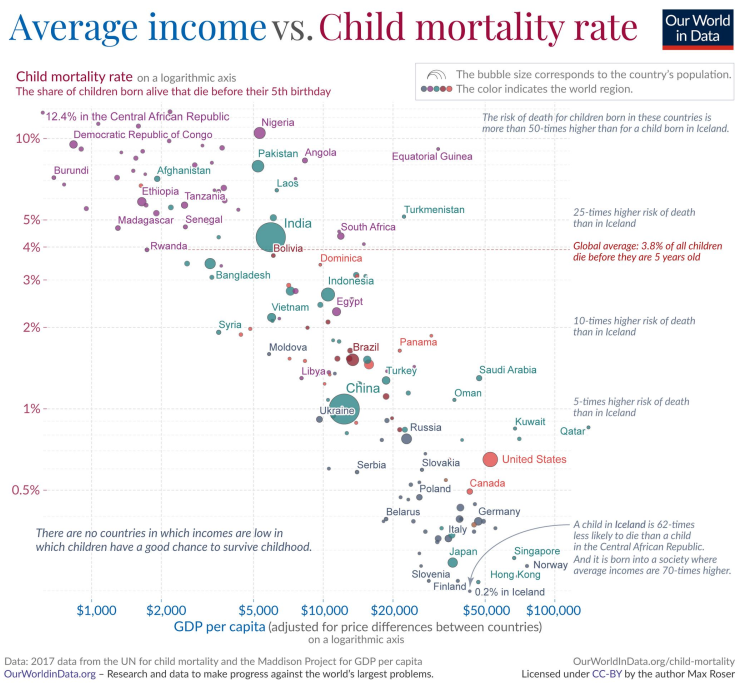

Figure 1. Average income vs Child mortality rate - showing a strong relationship between a strong economy and a strong healthcare system

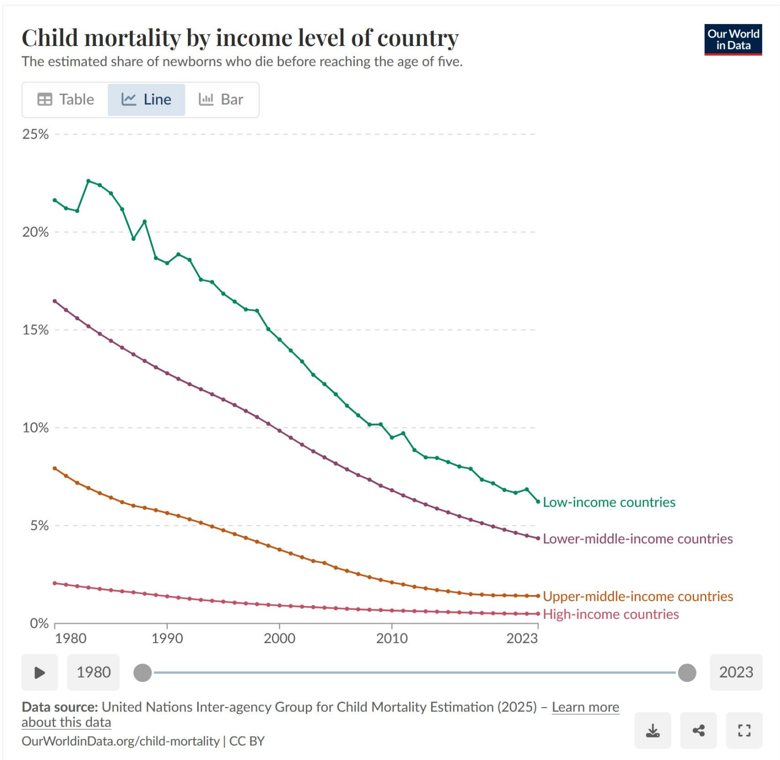

Figure 2. Child mortality by income level of country - steady drop across groups, especially on the lower end of the spectrum

The associated graph references income groups. They are useful when comparing countries because the differences can be immense. The world bank defines 4 development levels based on median income per day per person (including children) - expressed in 2024 dollars:

Income Group	Daily Income (per person)
Low-income countries	$\leq$ $3.10 per day
Lower-middle income countries	$3.10 – $12.20 per day
Upper-middle income countries	$12.20 – $38.00 per day
High-income countries	$\geq$ $38.00 per day

Table 1. Country income groups

2. Materials and Methods

The datasets used in this study are split in 2 parts:

Exploration text analysis per income groups UN debate corpus[2] Income groups - our world in data[3]

Predicting childhood mortality using economic factors: Independent variable datasets: GDP per capita: [4], Electricity access[5], Public healthcare spending as share of GDP[6]

Dependent variable dataset: Child mortality rate[7]

Despite SDG being formalized in the 2010s, health has always been recognized as a national priority and basic human right. We adjust the window of exploration to an age where the variances are large between income_groups, and where the dependent variable has seen very high shifts.

How income relates to health outcomes is what unites the two research questions:

Q1: Between 1993-1997, do certain income_group countries speak more or less about SDG3: Health and wellbeing ?

Reasoning: the author has an explicit assumption that countries on the low end are focused on access to clean food and water, while the high income have moved on to other goals.

Q2: Can a country's health be predicted by GDP per capita, Electricity access and Public healthcare expenditure ?

Reasoning: According to research made by Gapminder foundation[8], positive health outcomes are seen in countries which invest in a few areas:

Higher income levels give families the means to afford better nutrition, cleaner housing, and access to medical services. **Electricity** enables refrigeration of vaccines and medical supplies. **Clean water and sanitation** sharply reduce infectious diseases. **Female education** in particular is a powerful driver: educated mothers are more likely to vaccinate their children, use family planning, and adopt hygienic practices. They also promote uptake of modern medicine.[9]

2 https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/0TJX8Y
3 https://ourworldindata.org/world-bank-income-groups-explained
4 https://ourworldindata.org/grapher/access-to-electricity-vs-gdp-per-capita
5 https://ourworldindata.org/grapher/access-to-electricity-vs-gdp-per-capita
6 https://ourworldindata.org/explorers/global-health
7 https://ourworldindata.org/explorers/global-health
8 https://www.gapminder.org/
9 https://www.goodreads.com/book/show/34890015-factfulness

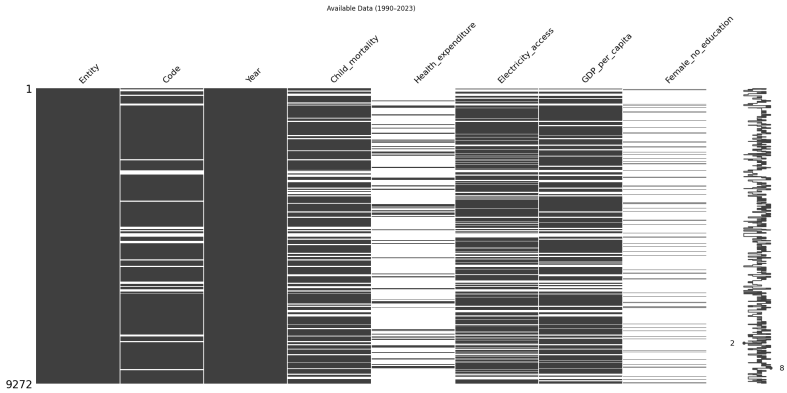

Figure 3. Available data '90-'23 Data scarcity was an issue for female_no_education and health_expenditure

2.1. Data preparation

The data was only scarce for a few fields, and our analysis was re-shaped to not use those fields. NaN values were investigated, leading the author to follow those countries, years, and fields for which there was data. Accurate imputation by machine learning or mean is a large undertaking, and it risks diluting the quality of our predictions.

Selecting the timeframe was key: a 5 year period for the explorative question was short enough for the income level of countries to not change, but large enough to see patterns beyond the noise in debate speech topics.

Filtering based on income group data: 106 countries have full income_group data.

The timeframe is different for the Predictive question, and so is the selection of countries used. This is further described in Results.

3. Results

3.1. EDA: UN Debate corpus - analysis with NLTK (lexical) and BERT (semantic)

The speech data covered speeches between 1946-2024. Countries are well represented. In a given 5-year period ('93–'97), 106 countries selected due to full income_group data. Leading to a total of 530 speeches - which means that nearly all of those countries had a speech in each of those years.

These 106 countries have a fair spread across income groups, with the lowest group dropping to 1/5 of total.

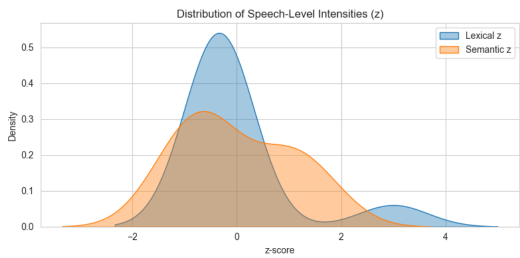

Figure 4. Distribution of Speeches covering SDG3 — NLTK (lexical) and BERT (semantic)

Note: the number of speeches analyzed above differs greatly; variance stays high.

3.1.1. NLTK - lexical analysis

A dictionary covering 32 health related terms (Univ. Toronto[10] was used to count how many occurences of this theme-related words were there per speech. SDG3 hits per

[10] https://sustainability.utoronto.ca/inventories/sustainable-development-goals-sdgs-keywords/

speech varied in the area of 5 per speech and then dropped to 1-2 when more countries or years were added. Number that vary this little relative to the size of the speech, meant that analysis was difficult.The non-contextual nature of the analysis meant the author had to look further.

3.1.2. BERT (all-MiniLM-L6-v2)

A more advanced semantic analysis was performed with a model from the BERT family. It was chosen due to its ability to run without fine-tuning (one-shot), to distinguish topics despite sentences changing themes suddenly. A more precise model is available if a run time is expanded to tens of minutes. The same dictionary was used as before.

The metric chosen was 'Count of sentences that cover SDG3 per 100 sentences'

The distribution is normal, so an ANOVA can be performed.

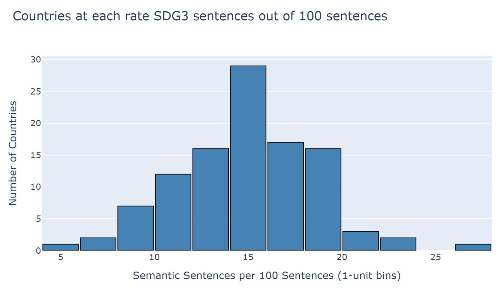

Figure 5. Normal distribution for the rate of SDG3 sentences

3.1.3. Visualizing variance in speech topics per group

The aim of including more years and speeches (lower signal-noise-ratio), is to account in case other pressing matters changing the themes of the speeches (wars, economic crises).

This helped increase the strength of the ANOVA test between groups, but the results are not conclusive

- One-way ANOVA: F = 1.3310, p = 0.263
- KruskalWallis: H = 4.5868, p = 2.047

A closer look at the variance explains why:

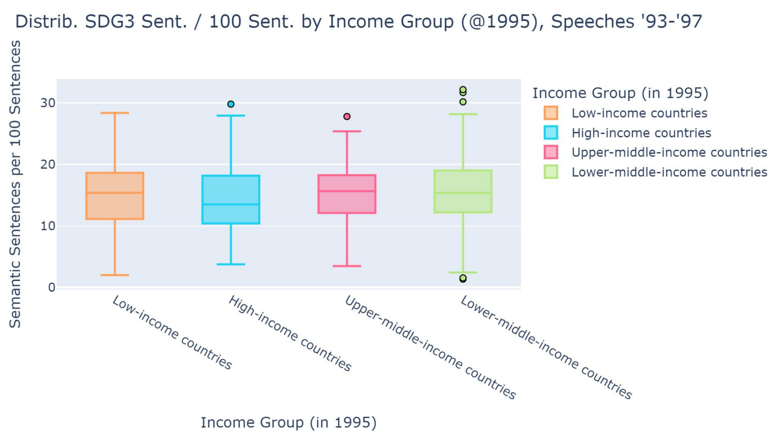

Figure 6. Large variance in SDG3 Sentence rate between income groups

Preliminary Q1 Answer: The debate corpus includes speeches which change topics within one sentence. They also focus on foreign policy and what countries have in common.

As the example sentences in the above analysis show, health topics can be used to describe internal as well as external missions, which is a result of these speeches being political and not scientific in nature.

This means that we require new sources of data to describe how well a country is doing on the health continuum, and what factors may be contributing to it.

It helps to delve behind the average or metric and see the data. Sentences ranked as SDG3 related are previewed below:

> *"mental health problems resulting from many years of war and bombings have no relief"*

> *"attempts to provide health care medicine immunization and so on are stymied by the destruction of so many hospitals and clinics"*

These sentences show the international, peace-minded and broad nature of the sentences. This explains why the initial assumptions of lower-middle and upper-middle income countries being focused on their internal healthcare was not observed.

There is simply too much variance. Even in the speeches of the same country, year to year.

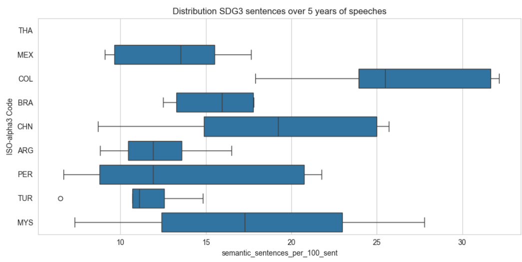

Figure 7. Large variance of SDG3 sentences rate over 5 speeches given by the same country

3.2. *PREDICTING: child mortality based on economic indicators*

3.2.1. Feature Engineering

The 3 features chosen as X variables were: **Health_expenditure**, **Electricity_access**, **GDP_per_capita**.

While the Y variable was: **Child_mortality** - as a proxy for resilience of the healthcare system.

A collinearity check for the 3 predictors returned a Variance inflation factor of 1.2 - 1.7, a low co-linearity. All 3 regression models keeping the coefficients close to each other confirms this fact.

Robust scaling was used to account for the large variance in GDP output.

Cross validation of 5 folds employed for the large dataset, increased the robustness of our analysis. By dividing the folds, we are left with hundreds of rows per fold.

3.2.2. Performance analysis

Ordinary least squares is a common model and would be used as a baseline. Lasso and Ridge have their strengths relating to equating a coefficient to 0, or performing well under mild collinearity.

Ensemble trees are useful in their interpretability. The fact that it is run more than once helps us confirm whether giving role to independent variables is done by more trees than one.

	Model	Split	R^2	RMSE	MAE	MAPE
0	OLS	Train	0.8045	0.4732	0.2964	0.4095
1	OLS	Validation	0.7893	0.4767	0.2984	0.4118
2	OLS	Test	0.8330	0.4720	0.2791	0.4086
3	Ridge	Train	0.8045	0.4732	0.2963	0.4092
4	Ridge	Validation	0.7893	0.4767	0.2983	0.4116
5	Ridge	Test	0.8330	0.4719	0.2790	0.4084
6	Lasso	Train	0.8043	0.4734	0.2928	0.3992
7	Lasso	Validation	0.7893	0.4769	0.2949	0.4017
8	Lasso	Test	0.8338	0.4709	0.2755	0.4001
9	RandomForest	Train	0.9691	0.1881	0.1241	0.1845
10	RandomForest	Validation	0.9017	0.3290	0.1857	0.2349
11	RandomForest	Test	0.9375	0.2888	0.1788	0.2440

Table 2. Model performance for all countries with . Rows: 1,478, Countries: 52, Years: 1990–2023

Results are robust despite 52 countries being represented, with a skew toward the high-income groups.

The spread in performance over test/train/validation is in the range of [0.79, 0.83] which means they are closely grouped and therefore there is no overfitting or underfitting.

The Random Forest is on the overfitting side, presumably incorporating noise. We discuss the consequences of training on a dataset skewed toward high-income countries in the coming section.

3.2.3. Interpreting coefficients

The negative value of these coefficients is expected, as a higher rate of wealth, investment or electricity access is associated with a lower mortality.

Lasso Regression (L1) kept electricity as a coefficient, suggesting that it does indeed serve as a differentiator between countries with the same GDP or health expenditure.

The coefficients are very closely related between the 3 regression types, suggesting low collinearity. Lasso L1 is given below:

Feature	Coefficient
GDP_per_capita	-0.2262
Health_expenditure	-0.2117
Electricity_access	-0.1827

Table 3. Lasso regression coefficients.

For the tree-based model, Public health expenditure is shrunk presumably because of its closely related nature to GDP. However, even after accounting for GDP_per_capita, Health_expenditure provides a small unique contribution to the regression models, which may suggest that in these decades the Health expenditure policies of similar countries had not yet converged.

Feature	Importance
GDP_per_capita	0.5923
Electricity_access	0.3772
Health_expenditure	0.0305

Table 4. Random Forest feature importances.

4. Discussion

4.1. Answer to question 1

Between 1993-1997, do certain income_group countries speak more or less about SDG3: Health and wellbeing ?

As the findings show - through variance analysis - the distribution of health related sentences simply varies too much for one country, or for one income group. The speeches have an international stance, and talk a lot about peace and collaboration. Moreover, wellbeing is described widely, therefore it falls in a large overlap with at least two other SDGs: Water and Peace.

The answer therefore has to be inconclusive: the speeches are simply too general and political.

These findings are confirmed with an LLM which was tasked in extracting topics - broadly, or specifically about SDG3: Health and wellbeing. The author also confims it by listening to at least 2 speeches. They are general and change topic from sentence to sentence.

4.2. Answer to question 2

Can a country's health be predicted by GDP per capita, Electricity access and Public healthcare expenditure ?

The results are robust, and generalize well. They support the introductory thesis that economic strength is one of the largest factors influencing the healthcare system of a country - but our analysis does not draw the cause-effect relationship.

For that, we return to the introductory text which states that among the key factors influencing healthcare quality are: Higher income levels give families the means to afford better nutrition, cleaner housing, and access to medical services. Electricity enables refrigeration of vaccines and medical supplies. Clean water and sanitation sharply reduce infectious diseases. Female education in particular is a powerful driver: educated mothers are more likely to vaccinate their children, use family planning, and adopt hygienic practices. They also promote uptake of modern medicine.[11]

The data is skewed toward countries on the high income spectrum. A weighing or under-sampling algorithm was tried, but it comes with the risk of averaging out over good data, or not using it at all.

An analysis using more informative X variables, as well as specific income groups was attempted, but the remaining dataset was simply too small - even without cross validation. Any predictions were unusable - no amount of creativity solved that. The author is therefore left to look for more data-dense sources, and focus on a few case studies.

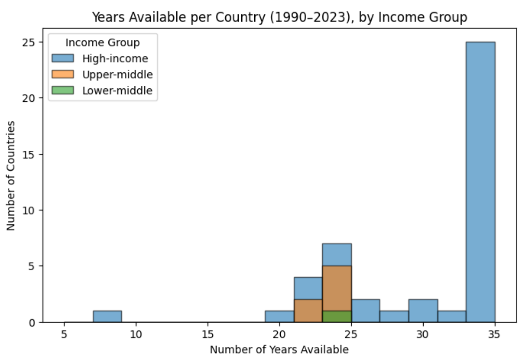

Figure 8. Years with data in all of the 3 X variables. Note: the code may use inferred income groups based on GDP

[11] https://www.goodreads.com/book/show/34890015-factfulness

4.3. Limitations

A more useful analysis would require us to zoom-in: countries which are developing (middle 2 income_groups), a shorter timeframe, especially the years where we see the largest variance in X and Y variables (1990-2005) - giving the model more resolution to work with.

A data-dense dataset for "female education" and "clean water" would greatly improve the predictive power and policy-making value of this analysis.

Finally, countries on the low end of the income spectrum are plagued by inequality. These statistics can be heavily skewed by the few "who have a lot". The author is confident that metrics like households with access to electricity and females with education would be good proxies to account for these inequalities, if the data were found.

In short, we would be able to detect cause-effect relationships, and highlight what worked for those who did it well.

However, it is a well known fact that the record is scarce, as Gapminder foundation states[12]:

> *"Most of our data are not good enough for detailed numeric analysis. They are only good enough to revolutionize people's worldview."*

5. Conclusion

The UN General Debate text was used to detect patterns between countries in different income groups. NLTK and BERT were used for Lexical and Semantical analysis. The speeches have great variability in the rate of sentences that talk about health, both within the same country, and within one income group. Three conomic indicators were investigated whether they are good predictors for childhood mortality in data covering 3 decades. The results are robust and generalize well, despite good data only being available for high-income countries. More dense data covering lower income groups is required for a followup panel that can uncover cause-effect links or prescribe policy.

12 https://www.gapminder.org/data/documentation/