

World News of Natural Sciences

An International Scientific Journal

WNOFNS 59 (2025) 306-323

EISSN 2543-5426

Evaluating Child Survival Predictors in Nigeria Using the Cox Proportional Hazard and Gaussian Accelerated Failure Time Models

**Yussuf Toheeb Akorede^{1,a}, Kalu Uchechukwu^{1,b}, Okoro Ndubuisi Obuka²,
Uzoamaka Katherine Emeaso³, Esther Sampson – Achonu⁴**

¹ Department of Mathematics and Statistics, Kwara State University Malete, Nigeria

² Department of Mathematics and Statistics, Federal Polytechnic Ilaro, Ogun State, Nigeria

³ Department of Microbiology, Kwara State University Malete, Nigeria

⁴ Department of Science Education, Nnamdi Azikiwe University Awka, Nigeria

^{a,b} E-mail address: akoredet@gmail.com , kaluuchechukwu45@gmail.com

ABSTRACT

This study assessed estimated child survival rates in Nigeria using data extracted from the 2018 Nigeria Demographic and Health Survey (NDHS). The main objectives is to stimulate and model child survival using semi parametric (Cox proportional hazards) and parametric (Gaussian accelerated failure time) survival models and to compare their clinical trails. letter survival. The study explores the influence of various factors, including geographic region, maternal education, household wealth, sanitation, water access, and gender, on child survival. Descriptive statistics reveal that the average age of first birth is approximately 19 years, with a mean age till death of 107 months. Key findings show that both models takes into account identify s factors such as regional disparities, maternal education, sanitation processes, and socioeconomic status as determinants of child survival. Children r in the Northeast and Northwest are at higher risk of death while children in the East, South-South, and Southwest regions are at lower risk of death. Additionally, lower parental education and economic hardship were linked to elevated child mortality rates, while higher education and wealth contributed to improved survival rates. Inadequate sanitation practices, such as using pit latrines and relying on water from trucks or wells, were associated with an increased risk of infant mortality. Gender differences were also observed, with males facing higher mortality risks than the female. Sanitation practices, including the use of pit toilets and reliance on tanker/cart or well water, increased the risk of infant death. To assess the models' evaluations, the study compared the Cox PH model with the Gaussian AFT model

using three criteria such as the likelihood ratio (-2LL), Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC). Results showed that the Gaussian AFT model (with -2LL of 641387.8, AIC of 641431.8, and BIC of 641646.5) outperformed the Cox PH model in terms of goodness-of-fit and overall model quality.

Keywords: Child survival, Cox Proportional Hazards Model, Gaussian Accelerated Failure Time Model, Wald Test, Likelihood Ratio Test

1. INTRODUCTION

Child survival has been a key area of international development and yet we still have large disparities in child mortality rates both within and across countries. This is an extraordinary health advance, especially because millions of children continue to die each year from preventable illness. This escalating crisis has stimulated existing solutions to be enhanced and new solutions to be developed to avoid further deaths of children and secure a bright future for them (Mirza et al. 2021).

Nigerian child survival is considered one of the critical health issues. Children die here, and the number significantly rises because of various socioeconomic, environmental, and health-related factors despite various healthcare improvements. Child mortality in Nigeria is still counted as one of the highest in the world according to the World Bank (2023) as a big proportion of death happens to children from rural areas, mostly low-income households. This indicates the seriousness of the problem and calls for more interventions and inquiries into the factors concerning child survival.

Antenatal care services have been noted to adversely affect child mortality along with maternal factors such as maternal education, women's employment, antenatal care, socioeconomic status, and the accessibility, availability, and quality of healthcare (Ademola & Alabi, 2023). For example, determinants of child survival include maternal education, antenatal care, and skilled delivery assistance (Ogundele et al., 2022). The availability of clean water, sanitation, and adequate nutrition can significantly impact child health outcomes (Ogunbanjo et al., 2021).

The identification of contributory mortality factors and their influence or influence over time is one of the last critical issues of child survival. Many past studies have pinpointed a broad range of determinants: maternal health, socioeconomic circumstances and environment. One example was the impact of maternal education on child mortality; it caused a reduction in child mortality through improved health behaviours (e.g., nutrition and professional care seeking during pregnancy and childbirth) (Ademola & Alabi, 2023). Additionally, access to clean water, sanitation, and adequate nutrition has been shown to have a substantial impact on child health outcomes (Ogunbanjo et al., 2021).

Recognizing the reasons behind child mortality and how these factors influence one another over time is crucial for enhancing child survival efforts, various determinants including maternal health, socio-economic conditions and environmental factors have been highlighted in earlier studies. For instance, higher maternal education decreases child mortality as women adopt healthier health-seeking behaviors like eating a balanced diet and accessing professional health care during pregnancy and childbirth (Ademola & Alabi, 2023).

Furthermore, the availability of clean water, sanitation and sufficient healthcare facilities are essential in averting common childhood illnesses which are a leading cause of death in Nigeria (Ogunbanjo et al., 2021).

Given the intricate nature of factors affecting child survival, employing advanced statistical models is essential to analyze their impacts. Survival analysis offers a comprehensive framework to understand the timing of child deaths and the influencing factors. The Cox Proportional Hazards (PH) model is a prevalent method in survival analysis, enabling the assessment of multiple covariates' effects on the hazard rate (Cox, 1972). However, this model assumes that the hazards between different groups remain proportional over time, which may not always be the case in real-world data, particularly when the survival times are not homogeneous.

To overcome this limitation, the Gaussian Accelerated Failure Time (AFT) model has become a useful alternative. This model directly focuses on the time until an event occurs and considers the cumulative impact of various factors on survival time, making it appropriate for situations where the proportional hazards assumption may not be valid (Wei, 1992). In examining child survival in Nigeria, the AFT model can provide valuable insights into how socio-economic and environmental factors affect the timing of mortality, potentially revealing different dynamics. The goal of this study is to assess the predictors of child survival in Nigeria by utilizing both the Cox Proportional Hazards and the Gaussian Accelerated Failure Time models. By comparing these approaches, this research seeks to enhance our understanding of the factors that contribute to child mortality in Nigeria, helping to identify key areas for intervention and policy enhancement.

2. MATERIALS AND METHODS

The data for this study was obtained from the 2018 Nigeria Demographic and Health Survey (NDHS). The study effectively employed the Semi-parametric (Cox proportional hazard model), the parametric model (Gaussian Accelerated Failure Time Model), as well as the Wald test and the Likelihood Ratio Test, including AIC and BIC.

2. 1. Cox Proportional Hazards Model

As stated by (Kalu et al., 2024) the Cox regression model is a statistical theory of counting processes that unifies and extends nonparametric censored survival analysis. The approach integrates the benefits of nonparametric and parametric approaches to statistical inferences. The Cox proportional hazards regression model relates covariates to the hazard function as follows:

$$h(t/x) = h_o(t)c(B^i x) \quad (1)$$

where: $h_o(t)$ is called the baseline hazard function, which is the hazard function for an individual for whom all the variables included in the model are zero $B^i = B_1 + B_2 + \dots + B_p$ is a parameter vector of regression coefficients of $X = X_1 + X_2 + \dots +$, is the value of the vector of explanatory variables for a particular individual, and $c \cdot$ is a fixed, known scalar function is the value of the vector of explanatory variables for a particular individual, and $c(.)$ is a fixed, known scalar function.

This is a semi-parametric model where the baseline hazard $h_o(t)$ is estimated non-parametrically, while the covariate effect is constrained by the parametric representation $c(B^i x)$, where $c(\cdot)$ takes an exponential form:

$$c(B^i x) = e^{(B^i x)} = e^{\sum_{j=1}^p B_{ij} x_{ji}} \quad (2)$$

Which assures that the hazard is non-negative and assumes that covariate effects on the hazard are multiplicative. Therefore,

$$h(t/x) = h_o(t)c(B^i x) = h_o(t)e^{(B^i x)} = h_o(t)e^{\sum_{j=1}^p B_{ij} x_{ji}} \quad (3)$$

The Cox model is called a proportional hazards model since the ratio of hazard rates of two individuals with covariate values x_1 and x_2 , at time t is

$$\frac{h(t/x_1)}{h(t/x_2)} = \frac{h_o(t)e^{(B^i x_1)}}{h_o(t)e^{(B^i x_2)}} = \frac{e^{(B^i x_1)}}{e^{(B^i x_2)}} = e^{(B^i (x_1 - x_2))} \quad (4)$$

The hazard ratio is time-independent as the ratio does not depend on t . Since the hazard function at t given covariate x is

$$h(t/x) = h_o(t)e^{(B^i x)} \quad (5)$$

Probability Density Function $f(t)$: The probability density function $f(t)$ is the frequency of events per unit time. The probability density function is related to the survival function. That is,

$$f(t) = \frac{dS(t)}{dt} \quad (6)$$

Cumulative Hazard $H(t)$: The cumulative hazard describes the accumulated risk up to time t . It is given by

$$H(t) = \int_0^t h(u)du \quad (7)$$

Survival Function $S(t)$: The survival (or survivor) function given by,

$$S(t) = P(T > t) = 1 - H(t), 0 < t < \infty \quad (8)$$

Simply indicates the probability that the event of interest has not yet occurred by time t ; thus, if T denotes time until death, $S(t)$ denotes the probability of surviving beyond time t .

Hazard Function $h(t)$: The hazard function, $(h(t))$, is the instantaneous rate at which events occur, given no previous events. That is, the hazard function is a conditional density, given that the event in question has not yet occurred before time. Mathematically given by:

$$h(t) = \lim_{\delta t \rightarrow 0} \frac{P(t \leq T \leq t + \delta t / T \geq t)}{\delta t} = \frac{f(t)}{S(t)} = \frac{d}{dt} \log S(t) \quad (9)$$

2. 2. Gaussian Accelerated Failure Time Model

The Gaussian Accelerated Failure Time (GAFT) model is a parametric survival analysis model that is a variation of the more commonly used Cox Proportional Hazards model. While the Cox model doesn't make specific assumptions about the shape of the baseline hazard function, the GAFT model assumes that the log of the survival time follows a Gaussian distribution (Molstad, A. J. and Hsu, L. and Sun, W., 2021). Mathematically, the GAFT model is expressed as follows:

$$\log(T) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon \quad (10)$$

where:

- $\log(T)$ represents the natural logarithm of the survival time, which is assumed to be normally distributed.
- $X_1, X_2, \dots, X_p$ are the covariates that can influence the survival time.
- $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ are the model parameters (coefficients) that determine the relationship between the covariates and the log of the survival time.
- ε is the error term, which follows a standard Gaussian distribution.

In this model, the relationship between the covariates and the log of the survival time is linear. The parameters $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ determine the strength and direction of the association between the covariates and the log of the survival time.

2. 3. Wald Test

In a regression model with a single explanatory variable, x , we consider the hypothesis $H_0: \beta = 0$ (x has no effect on the response variable). The Wald Test Statistic is then given by:

$$Z = \frac{\hat{\beta}}{se(\hat{\beta})} \quad (11)$$

where: $se(\beta)$ is the unrestricted standard error of β , it has an approximate standard normal distribution when H_0 is true with equivalent Z^2 which has a Chi-squared distribution with $df = 1$.

2. 3. 1. Confidence Interval for Estimate Effects

A Wald confidence interval for the effect of estimate in a proportional hazard Cox regression model is;

$$\hat{\beta} \pm Z_{\alpha/2} se(\hat{\beta}) \quad (12)$$

and can be transformed into a hazard scale by exponentiating the confidence interval, that is;

$$\exp [\hat{\beta} \pm Z_{\alpha/2} se(\hat{\beta})] \quad (13)$$

where: β is the effect estimate, $se(\beta)$ is the standard error estimate and α is the significance level.

2. 4. Likelihood Ratio Test

This denotes the maximized value of the likelihood function to be l_0 under $H_0 : \beta = 0$ and l_1 when β needs not equal 0. The model with β has only an intercept term and is called the null model. The likelihood ratio test statistic equals:

$$\log\left(\frac{l_1}{l_0}\right) = 2[\log(l_1) - \log(l_0)] = 2[(L_1 - L_0)] \quad (14)$$

where: $L_0 = \log l_0$ and $L_1 = \log l_1$ denote the maximized log-likelihood functions. Under $H_0 : \theta = 0$, with test statistic which is also an approximate chi-squared distribution with specified degree of freedom. This test measures the improvement of fit that the explanatory variables make compared to the null model.

2. 5. Models Adequacy and Model Evaluation

Alkaike Information Criterion (AIC) and Bayesian Information Criterion are used to compare models across different samples or to compare non-nested models that cannot be compared with LR-test. Hence, the model with the smaller AIC or BIC is considered a better-fitting model.

$$AIC = -2(\log\text{-likelihood function}) + 2p \quad (15)$$

$$BIC = -2(\log\text{-likelihood function}) + \ln(n) \times p \quad (16)$$

where: p is the number of parameters estimated (including the constant).

3. RESULTS AND DISCUSSION

3. 1. Data Presentation

Table 1. Overview of the datasets.

Agefbirth	Edu	Wealth	toilet	Resid1	Region1	water	Diagram	cages	sex	Censored diagram	alive
11	0	3	2	2	3	6	24	NA	2	24	1
11	0	3	2	2	3	6	180	15	2	60	0
11	0	3	2	1	3	2	84	7	2	60	0
11	0	3	2	1	3	2	120	10	2	60	0
11	0	3	2	1	3	2	132	11	2	60	0
...	...	...	...	...	...	...	...	...	...	...	...

...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...
43	3	3	2	2	4	2	36	3	2	36	1
43	3	3	2	2	4	2	60	5	2	60	1
43	2	3	3	1	5	2	12	1	2	12	1
45	3	3	3	1	5	6	0	0	2	0	1
45	3	3	3	1	5	6	48	4	2	48	1
48	3	3	3	2	4	2	12	1	2	12	1

Table 1 presents the overview of the datasets used in this study, the table depicts the first and last 6 observations of the total observations. Next, we have the description of the data.

Table 2. Variables in the datasets and their labels.

Variables	Description
Agefbirth	Age of respondents at first birth
Edu	Highest Educational Level: 0 = No Education, 1 = Primary, 2 = Secondary, 3 = Higher
Wealth	Wealth index combined: 1 = Poor, 2 = Middle, 3 = Rich
Toilet	Type of Toilet Facility: 1 = No facility/bush/field, 2 = pit-toilet/latrine, 3 = flush-toilet, 4 = Others
Resid1	Type of place of residence: 1 = Urban, 2 = Rural
Region1	Region: 1 = North Central, 2 = North East, 3 = North West, 4 = South East, 5 = South South, 6 = South West
Water	Source of drinking water: 1 = piped water, 2 = well or borehole, 3 = well, 4 = spring/river/dam/lake/ponds/stream/canal/irrigation channel, 5 = Rainwater, 6 = tanker or cart n others
Dagem	Age of child till death (months)
Cagec	The current age of the child
Csex	Sex of child: 1 = Male, 2 = Female
Censored diagram	Censored age of child till death in months
Alive	Child is alive: 0 = Yes, 1 = No

3. 2. Descriptive statistics for the data

Table 3. Summary of continuous variables.

Summary	Variables		
	Age at First Birth	The current age of the child	Age till death (months)
Min	11.00	0.00	0.00
Mean	18.79	10.08	107.4
Max	48.00	37.00	444.0

Table 3 displays the summary of the continuous variable in the dataset. It was seen that the mean age at first birth is approximately 19 years while the mean current age of the child is approximately 10 years old and the mean age till death (in months) is approximately 107 months.

3. 3. Visualizations for Categorical Variables in the Dataset

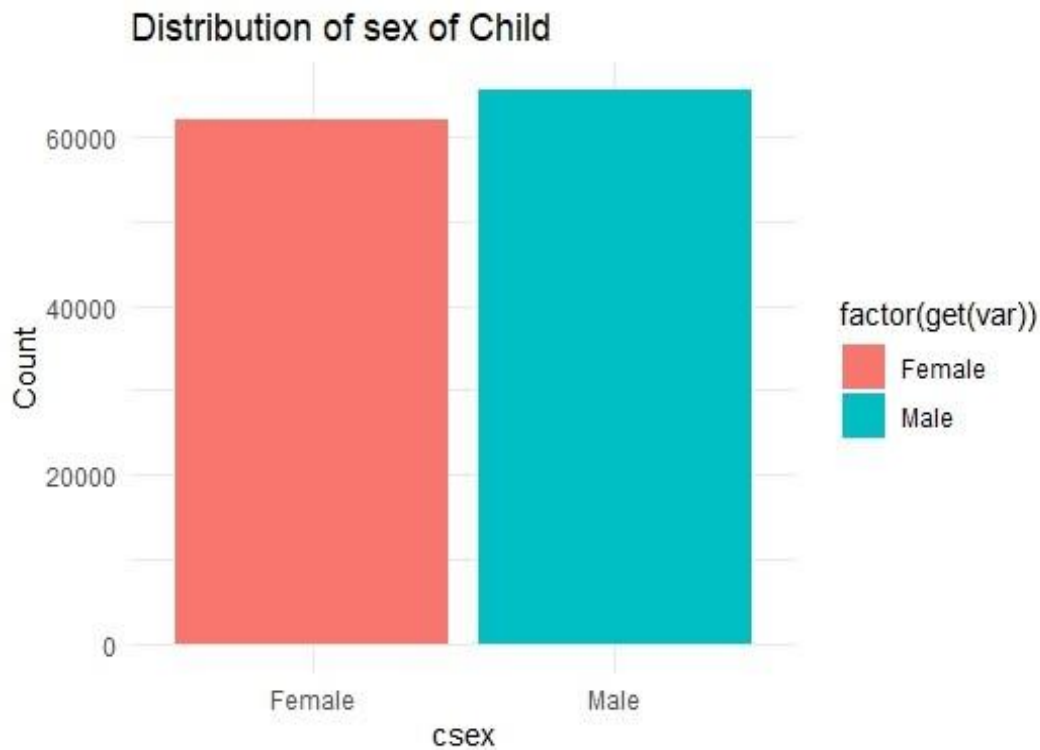


Figure 1. Distribution of respondents by Sex of Child.

Figure 1 shows the distribution of frequency by gender, A large percentage of the respondents are male.

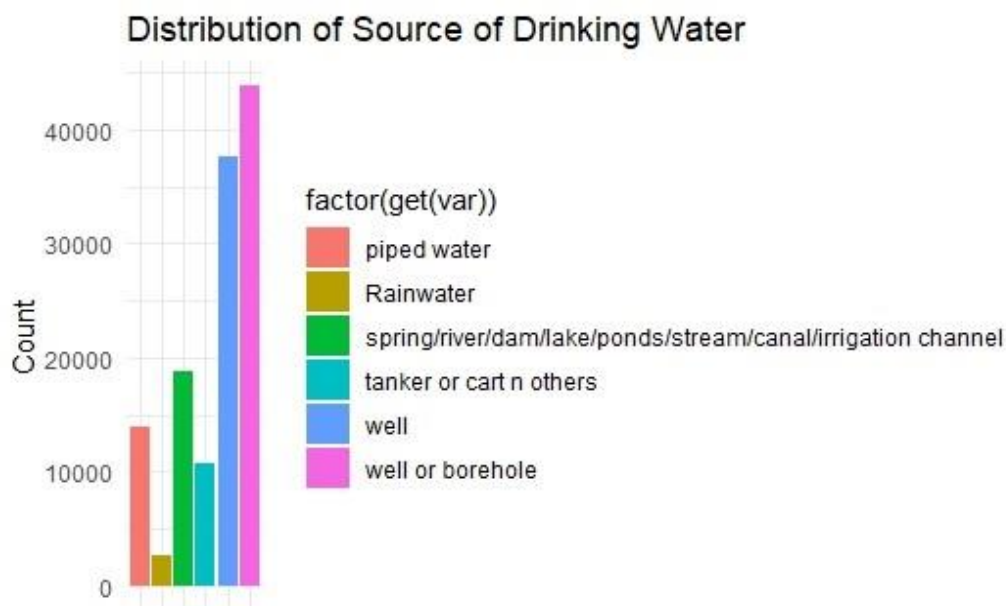


Figure 2. Source of Drinking Water.

Figure 2 shows the distribution of the source of drinking water of respondents. A large percentage of the respondents source their drinking water from a well or borehole while the lowest percentage drink from the rainwater.

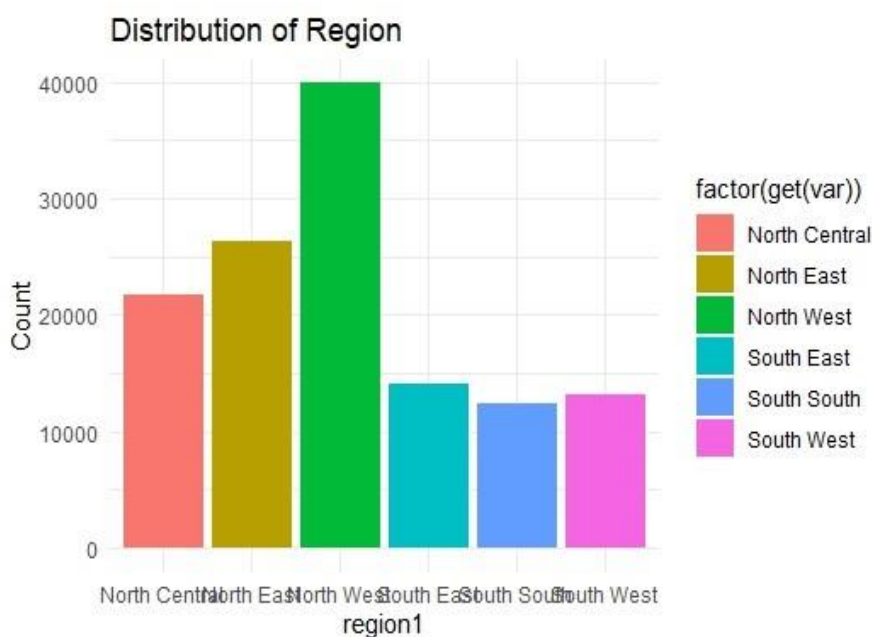


Figure 3. Distribution of Respondents by Religion.

Figure 3. displays the distribution of respondents by region, Most of the respondents were from the North West region, followed by the North East. The South-South region has the least number of respondents.

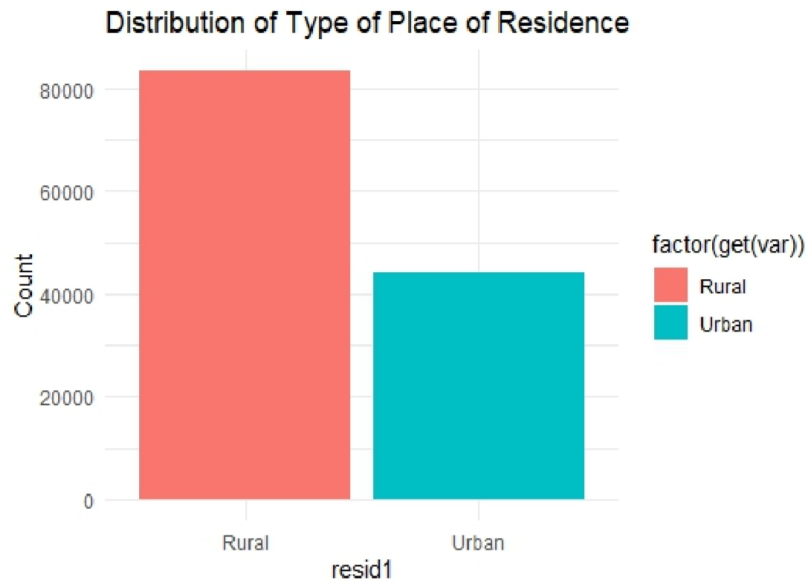


Figure 4. Distribution of respondents by Place of Residence.

Figure 4 depicts the distribution of the type of place of residence. A large percentage of the respondents were from the rural area.

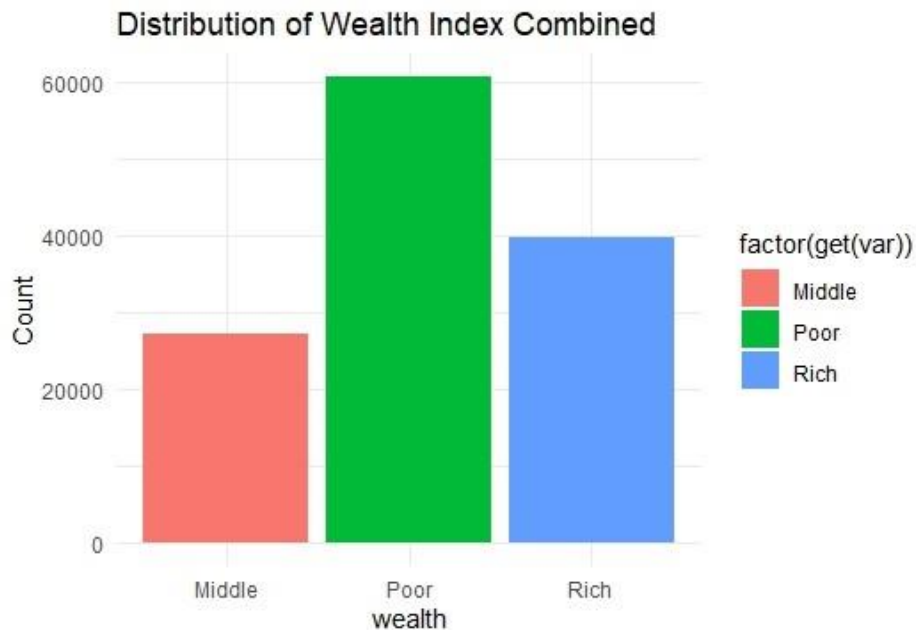


Figure 5. Distribution of Respondents by Wealth Index.

Figure 5 represents the distribution of wealth index of households. Most of the households fall in the poor category followed by the rich and the least category is the middle class.

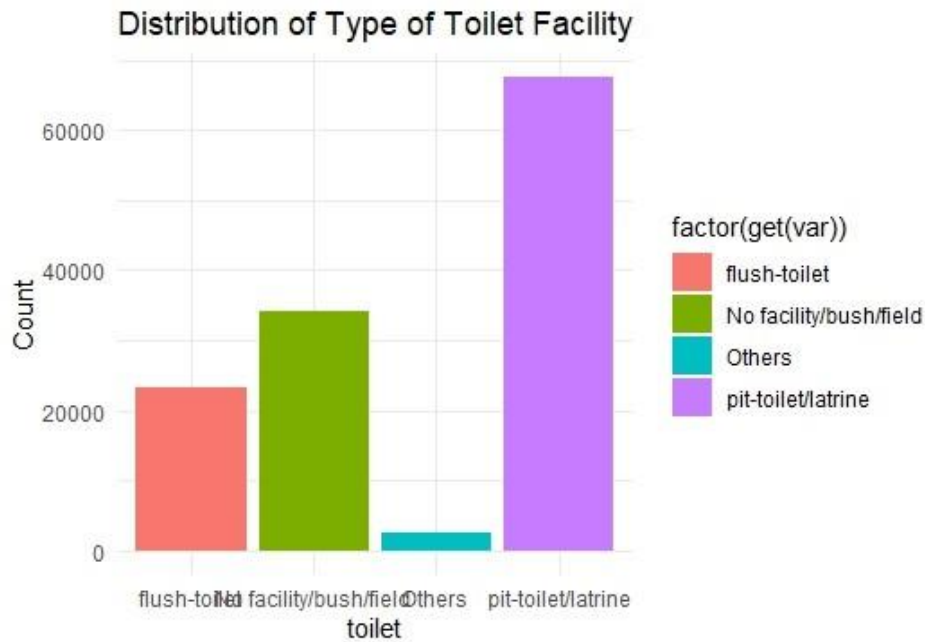


Figure 6. Distribution of Respondents by Usage of Toilet Facility.

Figure 6 is the representation of the distribution of the type of toilet facility used by respondents. Most of the respondents used pit toilets/latrines.

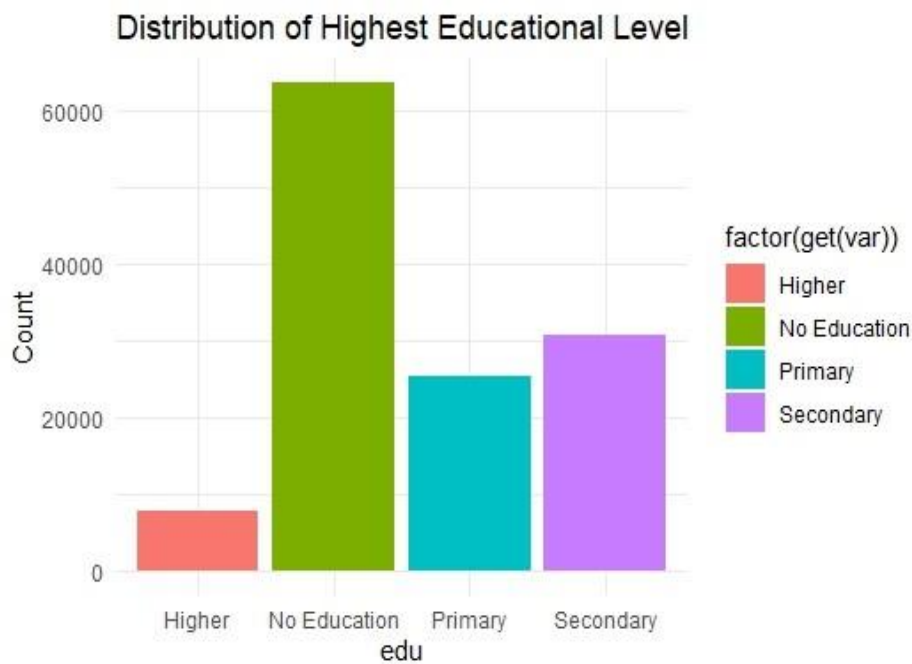


Figure 7. Distribution of respondents by Educational Level.

Figure 7 shows the bar representation of the Highest educational level of respondents. A few of the respondents had higher education qualifications while many of the respondents did not have any formal education.

3. 4. Semi Parametric Model

Cox Proportional Hazard Model.

Table 4. Result from Cox Proportional Hazard Model.

Variables	Coef	Hazard h(t)	se(coef)	Z	p-value	95% Conf. Interval	Significance Status
North-East	0.1068	1.1128	0.0158	6.7480	1.50×10^{-11}	[1.0788, 1.1478]	Significant
North-West	0.2461	1.2790	0.0154	15.9730	$< 2 \times 10^{-16}$	[1.2409, 1.3182]	Significant
South-East	-0.1148	0.8916	0.0184	-6.2450	4.23×10^{-10}	[0.8600, 0.9243]	Significant
South-South	-0.2002	0.8186	0.0192	-10.4060	$< 2 \times 10^{-16}$	[0.7883, 0.8501]	Significant
South-West	-0.1137	0.8925	0.0186	-6.1220	9.23×10^{-10}	[0.8606, 0.9256]	Significant
North-Central	Reference						
No Education	-0.4436	0.6417	0.0213	-20.7990	$< 2 \times 10^{-16}$	[0.6154, 0.6691]	Significant
Primary Education	-0.4610	0.6307	0.0216	-21.3100	$< 2 \times 10^{-16}$	[0.6045, 0.6580]	Significant
Secondary	0.0194	1.0195	0.0196	0.9850	0.3245	[0.9810, 1.0596]	Not Significant
Higher	Reference						
Poor	0.0923	1.0966	0.0126	7.3280	2.34×10^{-13}	[1.0699, 1.1240]	Significant
Rich	-0.1350	0.8737	0.0145	-9.2950	$< 2 \times 10^{-16}$	[0.8492, 0.8989]	Significant
Middle	Reference						
No Facility/Bush Field	0.0036	1.0036	0.0170	0.2120	0.8321	[0.9708, 1.0375]	Not Significant
Pit/Latrine	0.0387	1.0394	0.0151	2.5670	9.43×10^{-8}	[1.0092, 1.0706]	Significant
Others	0.1753	1.1916	0.0328	5.3370	0.0102	[1.1173, 1.2708]	Significant
Flush toilet	Reference						
Male	0.0213	1.0215	0.0087	2.4480	0.0144	[1.0042, 1.0391]	Significant
Female	Reference						
Rain Water	-0.0187	0.9815	0.0355	-0.5250	0.5994	[0.9155, 1.0523]	Not Significant
Spring/River/Lake	0.0224	1.0226	0.0190	1.1770	0.2393	[0.9852, 1.0615]	Not Significant
Tanker / Cart	0.0949	1.0996	0.0206	4.6080	4.07×10^{-6}	[1.0561, 1.1449]	Significant
Well	0.0397	1.0405	0.0162	2.4510	0.0142	[1.0080, 1.0740]	Significant

Borehole	0.0301	1.0306	0.0156	1.9280	0.0538	[0.9995, 1.0627]	Not Significant
Piped water	Reference						
Urban	-0.1030	0.9021	0.0113	-9.1460	$< 2 \times 10^{-16}$	[0.8824, 0.9223]	Significant
Rural	Reference						

Concordance = 0.563 (se = 0.001), Likelihood ratio test = 2559 on 20 df, $p = < 2e^{-16}$

Table 4 reveals the findings from the Cox Proportional Hazard Model, revealing a concordance value of 0.563 and a significant likelihood ratio ($p < 0.05$), indicating a robust fit for the model. The analysis highlights key factors influencing infant mortality risk. Regionally, children from the North East (hazard ratio of 1.1128) and North West (hazard ratio of 1.2790) face a higher risk of infant death compared to those from North Central, with a significant increase of 11.28% and 27.90% respectively. In contrast, children from South East (hazard ratio of 0.8916), South-South (hazard ratio of 0.8186), and South West (hazard ratio of 0.8925) experience a protective effect, with lower risks of infant mortality compared to those from North Central.

Social status factors also plays a crucial role in child survival. Children of mothers with no education (hazard ratio of 0.6417) or primary education (hazard ratio of 0.6307) have a significantly lower risk of child death, suggesting a protective effect for these groups compared to those whose mothers have tertiary education. Wealth is another critical factor: being poor (hazard ratio of 1.0966) increases the risk of infant mortality by 9.66%, while being wealthy (hazard ratio of 0.8737) reduces the risk by 12.68% compared to the middle class.

Living conditions, particularly sanitation and water sources, also influence survival outcomes. Children using pit/latrine (hazard ratio of 1.0394) or other types of toilets (hazard ratio of 1.1916) face a higher risk of child mortality compared to those using flush toilets. Similarly, using tanker/cart (hazard ratio of 1.0996) or well water (hazard ratio of 1.0405) increases the risk of child mortality compared to piped water sources. Gender differences also impact child survival: male children (hazard ratio of 1.0215) have a slightly higher risk of child mortality compared to females. Finally, children residing in urban areas (hazard ratio of 0.9021) experience a protective effect against child mortality, with a 9.79% lower risk compared to their rural counterparts. These findings highlight the complex interplay of regional, educational, economic, environmental, and gender-related factors in influencing infant survival.

3. 5. Parametric Model

Gaussian Accelerated Failure Time Model.

Table 5. Result from Gaussian Accelerated Failure Time Model.

Levels	Value	Standard Error	Z	p-value	Conf. Interval	Significance Status
Intercept	60.8724	0.91185	66.76	$< 2 \times 10^{-16}$	[59.0852, 62.6596]	Significant
North-East	-3.7446	0.50542	-7.41	1.3×10^{-13}	[-4.7352, -2.7540]	Significant

North-West	-8.0799	0.49417	-16.35	$< 2 \times 10^{-16}$	[-9.0485, -7.1113]	Significant
South-East	4.030	0.57737	6.98	2.9×10^{-12}	[2.8985, 5.1618]	Significant
South-South	6.2638	0.59665	10.50	$< 2 \times 10^{-16}$	[5.0944, 7.4332]	Significant
South-West	3.7506	0.58003	6.47	1.0×10^{-10}	[2.6138, 4.8874]	Significant
North-Central	Reference					
No Education	14.1105	0.69528	20.29	$< 2 \times 10^{-16}$	[12.7478, 15.4732]	Significant
Primary Education	14.2273	0.69176	20.57	$< 2 \times 10^{-16}$	[12.8715, 15.5832]	Significant
Secondary	-0.5370	0.63871	-0.84	0.40047	[-1.7889, 0.7148]	Not Significant
Higher	Reference					
Poor	-2.9469	0.40051	-7.36	1.9×10^{-13}	[-3.7319, -2.1619]	Significant
Rich	4.3877	0.45801	9.58	$< 2 \times 10^{-16}$	[3.4900, 5.2854]	Significant
Middle	Reference					
No Facility/Bush Field	0.1445	0.53330	0.27	0.78640	[-0.9007, 1.1898]	Not Significant
Pit/Latrine	-5.4096	0.47532	-2.54	3.5×10^{-7}	[-2.1379, -0.2747]	Significant
Others	-1.2063	1.06204	-5.09	0.01115	[-7.4912, -3.3280]	Significant
Flush toilet	Reference					
Male	-0.93531	0.27754	-3.37	0.00075	[-1.4793, -0.3913]	Significant
Female	Reference					
Rain Water	0.4504	1.09611	0.41	0.68113	[-1.6979, 2.5988]	Not Significant
Spring/River/Lake	-0.9449	0.60228	-1.57	0.11665	[-2.1254, 0.2355]	Not Significant
Tanker / Cart	-2.9880	0.65885	-4.54	5.8×10^{-6}	[-4.2793, -1.6967]	Significant
Well	-1.3550	0.51632	-2.62	0.00868	[-2.3671, -0.3431]	Significant
Borehole	-1.0247	0.49596	-2.07	0.0388	[-1.9968, -0.0527]	Significant
Piped water	Reference					
Urban	3.21422	0.35624	9.02	$< 2 \times 10^{-16}$	[2.5160, 3.9124]	Significant
Rural	Reference					
Log(scale)	3.76586	0.00353	1067.55	$< 2 \times 10^{-16}$		Significant

Loglik (model) = -320693.9, Loglik (intercept only) = -321965, Chisq = 2542.07 on 20 df

Table 5 highlights the comparison between the Gaussian Accelerated Failure Time (AFT) model and an intercept-only model, demonstrating that the model with covariates is a significantly better fit (log-likelihood: -320693.9 vs. -321965, Chi-squared = 2542.07,

$p < 2e^{-16}$). The log-scale parameter confirms the validity of the log-normal distribution assumption for survival times.

The analysis reveals several key findings:

- ✓ Geographical regions: Children from the North East (-3.74622) and North West (-8.07990) experience significantly shorter survival times compared to the North Central region. In contrast, children in South East (4.03013), South-South (6.26383), and South West (3.75060) regions have significantly longer survival times.
- ✓ Education: Children of parents with no education (14.11049) or primary education (14.22732) exhibit longer survival times compared to those with higher education, though secondary education does not significantly affect survival ($p = 0.40047$).
- ✓ Wealth status: Poor families (-2.94692) show shorter survival times compared to middle-class families, while children from richer families (4.38744) experience longer survival times.
- ✓ Sanitation: Children with access to pit/latrine toilets (-5.40960) or other types of toilets (-1.20631) have significantly shorter survival times compared to those using flush toilets.
- ✓ Gender: Male children (-0.93531) have shorter survival times than females, a statistically significant difference.
- ✓ Water source: Children using tanker/cart (-2.98800), well (-1.35508), or borehole (-1.02474) water sources show shorter survival times compared to those with access to piped water.
- ✓ Residential area: Children living in urban areas (3.21422) have significantly longer survival times compared to those living in rural areas.

3. 6. Model Evaluation

We then compare the two models (proportional hazard Cox model and Gaussian Aft model) using three criteria (Likelihood Ratio (-2LL), Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC)). Based on these statistics, the GAUSSIAN AFT model outperforms the Cox PH model in terms of goodness of fit and model quality. It has a lower -2LL, lower AIC, and lower BIC, indicating that it provides a more appropriate and parsimonious representation of the data, despite having more degrees of freedom. The summary of the statistics is shown in the table below.

Table 6. Model Summary.

Model	Observations	-2LL	Df	AIC	BIC
Cox PH	127545	1217380.6	20	1217421	1217598
GAUSSIAN AFT	127545	641387.8	22	641431.8	641646.5

Table 6 shows that to assess the effectiveness of different modeling approaches, the Cox PH model was compared with the Gaussian AFT model using three criteria: Likelihood Ratio (-2LL), Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC).

The Gaussian AFT model outperformed the Cox PH model in terms of goodness of fit and model quality, despite having more degrees of freedom.

4. CONCLUSIONS

The Cox Proportional Hazard model was fitted to examine the factors affecting child survival. The results revealed that the geographic region plays a significant role in child survival. Children in the North-Eastern region exhibited an 11.28% higher risk of experiencing infant death compared to those in the North-Central region. Children from the North-West had a 27.90% higher risk of infant death. South-Eastern and South-Southern regions exhibited protective effects, with an 11.84% and 18.14% lower risk of infant death, respectively. The South-West region also had a 10.75% protective effect. The Gaussian Accelerated Failure Time (AFT) model was utilized to further explore the factors affecting child survival. The results indicated a highly significant difference between models with and without covariates, supporting the use of the model. Geographic regions have a significant impact on child survival, with varying effects. Maternal education and socioeconomic status are important factors, where having "No Education" or "Primary Education" offers some protective benefits compared to higher education levels. The availability of sanitation facilities and drinking water sources also plays a crucial role in child survival, with certain facilities linked to increased risks. Gender and the type of residential area are key considerations, as males tend to face higher risks, while urban settings generally provide a protective advantage. To evaluate the effectiveness of different modeling techniques, the Cox Proportional Hazards (PH) model was compared to the Gaussian Accelerated Failure Time (AFT) model using three criteria: Likelihood Ratio (-2 LL), Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC). The Gaussian AFT model demonstrated better performance with a Likelihood Ratio of 641387.8, an AIC of 641431.8, and a BIC of 641646.5, surpassing the Cox PH model in terms of fit and quality, even with more degrees of freedom. This study sheds light on the various factors influencing child survival in Nigeria, emphasizing the importance of geographic region, maternal education, socioeconomic status, sanitation, water sources, gender, and residential area. Understanding these elements is vital for creating targeted interventions and policies to enhance child survival rates in Nigeria. The findings underscore the necessity for customized strategies to tackle regional disparities, improve education, and increase access to clean sanitation and safe drinking water, especially in rural communities. By concentrating on these factors and utilizing the Gaussian AFT model, we can lower infant mortality risk and secure a healthier future for children in Nigeria. The Gaussian AFT model, with its superior performance, offers a more fitting and streamlined representation of the data, guiding future research and policy initiatives.

References

- [1] Adeolu, M., Akpa, O., Adeolu, A., & Aladeniyi, I. (2016). Environmental and socioeconomic determinants of child mortality: Evidence from the 2013 Nigerian demographic health survey. *American Journal of Public Health Research*, 4(4), 134-141. <https://doi.org/10.12691/ajphr-4-4-3>

- [2] Adeyinka, D., Muhajarine, N., & Petrucka, P. (2020). Inequities in child survival in Nigerian communities during the sustainable development goal era: Insights from analysis of 2016/2017 multiple indicator cluster survey. *BMC Public Health*, 20, 1613. <https://doi.org/10.1186/s12889-020-09672-8>
- [3] Akinyemi, J., Bamgboye, E., & Ayeni, O. (2015). Trends in neonatal mortality in Nigeria and effects of bio-demographic and maternal characteristics. *BMC Pediatrics*, 15, 36. <https://doi.org/10.1186/s12887-015-0349-0>
- [4] Alkema, L., Chao, F., You, D., Pedersen, J., & Sawyer, C. C. (2014). National, regional, and global sex ratios of infant, child, and under-5 mortality and identification of countries with outlying ratios: A systematic assessment. *The Lancet Global Health*, 2(9), e521-e530. [https://doi.org/10.1016/S2214-109X\(14\)70280-3](https://doi.org/10.1016/S2214-109X(14)70280-3)
- [5] Bado, A., Susuman, A., & Nebie, E. (2016). Trends and risk factors for childhood diarrhea in sub-Saharan countries (1990-2013): Assessing the neighborhood inequalities. *Global Health Action*, 9, 30166. <https://doi.org/10.3402/gha.v9.30166>
- [6] Chao, F., You, D., Pedersen, J., Hug, L., & Alkema, L. (2018). National and regional under-5 mortality rate by economic status for low-income and middle-income countries: A systematic assessment. *The Lancet Global Health*, 6(5), e535-e547. [https://doi.org/10.1016/S2214-6381\(18\)30059-7](https://doi.org/10.1016/S2214-6381(18)30059-7)
- [7] Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(2), 187-220
- [8] Ezeh, O., Agho, K., Dibley, M., Hall, J., & Page, A. (2014). Determinants of neonatal mortality in Nigeria: Evidence from the 2008 demographic and health survey. *BMC Public Health*, 14, 521. <https://doi.org/10.1186/1471-2458-14-521>
- [9] Fagbamigbe, A. F., Kandala, N. B., & Uthman, O. A. (2020). Decomposing the educational inequalities in the factors associated with severe acute malnutrition among under-five children in low- and middle-income countries. *BMC Public Health*, 20(555), 1-14. <https://doi.org/10.1186/s12889-020-08635-3>
- [10] Horiguchi, M., Hassett, M., & Uno, H. (2019). How do the accrual pattern and follow-up duration affect the hazard ratio estimate when the proportional hazards assumption is violated? *The Oncologist*, 24(7), 867-871. <https://doi.org/10.1634/theoncologist.2018-0141>
- [11] Kuitunen, I., Ponkilainen, V., Uimonen, M., Eskelinen, A., & Reito, A. (2021). Testing the proportional hazards assumption in Cox regression and dealing with possible non-proportionality in total joint arthroplasty research: Methodological perspectives and review. *BMC Musculoskeletal Disorders*, 22(1), 489. <https://doi.org/10.1186/s12891-021-04379-2>
- [12] Mirza Balai, Hunter Wade York, Kam Sripada, Elodie Besnier, Hanne Dahl Vonen, & Aleksandr Aravkin. (2021). Parental education and inequalities in child mortality: A global systematic review and meta-analysis. *The Lancet*, 398(10300), P608-620

- [13] Molstad, A. J., Hsu, L., & Sun, W. (2021). Gaussian process regression for survival time prediction with genome-wide gene expression. *Biostatistics*, 22(1), 164-180. <https://doi.org/10.1093/biostatistics/kxz023>
- [14] Morakinyo, O. M., & Fagbamigbe, A. F. (2017). Neonatal, infant, and under-five mortalities in Nigeria: An examination of trends and drivers (2003-2013). *PLoS ONE*, 12(8), e0182990. <https://doi.org/10.1371/journal.pone.0182990>
- [15] Ntukidem, S., Chukwu, A., Oyamakin, S., James, C., & Habimana-Kabano, I. (2023). Trend analysis and determinants of under-5 mortality in Nigeria: A machine learning approach. *Asian Journal of Probability and Statistics*, 24, 10-23. <https://doi.org/10.9734/AJPAS/2023/v24i2520>
- [16] Olaitan, T., Okafor, I., Onajole, A., & Abosede, O. (2017). Ending preventable maternal and child deaths in western Nigeria: Do women utilize the lifelines? *PLoS ONE*, 12(5), e0176195. <https://doi.org/10.1371/journal.pone.0176195>
- [17] Royston, P., & Parmar, M. (2011). The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when the proportional hazards assumption is in doubt. *Statistics in Medicine*, 30(19), 2409-2421. <https://doi.org/10.1002/sim.4274>
- [18] Stensrud, M. J., & Hernán, M. A. (2020). Why test for proportional hazards? *JAMA*, 323(14), 1401-1402. <https://doi.org/10.1001/jama.2020.1267>
- [19] Yaya S, Bishwajit G, Okonofua F, Uthman OA (2018). Under five mortality patterns and associated maternal risk factors in sub-Saharan Africa: A multi-country analysis. *PLoS ONE* 13(10): e0205977. <https://doi.org/10.1371/journal.pone.0205977>
- [20] Zhang, Z., Rehkopf, D., & Kindig, D. (2017). Urban-rural differences in childhood mortality in the United States: Trends and determinants, 1980-2014. *American Journal of Public Health*, 107(6), 883-890. <https://doi.org/10.2105/AJPH.2017.303721>