Pakistan Journal of Life and Social Sciences
www.pjlss.edu.pk

https://doi.org/10.57239/PJLSS-2024-22.2.001388

RESEARCH ARTICLE

Determine the Diseases that Most Cause Mortality In Children Under The Age Of Five Years In Gezira State (Sudan) By Using Multiple Regression

Abdalla Ahmed Alkhalifa Abdalla*

Department of Statistics, Faculty of Science, University of Tabuk, Tabuk city, Saudi Arabia

ARTICLE INFO

Received: Oct 17, 2024

Accepted: Dec 9, 2024

Keywords

Child mortality
Diarrheal
Diseases
Malaria
Pneumonia
Meningitis
Measles
Multiple regression

***Corresponding Author**

dr.a.alkhalifa@gmail.com

ABSTRACT

The study was conducted in Gezira State in Sudan for the years 2003 - 2022 to determine the diseases that most cause mortality in children under the age of five (U5). The number of mortality of children (U5) was taken as a dependent variable and the number of infections with diarrheal diseases, malaria, pneumonia, meningitis, and measles were taken as independent variables. These diseases taken because are the most children diseases registered in the region. The multiple regression method was used, and the study showed that the most common diseases causing mortality among children (U5) in the region are meningitis and measles, although they recorded the lowest number of infections during the study period, which means that there is a weakness in controlling and treating these diseases, in addition to the fact that cases reach Hospitals are late due to poor health awareness and people use herbal treatment and local medicines for a period of time before coming to the hospital. The study recommended intensifying health awareness and training health personnel in dealing with these diseases with high efficiency.

INTRODUCTION

Children (U5) mean every child born alive from the moment of his birth until he reaches the age of five. "In order to address the specific health challenges and needs of young children more effectively, the first five years of life are further subdivided into the neonatal period (the first 28 days of life), infancy (the first year of life) and pre-school years (1-5 years)" (1). The Sustainable Development Goals (SDGs) adopted by the United Nations in 2015 were developed to promote healthy lives and well-being for all children. The SDG Goal 3.2.1 is to end preventable mortality of newborns and under-5 children by 2030. There are two targets: reduce newborn mortality to at least as low as 12 per 1000 live births in every country; and reduce children U5 mortality to at least as low as 25 per 1000 live births in every country. (2) . According to World Health Organization statistics, the mortality rate for children in Sudan in 2021is 62 per 1,000 live births (3), which is a high rate. Hence the objective of this study to find out the most important diseases that cause child mortality in Gezira State in Sudan. Children are one of the most important determinants of the progress and development of human societies. Accordingly, child mortality and its high rates have a negative impact on the progress of these societies. Therefore, it is necessary to study the causes that lead to child mortality, and from here comes the importance of the study.

MATERIAL AND METHODS:

Model: The multiple linear regression method was used to determine which of these diseases is the most common cause. In mortality.

Multiple linear regression model:

The multiple linear regression model is used to describe the relationship between the dependent variable y consisting of (n) Of the observations and independent variables (x_1, x_2, ..,x_n), this relationship can be expressed with the following equation **(4)**

$Y_i = {}^{\beta}{}_0 + {}^{\beta}{}_i x_{i1} + + {}^{\beta}{}_i x_{i2} + \ldots\ldots + {}^{\beta}{}_k x_{ik} + \varepsilon_i$

$Y_i = {}^{\beta}{}_0 + \sum {}^{\beta}{}_j x_{ij} + \varepsilon_{ij-1}$ ..... (1)

.i = 1,2,....,n

Where (β, β0, β1, ,...., βk) expresses the regression coefficients, εi expresses the random error of observation

Number i, i=1, 2, … , n. Since the number of observations is n, we have n equations that can be formulated in matrix form as follows:

$$\begin{pmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & X_{11}X_{12} \ldots\ldots\ldots X_{1k} \\ 1 & X_{21}X_{22} \ldots\ldots\ldots X_{2k} \\ . & . \quad . \quad . \\ . & . \quad . \quad . \\ . & . \quad . \quad . \\ 1 & X_{1n}X_{2n} \ldots\ldots\ldots X_{nk} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ . \\ . \\ \beta_n \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ . \\ \varepsilon_n \end{pmatrix}$$

In compact form is Y=XB+U

whereas:

Y: expresses the vector of dependent observations, and it is of degree (1×n), and the number i element in this vector is yi

X: The independent view matrix (clarification) is represented, which is from the staircase (nx(k+1))

The number i in this matrix is: (1 Xi1 Xi2 Xik)

B: It expresses the vector of regression coefficients, and it is of degree (1× (k+1))

U: expresses the vector of random errors, and it is of degree (1×n), and the number i is the random error εi

The multiple regression model in equation (2) based on several assumptions:

1- The matrix of independent variables is determined and fixed.

2- There is no complete or almost complete linear relationship between the independent variables, meaning that (Rank(X) = (k+1) < n

3- There is statistical independence between the independent observations (Xi1,Xi2, … , Xik) and the random error εi. That is, the columns of the matrix X is linearly independent of the random error vector U. Mathematically, it is as follows:

Cov (X, U) = E (X'U)- [E(X)]' [E (U)] =0

4- The random error εi, I = 1, 2, … , n has a normal distribution with a mean of zero, and a variance of σ^2 is fixed from observation to another i.e. εi ~ N(0, σ^2), , and it is also assumed that the errors are Statistically independent, and this is expressed mathematically as follows:

E(εi) =0

Cov(εi, εj)= E (εi εj)={ σ^2 if i = j and 0 if i≠j)

That is, the error vector U follows a multiple normal distribution with a mean of zero and has a covariance matrix Σ, that is U~ N (0,Σ)

And the matrix Σ is a symmetric matrix of degree (n × n), it is expressed as follows:

$$\Sigma = \begin{pmatrix} \sigma_u^2 & 0 \dots \dots \dots 0 \\ . & \sigma_u^2 & . \\ . & . & . \\ . & . & . \\ 0 & 0 & \sigma_u^2 \end{pmatrix} = \sigma^2 \begin{pmatrix} 1 & 0 \dots \dots \dots 0 \\ 0 & 1 \dots \dots \dots 0 \\ . & . & . \\ . & . & . \\ 0 & 0 & 1 \end{pmatrix}$$

And the estimates obtained by using ordinary least square method are given by **(5)**

$$\hat{\beta} = \begin{pmatrix} n & \sum_{i=1}^{n} X_{1i} & \sum_{i=1}^{n} X_{2i} & \dots\dots & \sum_{i=1}^{n} X_{ki} \\ \sum_{i=1}^{n} X_{1i} & \sum_{i=1}^{n} X_{1i}^2 & \sum_{i=1}^{n} X_{1i}X_{2i} & \dots\dots & \sum_{i=1}^{N} X_{1i}X_{ki} \\ \sum_{i=1}^{n} X_{2i} & \sum_{i=1}^{n} X_{2i}X_{1i} & \sum_{i=1}^{n} X_{2i}^2 & \dots\dots & \sum_{i=1}^{n} X_{1i}X_{ki} \\ . & . & . & & . \\ . & . & . & & . \\ . & . & . & & . \\ \sum_{i=1}^{n} X_{ki} & \sum_{i=1}^{n} X_{ki}X_{1i} & \sum_{i=1}^{n} X_{ki} & \dots\dots & \sum_{i=1}^{n} X_{ki}^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{i=1}^{n} y_{1i} \\ \sum_{i=1}^{n} X_{1i}y_i \\ . \\ . \\ . \\ . \\ . \\ \sum_{i=1}^{n} X_k y_i \end{pmatrix}$$

The definitions for the variables in this study are:

Yi is the number of mortality of children U5

X_{i1} is the number of infections by diarrheal in the year i

X_{i2} is the number of infections by malaria. in the year i

Xi_3 is the number of infections by pneumonia. in the year i

X_{i4} is the number of infections by meningitis. in the year i

X_{i5} is the number of infections by measles. in the year i

Data: The source of data is Statistics Office at Wad Madani Children's Teaching Hospital in the Gazera state in Sudan.

Table (1): Number of mortalities and infections in Gazera state (Sudan) 2003-2022

Number of infections by					Number of mortalities	
Measles	**Meningitis**	**Pneumonia**	**Malaria**	**Diarrhea**		**Year**
139	117	3683	2423	2189	251	2003
449	98	4711	2859	3392	224	2004
14	190	5687	3595	3130	200	2005
135	123	4132	3157	3091	198	2006
109	117	3791	3303	3113	185	2007
25	114	3851	4011	3571	151	2008
0	145	4903	3446	4385	178	2009
3	93	2851	2596	2972	134	2010
46	69	874	3019	2504	131	2011
39	73	1719	5343	2893	137	2012
37	105	1588	5400	3450	167	2013
37	93	2124	5486	3010	158	2014
37	103	2966	5559	3103	165	2015
35	116	2599	5464	3240	163	2016
37	106	2905	5892	2995	163	2017
38	87	2859	5927	2827	145	2018
39	99	1709	5569	3403	160	2019
35	115	1552	5949	2832	147	2020
38	91	2666	5811	2979	167	2021
39	98	1724	5902	2891	152	2022
67	108	2945	4536	3099	169	Average
44	26	1286	1331	439	30	Standard deviation

Statistical package SPSS version 28 used to analyze data and obtained the results.

Analysis and results: From table (1), the average monthly death rate is 14 children, and that the average monthly incidence of diarrheal diseases, malaria, pneumonia, pneumonia, and measles is 209, 252, 73, 6, and 4, respectively, which indicates that malaria is the disease that most affects children (U5) in the region. This is due to the nature of the agricultural area and the breeding of mosquitoes that transmit malaria. The least common diseases are meningitis and measles, and this is due to adherence to their vaccination doses. Based on the data, the researcher assumes that malaria and pneumonia are the most common causes of death among children (U5) due to the high number of infections compared to other diseases.

The test of significant difference within each variable showed zero P-Value for all variables (less than 0.05) indicating that there is high significant difference within each variable (11), this means that there is a difference in the values of the variables from one year to another and that there is fluctuation in the general trend that varies from one variable to another.

Model results showed that $R^2 = 0.834$ (R^2 Show the variations in the dependent variable explained by the independent variables) (7) i.e. the number of infections with the six diseases explains the changes in the number of mortalities by a percentage of 84.3%. And the Durbin -Watson test is 1.349 which lies between 0 and 4 indicating that the model is free of autocorrelation (8). The variance inflation

factor (VIF) for all variables is less than 5 (table 2) indicating that there is no problem of multicollinearity (9). The result of analysis of variance (ANOVA) (F value) showed zero P-Value which means the fitness of the model. (10)

The coefficient table is:

Table (2): Coefficient results

	B	**t**	**Significant value**	**VIF**
(Constant)	83.602	2.360	.033	
Diarrhea	-.003-	-.338-	.740	1.564
Malaria	.001	.291	.775	1.723
Pneumonia	.005	.965	.351	4.214
Meningitis	.473	2.229	.043	2.776
Measles	.483	4.900	.000	1.719

From table (2), the model can be written as follows:

$Yi = 83.602 - 0.003\ (X_{i1}) + 0.001\ (_{xi2)} + 0.005\ (X_{i3}) + 0.473\ (X_{i4}) + 0.483\ (X_{i5})$

And the significance value for diarrheal, malaria, and pneumonia is greater than 0.05, which means that they are not significant That is, they have a weak effect on the mortality of children U5 in the region, while the significance value for meningitis and measles is less than 0.05, which means that they are significant and have a high impact on the mortality of children U5 in the region.

Figure (1) explain that the residuals are normally distributed, indicating the validity of the model. (11)

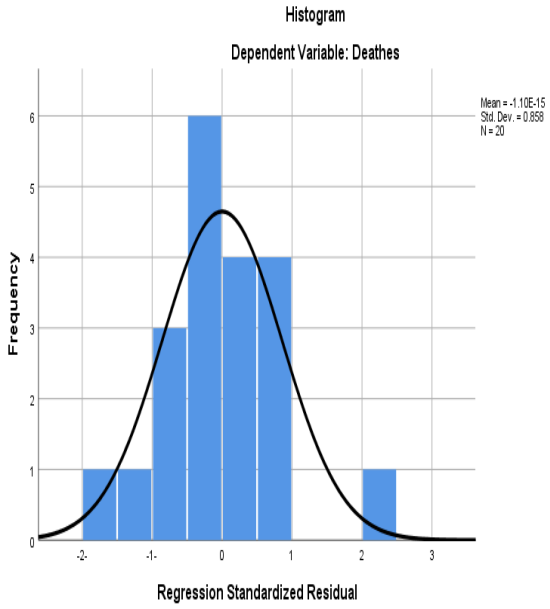

Figure (1): Histogram of residuals

Figure (2) shows that the residuals values are closer to the line, which is another evidence for the validity of the model.

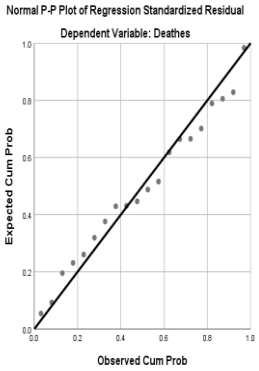

Figure (2): Residual distribution

reanalyze data with only the significant independent values (meningitis and measles) showed R2 = 0822 and zero significant value indicating that they are highly contributing to mortality of (U5) children.

CONCLUSION:

From the above discussion and analysis, it can be concluded that meningitis and measles are highly contributing (82.2%) to mortality of (U5) children in Gazera region (Sudan), although the number of infections during the year is small compared to other diseases. The result proved that the hypothesis that malaria and pneumonia are the most common causes of death in children (U5) is incorrect, which means that there is a weakness in controlling and treating these diseases, in addition to the fact that cases reach Hospitals are late due to poor health awareness and people use herbal treatment and local medicines for a period of time before coming to the hospital.

The study recommends that the competent authorities ensure and follow up on vaccinating all children, as well as providing intensive care in case of infection, and educating citizens, especially in the countryside, by going immediately after the infection occurs to the nearest hospital and not wasting time on folk treatment.

REFERENCES:

(1) Ehab Ahmed M. Frah*, Abdalla Ahmed Alkhalifa, Using Factor Analysis to Determine the Most Important Disease Causes: Under-Five Mortality in Tabuk Region, 2017, International Journal of Statistics and Applications 2021, 11(3): 70-77

(2) World Health Organization, Children: improving survival and well-being, Global response: Sustainable Development Goal 3.2.1, 8 September 2020 (https://www.who.int/news-room/fact-sheets/detail/children-reducing-mortality)

(3) World Health Organization, Infant mortality rate per 1000 live births (https://rho.emro.who.int/Indicator/TermID/48).

(4) Alvin C. Rencher and G. Bruce Schaalje, Linear Models In Statistics, John willy & Sons, New Jersey, United States, (2008).

(5) Muhammad Abdel Rahman Ismail., Linear regression analysis, Institute of Public Administration - Research Center - Riyadh - Saudi Arabia (2001).

(7) Damodar N. Gujarati & Dawn C. Porter, Basic Econometrics, fifth edition, The McGraw-Hill, New York City, United States, 2009

(8) Phil Crewson, Applied Statistics iBook (2017) file:///C:/Users/HP/Downloads/AppliedStatistics.pdf

(9) Multiple regression model to analyze the total LOS for patients undergoing laparoscopic appendectomy et.al., Multiple regression model to analyze the total LOS for patients undergoing laparoscopic appendectomy, BMC Medical Informatics and Decision Making, (2022) 22:141 (https://doi.org/10.1186/s12911-022-01884-9).

(10) Thoms Andren, Econometrics, Bookboon.com, United Kingdom &Denimark92007)

(11) Nasser Hasan, Multiple Regression Using SPSS, University of Miami, United States (2020) https://sites.education.miami.edu/statsu/wp-content/uploads/sites/4/2020/06/Regression-Webinar.pdf