THE IMPACT OF SOCIOECONOMIC DETERMINANTS ON INFANT MORTALITY: AN ANALYTICAL APPROACH

Violeta Cvetkoska
Professor at Ss. Cyril and Methodius University in Skopje,
Faculty of Economics – Skopje
violeta.cvetkoska@eccf.ukim.edu.mk

Filip Peovski
Teaching and Research Assistant at Ss. Cyril and Methodius University in Skopje,
Faculty of Economics – Skopje
peovski@eccf.ukim.edu.mk

Damjan Stojkovski
Ss. Cyril and Methodius University in Skopje, Faculty of Economics – Skopje
stojkovski.damjan@outlook.com

Mihael Joshua Vlaisavljevikj
Ss. Cyril and Methodius University in Skopje, Faculty of Economics – Skopje
mihaelvlaisavljevic@yahoo.com

Jasna Joanidis
Ss. Cyril and Methodius University in Skopje, Faculty of Economics – Skopje
jasnajoanidis02@gmail.com

Anastasija Kopcharevska
Ss. Cyril and Methodius University in Skopje, Faculty of Economics – Skopje
kopcharevskaa@gmail.com

Nemanja Marikj
Ss. Cyril and Methodius University in Skopje, Faculty of Economics – Skopje
marikjnemanja@gmail.com

ABSTRACT
Investigating the impact of socioeconomic factors on infant mortality rates (IMRs) is of key importance to sustainable economic development. The lack of global studies is notable, with previous publications exploring these relationships only in specific regions or countries. Utilizing data from the World Bank and the World Health Organization (WHO), we fit a multiple linear regression model to examine the impact of individual variables such as the number of doctors per 1000 people, female literacy rate over 15 years, corruption index, and health expenditure per capita. Our findings reveal significant relationships between the natural logarithm of IMR and the natural logarithm of health expenditure per capita, as well as the number of doctors, suggesting that higher levels of health care expenditure and greater availability of medical workers significantly influence the level of infant mortality rates globally. This research deepens the understanding of the multifaceted determinants of IMR and highlights the importance of targeted interventions to improve health care. Consequently, policymakers and stakeholders can develop more effective strategies to promote long-term sustainable economic development and improve infant health outcomes.

10.47063/EBTSF.2024.0024
http://hdl.handle.net/20.500.12188/31970

Keywords: *Regression, Infant mortality, Socioeconomic factors, Health care system, Sustainable development, Health expenditure.*

JEL classification: *I15, C20.*

1. INTRODUCTION

Infant mortality, defined as the death of a child under the age of one, is one of the most significant indicators of a country's healthcare system, development, and socioeconomic conditions. Despite substantial advancements in healthcare and medical technology in recent decades, there are still notable disparities among countries and regions when it comes to infant mortality rates (IMRs). A complex interplay between socioeconomic and environmental factors occurs, which deserves a higher focus of the scientific work globally. Namely, economic development can be observed through the national healthcare system, with the health sector and economic conditions being mutually dependent. Although similar studies on the determinants of infant mortality can be found, none focus on the specific choice of factors nor analyses the global context with the latest data. We aim to fill this gap in the academic literature.

This paper aims to uncover the impact of various socioeconomic factors on infant mortality rates as a proxy variable for the quality of the national healthcare system. This is achieved through secondary cross-sectional data for 2023 obtained mainly through the World Bank and the World Health Organization (WHO) databases. A total set of 227 countries is analysed, segregated into several geographical regions. We looked at external, socioeconomic factors to see if they are significant in determining the national infant mortality rates. For instance, the increased government spending in the health sector affects health infrastructure and the overall quality of health, which could lead to a theoretical decrease in the mortality of newborns. When deliveries are performed by professional medical staff, such as adequately qualified doctors, nurses, and midwives, they contribute to better medical care, monitoring, and intervention that reduce the risk of complications and death of the baby or the mother. We also observe the impact of corruption, as it can lead to the inappropriate allocation of resources, including funds, medical supplies, and personnel that will later result in a reduction in the quality and availability of medical care, potentially contributing to increased infant mortality. We hypothesize that specific healthcare and socioeconomic metrics will have a direct impact on infant mortality rates (IMR), with clear statistical correlations between a combination of key variables and IMR on a global scale. This analysis investigates which metrics are most effective in improving IMR through rigorous statistical methods.

Through a multiple linear regression model, we observe the impact of corruption, the number of doctors per 1000 people, health expenditures per capita, and female literacy rates as independent variables on the infant mortality rates. General results confirm the theoretical expectations of lower infant mortality rates in countries with higher health expenditure per capita, as well as a higher number of available medical staff. The conducted machine learning regression also proves that GDP per capita, the available medical staff, availability of health insurance are significant determinants of IMRs as well as confirming the regional disparities in the world.

The study is structured as follows. After the brief introduction to the topic, the paper highlights the key concepts in the global literature review in Section 2. Next, the research provides a detailed exploratory data analysis followed by the methodological approach. In Section 4, the study elaborates the obtained results and provides key insights into the topic. Finally, an adequate conclusion is given.

2. LITERATURE REVIEW

Healthcare and economics are intertwined as there is no healthy economy without a healthy nation. As stated by Erdoğan *et al.* (2013), 'health is one of the necessary elements to socially develop a society'. Global literature has discussed this topic extensively over the years, focusing on various aspects deemed important. Most research on this topic has focused on the medical part of the equation, concentrating on genetic predispositions, medical conditions, and other, risk factors (see Baraki *et al.*, 2020; Satti *et al.*, 2023; Fisher *et al.*, 2024, p. 14). However, we conducted this research to examine the influence of relevant socioeconomic determinants on the quality of health in the country, reflected through infant mortality, which in turn is proven to be in a long-term relationship with the business cycles (Ogburn and Thomas, 1922; Schady and Friedman, 2007). Subsequently, this has also been confirmed by Maruthappu *et al.* (2010).

Infant mortality can represent a major economic burden. In developing countries, this burden is borne by a large number of poor families, while in developed countries such as the United States, the burden is reflected in high health costs (Okobi *et al.*, 2023). It is worth noting that some of the potential determinants of IMRs have been previously studied, however on a smaller scale. For instance, Bexson *et al.* (2021) studied the effects of the number of doctors per 1000 citizens in Brazil. They concluded that increasing their number could only affect the underdeveloped areas of the country, in turn signaling a lower marginal utility in the developed Brazilian regions.

Less developed countries are more commonly studied in the global academic literature. For instance, Jayachandran and Jarvis (1986) found that good nutrition and midwife availability are more significant determinants of infant mortality rates than formally trained healthcare staff in 60 less-developed economies. Stockwell and Wicks (1984) found a strong inverse relationship between low-income families and the IMRs, noting that socioeconomic disparities may hinder medical advancements in developed regions such as metropolitan Ohio.

Another study (Salariak *et al.*, 2009) found that female literacy rates, welfare, and access to high-quality healthcare are major determinants of infant mortality rates. For similar findings, we suggest Ahmed *et al.* (2011) as well as Kim and Saada (2013) which provide a systematic literature review on the topic. Additionally, it is worth noting that the work of Prisco *et al.* (2015) shows that on the European level, higher GDP and female education levels are strongly associated with lower rates of infant mortality, unlike the income inequality measured through the Gini index which does not show statistical significance. This is considered opposite to the findings of Waldmann (1992) whose study highlights that high infant mortality rates are associated with income inequality.

On the contrary, Younger (2001) argues that economic growth, expressed through GDP per capita, does not consistently influence infant mortality, but rather variables like primary school enrolments and DPT vaccination rates for infants do so. Moreover, the availability of healthcare measured through the number of doctors, nurses, and hospital beds per 1000 inhabitants has no impact on the reduction of IMRs. Using a country fixed-effects models, Conley and Springer (2001) find that independent of economic development, higher public health spending significantly lowers the rates of infant mortality.

It is worth noting that machine learning (ML) is becoming increasingly popular among studies focusing on developing predictive models and unraveling non-linear relationships, especially in the field of pathology (Islam *et al.*, 2020; Dahiwade *et al.*, 2019; Jing *et al.*, 2023). Some common ML algorithms include decision trees, support vector machines (SVM), k-nearest neighbors (KNN), and random forests. For larger datasets in which interpretability of the results is not paramount, neural networks are used (Ahmed *et al.*, 2019; Cabaneros *et al.*, 2019; Jiang and Luo, 2022).

Our study addresses the gap in the literature by conducting a large-scale global study on the latest data, focusing on the staff sufficiency of the healthcare systems, health expenditures, literacy rates in females, GDP per capita, and even the country's corruption level. Additionally, Microsoft Power BI's Key Influencers ML tool was utilized to allow for identifying key factors in a supervised learning framework without extensive manual data preprocessing, making it accessible for large-scale data analysis. With this, existing literature is complemented while potentially raising questions for further broadening of the study.

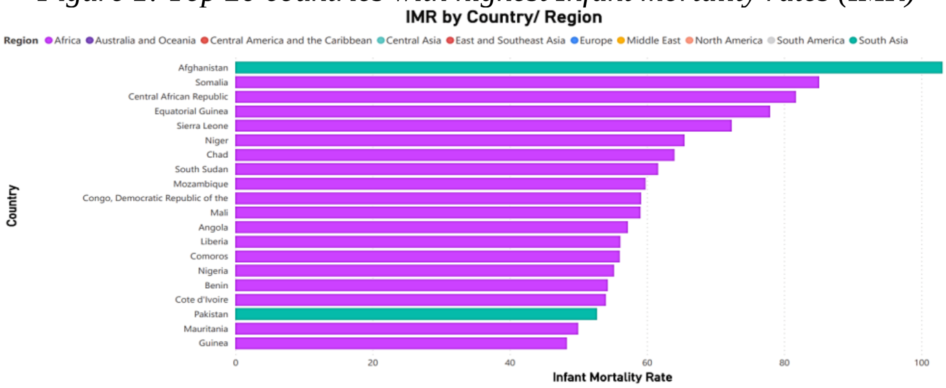
3. METHODOLOGICAL APPROACH

3.1. Exploring data characteristics

Secondary cross-section data for 227 countries in 2023 was utilized and extracted from various sources, such as the World Health Organization (WHO), the World Bank, CIA (Central Intelligence Agency), and the PRB (Population Reference Bureau). The main dependent variable in our study is the infant mortality rate (IMR) measured as the number of deaths of children under the age of one, per 1,000 live births.

In Figure 1, we can see that the dominant regions in terms of high infant mortality are Africa and South Asia, in countries that are still underdeveloped. Afghanistan stands out with the highest death rate per 1,000 infants in 2023 or 103.1. On the other hand, the lowest death rate is found in the European countries. For instance, Singapore and Slovenia had the joint-lowest rate of 1.5 which is far below the rates found in African countries. Low IMRs are also characteristic for countries in North America and East and Southeast Asia.

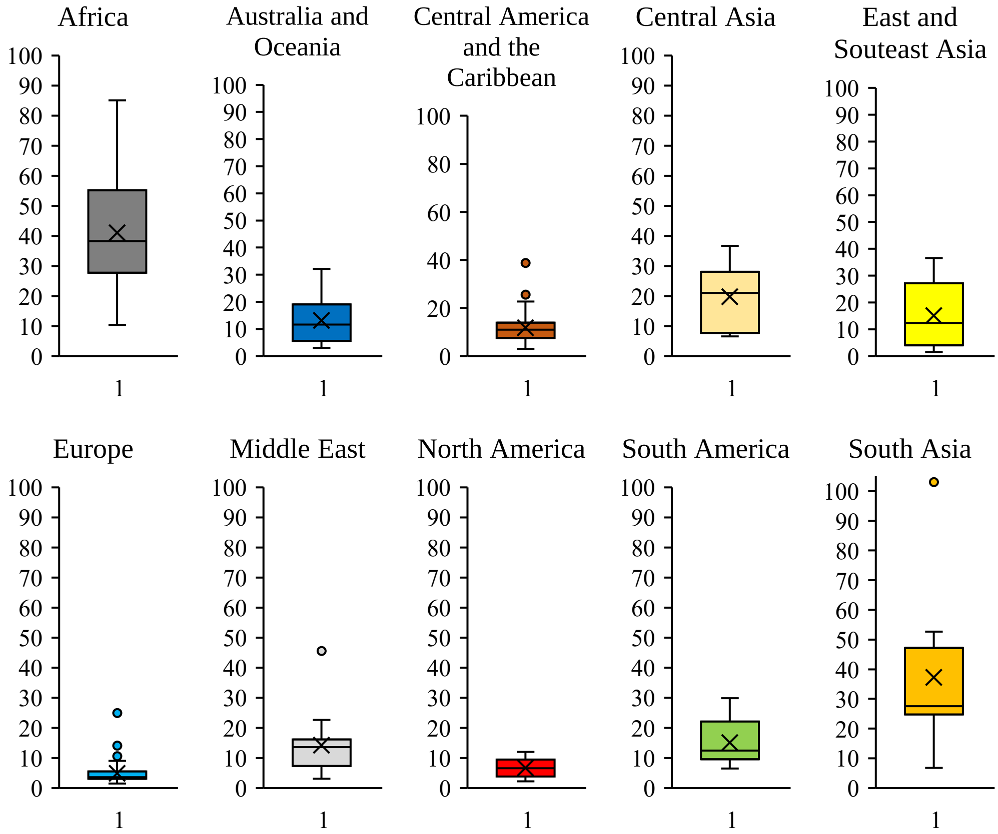
Figure 1: Top 20 countries with highest Infant mortality rates (IMR)



(Source: World Health Organization; Authors' depiction)

The box and whisker charts show that there are huge differences in IMR across the world and by region. Africa has the highest median IMR with large IQR and outliers meaning high variability and higher IMR. On the other hand, the median IMRs of Europe, North America, Australia and Oceania are much lower with a smaller IQR, implying that the mortality rates in these areas are more stable and lower. Central America and the Caribbean, Central Asia, and East and Southeast Asia have moderate median IMRs, with their corresponding IQR and outliers to indicate the ranges and disparities within the regions. The Middle East and South Asia have higher median IMRs than the more developed regions, although not as high as Africa; South Asia also has some very high outliers. In general, these charts show the contrast of the economically developed regions with the regions where the health of infants remains a crucial problem.

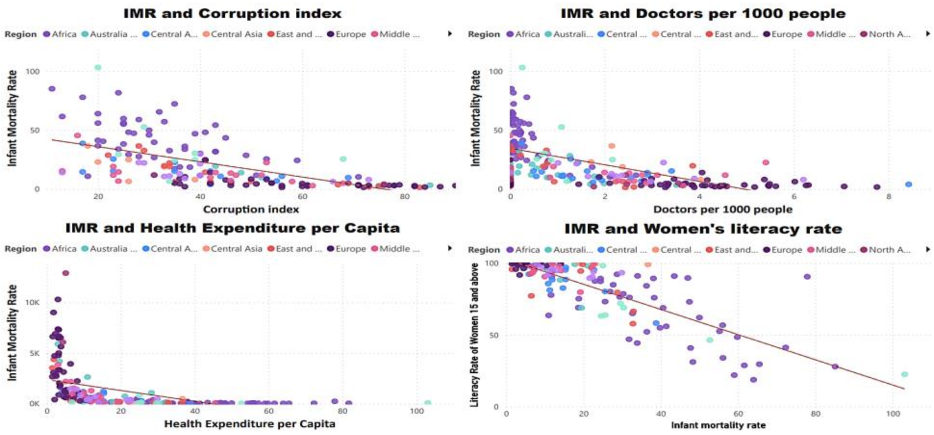
Figure 2: Boxplot of infant mortality rates (IMRs) by regions



(Source: World Health Organization; Authors' depiction)

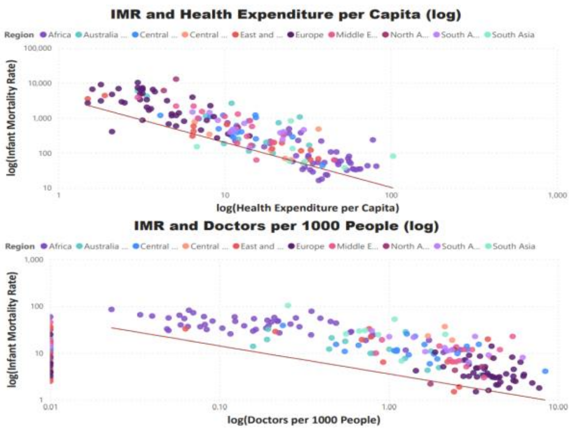
To check for the direction and the behaviour of the relationship across data points, we construct several scatter plots, with variable pairs depicted in Figure 3. There is a notable inverse relationship between the corruption index and the infant mortality rates. This is important because corruption translates through inefficient and flawed institutions, lack of ethical criteria as well as inefficient allocation of resources – mainly in the health sector. Furthermore, we have to note that a higher corruption index is desirable and means a lower corruption level for the country. Data points on the left-hand side of Figure 3a data show higher variability, as infant mortality rates are naturally more dispersed in those countries due to the influence of several other factors. Next, there is a clear inverse relationship between the IMRs and the number of doctors available, measured per 1000 inhabitants. The availability of qualified medical staff is most notable in European countries, ranging between 3 and 6 doctors per 1000 people. On the contrary, the lack of such work profiles in African countries is highly correlated with the substantially higher IMRs. A slightly non-linear curve, even though much more prominent than the one for the previous case, maybe a better fit for the relationship between IMRs and health expenditures per capita. That being said, the logarithmic form, depicted in Figure 4, is considered to pronounce the linear relationship between the variables of interest. Nevertheless, there is a strong inverse relationship which is expected in theory, given that higher investments in medical staff, infrastructure, and technological investments are indeed associated with lower rates of infant mortality. In Figure 3d, we can observe the relationship between women's literacy rate and infant mortality. That being said, educated women have better knowledge about health, hygiene, nutrition, and overall healthcare utilization. Women in less developed regions show a lower tendency to seek medical attention during pregnancy and childbirth, thus resulting in higher infant mortality rates.

Figure 1a, 3b, 3c, and 3d: Scatter plots of Infant Mortality Rate (IMR) and Corruption Index, Doctors per 1000 people, Health Expenditure per capita, and Female Literacy Rate across countries



(Source: Authors' depiction)

Figure 4a and 4b: Infant Mortality Rate and logarithms of Health Expenditure per capita, and Doctors per 1000 People



(Source: Authors' depiction)

The correlation matrix with all the variables analysed prior to the modeling phase is presented in Table 1. It is worth noting that some of the variables were selected to check the consistency of the data. This is achieved by examining the correlation between infant mortality and similar indicators such as mortality rate, birth rate, and perinatal mortality rate. From the variables that were of interest to the research, those that have a significant degree of correlation and of which there are possible mechanisms through which they would theoretically influence IMRs were selected. The variables in this research are divided into 3 groups: I) Variables that are defined by the same or similar metrics as IMR (variables 2, 3, 19 from Table 1); II) Socioeconomic and healthcare variables (metrics) that have previously mentioned theoretical grounds to influence IMR (var. 1, 4-13, 15-18, 20, 21 from Table 1); III) IMR as the primary variable of the study (var. 14 from Table 1). It is important to note that there is a large amount of data for individual regions, which is why wrong conclusions can be drawn when analysing data by region. Some control variables were also implemented. For instance, we observe a large positive correlation

(0.8) between the IMRs and maternal mortality ratio, which is theoretically expected due to the complexity of the delivery process. A significantly large positive correlation is also noted between the IMRs and birth rates (0.85) which can be self-explanatory and spur multicollinearity in the modelling process afterwards. A moderate inverse relationship is observed between the IMRs and the number of doctors per 1000 citizens (-0.65), while a slightly lower correlation is noted with the health expenditures per capita for 2021 (-0.47). A moderate inverse relationship is noted between the infant mortality rates and the country's corruption index or -0.61. Surprisingly, real GDP per capita for 2022 does not seem to exhibit any correlation with the infant mortality rates in 2023, which may indicate that non-economic factors may be more significant for the analysis itself.

Table 1: Correlation matrix

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1	1.00	-0.09	0.13	0.21	0.24	0.11	-0.03	-0.07	0.26	0.03	0.05	0.00	-0.08	-0.05	0.02	0.20	0.24	-0.05	0.10	0.09	-0.05
2	-0.09	1.00	-0.29	-0.57	-0.72	-0.04	-0.77	0.16	-0.53	0.46	-0.58	-0.27	-0.46	0.85	0.76	0.21	-0.45	-0.33	-0.81	-0.55	-0.79
3	0.13	-0.29	1.00	0.51	0.41	0.09	0.03	-0.17	0.29	0.01	0.20	0.26	0.19	-0.07	-0.01	-0.24	0.18	0.31	0.05	0.11	0.07
4	0.21	-0.57	0.51	1.00	0.61	0.03	0.47	-0.21	0.58	-0.11	0.37	0.27	0.55	-0.51	-0.49	-0.23	0.30	0.27	0.48	0.27	0.55
5	0.24	-0.72	0.41	0.61	1.00	0.08	0.56	-0.21	0.75	-0.15	0.62	0.24	-0.21	-0.65	-0.59	0.35	0.58	0.41	0.71	0.52	0.61
6	0.11	-0.04	0.09	0.03	0.08	1.00	-0.05	0.04	0.09	0.04	0.09	0.00	-0.06	0.01	0.13	0.18	0.16	0.09	0.02	0.02	0.50
7	-0.03	-0.77	0.03	0.47	0.56	-0.05	1.00	-0.06	0.45	-0.41	0.52	0.24	0.13	-0.77	-0.80	0.24	0.31	0.34	0.69	0.56	0.78
8	-0.07	0.16	-0.17	-0.21	-0.21	0.04	-0.06	1.00	-0.09	-0.32	-0.14	-0.16	0.39	0.18	0.02	-0.56	-0.21	-0.13	-0.21	-0.11	0.01
9	0.26	-0.53	0.29	0.58	0.75	0.09	0.45	-0.09	1.00	-0.02	0.76	0.18	0.01	-0.52	-0.50	0.40	0.77	0.40	0.62	0.47	0.57
10	0.03	0.46	0.01	-0.11	-0.15	0.04	-0.41	-0.32	-0.02	1.00	-0.11	0.04	-0.13	0.49	0.38	0.41	0.15	0.08	-0.20	-0.03	-0.58
11	0.05	-0.58	0.20	0.37	0.62	0.09	0.52	-0.14	0.76	-0.11	1.00	0.19	-0.09	-0.61	-0.48	0.59	0.76	0.48	0.69	0.56	0.47
12	0.00	-0.27	0.26	0.27	0.24	0.00	0.24	-0.16	0.18	0.04	0.19	1.00	-0.18	-0.27	-0.23	-0.03	0.13	0.18	0.23	0.20	0.22
13	-0.08	-0.46	0.19	0.55	-0.21	-0.06	0.13	0.39	0.01	-0.13	-0.09	-0.18	1.00	-0.12	0.06	-0.29	-0.08	0.04	0.08	0.23	-0.80
14	-0.05	0.85	-0.07	-0.51	-0.65	0.01	-0.77	0.18	-0.52	0.49	-0.61	-0.27	-0.12	1.00	0.80	-0.68	-0.47	-0.25	-0.84	-0.58	-0.80
15	0.02	0.76	-0.01	-0.49	-0.59	0.13	-0.80	0.02	-0.50	0.38	-0.48	-0.23	0.06	0.80	1.00	-0.49	-0.33	-0.23	-0.79	-0.54	-0.77
16	0.20	0.21	-0.24	-0.23	0.35	0.18	0.24	-0.56	0.40	0.41	0.59	-0.03	-0.29	-0.68	-0.49	1.00	0.40	0.37	0.59	0.29	0.68
17	0.24	-0.45	0.18	0.30	0.58	0.16	0.31	-0.21	0.77	0.15	0.76	0.13	-0.08	-0.47	-0.33	0.40	1.00	0.51	0.63	0.51	0.45
18	-0.05	-0.33	0.31	0.27	0.41	0.09	0.34	-0.13	0.40	0.08	0.48	0.18	0.04	-0.25	-0.23	0.37	0.51	1.00	0.40	0.57	0.24
19	0.10	-0.81	0.05	0.48	0.71	0.02	0.69	-0.21	0.62	-0.20	0.69	0.23	0.08	-0.84	-0.79	0.59	0.63	0.40	1.00	0.64	0.66
20	0.09	-0.55	0.11	0.27	0.52	0.02	0.56	-0.11	0.47	-0.03	0.56	0.20	0.23	-0.58	-0.54	0.29	0.51	0.57	0.64	1.00	0.56
21	-0.05	-0.79	0.07	0.55	0.61	0.50	0.78	0.01	0.57	-0.58	0.47	0.22	-0.80	-0.80	-0.77	0.68	0.45	0.24	0.66	0.56	1.00

(Source: Authors' calculations)

Note: 1 – (immigration-emigration)/1000 citizens; 2 – Birth rate; 3 – Death rate; 4 – Hospital beds/1000 citizens; 5 – Doctors/1000 citizens; 6 – Real GDP per capita (2022); 7 – Births attended by skilled staff; 8 – Low birthweights babies; 9 – Nurses and midwives/1000 citizens; 10 – Income per annum; 11 – Corruption index; 12 – Health insured population (in %); 13 – Hospital stay; 14 – Infant mortality rate; 15 – Maternal mortality deaths/1000; 16 – Natural deliveries per 1000 births; 17 - Health expenditure per capita (2021); 18 – Total health expenditure rate (2021); 19 – Life expectancy; 20 – Government expenditures on health (in %); 21 – Literacy rate in females aged above 15.

3.2. Methodology

We conduct three different regression analyses through the Eviews and Microsoft Power BI software tools. First, in order to select the independent variables that could have an impact on the mortality of newborns, a simple linear regression was performed with each of the possible factors. Subsequently, we observed the coefficient of determination (R^2) as an indicator of the explanatory power of each variable. The method used to estimate the linear regression model is the ordinary least squares (OLS) method. The linear regression model is defined as

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

where Y is the dependent variable i.e., infant mortality rate in our case, X is the independent variable, while β_0 and β_1 are intercepts and slope parameters to be estimated. With ε we denote the residual variability in the model.

Second, we conducted a multiple linear regression model to find the strongest relationships between a combination of factors and the infant mortality rates. Simple transformations of variables were applied to the most significant regression models to meet the theoretical assumptions for regression of cross-sectional data: linearity, homoscedasticity, normality (normal distribution of residuals), and independence (absence of autocorrelation in residuals). The models were selected by comparing several statistical indicators: the statistical significance of each coefficient separately, the statistical significance of the model (F-statistic), the coefficient of determination (R^2 and the adjusted R^2 for comparison between models with different numbers of variables) as well as different information criteria (i.e., Akaike, Bayesian). The general form of the models that showed the best compromise between their fit and complexity is

$$Y = \beta_0 + \beta_1 X + \beta_2 X_2 + \cdots + \beta_i X_i + \varepsilon$$

or

$$Y = \beta_0 + \sum_{i=1}^{k} \beta_i X_i + \varepsilon$$

The slope coefficients indicate a linear relationship at a marginal change in the corresponding independent variable, all else being equal. By transforming all variables in a given model, into their logarithm form, we explore a relationship in which the effect of each variable on the outcome is proportional to its natural logarithm.

Finally, the Key Influencers tool from Microsoft Power BI software was used to capture associations between a larger number of variables that could potentially have an impact on increasing or decreasing infant mortality. Key Influencers determine which variables have the most pronounced influence on the target variable. This tool uses machine learning algorithms to analyse statistical measures like correlation, performs linear regressions, logistic regressions (for categorical, qualitative data like region in our case), decision tree analyses, and association rule learning). We specifically utilize it to solve a regression problem through a machine learning setting, thus introducing a new method of modelling besides the ones already elaborated in the global literature. The analyses help us assess theoretically proposed causal relationships between certain socioeconomic indicators and infant mortality.

4. ANALYSIS OF RESULTS

In the first step, we performed several regression analyses to examine the assumptions for modeling cross-section data as well as the explanatory power of each variable, observed through the R^2 in the models. We found that most of the associations of the independent variables with the dependent variable are heteroscedastic, that is, ten out of the total thirteen samples, with one of them being excluded. The analyses that showed heteroscedasticity included the following independent variables: state expenditures on health; birth rate; deliveries performed by skilled and qualified employees; corruption; doctors per 1000 people; health expenditure per capita; life expectancy; maternal mortality rate; and nurses per a thousand people.

Out of the conducted regression analyses (with IMR as the dependent variable), the ones containing the following independent variables exhibited homoscedasticity: hospitals per capita, natural births per thousand births, female literacy rate, and hospital beds per one thousand people. The last was considered homoscedastic due to the Breusch-Pagan-Godfrey test showing values near the critical values. We then examined the normality of the individual models, where we found that none were normal, with a small exception. The regression analysis where the maternal mortality rate was the independent variable, although de facto it had no normality at a 95% confidence level, the distribution of its residuals was characterized by only a slight leptokurcy. However, after the in-depth examination of the regression analyses, we can still conclude that the heteroscedasticity, as well as the normality of the residuals, is not as important due to the nature and objectives of the research, especially since the results of the real data showed similarity with the transformed data in the further multivariate analyses.

In addition to this, we also looked for correlation in all samples, from which we speculated the risk of multicollinearity due to certain high correlations between independent factors, which can increase errors in multifactor analysis. Hence, medium/high correlations between two independent variables that we should consider are about four, that is, the correlation between doctors per thousand people in relation to corruption, which is 0.61; the correlation of 0.58 between doctors per thousand people and state expenditures on health; doctors per one thousand people in relation to the literacy rate of women (0,6); corruption and public expenditure on health, where the correlation is quite strong with 0.75. In our analysis there were more independent variables with strong correlations, but we stuck to the four above, because they are part of our two main models, which we used to study IMR. Furthermore, we obtained the coefficient of determination (R^2) for each regression, i.e. we found out which of the independent variables explain our dependent variable the best. We found that birth rate with 72%, life expectancy rate with 70%, maternal mortality rate per thousand births with 64% and female literacy rate also with 64% statistically explain the variability of infant mortality rate.

Table 2: Simple linear regression between the IMRs and chosen factors.

Variable	*Coefficient*	R^2	*Heteroskedasticity*	*Normality*	*Obs.*
Govt. Expenditure on healthcare in %	-0.2415	0.3392	Heteroskedastic	Not normal	166
Birth rate	1.7784	0.7275	Heteroskedastic	Not normal	227
Births attended by skilled staff	-0.8829	0.5933	Heteroskedastic	Not normal	176
Corruption index	-0.6420	0.3702	Heteroskedastic	Not normal	163
Doctors per 1000 citizens	-6.9900	0.4286	Heteroskedastic	Not normal	181
log(Govt. Expenditure on healthcare in %)	-8.6633	0.5740	Heteroskedastic	Not normal	178
Hospital beds per 1000 citizens	-4.3349	0.2616	Homoskedastic	Not normal	174
log(Hospitals per capita)	-5.9210	0.1195	Homoskedastic	Not normal	123
Life expectancy	-2.1425	0.7098	Heteroskedastic	Not normal	182
Maternal mortality per 1000	0.0740	0.6417	Heteroskedastic	Normal	186
Natural deliveries per 1000 births	-0.0158	0.4636	Homoskedastic	Not normal	31
Nurses and midwives per 1000 citizens	-2.2034	0.2713	Heteroskedastic	Not normal	188
Literacy rate in females aged above 15	-0.7299	0.6406	Homoskedastic	Not normal	128

(Source: Authors' calculations)

After conducting the multiple linear analysis, about 69% of the variability IMR in 112 countries is explained by the percentage of literacy among women over 15 years and the index of corruption. The linear association is represented by the coefficients of the corresponding (explanatory) variables ceteris paribus: for each percentage point increase in the literacy rate among women, infant mortality decreases by 0.63 and for every point of increase in the corruption index, mortality decreases by 0.36. The residuals do not have a normal distribution and according to the Durbin-Watson statistic (2.6969), there is autocorrelation in the residuals. Although the results indicate that the model is heteroscedastic, for research purposes, we consider it to be homoscedastic, as the probabilities of the Chi-square and F statistics from the conducted Breusch-Pagan-Godfrey heteroscedasticity test on this specific model are close to the critical values (0.05). In order to meet the assumptions of cross-sectional data regression, a model was estimated in which all variables were transformed with a natural logarithm. The transformation of the variables solves all the problems with the fulfillment of the criteria (Table 4). Akaike and Schwartz's (Bayesian) information criterion indicate that the newly obtained model is more suitable for use, but the reduced coefficient of determination (54.6%) indicates that the phenomenon is explained in a smaller percentage by the transformed data. However, the non-linear relationship shown in this model is significant and provides insight into the interdependency of the variables. According to the obtained result, about 60% of the dependent variable is explained through the phenomena of the rate of doctors and healthcare expenditures per capita. The linear relationship indicates a decrease in mortality per 1000 newborns of -2.07 for each doctor per 1000 population on average, *ceteris paribus*, and a decrease of about 7 on average, *ceteris paribus* in the dependent variable for each natural logarithm unit of a dollar per capita spent on health. The variable was logarithmized due to large differences between series values. We believe that the logarithmic series is still a representative indicator of health spending in individual states. There is no autocollinearity, but the model is heteroskedastic and the distribution of residuals is not normal as per Table 4. The transformations in the final model contribute to the fulfillment of all the necessary assumptions for such a model confirmed by the heteroscedasticity, multicollinearity and normality tests (Table 4). In this model, the Infant Mortality Rate (IMR), health expenditure per capita, and the number of doctors per 1000 people were log-transformed using the natural logarithm. According to the information criteria, this model compared to all the others is by far the most suitable and according to the coefficient of determination, it is explained with the highest percentage (78.568%). All previous models were considered solely because of the economic significance of the variables and results.

Table 3: Modelling results of infant mortality rates.

	Model 1	**Ln Model 2**	**Model 3**	**Ln Model 4**
Intercept	89.4818^{***} (4.3008)	11.1806^{***} (0.7431)	66.7947^{***} (4.4031)	5.0095^{***} (0.2466)
Female literacy rate	-0.6310^{***} (0.0561)			
Corruption index	-0.3604^{***} (0.0858)			
ln(Female literacy rates)		-1.1226^{***} (0.1865)		
ln(Corruption index)		-0.9714^{***} (0.1767)		
Doctors per 1000 citizens			-2.0771^{**} (0.8214)	
ln(Health expenditures per capita)			-6.9788^{***} (0.9178)	-0.4074^{***} (0.0410)
ln(doctors per 1000 citizens)				-0.1894^{**} (0.0477)
Included Obs.	112	112	164	164
R-squared	0.6902	0.5463	0.6000	0.7857
Adjusted R-squared	0.6846	0.5380	0.5950	0.7830
S.E. of regression	11.2121	0.6262	12.2395	0.4836
SSR	13,702.62	42.7458	24,118.46	37.6545
Log likelihood	-428.1044	-104.9801	-641.9570	-112.0503
F-statistic	121.4397^{***}	65.6194^{***}	120.7401^{***}	295.1107^{***}
AIC	7.6983	1.9282	7.8653	1.4031
BSC	7.7711	2.0010	7.9920	1.4598
HQ	7.7278	1.9576	7.8884	1.4261
DW statistic	2.6969	2.1683	2.0145	1.9778

Note: *,**,*** indicate statistical significance at the 10%, 5%, and 1%, respectively, based on the conducted t-test.

(Source: Authors' calculations)

Table 4: Heteroscedasticity and multicollinearity test results.

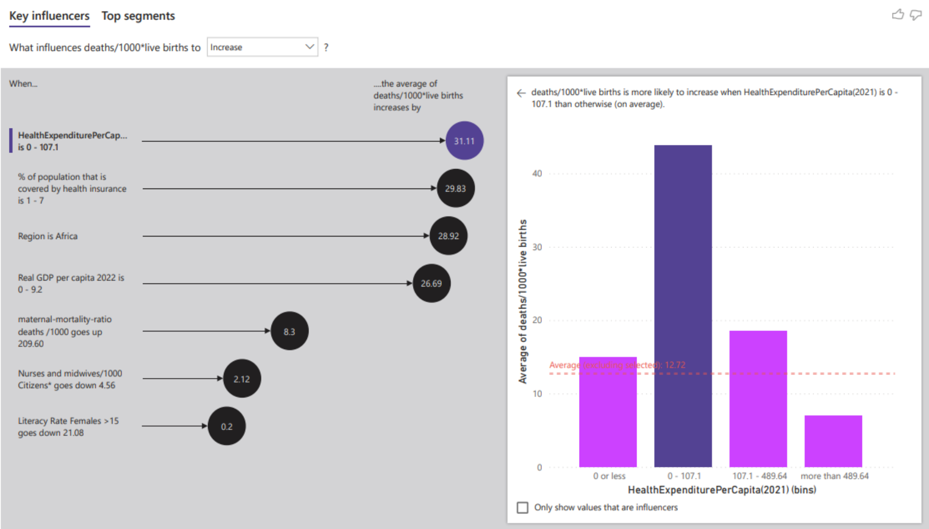
Model num.	Model 1	Ln Model 2	Model 3	Ln Model 4
Breusch-Pagan-Godfrey heteroscedasticity test				
F-stat	3.3244	2.069194	4.6819	0.259401
Prob. F	0.0397**	0.1312	0.0106**	0.7718
Obs*R-sq.	6.4391	4.096748	9.014	0.526772
Prob. Chi-Sq.(2)	0.04**	0.1289	0.011**	0.7684
Scaled explained SS	19.074	4.075332	37.481	0.64675
Prob. Chi-Sq.(2)	0.0001***	0.1303	0.0000***	0.7237
Variance Inflation Factors multicollinearity test				
Centered VIF	1.2869	1.323502	2.5625	3.277629
Centered VIF	1.2869	1.323502	2.5625	3.277629

*Note: *,**,and *** statistical significance at the 10%, 5%, and 1% respectively.*

(Source: Authors' calculations)

From the conducted machine learning, we see that the greatest influence on the reduction of infant mortality was the level of GDP per capita (according to purchasing power parity) ranging between 21.3 and 139.1 USD. The relationship with this indicator, however, isn't very precise due to the large range it was analysed in. Furthermore, the region in which the country is located has a similar effect on the investigated indicator - if that region is Europe, on average, the indicator decreases by as much as 18.26. The percentage of health insurance among the population is another important indicator. In countries with over 90% health insurance coverage, there is an average reduction in mortality per 1000 newborns by 15.46. The number of nurses and midwives per 1000 inhabitants and the literacy of women over 15 years old were also found to be significant to a small extent. On average, they reduce IMR by 2.12 and 0.2 respectively. More significant results were obtained when the Key Influencers tool analysed the top contributors to increases in infant mortality rates. On average, mortality increases by more than 26% when there are minimal health expenditures per capita (below $107.1), minimal percentage of the population covered by health insurance (up to 7%), when the country's region is Africa and when GDP per capita (according to purchasing power parity) is between $0 and $9.2 respectively.

Figure 5a and 5b: Key influencers of infant mortality rates in 2023



(Source: Authors' calculations)

The analysis of key influencers on infant mortality rates, particularly their increase and decrease, highlights significant economic implications. When observing the decrease in infant mortality rates, a higher real GDP per capita implies better economic conditions, which typically correlate with improved access to healthcare, better nutrition, and enhanced living conditions. These factors contribute to the reduction of infant mortality, as they directly impact the overall health and well-being of both mothers and infants. Additionally, the availability and quality of healthcare services, for example, the presence of skilled nurses and midwives, play a crucial role in ensuring safer childbirth and postnatal care, thus reducing IMRs.

On the other hand, the increase in infant mortality rates is often associated with low health expenditure per capita. This scenario indicates underfunded healthcare systems that struggle to provide adequate medical services and facilities (including up-to-date technological capabilities), leading to higher infant mortality. Economic constraints can also result in insufficient healthcare infrastructure, lack of essential medical supplies as well as inadequate or exclusive training for healthcare professionals, further exacerbating the situation. Furthermore, regions with low economic development, such as certain areas in Africa, are more vulnerable to higher infant mortality rates due to the compounded effects of poverty, malnutrition, limited access to healthcare, and low quality of life overall.

In the general context, the interplay between national wealth, healthcare investment, and IMRs emphasizes the critical importance of economic policies that prioritize healthcare funding. Investment in healthcare not only improves health outcomes but also promotes economic stability and growth by fostering a healthier population. Addressing economic disparities on both central and local levels, while improving healthcare infrastructure, are some of the most important strategies for reducing infant mortality rates and improving the overall public health of a nation.

5. CONCLUSION

In this study, we aimed to investigate the impact of various socioeconomic and healthcare factors on infant mortality rates (IMR) across 227 countries, hypothesizing that specific metrics would have a direct impact on IMR. This hypothesis was successfully confirmed, identifying particular key factors. Our findings indicate that several variables, including state expenditures on health, birth rate, deliveries performed by skilled staff, corruption, doctors per 1000 people, health expenditure per capita, life expectancy, maternal mortality rate, and nurses per 1000

people, exhibit heteroscedastic associations with IMR. Despite the presence of heteroscedasticity and non-normal residuals in most models, the overall explanatory power of these variables remains significant. The most significant relationships proved to be the ones between natural logarithms of the variables, which were used to stabilize the model. The regression analyses reveal that the birth rate, life expectancy, maternal mortality rate, and female literacy rate are particularly strong predictors of IMR, explaining 64-72% of its variability. In our multiple linear analysis, about 69% of the variability in IMR is explained by female literacy rates and the corruption index. Specifically, an increase in female literacy by one percentage point results in a 0.63 decrease in IMR, while a one-point increase in the corruption index decreases IMR by 0.36. The model that proved to be the most accurate (R^2=78,57%) and fitting (lowest information criteria values), while meeting all theoretical assumptions for cross-sectional regression analysis, was the one in which the natural logarithm of IMR was the dependent variable and the logarithms of doctors per 1000 people and health expenditure per capita were independent variables. The last model (Ln Model 4) may have further positive implications for the efficiency of future budgets and policies regarding IMR. In other words, budgets focused on public health and policies focused on producing more doctors per 1000 citizens may be the most productive way to improve IMR. However, our study faced several limitations. First, the presence of heteroscedasticity and non-normal residuals in many models could affect the robustness of our findings. Although we transformed variables using natural logarithms to address these issues, the transformed model's reduced explanatory power (54.6%) indicates a trade-off between meeting statistical assumptions and retaining predictive strength. Second, some papers divide their studied regions by income levels, finding that some policies are more effective based on income levels (Bexson *et al.*, 2021). This is one example of how our study could be improved if we were to go in depth regarding the classification of the categorical data types. Still, findings from more recent papers align with the results from our wider-encompassing study (Kammerlander and Schulze 2023; Salariak et al, 2009; Prisco *et al.*, 2015).
There are important implications for the overall success of the economy, which, among other things, is necessary for improving the healthcare system through various mechanisms. The quality of the health care system could improve if there was improvement in the most significant determinants that we mentioned above. Namely, through the aging of the population, without having an increase in the birth rate (or a decrease in the IMR), the old age dependency ratio would continue to increase, impacting the sustainability of pension funds and destabilizing the labour market. Future research should focus on refining these models by exploring alternative transformations or advanced statistical techniques to mitigate heteroscedasticity and multicollinearity. Additionally, expanding the dataset to include more countries and incorporating time-series data estimation in a panel format (Sari and Prasetyani, 2021) could provide a more comprehensive understanding of the dynamic relationships between socio-economic factors and IMR. Investigating the impact of other potential influencers, such as healthcare quality, access to education, and social policies, could further enhance the explanatory power of the proposed models.

REFERENCES

Ahmed, N., Yigit, A., Isik, Z. and Alpkocak, A. (2019), "Identification of leukemia subtypes from microscopic images using convolutional neural network", *Diagnostics*, Vol. 9, 104. https://doi.org/10.3390/diagnostics9030104

Ahmed, T., Ullah, S., Jabeen, T., and Sabir, S. (2011), "Socio-economic determinants of infant mortality in Pakistan", *[Thesis]*.

Baraki, A.G., Akalu, T.Y., Wolde, H.F., Lakew, A.M., and Gonete, K.A. (2020), "Factors affecting infant mortality in the general population: Evidence from the 2016 Ethiopian demographic and health survey (EDHS); a multilevel analysis", *BMC Pregnancy Childbirth,* Vol. 20, p. 1-8. https://doi.org/10.1186/s12884-020-03002-x

Bexson, C., Millett, C., Santos, L.M.P., De Sousa Soares, R., De Oliveira, F.P. and Hone, T. (2021), "Brazil's more doctors programme and infant health outcomes: A longitudinal analysis", *Human Resources Health,* Vol. 19, p. 1-10. https://doi.org/10.1186/s12960-021-00639-3

Cabaneros, S.M., Calautit, J.K. and Hughes, B.R. (2019), "A review of artificial neural network models for ambient air pollution prediction", *Environmental Modelling & Software,* Vol. 119, pp. 285-304. https://doi.org/10.1016/j.envsoft.2019.06.014

Conley, D. and Springer, K. W. (2001), "Welfare state and infant mortality", *American Journal of Sociology,* Vol. 107 No. 3, pp. 768-807.

Dahiwade, D., Patle, G. and Meshram, E. (2019), "Designing disease prediction model using machine learning approach", in: *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC),* pp. 1211–1215. https://doi.org/10.1109/ICCMC.2019.8819782

Erdoğan, E., Ener, M., and Arıca, F., (2013), "The strategic role of infant mortality in the process of economic growth: An application for high income OECD countries", *Procedia - Social and Behavioral Sciences,* Vol. 99, pp. 19-25. https://doi.org/10.1016/j.sbspro.2013.10.467

Fisher, E., Keeble, E., Cheung, R., Hargreaves, D., Wortley, E., and Elias, L. (2024), "Understanding differences in infant mortality rates across local areas", *Research Report, Nuffield Trust.*

Islam, T., Kundu, A., Islam Khan, N., Chandra Bonik, C., Akter, F. and Jihadul Islam, M. (2022), "Machine learning approaches to predict breast cancer: Bangladesh perspective", in Karuppusamy, P., García Márquez, F.P., Nguyen, T.N. (Eds.), *Ubiquitous Intelligent Systems,* Springer Nature, Singapore, pp. 291-305. https://doi.org/10.1007/978-981-19-2541-2_23

Jayachandran, J. and Jarvis, G. (1986), "Socioeconomic development, medical care, and nutrition as determinants of infant mortality in less-developed countries", *Social Biology,* Vol. 33 No. 3-4, pp. 301-315.

Jiang, W. and Luo, J. (2022), "Graph neural network for traffic forecasting: A survey", *Expert Systems with Applications,* Vol. 207, 117921. https://doi.org/10.1016/j.eswa.2022.11792

Jing, F., Ye, Y., Zhou, Y., Ni, Y., Yan, X., Lu, Y., Ong, J., Tucker, J.D., Wu, D., Xiong, Y., Xu, C., He, X., Huang, S., Li, X., Jiang, H., Wang, C., Dai, W., Huang, L., Mei, W., Cheng, W., Zhang, Q., Tang, W. (2023), "Identification of key influencers for secondary distribution of HIV self-testing kits among Chinese men who have sex with men: Development of an ensemble machine learning approach", *Journal of Medical Internet Research,* Vol. 25, e37719. https://doi.org/10.2196/37719

Kammerlander, A. and Schulze, G.G. (2023), "Local economic growth and infant mortality", *Journal of Health Economics,* Vol. 87, 102699.

Kim, D. and Saada, A. (2013), "The social determinants of infant mortality and birth outcomes in western developed nations: A cross-country systematic review", *IJERPH,* Vol. 10, pp. 2296-2335. https://doi.org/10.3390/ijerph10062296

Maruthappu, M., Watson, R.A., Watkins, J., Zeltner, T., Raine, R., and Atun, R. (2017), "Effects of economic downturns on child mortality: a global economic analysis 1981–2010", *BMJ Glob. Health 2,* e000157. https://doi.org/10.1136/bmjgh-2016-000157

Ogburn, W.F. and Thomas, D.S. (1922), "The influence of the business cycle on certain social conditions", *Journal of the American Statistical Association*, Vol. 18, pp. 324-340. https://doi.org/10.1080/01621459.1922.10502475

Okobi, O.E., Ibanga, I.U., Egbujo, U.C., Egbuchua, T.O., Oranu, K.P., and Oranika, U.S., (2023), "Trends and factors associated with mortality rates of leading causes of infant death: A CDC wide-ranging online data for epidemiologic research (CDC WONDER) database analysis", *Cureus*. Vol. 15 No. 9, e45652. https://doi.org/10.7759/cureus.45652

Prisco, G., Pennazio, R., Serafini, A., Russo, C. and Nante, N. (2015), "Infant mortality trend in Europe: socio-economic determinants: Gabriella Prisco", *The European Journal of Public Health*, Vol. 25(suppl_3).

Salarilak, S.H., Khalkhali, H.R., Entezarmahdi, R., Pakdel, F.G., and Faroukheslamloo, H.R., (2009), "Association between the Socio-Economic Indicators and Infant Mortality Rate (IMR) in Iran", *Iranian Journal of Public Health*, Vol. 38, pp. 21-28.

Sari, V.K., Prasetyani, D. (2021), "Socioeconomic determinants of infant mortality rate in Asean: A panel data analysis", *JAS (Journal of ASEAN Studies),* Vol. 9. https://doi.org/10.21512/jas.v9i1.7280

Satti, M.I., Ali, M.W., Irshad, A. and Shah, M.A. (2023), "Studying infant mortality: A demographic analysis based on data mining models", *Open Life Sciences*, Vol. 18, 20220643. https://doi.org/10.1515/biol-2022-0643

Schady, N. and Friedman, J. (2007), "Infant mortality over the business cycle in the developing world", *World Bank Policy Research Working Paper,* (4346).

Stockwell, E. G. and Wicks, J. (1984), "Patterns and variations in the relationship between infant mortality and socioeconomic status", *Social Biology,* Vol. 31 No. 1-2, pp. 28-39.

Waldmann, R. J. (1992), "Income distribution and infant mortality", *The Quarterly Journal of Economics,* Vol. 107 No. 4, pp. 1283-1302.

Younger, S. D. (2001), "Cross-country determinants of declines in infant mortality: A growth regression approach", *Cornell Food and Nutrition Policy Program, Working Paper,* Vol. 130.