Exploring Determinants of Child Mortality: The Role of Vaccination, Healthcare Spending, and Nutrition*

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

Junbo Li

November 21, 2024

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, focusing on the effects of vaccination coverage, health expenditure per capita, and food production indices. Using a Bayesian 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 critical need for targeted interventions in high-mortality regions, emphasizing 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

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

In this study, we examined the under-five mortality rate for each country in 2022, focusing on three key factors: food production index, DPT vaccine coverage, and per-capita health expenditure. These factors were chosen because they reflect critical components influencing child mortality. Infectious diseases, such as pneumonia, diarrhea, and malaria, remain leading causes of under-five deaths, and these are directly linked to vaccination coverage and healthcare spending. Similarly, maternal health during childbirth is influenced by healthcare accessibility and quality, which is often tied to national health expenditures (World Health Organization (2024)). While food production does not directly address infectious disease or maternal health, it reflects a nation's ability to meet nutritional needs, which is critical for preventing malnutrition—a major underlying cause of child deaths (add supporting explanation and citation).

Our findings reveal that increasing vaccine coverage and per-capita health expenditures significantly reduces under-five mortality rates. However, food production index does not exhibit a direct impact, suggesting that its role is mediated by other factors such as distribution systems and food quality. Despite progress in reducing global mortality, 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)).

Furthermore, the distribution of under-five mortality rates 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 (Arel-Bundock (2022)). This disparity highlights the urgent need for targeted interventions in high-mortality regions. The factors studied in this paper, such as vaccination and healthcare spending, are actionable and can be effectively improved in these regions through foreign aid and international collaboration.

The remainder of this paper is organized as follows. Section 2 describes the data sources, key variables, and preprocessing steps, including transformations applied to address skewness and improve model validity. Section 3 outlines the modeling approach, including the choice of predictors, justification of the model structure, and Bayesian implementation. Section 4 presents the results of the analysis, highlighting the effects of each predictor on under-five mortality. Finally, Section 5 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.

1.1 Estimand

The primary 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.

2 Data

2.1 Overview Check packages

For this analysis, we combined four data sets all together into one. All these four data sets comes from **Worldbank** open data platform (Arel-Bundock (2022)). 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** ((**knitr?**)), **arrow** (Richardson et al. (2024)), **here** (Müller (2020)), and **dpylr** (Wickham et al. (2023)). The paper is outlined in github using starter folder provided in **Telling Stories With Data** (Alexander (2023)).

2.2 Measurement

The data used in this study were collected from established international organizations and reflect key metrics in health and socioeconomic development. The under-five mortality rate, the primary response variable, is measured as the probability per 1,000 live births that a child will not survive to their fifth birthday, given age-specific mortality rates of the specified year. This indicator is developed by the UN Inter-agency Group for Child Mortality Estimation and leverages statistical models to reconcile differences among multiple data sources.

Predictors include:

Food Production Index: This metric, provided by the Food and Agriculture Organization, captures the aggregate volume of food production normalized to the 2014-2016 base period. It excludes non-nutritive edible items such as coffee and tea and accounts for deductions of intermediate agricultural inputs like seed and feed.

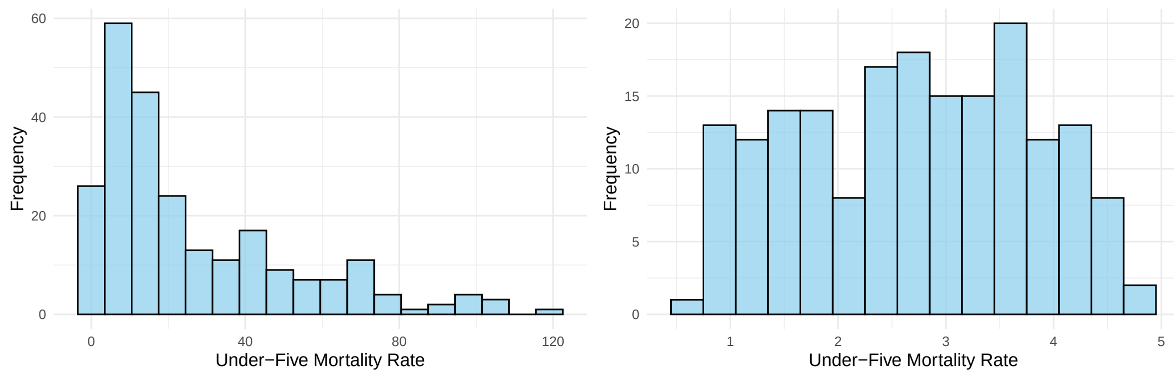
Health Expenditure per Capita: Sourced from the World Health Organization's Global Health Expenditure database, this indicator measures current health spending per capita in current US dollars, including goods and services consumed annually.

DPT Vaccine Coverage: This is the percentage of children aged 12-23 months who received three doses of the diphtheria, pertussis, and tetanus vaccine, based on estimates derived from national administrative data and household surveys.

2.3 Outcome variables

The response variable in our analysis is the mortality rate below 5 year-old children for all countries. Here Mortality rate is calculated based on number of deaths of children under 5 years old per 1000 person. The histogram of this variable is shown in Figure 1a. We observe that the data ranges from 0 to 120, with a peak at 10-15.

However, this data is extremely right skewed and have several outlyers, thus we decided to perform log transformation to stabilize the variance and in-proving the normality of our dataset for a better regression model latter. The distribution after we perform log transformation is shown in Figure 1b, here the distribution is approximately normal ranging from 2 to 5 with a peak at 3.



(a) Dsitribution before log tranformation

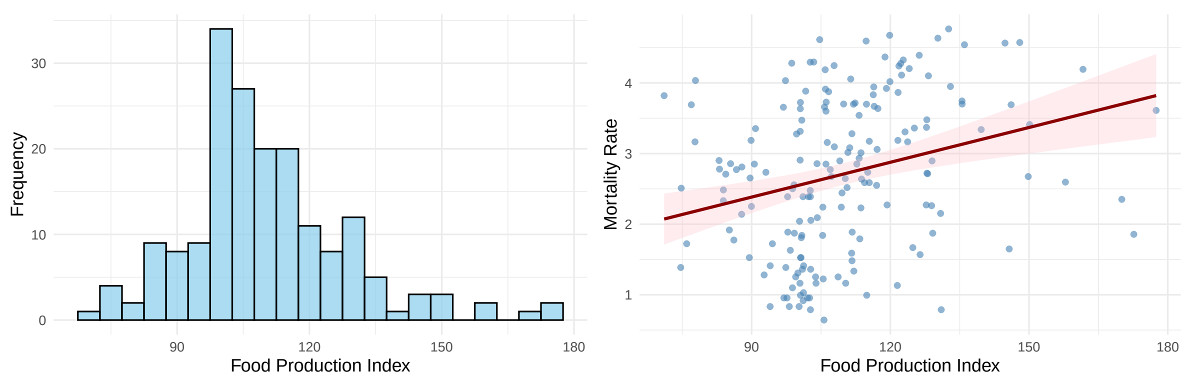
(b) Distribution after log transformation

Figure 1: Data Analysis for Mortality Rate of Different Countries

2.4 Predictor variables

In our mdoel, we have three Preditor variables, namely, Food production index, which is the relative level of agricultural production for each nation compared with the base period 2014-2016 (Arel-Bundock (2022)); Current health expenditure per capita (current US$); and DPT vaccine percentage.

We will first look at the food production index. From Figure 2a, we see that it ranges from 60 to 180, with a center at 110, and the shape is approximated normal distributed. Figure 2b is the scatter plot between Food production index and log of Mortality rate, with a best line of fit and standard error. We obsereve a slightly positive linear relationship between these two variables. Detailed relationship will be studied in Section 3.

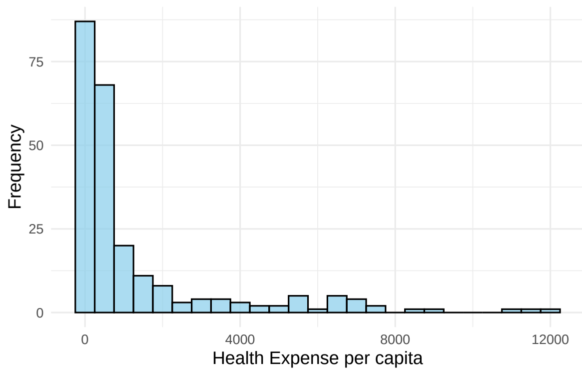


(a) Histogram of each countries' food production index

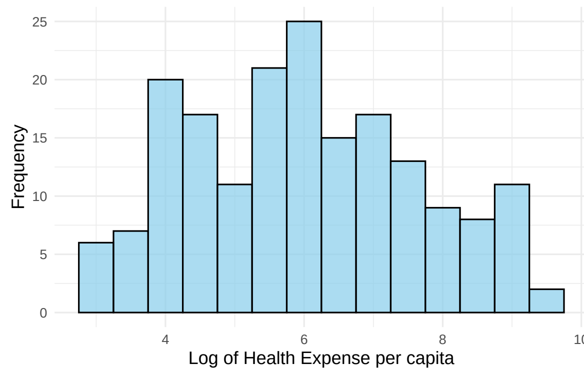
(b) Scatter plot of each countries' Food production index VS. Log of Mortality

Figure 2: Data Analysis of Predictor Food Production Index

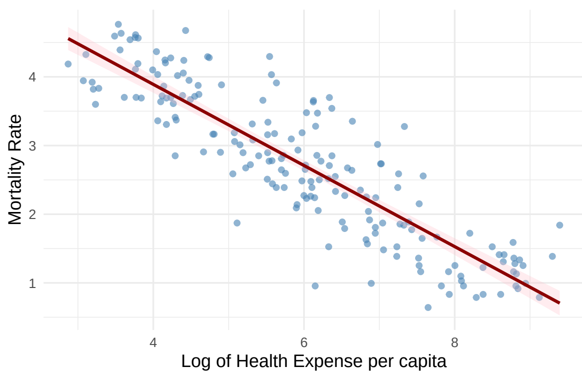
Next is the Current health expenditure. From Figure 3a, we see that it ranges from 0 to 12000, but the shape of the distribution is extremely skewed to the right. Thus we decided to perform a log transformation to stabilize the normal shape. Figure 3b is the histogram after the log transformation, we see that it now ranges from 0 to 10 with a center at 6 and the shape a approximately normal now. Figure 3c is a scatter plot between Log of Mortality rate and Log of Current health Expenditure, with a best line of fit and standard error. We observe a significant negative linear relationship between these two variables. Detailed relationshape will be studied in Section 3.



(a) Histogram of each countries' Health Expenditure Per Capita, Measured in US$. Before log transformation



(b) Histogram of each countries' Health Expenditure Per Capita, Measured in US$. After log transformation

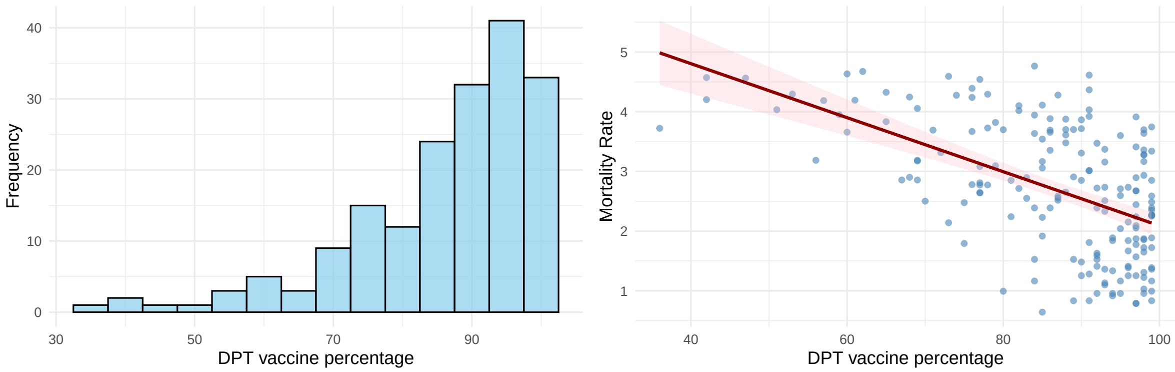


(c) Scatter plot of each countries' Health Expenditure Per Capita, Measured in US$ VS. Log of Mortality

Figure 3: Data Analysis of Predictor Current Health Expenditure

Finally is the DPT vaccine percentage. From Figure 4a, we see that it ranges from 30 to 100, but the shape of the distribution is skewed to the left, we will discuss this in Section 4. Here we do not have efficient tecnics to stablize the distribution. Figure 4b is a scatter plot between DPT vaccine percentage and Log of Mortality rate, with a best line of fit and standard error.

We observe a significant negative linear relationship between these two variables. Detailed relationship will be studied in Section 3.



(a) Histogram of each countries' DPT vaccine percentage

(b) Scatter plot of each countries' DPT Vaccine Percentage VS. Log of Mortality

Figure 4: Data Analysis of Predictor DPT Vaccine Percentage

2.5 Missing Data and Time Inconsistancy

In our file, the varible Health Expendure is collect for year 2021 while the others is collected in year 2022. We done this is because there is n data avliable for year 2022 of this data. We believe that these data are recent and so the data collected in year 2021 could still do a good job associating data in 2022. Also there are a few data missing for some minor countries. We made the decision to drop them and continue our modeling. We will discuss further implications of these two in Section 4

3 Model

The goal of our modeling strategy is to investigate how the DDT vaccine coverage, food production index, and current health expenditure per capita (current US$) relate to the under-five mortality rate for each nation.

We aim to use this model to understand how the above mentioned three factors' impact on child mortality and identify opportunities for improving health outcomes, especially for high mortality rate nations. Background details and diagnostics are included in Appendix B.

3.1 Model set-up

Let y_i be the logrithm of 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}: Logrithm of current health expenditure per capita (current US$) for nation i. The model is specified as:

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

We implement this model in R (R Core Team 2023) using the rstanarm package (Goodrich et al. 2022), employing its default priors for predictors and a distribution of Normal(_i,).

3.1.1 Model justification

We use a multivariate linear model to capture the relationship between the predictors and the response variable. This choice is justified by the linear trends observed in the data (see Section 2). The logarithmic transformation of the under-five mortality rate and current health expenditure per capita is applied to stabilize variance and linear relationships, ensuring model validity.

The Bayesian framework is employed due to its ability to incorporate prior knowledge, improve uncertainty quantification, and handle small sample sizes effectively.

4 Results

Our results are summarized in **?@tbl-modelresults**.

##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). This serves as a baseline for interpreting the effects of the predictors.

##Food: 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 naïvely 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, and food production is not strongly related to overall under-five mortality rates. This finding suggests that while food production is essential for societal well-being, its immediate impact on reducing child mortality might depend on other factors such as food access, distribution systems, and nutritional quality.

##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.

This highlights the critical role of vaccination programs in reducing child mortality. For example, increasing vaccine coverage by 10 percentage points could reduce the mortality rate by approximately 13%, emphasizing the importance of robust immunization initiatives.

##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.

Term	Estimate	Std. Error	2.5% CI	97.5% CI
(Intercept)	6.643	0.363	6.047	7.222
Food	0.003	0.002	0.000	0.007
Vacinne	-0.013	0.003	-0.019	-0.008
Health_expense	-0.529	0.029	-0.578	-0.482

5 Discussion

5.1 implications

The UN's Sustainable Development Goal 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 (n.d.)). While substantial progress has been made, achieving this target within the next six years presents significant challenges. This study confirms that while the global trend shows a decrease, the pace of this reduction has slowed since 2015. Furthermore, in Section 2 we found significant disparities between countries. Some nations already meet the SDG target, but others exhibit mortality rates four times higher than the threshold, predominantly in low-income regions such as sub-Saharan Africa.

To meet the SDG targets, urgent and focused international collaboration is required. Our findings emphasize two critical areas for intervention: vaccine coverage and healthcare spending. Vaccination programs have a clear and significant impact on reducing child mortality. Therefore, international aid should prioritize making newborn vaccines affordable, ensuring their availability, and facilitating universal access in high-mortality regions.

Healthcare spending presents a more complex challenge. It encompasses multiple factors, including a country's economic status, infrastructure development, and access to hospital services. Addressing these requires a sustained, long-term commitment. Developed countries can play a pivotal role by providing financial support, sharing technological advancements, and fostering knowledge transfer to enhance healthcare systems in developing nations. Without addressing these systemic issues, achieving equitable child survival rates will remain elusive.

5.2 limitation and next steps

This study has several limitations that should be addressed in future research. First, as discussed in Section 2, we lack data from several smaller countries. While their populations are minimal, excluding these nations raises potential biases and ethical concerns by not representing their unique challenges. Second, the most recent data available is from 2022. More current datasets would enable a more accurate analysis, reflecting recent developments in healthcare and child mortality trends.

Additionally, while this study focused on three predictors, the dataset contains other indicators that could potentially enrich the analysis. However, many of these indicators suffer from missing or outdated data, limiting their utility. Future research should prioritize ensuring data completeness and timeliness to enhance model robustness.

Lastly, advanced modeling techniques and the incorporation of additional predictors could yield deeper insights. Future studies should explore innovative models to capture complex relationships and interactions, thereby providing more nuanced recommendations for policy interventions.

Appendix

A Additional data details

A.1 Challenges in Observational Data and Sampling

The data used in this study, particularly under-five mortality rates, are largely derived from surveys, censuses, and observational data rather than complete vital registration systems. Observational data present unique challenges, such as recall bias in surveys, underreporting in regions with weak administrative systems, and inconsistencies in data collection methodologies across countries. These issues are particularly acute in low- and middle-income countries (LMICs), where infrastructure limitations impede accurate data capture.

Sampling Strategies To mitigate these challenges, the UN Inter-agency Group for Child Mortality Estimation employs a combination of direct and indirect estimation techniques. Direct methods rely on household surveys, such as Demographic and Health Surveys (DHS) and Multiple Indicator Cluster Surveys (MICS), which use stratified random sampling to ensure representativeness. These surveys ask respondents to recall birth and death histories, which introduces recall bias, particularly for events occurring further in the past.

Indirect methods use model life tables and historical data to extrapolate mortality estimates in the absence of recent or reliable survey data. These methods rely on assumptions about fertility, mortality trends, and the demographic structure of the population, which may not hold true in all contexts, particularly in regions experiencing rapid epidemiological transitions or conflict.

A.2 Linkages to Literature

The challenges and methodologies described align with findings in the literature. Alkema and New (2014) emphasized the importance of integrating multiple data sources to improve the reliability of child mortality estimates. They highlighted the use of Bayesian hierarchical models to synthesize survey and vital registration data, which is a key component of the UN Inter-agency Group's methodology.

Moultrie et al. (2013) detailed the limitations of indirect estimation methods, particularly their reliance on model assumptions that may not hold in regions undergoing rapid social or economic change. Their findings support the use of simulation studies, such as ours, to assess the robustness of mortality estimates under varying data quality scenarios.

A.3 Implications for Policy and Future Research

The limitations of survey-based and observational data highlight the need for:

Investment in Vital Registration Systems: Increasing the coverage and quality of birth and death registrations in LMICs would reduce reliance on indirect methods and improve the accuracy of mortality estimates. Improved Survey Methodologies: Incorporating digital data collection tools and cross-validation with other data sources can reduce recall bias and enhance data reliability. Advanced Statistical Techniques: Bayesian hierarchical models and machine learning approaches can integrate diverse data sources and account for biases more effectively than traditional methods. Future research should focus on combining survey data with novel data sources, such as satellite imagery and mobile health records, to address data gaps and improve mortality estimation in real time. These approaches could revolutionize the monitoring of global health indicators, enabling more targeted and timely interventions. (Add refer)

B Model details

B.1 Posterior predictive check

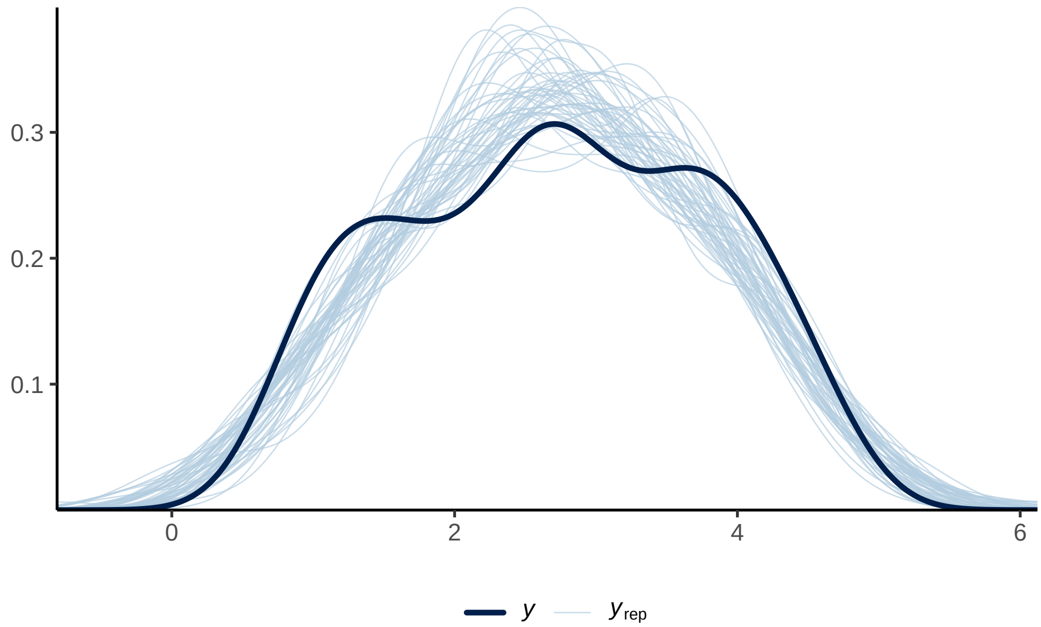
We conducted a posterior predictive check to evaluate how well the model predicts under-five mortality rates. Figure 5a illustrates the predictive distribution compared to the observed data. The close alignment between observed and predicted values indicates that the model captures the main trends in the data effectively.

In Figure 5b 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.

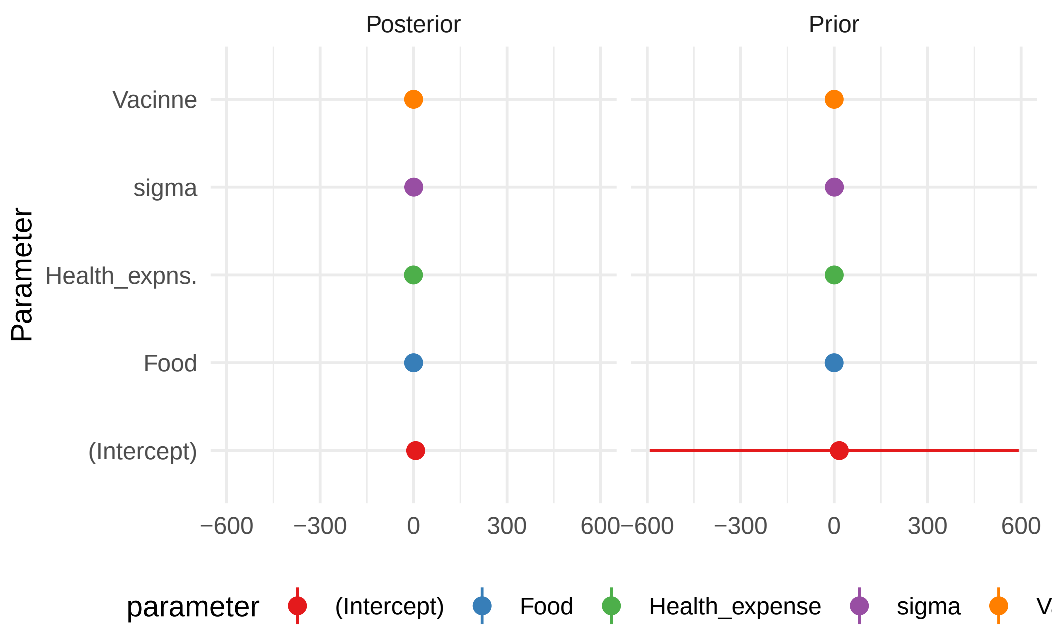
B.2 Diagnostics

Figure 6a 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 6b 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.

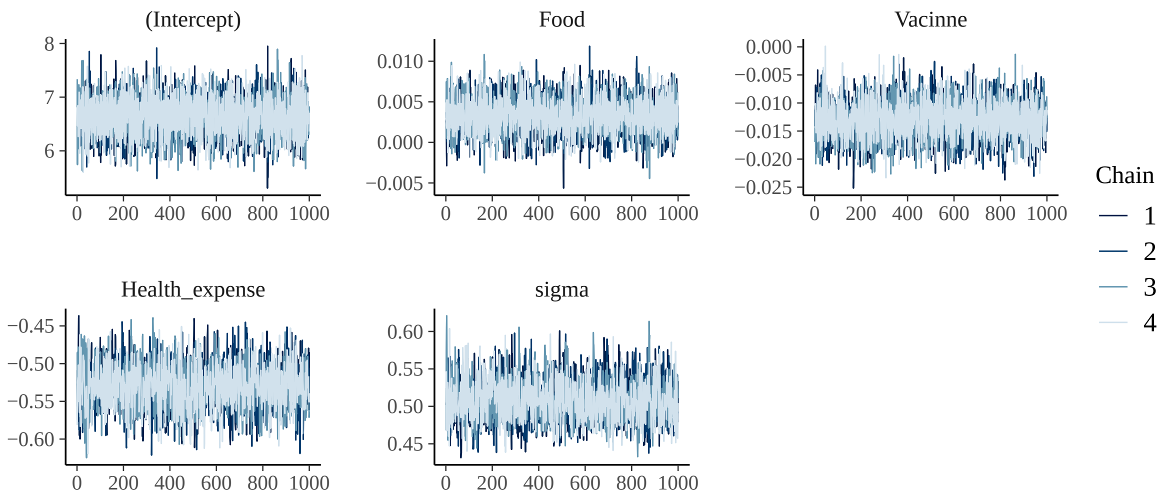


(a) Posterior prediction check

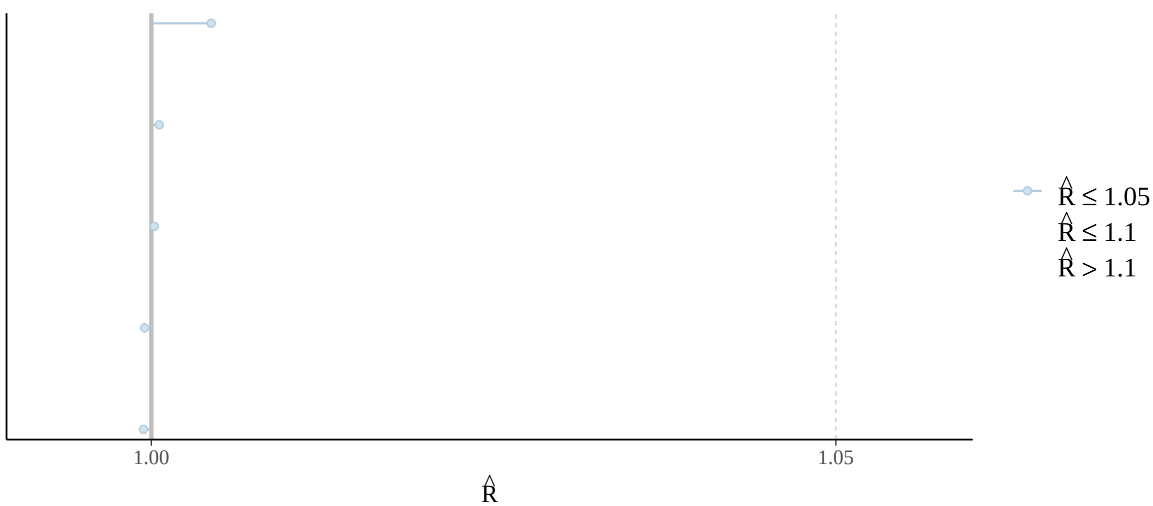


(b) Comparing the posterior with the prior

Figure 5: Examining how the model fits, and is affected by, the data



(a) Trace plot



(b) Rhat plot

Figure 6: Checking the convergence of the MCMC algorithm

References

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

Arel-Bundock, Vincent. 2022. *WDI: World Development Indicators and Other World Bank Data*. https://CRAN.R-project.org/package=WDI.

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

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

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

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.

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. 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 Health Organization. n.d. "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.

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