Exploring Determinants of Child Mortality Globally (2022): The Role of Vaccination, Healthcare Spending, and Nutrition*

Findings Reveal Vaccination and Healthcare Spending as Key Drivers of Mortality Reduction

Junbo Li

April 14, 2025

Child mortality has significantly decreased over the past three decades, but progress remains uneven across regions. This study examines the under-five mortality rates of 2022 for each country, focusing on the effects of DPT vaccination coverage, health expenditure per capita, and food production indices. Using a Bayesian multi-linear modeling approach, we found that increased vaccination coverage and healthcare spending significantly reduce child mortality, while food production had no direct impact. These findings underscore the need for targeted interventions in high-mortality regions, suggesting the importance of vaccination programs and healthcare funding in achieving global child survival goals.

Table of contents

*Code and data are available at: https://github.com/Junbo345/Mortality_analysis.

1 Introduction

Strong lensing occurs when a massive foreground object, such as a galaxy cluster, bends light from a more distant source, producing distinctive arcs or rings in images. Despite their significance for cosmology, these events are extremely rare, making up only a small fraction of astronomical observations. Traditional supervised learning methods require large quantities of labeled data, which is unfeasible in this context. The primary challenge, therefore, is to develop a system capable of identifying potential lensing candidates with minimal labeled data.

Thus we will use the state of art active learning strategy, BADGE, which aim to selects the most informative data points for human annotation and thus would improve training efficiency of the Machine Learning model (**cite?**). This strategy have already had promising outcomes in many lacking labeled classification tasks and is of the most effective ways in active learning (**cite?**).

However, we are still facing another problem which is the imbalanced data. the strong lensing ones in our dataset only consists 8% of the total data. And thus will make the evaluation and traing very trick. So we will incorporate better evaluating metrics and training loss.

The remainder of this paper is organized as follows. Section 2 describes possible losses functions we incorporate into our model. **?@sec-metric** introduces the new metrics we incorporated Section 7 presents the results of new losses and evaluation metric. Finally, Section 8 discusses the implications of the findings, the limitations of the study, and recommendations for future research in finding strong lensing data.

2 Method

To test the new metric and loss functions, we use the available strong lensing data for regression. We concatenate the features extracted from the CNN and contatenate them with the expert labels for regression. The result data has a total of ~ and ~ strong lenseing ones.

2.1 New loss function

Here we will investigate two loss functions, ranknet loss and focal loss. Ranknet loss is sepcified as follow: Given training data x_1, x_2, ... , x_n, and their rankings, y_1, y_2, ..., y_n, we will investigate all combinations of ordered pairs (x_i, x_j) where their labels y_i and y_j are different. The total loss is:

$$L = \sum_{i<j} -y_{ij} \log P(i > j) - (1 - y_{ij}) \log(1 - P(i > j))$$

Where $y_i j$ is 1 if x_i is ranked higher than y_j, 0 if x_i is ranked lower than y_j (here we will assume no x_i can have the same ranking as x_j if i is not equal to j). $P(i > j) = \frac{1}{1+e^{-(s(x_i)-s(x_j))}}$

is the probability that x_i is ranked higher than x_j. s() is the ranknING score assigned to i, higher ranking cores means higher ranking. This loss aims to peanalize the pairs of data x_i and x_j where x_i should be ranked higher than x_j but instead x_i is ranked lower by the model for x_j.

The focal loss is specified as follows:

$$FL(p_t) = -\alpha_t(1-p_t)^\gamma \log(p_t)$$

Here p_t is the model's predicted probility if the true label is strong lensing, one minus the predicted probbility if the true label is not strong lenseing. a_t is 0.75 if the true lavel is strong lensing and 0.25 if the true label is not strong lensing. The aim of this loss function is to avoid easy labels and penalize the model most if it incorrectly label a strong lensing one as non-lensing.

2.2 New evaluating matrics

Despite the fact that we need new loss functions, we also need new evaluation metrics to better evaluate the performance of model on biased labels. As said, the labels of strong is very rare relative to non lensing. So if we use the default prediction accuracy then by just predicting everything by strong lensing the model will get a very decent performance while did not contribute much to the final goal of our research purpose.

Thus we will propose a practical and better to understand metric, which is the # of correct labels in the top 10% of the highest ranked data. This is more to common sense and align the final purpose of our research.

3 Results

To test the two loss functions useing the new metric we build a one layer neuronetwork with 40 input neurons, 128 hidden neurons and one output which is the score of the input data in pytorch. We split the dataset into training and testing parts with the ratio of 7:3. To gain efficiency in training, we split the training data into minibatches of size 128 and after finetuning set the training epohs be 100 to avoid overtraining. We also used a neuronet work with crossentropy loss as a conterpart. We ran 10 independent groups and the result is the following:

Figure 1 show the performance of two new proposed loss functions compared to crossentropy loss. They are evaluated based on average number of correct labels in the top 100 labels predicted by the neuronetwork, roughly 10% of all the testing data. The performace of three losses are quite th same, centered around 40 labels. We see that rank loss does not outperform crossentropy loss while focal loss did but there is no promising beatens since the error bars contain each other's mean.

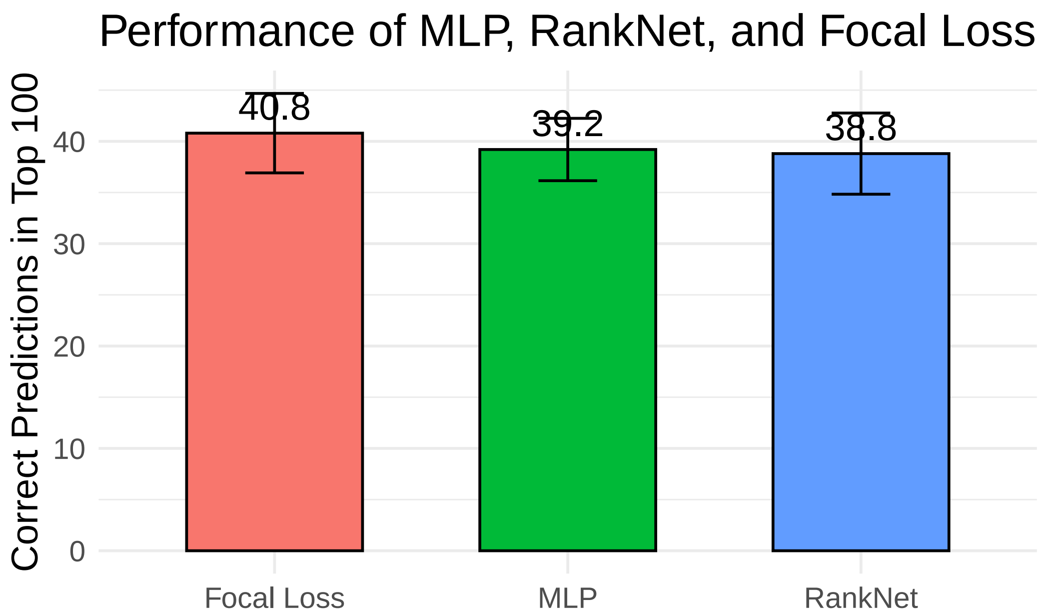


Figure 1: Performance of two new proposed loss functions compared to crossentropy loss. They are evaluated based on number of correct labels in the top 100 labels predicted by the neuronetwork, roughly 10% of all the testing data. We see that rank loss does not outperform crossentropy loss while focal loss did but there is no promising beatens since the error bars contain each other's mean.

4 Dsicussion

4.1 Conslusion and limitation

To our experiment, we see that the Focal loss would be the best choice, but in reality it will requires more computational complexity in training and espically in BADGE where we need to find the gredient wrt the second last layer of the neuronet work. With cross entropy, the gredient has a nice closed form, which is ... However, using focal loss the gredient is much mroe computational cost than cross entropy becasue of the a_t and γ term added. Thus we still need to either finetuning the hyperparameters of the loss to improve its performance. As for ranknet loss, since we need the labels of two diff data be diff, and the case is very rare in each batch and thus will result some training batches yeild non pairs.

4.2 Future work

Although we encountered certain limitations during the project, our findings open the door to several promising avenues for future work. These include fine-tuning the CNN feature extractor that supplies input to our model, enhancing the loss function or exploring alternative ranking approaches, and refining the evaluation criteria to better align with real-world requirements. Additionally, incorporating active learning strategies to gather more labeled data could significantly improve model performance and generalizability.

Over the past 30 years, advancements in society and technology have significantly reduced global mortality rates, particularly for children under the age of five (child mortality). The child mortality rate declined from 93 deaths per 1,000 live births in 1990 to 37 deaths per 1,000 live births in 2022 (World Bank 2024d). This progress represents millions of children who now have better chances of survival, marking a substantial achievement in global health.

Despite this progress, the pace of reduction has slowed considerably since 2015, decreasing from an annual rate of 3.8% during the Millennium Development Goal (MDG) era (2000–2015) to 2.1% during the Sustainable Development Goal (SDG) era (2015–2022) (UNICEF 2022). In 2022 alone, 4.9 million children under the age of five died, equating to approximately 13,400 deaths per day (UNICEF 2022). Moreover, the distribution of mortality rates under five remains highly uneven. Developed countries such as Canada and the United States report mortality rates as low as 10 per 1,000 live births while developing nations—particularly in sub-Saharan Africa—experience rates exceeding 100 per 1,000 live births (World Bank 2024d). This stark disparity underscores the urgent need for targeted interventions in high-mortality regions.

Infectious diseases, such as pneumonia, diarrhea, and malaria, remain the leading causes of under-five deaths, closely tied to vaccination coverage and healthcare spending. Maternal health is also influenced by healthcare accessibility and quality, which often correlates with national health expenditures (World Health Organization 2024a). While food production does not directly address these issues, it reflects a nation's ability to meet nutritional needs, a critical factor in preventing malnutrition—a major contributor to child mortality.

In this study, we examined the child mortality rate for each country in 2022, focusing on three key factors: food production index, DPT vaccine coverage, and per-capita health expenditure. Our findings reveal that increasing vaccine coverage and per-capita health expenditures significantly reduce under-five mortality rates. However, the food production index does not exhibit a direct impact, suggesting its role is mediated by factors such as food quality and distribution systems.

The estimand of this study is the effect of three predictors—food production index, vaccination coverage (DPT vaccine percentage), and per-capita health expenditure (log-transformed)—on the logarithm of the under-five mortality rate for each country in 2022. This analysis seeks to quantify the extent to which these factors contribute to variations in mortality rates across nations. By addressing these relationships, the study aims to identify potential intervention points for reducing preventable child deaths and achieving progress toward the Sustainable Development Goals for child survival.

The remainder of this paper is organized as follows. Section 5 describes the data sources, key variables, and preprocessing steps, including transformations applied to address skewness and improve model validity. Section 6 outlines the modeling approach, including the choice of predictors, justification of the model structure, and Bayesian implementation. Section 7 presents the results of the analysis, highlighting the effects of each predictor on under-five mortality. Finally, Section 8 discusses the implications of the findings, the limitations of the study, and recommendations for future research and policy interventions aimed at reducing child mortality globally.

5 Data

5.1 Overview

For this analysis, we used four datasets: child mortality rate, food production index, DPT vaccine coverage rate, and health expense per capita. All four datasets are downloaded from **Worldbank Open Database** (The World Bank 2024) through its API (Arel-Bundock 2022b). World Bank Open Database is a comprehensive resource offering free and publicly accessible development data from nearly every country. This platform provides an extensive range of indicators across sectors such as health, agriculture, and education, enabling researchers to access globally standardized data for comparative analysis. The platform ensures transparency and usability by offering tools for visualization, data export, and API integration, making it a reliable and versatile resource for exploring global trends.

We employed **R** (R Core Team 2023), a coding platform to download, clean, and conduct statistical analysis. Besides, we also utilized R packages **tidyverse** (Wickham et al. 2019), **rstanarm** (Goodrich et al. 2022), **ggplot2** (Wickham 2016), **knitr** (Xie 2014), **arrow** (Richardson et al. 2024), **here** (Müller 2020), **modelsummary** (Arel-Bundock 2022a), **testthat** (Wickham 2011) and **dpylr** (Wickham et al. 2023). The paper is outlined in Git Hub using the starter folder provided in **Telling Stories With Data** (Alexander 2023).

5.2 Measurement

The datasets of this analysis are collected and prepared by various UN-affiliated organizations. Below, we provide a brief overview of the measurement and transformation processes for these data sets. Detailed descriptions and methodological specifics can be found in Section A.

5.2.1 Response variable:

Mortality Rate: The child mortality rate is estimated by the UN Inter-agency Group for Child Mortality Estimation (IGME). The primary data source is each country's vital registration system. In countries where such a system is lacking, household surveys are conducted and then processed using a statistical model developed collaboratively by UN agencies and academic institutions to derive an estimate of the mortality rate (World Bank 2024d).

5.2.2 Predictors:

Food Production Index: This index, prepared by the Food and Agriculture Organization (FAO) of the United Nations, measures aggregate food production relative to the 2014–2016 baseline period. It is calculated yearly using price-weighted data on agricultural products for each country. (World Bank 2024b).

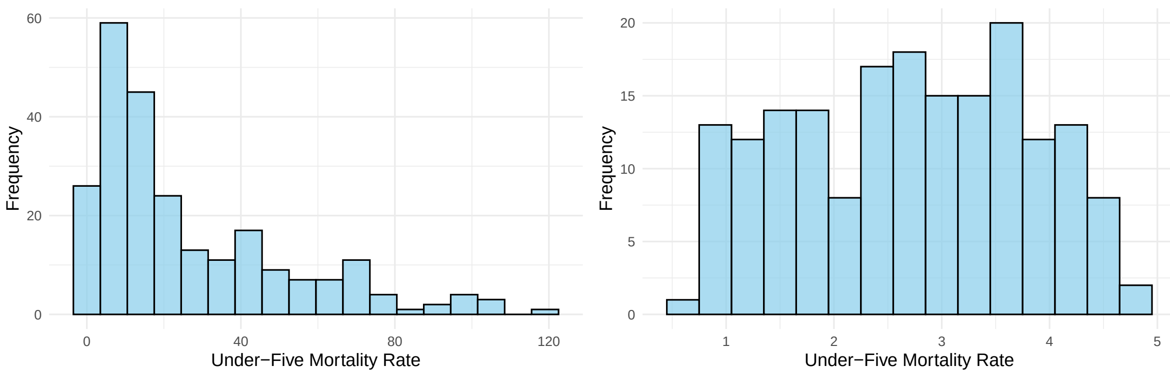
Health Expenditure per Capita: Data on health expenditure per capita are produced by the World Health Organization (WHO) following the System of Health Accounts 2011 (SHA 2011) framework. This framework tracks all health-related spending in a country over a defined period (World Bank 2024a).

DPT Vaccine Coverage: DPT vaccine coverage is jointly estimated by the WHO and UNICEF. This metric reflects the percentage of children aged 12–23 months in each country who have received diphtheria, pertussis (whooping cough), and tetanus (DPT) vaccine. Estimates are derived from reports by vaccine service providers and household surveys of vaccination history (World Bank 2024c).

5.3 Outcome variables

The response variable in this study is the child mortality rate in each country in 2022, measured as the number of deaths of children under five years old per 1,000 live births in 2022.

The Original data range from 0 to 120, with a peak between 10 and 15 deaths per 1,000 live births (Figure 2a). Due to extreme values and significant right skewness, a log transformation was applied to stabilize variance and improve normality. The transformed data range from 2 to 5, with a peak near 3, and the shape is approximately normal. (Figure 2b).



(a) Distribution of Mortality rate before log transformation

(b) Distribution of Mortality rate after log transformation

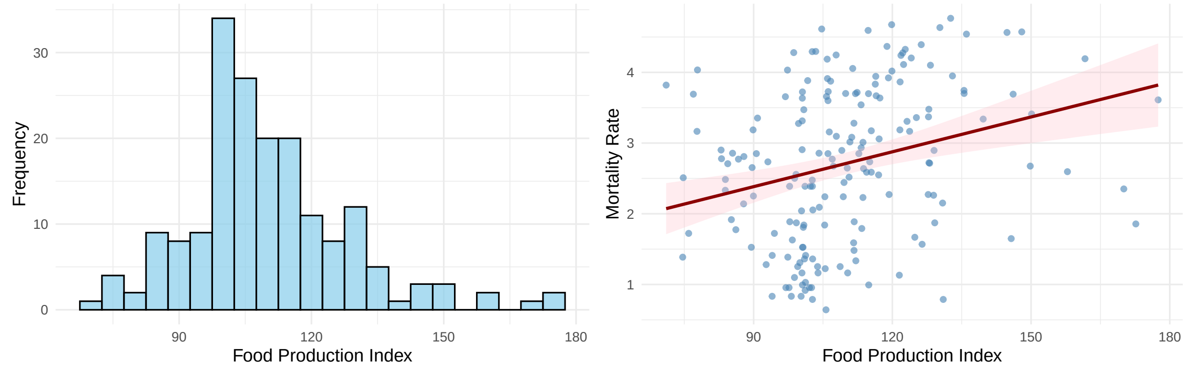
Figure 2: Data Analysis for Mortality Rate for children under five in different countries. The histogram on the top left shows that the original distribution is skewed to the right. The histogram on the top right is the distribution after log transformation and is approximately normal.

5.4 Predictor variables

The three predictors analyzed are:

Food Production Index: The Food Production Index measures the relative level of agricultural production for each country in 2022, using the 2014–2016 baseline as a reference point (Arel-Bundock 2022b). This index is price-weighted and reflects agricultural productivity across countries. As illustrated in Figure 3a, the index ranges from 60 to 180, with a mean of around 110 and an approximately normal distribution. The scatter plot between the food production index and log of child mortality in Figure 3b shows a slightly positive linear relationship. This relationship is examined in greater detail in Section 7.

Current Health Expenditure: The Current Health Expenditure per Capita captures health spending in US dollars for each individual in 2021, calculated under the System of Health Accounts (SHA 2011) framework. The raw data range from \$0 to \$12,000 and exhibit extreme



(a) Counts of Food production index for each country, measured as a basis of 2014-1016

(b) Relationship of each countries' Food production index VS. Log of Mortality rate

Figure 3: Data Analysis of Predictor Food Production Index. The histogram on the left shows an approximately normal distribution. The scatterplot on the right indicates there is a slight positive relationship between food production index and mortality rate.

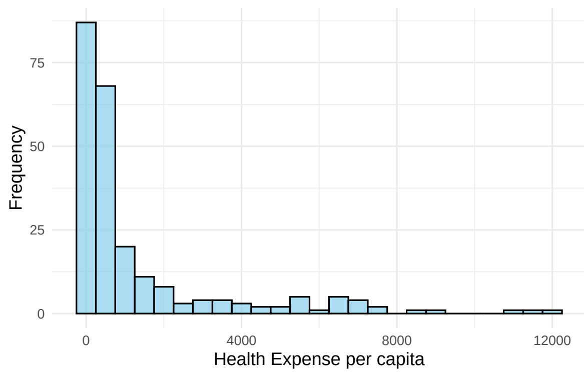
right skewness, as shown in Figure 4a. To address this, a log transformation was applied, resulting in a normalized range from 0 to 10, with a central tendency near 6 (Figure 4b). Analysis of the relationship between log-transformed health expenditure and log mortality rate (Figure 4c) demonstrates a significant negative linear trend, indicating that higher health expenditures are associated with lower mortality rates. This relationship is examined in greater detail in Section 7.

DPT Vaccine Coverage: DPT vaccine coverage rate reflects the percentage of children aged 12–23 months who received vaccinations against diphtheria, pertussis, and tetanus in 2022. Coverage ranges from 30% to 100%, with a left-skewed distribution, as depicted in Figure 5a. Analysis of the association between vaccine coverage and mortality rates (Figure 5b) shows a strong negative relationship, suggesting that increased vaccine coverage is correlated with lower mortality rates. This finding will be explored further in Section 6.

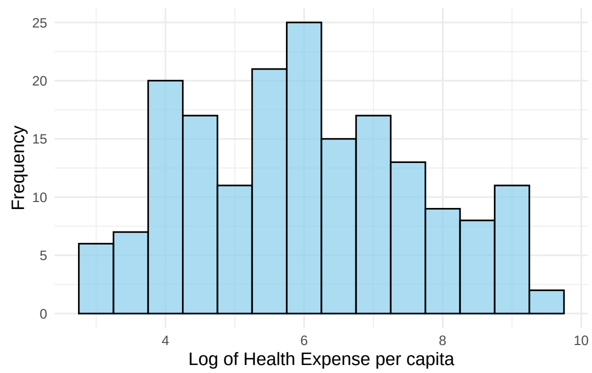
5.5 Missing Data and Time Inconsistency

The health expenditure variable was collected in 2021, while other variables reflect 2022 data. This discrepancy is due to the unavailability of more recent health expenditure data. These 2021 values are assumed to approximate 2022 conditions sufficiently for analysis.

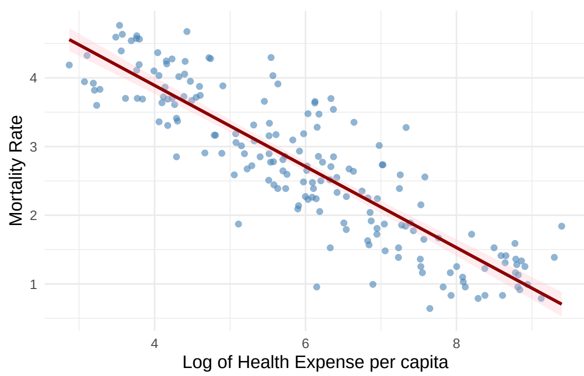
Additionally, some countries' data were not collected and were excluded from the analysis. The implications of these omissions and time inconsistency are addressed in Section 8.4.



(a) Counts of Health Expenditure Per Capita for each country (measured in US$) before log transformation

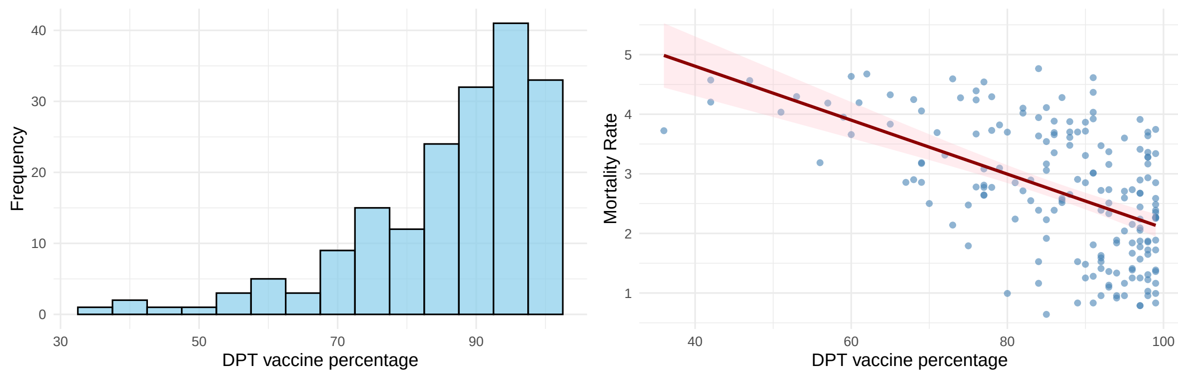


(b) Counts of each country's Health Expenditure Per Capita, (measured in US$) after log transformation



(c) Relationship of each countries' Health Expenditure Per Capita, Measured in US$ VS. Log of Mortality

Figure 4: Data Analysis of health expenses per capita. The histogram on the top left shows that the original distribution is skewed to the right. The histogram on the top right is the distribution after log transformation and is approximately normal.The scatterplot on the bottom left indicates there is a negative relationship between log of health expenses and Mortality rate.



(a) Counts of each country's DPT vaccine percentage

(b) Relationship of each country's DPT Vaccine Percentage VS. Log of Mortality

Figure 5: Data Analysis of DPT Vaccine Coverage. The histogram on the left shows that the distribution of DPT vaccine coverage is skewed to the left. The scatterplot on the right indicates there is a negative relationship between DPT vaccine coverage and Mortality rate.

6 Model

Our modeling strategy aims to explore the relationships between DPT vaccine coverage, the food production index, and current health expenditure per capita (in current US$) with the under-five mortality rate across nations. By analyzing these factors, we seek to identify opportunities to improve child health outcomes, particularly in countries with high mortality rates. Supporting model diagnostics are provided in Appendix B.

6.1 Model set-up

Let y_i be the logarithm of the under-five mortality rate for nation i. We define the following predictors:

x_{1i}: DDT vaccine coverage for nation i. x_{2i}: Food production index for nation i. x_{3i}: Logarithm of current health expenditure per capita (current US$) for nation i. The model is specified as:

$$
\begin{aligned}
y_i \mid \mu_i, \sigma &\sim \text{Normal}(\mu_i, \sigma) && (1)\\
u_i &= \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} && (2)\\
\alpha &\sim \text{Normal}(2.5, 1) && (3)\\
\beta_1 &\sim \text{Normal}(0, 2.5) && (4)\\
\beta_2 &\sim \text{Normal}(0, 2.5) && (5)\\
\beta_3 &\sim \text{Normal}(0, 2.5) && (6)\\
\sigma &\sim \text{Exponential}(1) && (7)
\end{aligned}
$$

We implement this model in **R** (R Core Team 2023) using the **rstanarm** (Goodrich et al. 2022) and **arrow** (Richardson et al. 2024) package.

6.1.1 Model justification

A Bayesian multivariate linear model was chosen to capture the relationships between the predictors (DPT vaccine coverage, food production index, and current health expenditure) and the response variable (under-five mortality rate). This choice is supported by observed linear trends between each predictor and response in Section 5. Logarithmic transformations were applied to the under-five mortality rate and current health expenditure per capita to normalize the distributions and stabilize their linear relationships with the predictors.

Default Bayesian priors were used for the predictors, as they provide a robust starting point. For the intercept term, a Normal(2.5, 1) prior was chosen, reflecting the observed distribution of the response variable (Figure 2a).

6.1.2 Model weakness

Several limitations are noted in this model:

1. Skewed Predictor Distribution: The distribution of DPT vaccine coverage is skewed left, which may introduce bias. Despite attempts to address this, an ideal solution remains elusive.
2. Potential Collinearity: While each predictor shows an individual linear relationship with the response variable, dependencies between predictors may affect model accuracy.
3. Simplified Assumptions: The model assumes linearity and does not capture potential non-linear interactions or regional variations that could influence under-five mortality rates.

6.2 Model Validation

The posterior and prior predictive checks are presented in Figure 6a. We observe an excellent fit between the posterior distribution from our model and the actual child mortality data, indicating that the model effectively captures the underlying trends in child mortality rates. In Figure 6b, we compare the posterior distribution with the prior. No significant deviations are observed, suggesting alignment between our prior beliefs and the posterior distribution.

Figure 7a displays the trace plot check, where no irregularities or concerning patterns are detected. Finally, the R-hat plot (Figure 7b) shows all values close to one, confirming the convergence of the Markov chains. Detailed information could be found in Section B.

7 Results

7.1 Model outcome

Our model results are summarized in Table 1. We will discuss them one by one.

7.1.1 Intercept:

The intercept estimate of 6.643 represents the expected logarithm of the under-five mortality rate when all predictors are held constant at their reference or baseline values (e.g., average food production, average vaccine coverage, and average health expenditure).

7.1.2 Food Production Index:

The coefficient for the food production index is 0.003, indicating a very slight positive relationship between food production and the under-five mortality rate. This naively suggests that reducing food production could also reduce mortality, which is counter-intuitive.

However, a closer examination of the confidence interval for this predictor shows that it includes 0. This implies that the relationship is not statistically significant, so we conclude that food production is not strongly related to overall under-five mortality rates.

7.1.3 Vaccine Coverage:

The coefficient for vaccine coverage is -0.013, meaning that for every 1 percentage point increase in vaccine coverage, the logarithm of the under-five mortality rate decreases by 1.3%, holding other variables constant.

Table 1: Explanatory model of Mortality rate for each country based on food production index, DPT vaccine coverage, and health expense.

	Mortality rate model
(Intercept)	6.6407
	(5.9254, 7.3233)
Food	0.0034
	(−0.0009, 0.0075)
Vacinne	−0.0131
	(−0.0193, −0.0068)
Health_expense	−0.5294
	(−0.5810, −0.4771)
Num.Obs.	182
R2	0.787
R2 Adj.	0.784
Log.Lik.	−133.072
ELPD	−136.8
ELPD s.e.	10.7
LOOIC	273.5
LOOIC s.e.	21.5
WAIC	273.5
RMSE	0.50

7.1.4 Health Expenditure (Log-Transformed):

The coefficient for the log-transformed health expenditure is -0.529, indicating that a 1% increase in health expenditure per capita is associated with a 0.529 unit decrease in the log of the under-five mortality rate, holding all other predictors constant.

8 Discussion

8.1 Overview

The United Nations' Sustainable Development Goal (SDG) 3.2 aims to reduce the mortality rate of children under five to fewer than 25 per 1,000 live births by 2030 (World Health

Organization 2024b). While substantial progress has been achieved globally, meeting this ambitious target within the remaining six years presents significant challenges. Although global trends indicate a decline in child mortality rates, the pace of this reduction has slowed since 2015 (World Health Organization 2024b).

A stark disparity persists across countries, as illustrated in Figure 2a. While some nations, especially high-income ones, have already surpassed the SDG target, others—predominantly in low-income regions such as sub-Saharan Africa—experience child mortality rates that are four times higher than the global goal.

To achieve the SDG targets, concerted efforts are essential, particularly through targeted support from international organizations like the World Health Organization (WHO). However, this raises critical questions: Where should resources be allocated most effectively? How can countries with high child mortality rates achieve the greatest impact with limited resources?

8.2 Correlation V.S. Causality

Numerous studies have explored the determinants of child mortality, consistently finding a negative correlation between mortality rates and macroeconomic indicators such as GDP (O'Hare et al. 2013) or education levels (Sartorius and Sartorius 2014). However, many of these factors are not direct causes of child mortality.

For instance, during the Great Recession, the United States experienced a reduction in mortality rates due to decreased air pollution following widespread factory closures (Finkelstein et al. 2024). This demonstrates the importance of identifying underlying causes rather than focusing solely on broad economic indicators. In low-income countries, the primary drivers of high child mortality rates are often basic issues such as malnutrition and inadequate access to primary healthcare (Doerr and Hofmann 2022). Addressing such root causes is crucial to achieving meaningful and sustainable improvements.

Our analysis identifies two factors with strong negative correlations to child mortality rates: DPT vaccine coverage and healthcare spending. However, it remains essential to determine whether these factors exert a direct influence or are proxies for other underlying variables.

The DPT vaccine protects against three deadly diseases: diphtheria, pertussis (whooping cough), and tetanus (lockjaw). Administered in five doses starting at two months of age, this vaccine provides critical immunity for newborns (Centers for Disease Control and Prevention 2024).

In 1990, diphtheria, pertussis, and tetanus collectively caused approximately 160,000 deaths among children aged 1–4 years. By 2019, this number had decreased to 60,000, demonstrating the vaccine's impact (Global Burden of Disease Collaborative Network 2020). Nonetheless, disparities in vaccination rates persist, leaving many children vulnerable to preventable diseases. Studies indicate that children receiving zero doses of vaccines are at significantly higher risk of mortality (Karlsson et al. 2024). Expanding vaccination coverage is a direct and cost-effective intervention to reduce child mortality and improve public health.

Healthcare spending encompasses all expenditures on health-related goods and services within a given year, including both private and public spending (World Bank 2024a). Research suggests that public healthcare spending has the most pronounced effect on reducing child mortality (Ray and Linden 2020), primarily by increasing access to trained healthcare professionals and improving maternity and neonatal care.

Public healthcare systems are often funded through tax revenues, which are tied to broader economic performance and governance structures (Ray and Linden 2020). While boosting overall tax revenues may be challenging, reallocating existing resources to prioritize public health can yield significant benefits. This approach is particularly relevant in countries where public spending on healthcare remains disproportionately low.

8.3 Conclusion

Our study shows that reducing the child mortality rate requires a dual focus on expanding vaccination coverage and optimizing healthcare spending. Vaccination programs, particularly for diseases like diphtheria, pertussis, and tetanus, represent a direct strategy for reducing child mortality. Simultaneously, enhancing public healthcare spending can address systemic issues such as inadequate maternal care and limited access to trained professionals.

On the other hand, efforts must be tailored to the specific needs and capacities of individual countries. International organizations can play a pivotal role by providing financial support, technical expertise, and policy guidance to countries with high child mortality rates. By prioritizing interventions with clear and measurable impacts, the global community can move closer to the goal of ensuring that no child dies from preventable causes.

8.4 Limitation and next steps

This study has several limitations that warrant consideration in future research. First, as discussed in Section 5.5, data from several countries is unavailable. While these nations typically have smaller populations, their exclusion may introduce biases and raise ethical concerns by neglecting their unique challenges. Second, the most recent data used in this study is from 2022, and there are temporal inconsistencies among the predictors. This is particularly important because both the predictors and the response variable are time-sensitive. Access to more current and synchronized datasets would enable a more accurate and relevant analysis, reflecting recent developments in healthcare and child mortality trends.

Additionally, while this study focused on three primary predictors, the dataset contains other indicators that could potentially enhance the analysis. However, many of these indicators are affected by missing or outdated values, limiting their utility. Future research should prioritize addressing data completeness and timeliness to improve model robustness and reliability.

Finally, incorporating advanced modeling techniques and additional predictors could provide deeper insights. Future studies should explore innovative models to better capture complex relationships and interactions, enabling more nuanced and actionable recommendations for policy interventions.

Appendix

A Sampling Methodology Overview and Evaluation

A.1 Overview

This study explores the relationship between under-five mortality rates and three key predictors: food production index, DPT vaccine coverage, and per capita health expenditure. The under-five mortality rate is derived from a combination of vital registration systems, household surveys (e.g., Demographic and Health Surveys and Multiple Indicator Cluster Surveys), and indirect estimation models developed by UN agencies and collaborating academic institutions (World Bank 2024d). These data sources aim to address disparities in coverage, particularly in regions where vital registration is incomplete.

The food production index, developed by the Food and Agriculture Organization of the United Nations (FAO), quantifies national agricultural output relative to a baseline period (2014–2016) using official national statistics and supplementary surveys (World Bank 2024b). Per capita health expenditure is sourced from the World Health Organization (WHO) using the System of Health Accounts (SHA 2011), which systematically tracks all public and private spending on health services, goods, and infrastructure (World Bank 2024a). Lastly, DPT vaccine coverage is estimated through a combination of administrative data from immunization service providers and household surveys, reflecting the proportion of children aged 12–23 months who received three doses of the vaccine (World Bank 2024c).

A.2 Population sample frame, and sampling approach

The target population for the response variable, under-five mortality rate, includes all children under the age of five globally in 2022. This metric captures the probability per 1,000 live births that a newborn will die before reaching five years of age, providing a direct measure of child survival across regions and contexts. The predictors each represent distinct populations. The food production index measures agricultural output for entire national populations, reflecting food availability and potential nutritional impacts. Similarly, per capita health expenditure pertains to the entire population within a country, as healthcare spending influences system-wide access and quality. DPT vaccine coverage, by contrast, specifically targets the population of children aged 12–23 months, providing insights into immunization rates and public health outreach.

The sampling frame for each variable aligns with its respective population. Under-five mortality rates rely on vital registration systems, which are designed to record all births and deaths. In regions where these systems are incomplete, surveys and statistical models supplement the data to ensure national representativeness. The food production index uses national agricultural statistics as the primary source, with secondary data from government ministries and international organizations filling gaps (World Bank 2024b). For DPT vaccine coverage, data are collected through administrative immunization records verified by household surveys (World Bank 2024c). Per capita health expenditure is derived from national accounts and follows the standardized SHA 2011 framework (World Bank 2024a).

The sampling approach for these variables combines census-based data collection with stratified sampling in regions lacking comprehensive infrastructure. While census methods aim to capture complete population-level data, resource constraints in low- and middle-income countries necessitate reliance on surveys and indirect estimation methods. These approaches use stratification to ensure representativeness while accounting for regional disparities.

A.3 Strength and weakness

The methodological framework demonstrates several strengths. First, the primary reliance on census-based or near-census-level data ensures comprehensive coverage for most variables. This is particularly evident in high-income regions, where systems such as vital registration achieve over 90% completeness (World Health Organization 2024c). Second, the use of supplementary data sources, such as surveys or administrative records, mitigates gaps in the primary datasets, enhancing data quality. For instance, the food production index excludes non-nutritive products such as coffee and tea, ensuring a focus on agricultural outputs relevant to nutrition (World Bank 2024b). The use of standardized frameworks, such as SHA 2011 for health expenditure, further supports consistency and comparability across countries.

However, notable weaknesses persist. In regions with incomplete infrastructure, such as Sub-Saharan Africa, vital registration systems capture only 44% of deaths, leading to substantial data gaps that require imputation through statistical models (World Health Organization 2024c). These models rely on assumptions that may not fully reflect local demographic trends, introducing potential biases. Moreover, reliance on household surveys for DPT vaccine coverage and mortality data introduces recall bias, as respondents may inaccurately report past events (Preston 1984). Even census-based data are susceptible to misreporting, undercounting, or misclassification, which can distort results. Furthermore, temporal mismatches among variables (e.g., health expenditure data lagging behind mortality rates) may complicate causal inference and trend analysis.

B Model details

B.1 Posterior predictive check

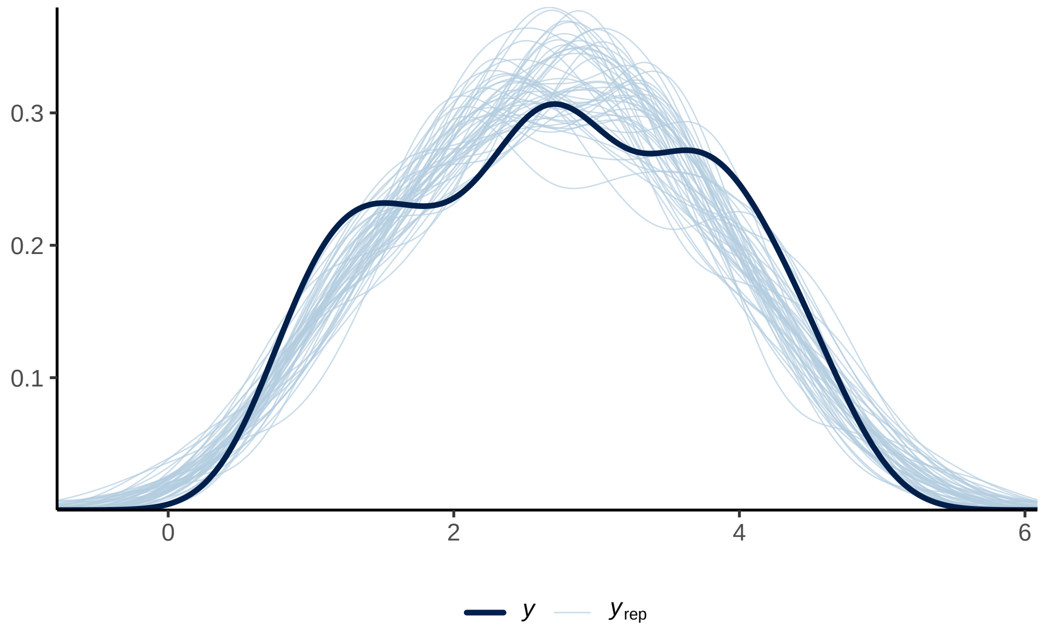
We conducted a posterior predictive check to evaluate how well the model predicts under-five mortality rates. Figure 6a illustrates the predictive distribution compared to the observed data. The close alignment between observed and predicted values indicates that the model captures the main trends in the data effectively.

In Figure 6b we compare the posterior with the prior. This shows that the data significantly updates the prior beliefs for all key parameters. The posterior distributions of predictors—vaccine coverage, health expenditure, and food production index—highlight the robustness of their effects in the model.

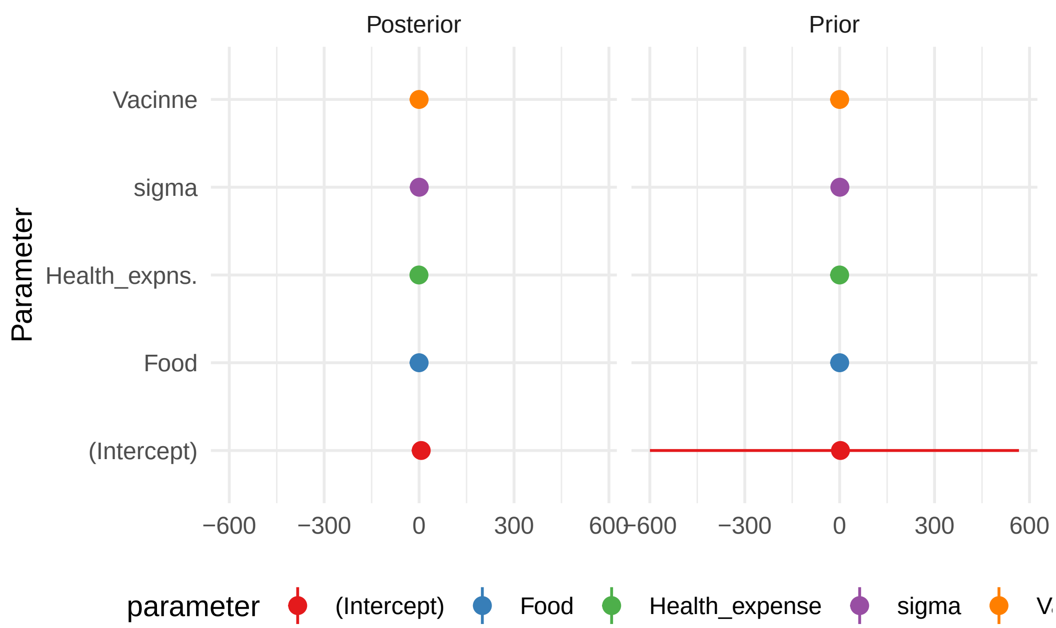
B.2 Diagnostics

Figure 7a is a trace plot. It shows adequate mixing of the Markov Chain Monte Carlo (MCMC) chains, indicating proper convergence. Additionally, the Rhat values for all parameters (Figure 6b) are below 1.05, further confirming convergence and the reliability of the parameter estimates.

Figure 7b is a Rhat plot. It shows no significant patterns, indicating that the model assumptions hold. Additionally, the Bayesian framework effectively quantifies uncertainty, with credible intervals for the predictors providing insight into their respective influences on mortality.

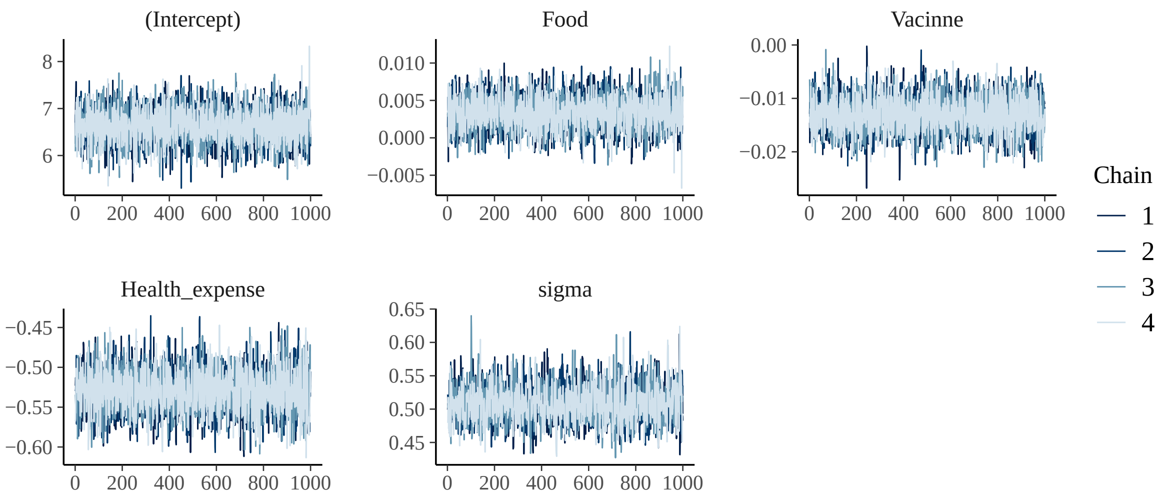


(a) Posterior prediction check

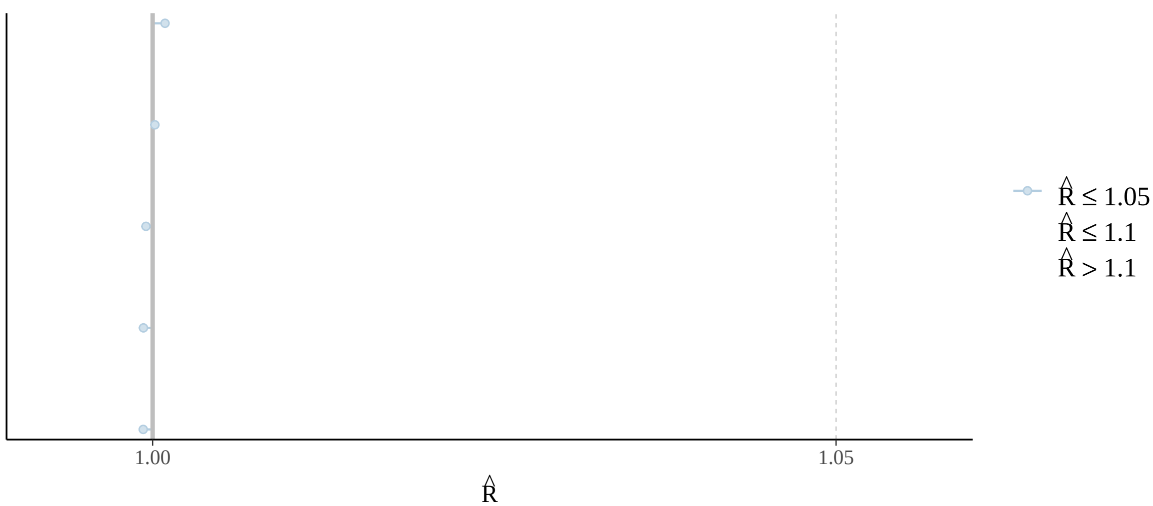


(b) Comparing the posterior with the prior

Figure 6: Examining how the model fits, and is affected by, the data



(a) Trace plot



(b) Rhat plot

Figure 7: Checking the convergence of the MCMC algorithm

References

Alexander, Rohan. 2023. *Telling Stories with Data*. Chapman; Hall/CRC. https://tellingstorieswithdata.com/.

Arel-Bundock, Vincent. 2022a. "modelsummary: Data and Model Summaries in R." *Journal of Statistical Software* 103 (1): 1–23. https://doi.org/10.18637/jss.v103.i01.

———. 2022b. *WDI: World Development Indicators and Other World Bank Data*. https://CRAN.R-project.org/package=WDI.

Centers for Disease Control and Prevention. 2024. "Diphtheria, Tetanus, and Pertussis (DTaP) Vaccine: What You Need to Know." https://www.cdc.gov/vaccines/hcp/vis/vis-statements/dtap.html#why-vaccinate.

Doerr, Sebastian, and Boris Hofmann. 2022. "Recessions and Mortality: A Global Perspective." *Economics Letters* 220: 110860. https://doi.org/https://doi.org/10.1016/j.econlet.2022.110860.

Finkelstein, Amy, Michael J. Schilbach, Matthew F. Notowidigdo, and Jintao Zhang. 2024. "Lives Vs. Livelihoods: The Impact of the Great Recession on Mortality and Welfare."

Global Burden of Disease Collaborative Network. 2020. "Global Burden of Disease Study 2019 (GBD 2019) Results." Seattle, United States: Institute for Health Metrics; Evaluation (IHME). https://ghdx.healthdata.org/gbd-2019.

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. "rstanarm: Bayesian applied regression modeling via Stan." https://mc-stan.org/rstanarm/.

Karlsson, Omar, Sunil Rajpal, Mira Johri, Rockli Kim, and SV Subramanian. 2024. "Prevalence and Trends of Not Receiving a Dose of DPT-Containing Vaccine Among Children 12–35 Months: An Analysis of 81 Low- and Middle-Income Countries." *Journal of Epidemiology and Global Health* 14 (3): 460. https://doi.org/10.1007/s44197-024-00294-6.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. https://CRAN.R-project.org/package=here.

O'Hare, Bernadette, Innocent Makuta, Levison Chiwaula, and Nathan Bar-Zeev. 2013. "Income and Child Mortality in Developing Countries: A Systematic Review and Meta-Analysis." *Journal of the Royal Society of Medicine* 106 (10): 408–14. https://doi.org/10.1177/0141076813489680.

Preston, Samuel H. 1984. "Children and the Age Pattern of Mortality." *Population and Development Review* 10 (1): 69–76. https://doi.org/10.2307/1171470.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Ray, Devdatta, and Mikael Linden. 2020. "Health Expenditure, Longevity, and Child Mortality: Dynamic Panel Data Approach with Global Data." *International Journal of Health Economics and Management* 20 (1): 99–119. https://doi.org/10.1007/s10754-019-09272-z.

Richardson, Neal, Ian Cook, Nic Crane, Dewey Dunnington, Romain François, Jonathan Keane, Dragoș Moldovan-Grünfeld, Jeroen Ooms, Jacob Wujciak-Jens, and Apache Arrow. 2024. *Arrow: Integration to 'Apache' 'Arrow'*. https://CRAN.R-project.org/package=arrow.

Sartorius, Benn K., and Kurt Sartorius. 2014. "Global Infant Mortality Trends and Attributable Determinants – an Ecological Study Using Data from 192 Countries for the Period 1990–2011." *Population Health Metrics* 12 (1): 29. https://doi.org/10.1186/s12963-014-0029-6.

The World Bank. 2024. "World Bank Open Data." https://data.worldbank.org/.

UNICEF. 2022. "Under-Five Mortality." https://data.unicef.org/topic/child-survival/under-five-mortality/#:~:text=The%20under-five%20mortality%20rate%20refers%20to%20the%20probability,age%20of%205%20dying%20every%20day%20in%202022.

Wickham, Hadley. 2011. "Testthat: Get Started with Testing." *The R Journal* 3: 5–10. https://journal.r-project.org/archive/2011-1/RJournal_2011-1_Wickham.pdf.

———. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

World Bank. 2024a. "Current Health Expenditure (." https://data.worldbank.org/indicator/SH.XPD.CHEX.GD.ZS?locations=US.

———. 2024b. "Food Production Index (2014-2016 = 100)." https://data.worldbank.org/indicator/AG.PRD.FOOD.XD.

———. 2024c. "Immunization, DPT (." https://data.worldbank.org/indicator/SH.IMM.IDPT.

———. 2024d. "Mortality Rate, Under-5 (Per 1,000 Live Births)." https://data.worldbank.org/indicator/SH.DYN.MORT.

World Health Organization. 2024a. "Child Mortality and Causes of Death." https://www.who.int/data/gho/data/themes/topics/topic-details/GHO/child-mortality-and-causes-of-death.

———. 2024b. "Children: Reducing Mortality." https://www.who.int/news-room/fact-sheets/detail/children-reducing-mortality#:~:text=There%20are%20two%20targets%3A%201%20Reduce%20newborn%20mortality,25%20per%201000%20live%20births%20in%20every%20country.

———. 2024c. "Civil Registration and Vital Statistics." https://www.who.int/news-room/fact-sheets/detail/civil-registration-and-vital-statistics.

Xie, Yihui. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in R." In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC.