Periodicals of Engineering and Natural Sciences ISSN 2303-4521
Vol. 8, No. 1, March 2020, pp.197-204

Identifying the most influential risk factors of frequent infant mortality in Iraq

Jassim N. Hussain
Department of Statistics, Administration and Economics College, University of Kerbala

ABSTRACT

Nowadays, frequent Infant Mortality for the married women are prevailing in Iraq. It may be also considered as a risky phenomenon in developing countries. It is the loss of an infant after viability. The estimated number is 2.5 million deaths occurring in the first month of life in 2017 alone. On the other hand, the survival rates of newborn infants reflect the quality of pregnancy care provided during work and viability and the infrastructure for the infant care in the different regions and countries of the world. Numerous risk factors are causing this phenomenon. Logistic regression is a statistical technique can be used to express the association between the number of infant mortality and the risk factors cause it. It helps to select the most influential risk factors for this case. The aim of this study is to use logistic regression to examine the association between biological, behavioral and lifestyle risk factors and the number of infant mortality, and to identify the most influential risk factors affected it. A simple random sample is drawn with size of 200 persons that consists of all mothers who visit the primary health care centers in Babylon province in year 2018. Seventeen risk factors are representing biological, behavioral and lifestyle factors of women under the study. The results of fitting binary and ordinal logistic regressions with all seventeen risk factors show that four risk factors show a significance effect on the dependent variable. Consequently, a stepwise logistic regression was fitted, and ordinal logistic regression model has fitted. Nevertheless, there are no much differences between the results of these models with different methods of fitting. All results show that husbands working has two times more likely to exhibit infant mortality than Husbands not working. Increasing age at marriage and woman weight were associated with an increased likelihood of exhibiting infant mortality, but increasing number of hours of women's sleeping was associated with a reduction in the likelihood of exhibiting infant mortality.

Keywords: Binary logistic regression, ordinal logistic regression, infant mortality, Influential risk factors

Corresponding Author:

Jassim N. Hussain
Department of Statistics, Administration and Economics College
University of Kerbala
E-mail: Jasim.nasir@uokerbala.edu.iq
Phone No: 009647734544055

1. Introduction

The infant mortality rates reflect the quality of pregnancy care provided during labor and viability and the infrastructure for the infant’s care in the different countries of the world. The rate of infant mortality is measured by the number of deaths of children under one year old as compared to the total number of births. Studies in [1-2], reported that the infant mortality rates in Iraq is 50 per 1000 living newborns until 1990, 101 in 1999, 35 in 2006 and was decreased to 33 per 1000 living newborns in 2011. The increase in child mortality rate in 1999 is a deflection that cannot be explained by demographic factors. It rather reflects the results of economic sanctions and health deterioration that added to the problems of the population situation during the 1990s. Despite the decay in infant mortality rate to reach to 33 deaths per 1000 live births in 2011, it is still highest compered to neighboring countries such as 11 per 1000 in Kuwait, 26 per 1000 in Arabia Saudi and Jordan, 15 per 1000 in Syria, 22 per 1000 in Palestine and 22 per 1000 in Turkey [3].

Meanwhile, globally, United Nations draw an image about this case. They stated an estimated 6.3 million children and youths died, mostly from preventable reasons [4] in 2017. Among them, 2.5 million deaths occurring in the first month of life. We can conclude that about 40 percent of them were died in the first month of life. Consequently, infant mortality rate is one of the most expressive indicators of development in all countries. This importance has taken form the fact that the decrease in newborn mortality is only possible with an improvement in living conditions of a most people [5]. Hence, a high infant mortality can indicate that the health needs unmet and the environmental factors unsuitable.

Results of similar studies in different countries have indicated that there is a gap between rural and urban regions in infant mortality rates because of the contribution of unobserved heterogeneity at the household and the community level [6]. The authors of [7], stated that according to the world health organization (WHO) reports, the leading causes of 80 percent of all newborn death include prematurity, low birth weight, infections, birth suffocation, and birth disturbance. The study in [3], stated that even with development over the past twenty years of the last century, millions of newborns, children and young youths die every year. These deaths reflect the limitations of children and communities to access basic health involvements. Therefore, they considered mortality rates among children as key indicators for child happiness; moreover, for supportable social and economic development.

The highest ratio of mortality of infants in recent years in Iraq is due to the wars and socioeconomic factors. Consequently, studying of infant mortality plays an important role in health planning in order to reduce the previous rate in children. Thus, the main objective of this study is to quantify the effect of risk factors for infant mortality and to determine causes of infant mortality and factors associated with it in the Babylon province in 2018 as a sample of Iraq population. Also, to compare the performance of binary logistic and ordinal logistic regression to find out the relationship between the infant mortality (number of death of newborn per woman) besides biological, behavioral and lifestyle factors which include the age of women, education, occupation, socioeconomic status, residency, age at marriage and other factors.
The plan of this paper is as follows: In Section 2, we present the methods and material include data collection and statistical analysis. Results and discussion of the results are presented in Section 3. Finally, some summarizing remarks are placed in Section 4.

2. Methods and material

2.1 Data collection

A simple random sample is drawn with size (200) consists of all women who visit the primary health care centers in Babylon province in 2018. Biological, behavioral and lifestyle factors stand for the characteristics of women under the study. The information was recorded from the records of the women who visited this center during 2018. The response variable (Y) which is the number of newborn death (0, 1, 2, and 3) during the marriage life. The risk factors are affecting this variable which are including X1 for age of women in years; X2 for age of women at marriage in years; X3 for women’s educational attainment years; X4 for husband’s educational attainment in years; X5 for women’s weight in Kg; X6 using contraception (1=yes, 2=No); X7 for Smoking (1=yes, 2=No); X8 for husband's age in years; X9 for husband job (1=Not work, 2= work); X10 for marriage period in years; X11 for number of children born; X12 for exercise per week in hours; X13 for thyroid disease (1=yes, 2=No); X14 for sleeping per day in hours; X15 is about taking medications (1=yes, 2=No); X16 is about duration of breastfeeding in months; and finally X17 is about mother's job (1=Not work, 2= work).

2.2 Statistical analysis

Descriptive statistics were obtained for different risk factors. The descriptive statistics include tables, bar-charts to describe the percentages of the risk factors and mean in addition to standard deviation for the continuous variables in the study. Since the response variable represents number of newborn death, the appropriate inferential statistical tools can be binary logistic regression model designed to fit a regression model in which the dependent variable Y consists of two cases (0= no death and 1= one death or more) [8]:

$$logit(\pi) = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + U \quad (1)$$

where, X_j is the covariates and β_j; $(j = 0,1, \ldots, p)$ are the unknown parameters of the model which are estimated by using one of the estimation methods. It describes the relationship between a dichotomous response variable and a set of explanatory variables.

On the other hand, the original data consist of the dependent variable Y with ordinal counts includes four ordered groups: $(0 = \text{ } no\ death, 1 = one\ death, 2 = two\ deaths, 3 = three\ deaths)$. Therefore, the ordinal logistic regression model which is designed to fit a regression model for the ordinal dependent variable can be used. The most well-known of these ordinal logistic regression methods is called the proportional odds model (POM). The model of ordinal logistic regression is [9]:

$$lin(\theta_j) = \alpha_j - \beta_1 X_1 + \cdots + \beta_p X_p \quad (2)$$

where, j goes from 1 to the number of categories of response variable minus 1. The general form of ordinal logistic model is the cumulative logit model (CLM). CLM assumes that the odds of response below a given response level are constant regardless of which level we have chosen. This model allows separated intercepts for the cumulative logit, but restricted the parameter sets for the predictors to be the same across all logits. On the other hand, POM is a special case of the CLM. Furthermore, logistic regression analysis is statistical technique that examines the influence of various risk factors on a categorical outcome by estimating the probability of the event's occurrence and the odds ratios for each risk factors. Consequently, there are many criteria used to assess the goodness of fit of the model such as likelihood ratio (LR) test which is used to assess the hypothesis that all $\beta's$ are equal to zero. On the other words, it measures how well the independent variables affect the response variable [10]. The second test is Pseudo R^2 for logistic regression to assess the proportion of accounted variation in the dependent variable by the risk factors. Both [11] and [12] cautioned against relying on LR measures in logistic regressions, but Spicer suggested that the two different R^2 measures might be viewed as cautious indicators of the range within which the actual influence of the independent variables on the dependent variable lies. The third criterion is Hosmer and Lemeshaw (H-L) test to assess the overall goodness of fit of the model [8]. This test is used to test the null hypothesis that the fitted logistic regression model is the correct model. The last criterion is Wald test to assess the significant of the coefficients of the independent variables in the model [13].

3. Results and discussions

The data included a sample of 200 mothers, who were in age of 17-49 in Babylon government. The response variable considered in this study was the number of new born death which is in two cases. The first one is dichotomous dependent variable (0=no death, 1=one or more deaths) pear each mother. The second case is the ordinal dependent variable which has four cases as mentioned above. The main goal of the study is to assess and identify the most influential risk factors that influence the number of the new born death. The result of descriptive and inferential statistics are summarized and described in this section. The data were analyzed using SPSS version 24. The descriptive statistics include tables, bar-charts to describe the frequency distribution and percentage of the variables in the study. Table 1 and Figure 1 show the frequencies of the dependent variable in the two cases when the dependent variable is binary and ordinal.

Table 1. Descriptive statistics of dependent variable

		n	%			n	%
Binary response variable	0= no death	109	54.5	Ordinal response variable	0= no death	109	54.5
	1= one death or more	91	45.5		1= one death	67	33.5
					2= two deaths	17	8.5
					3= three deaths	7	3.5

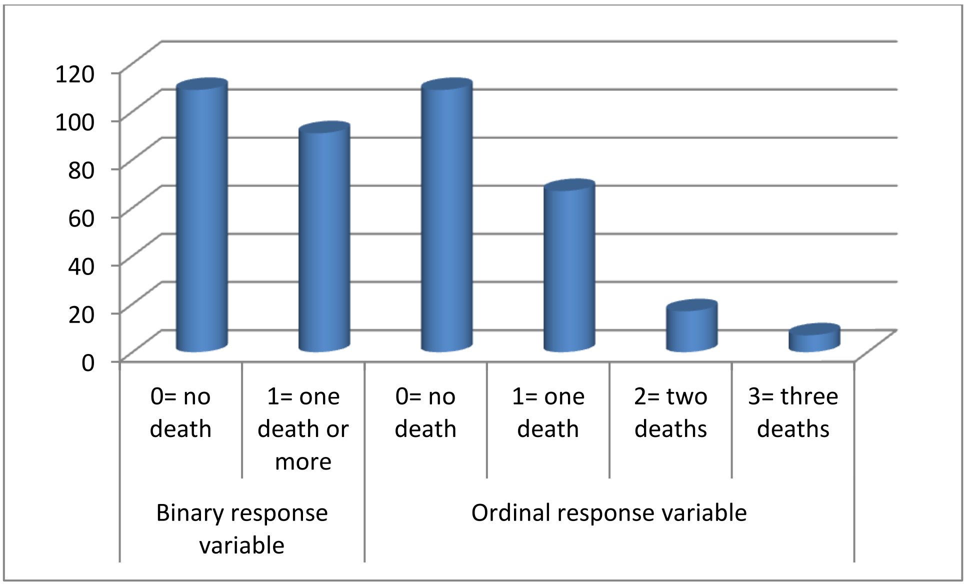

Figure 1. The frequencies of ordinal and binary dependent variable

The dataset consists of two groups of explanatory variables. The first group is a quantitative variables with the following descriptive statistics as in Table 2.

Table 2. Descriptive statistics of the continuous variables

Covariates	n	Min.	Max.	Mean	Std. deviation
Age Women (year)	200	17	49	31.03	8.362
Age at marriage (year)	200	14	26	18.64	3.203
W Education (year)	200	0	18	9.72	5.417
H Education (year)	200	0	18	11.66	4.163
Woman Weight (Kg)	200	52	90	68.52	8.913
Husband Age (year)	200	19	53	34.17	8.266
Marriage Period (year)	200	2	29	12.39	8.812
Exercise (hour)	200	0	9	3.85	2.903
Woman's sleep (hour)	200	6	9	8.09	.738
Breastfeeding (month)	200	0	25	23.21	2.750

The second group in the data set is the qualitative factors. Table 3 explains the descriptive statistics of them. Tables 1, 2 and 3 and Figures 1and 2 show a description of the variables in the study.

Table 3. Descriptive statistics of the risk factors

Factor	Levels	n	%
contraception	Not using	24	12
	Using	176	88
Smoking	Not smoking	76	38
	Smoking	124	62
Husband's job	Not working	110	55
	Working	90	45
Thyroid	Not using	26	13
	Using	174	87
Medications	Not Taking	60	30
	Taking	140	70
Mother's job	Not Working	170	85
	Working	30	15

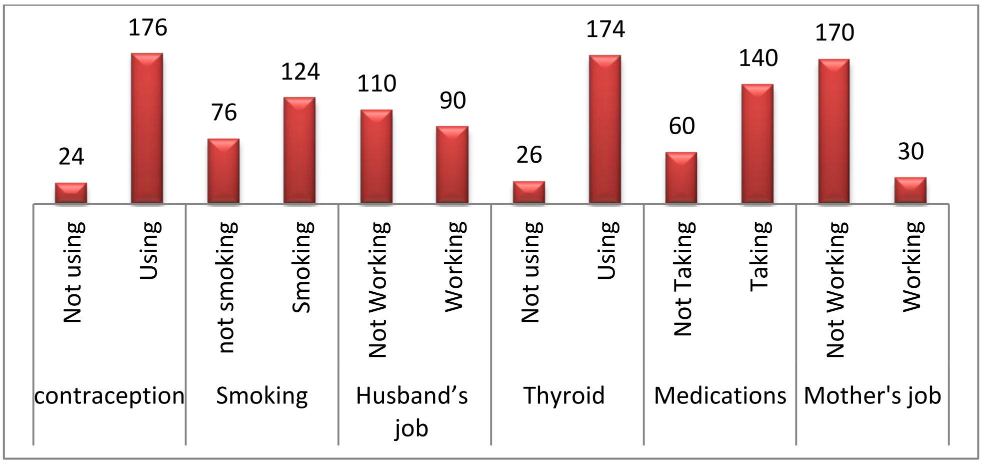

Figure 2. Descriptive Statistics of risk factors

On the other side, inferential statistics is the procedure by which we reach a conclusion about a population based on the information in the sample drawn from that population. Therefore, among inferential statistics, we would use binary logistic and ordinal logistic regression to identify the most influential risk factors that are affecting the response variable (the number of infant mortality). The results of fitting the binary logistic regression with all seventeen risk factors are given in Table 4. Classification in Table 4 provides a lot of important information about binary logistic regression, including, overall percentage accuracy in classification (PAC) was 82.0%, indicating that the model was good in classifying subjects.

Table 4. Classification table

	Predicted infant mortality			Percentage Correct
		No death	One death or more	
Observed infant mortality	No death (0)	97	13	PPV = 88.2 %
	One death or more (1)	23	67	NPV = 74.4%
Overall Percentage		Sensitivity = 80.8%	Specificity = 83.8%	PAC = 82.0%

The model's Sensitivity was 80.8% indicating that it was able to identify subjects belong to the one death or more group in 80.8% of the cases. Specificity was 83.8%, indicating that the model was able to truly identify subjects belong to the no death group in 83.8% of the cases. The positive predictive value (PPV) was 88.2%, while the negative predictive value (NPV) was 74.4%. Namely, 88.2% of subjects whom the model classified based on one death or more group were actually in that group. The 74.4% of subjects classified based on no death group were actually in that group. Secondly, the results of fitting binary logistic regression with all seventeen risk factors show that Nagelkerke R^2 which is indicating that the explained variation in the dependent variable based on our model was 42.6%. A fit model would reveal H- L test (8) = 31.270 with *p*-value less than 0.05 for test. Namely, it means that the goodness of fit for the model was not significant. Meanwhile, for $\chi 2(16) = 76.8,\ \mathrm{p-value} < 0.0001$, this p-value for the test for all slopes are zero. The low p-value indicates that the relationship between the response variable and the explanatory variables is statistically significant. Also, the overall correctly classified percentage equal to 82% in the classification table. Four risk factors only show a significance effect on the dependent variable (age at marriage (X2), husband Education (X4), women weight (X5) and number of hours of women's sleeping (X14)) which can considered them as the most influential risk factors and the model is:

$$logit(\pi_i) = -7.086 + 0.374X_2 + 0.214X_4 + 0.069X_5 - 0.780X_{14} \qquad (3)$$

Consequently, A Forward Stepwise (Wald) logistic regression was performed to confirm the most influential risk factors from the seventeen risk factors on the dependent variable frequent infant mortality. The model was statistically significant because $\chi 2(4) = 60.51,\ \mathrm{p} < 0.0005\ and\ H - L = 10.8,\ p - value = 0.213$ are agree which are means the overall fitting of the model was significant. The model explained 35% (Nagelkerke R^2) of the variance in infant mortality and correctly classified 71% of cases, the model's Sensitivity was 72.8% ; Specificity was 68.6%, the positive predictive value (PPV) was 75.5% while the negative predictive value (NPV) was 65.6%. The most influential risk factors (Age at marriage (X2), Women weight (X5), Husband job (X9) and number of hours for women sleeping (X14)) which are having a significant effect on the infant mortality based on:

$$logit(\pi_i) = -5.847 + 0.308X_2 + 0.095X_5 - 0.814X_9 - 0.777X_{14} \qquad (4)$$

Increasing in age at marriage and woman weight were associated with an increased probability of exhibiting Infant Mortality. But, increasing number of hours of women's sleeping and not working husbands were associated with a reduction in the probability of exhibiting infant mortality. In spite of the overall goodness of fit of this model, the statistical criteria are less than in the previous model. Also, there is changing in the influential risk factors where inter new risk factor (husband's Job) and get out (Husband's education). Therefore, we use the Backward Stepwise (Wald) logistic regression to fit the model. The model was statistically significant because $\chi 2(6) = 70.1,\ \mathrm{p-value} < 0.0001\ and\ H - L = 14.2,\ p - value = 0.08$ are showing that the overall fitting of the model was significant. The model explained 40% (Nagelkerke R^2) of the variance in infant mortality and correctly classified 80.0% of cases. The model's Sensitivity was 79.7% . Specificity was 80.5%. The positive predictive value (PPV) was 85.5%, while the negative predictive value (NPV) was 73.3%. The most influential risk factors (age at marriage (X2), women education (X3), husband education (X4), women's weight (X5), husband job (X9) and number of hours of women's sleep (X14)) have significant effects on the infant mortality based on:

$$logit(\pi_i) = -5.936 + 0.350X_2 - 0.121X_3 + 0.177X_4 + 0.094X_5 - 1.216X_9 - 0.950X_{14} \qquad (5)$$

Increasing in age at marriage, woman weight and husband's education were associated with an increased probability of exhibiting infant mortality. But, increasing number of hours of women's sleeping, not working husbands and not educated women were associated with a reduction in the possibility of exhibiting infant mortality.
Three methods of fitting binary logistic regression are giving different models with different risk factors and statistical criteria. We fit the ordinal logistic regression model when the response categories are ordered and fitting a binary logistic regression model may be throwing away information about the ordering. An ordinal logistic regression model preserves that information, but it is slightly more involved. The results of fitting ordinal logistic regression model show that the effect of the predictors on the odds of an event occurring in every subsequent category is the same for every category. This is an assumption of the model that it must be checked. The $\chi^2_{(30)} = 34.688$; $P - value = 0.254$, which means that null hypothesis is accepted. The proportional odds model is appropriate to fit the data. All the goodness of fit tests (likelihood chi-square, Person chi-square and deviance) have P-value <0.0001 which means the model is reliable to use for predicting.
Then, we have three fitted models with different intercepts but with same parameters of the explanatory variables. The most influential risk factors are (age at marriage (X2), husband's education (X4), women's weight (X5), husband's job (X9) and number of hours of women's sleeping (X14)).

$$logit(\theta_j) = \begin{bmatrix} 8.022 \\ 10.232 \\ 11.619 \end{bmatrix} + 0..260X_2 + 0.152\,X_4 + 0..051X_5 + 0.940X_9 - 0.605X_{14} \qquad (6)$$

Consequently, the event being modeled in the POM is not having an outcome in a single category as it is done in the binary models. Reasonably, the event being modeled is having an outcome in a particular category or any previous category. Therefore, the model has three values for intercept (α_j). All of them are statistically significant with P-value < 0.05. Finally, there are no much differences between the results of ordinal logistic regression model and the binary logistic regression with different methods of fitting. All the models represent that the most influential risk factor are age at marriage (X2), husband's education (X4), women's weight (X5), husband's job (X9) and number of hours of women's sleeping (X14)

4. Conclusion

We found from the above results that all the models indicated that the most influential risk factors of infant mortality are age at marriage of women and women's weight. The results show that any increasing of them leads to increasing in the probability of infant mortality. Also, the results show that not working husbands were associated with a reduction in the possibility of exhibiting infant mortality. Furthermore, increasing in the number of hours of women's sleeping has associated with a reduction in the possibility of exhibiting infant mortality. Therefore, more studies are needed to confirm this observed association between these factors and the infant mortality in the other cities in Iraq.

References

[1] H. Becher, O. Müller, A. Jahn, A. Gbangou, G. Gisela Kynast-Wolf and B. Kouyaté, Risk factors of infant and child mortality in rural Burkina Faso, *Bulletin of the World Health Organization,* vol.82, no.4,2004.

[2] Iraq National Population Commission (INPC), Iraq Population Situation Analysis- PSA, The Second National Report on the State of Iraq Population in the Context of the ICPD and MDGs; Supported by UNFPA-Iraq CO, 2012.

[3] United Nations (UN), Department of Economic and Social Affairs, Population Division. World Population Policies. New York, 2013. Retrieved on 30/8/2019: https://www.un.org/en/development/desa/population/publications/pdf/policy/WPP2013/wpp2013.pdf.

[4] United Nations Inter-agency Group for child mortality estimation (UN IGME), 2018, Levels and Trends in child mortality report, 2018.

[5] G.R. Sharifzadeh, K. Namakin, and H. Mehrjoofard, An Epidemiological Study on Infant Mortality and Factors Affecting it in Rural Areas of Birjand, Iran, Iran Journal Pediatr; vol.18,no.4, pp.335-342,2008.

[6] E. Van de Poel, O. O'Donnell and E.V. Doorslaer , What explains the rural-urban gap in infant mortality household or community characteristics; Retrieved on 30/8/2019, https://www.researchgate.net/publication/41086937,2009.

[7] N. Ghotbi, M. Zokai, K. Rahmani, F.Z. Vakili, S. Zandi and N. Asadi, Risk Factors Related to the Neonatal Mortality in Kurdistan Province, Iran: A Population-Based Case-Control Study, *Shiraz E-Med Journal*, vol.18, no.3, pp.1-9,2017.

[8] D.W. Hosmer and S. Lemeshow, (2000), Applied Logistic Regression, 2nd ed., John Wiley & Sons, Inc.

[9] O.C. Reddy and F. Alemayehu, Ordinal logistic regression analysis to assess the factors that affect health status of students in Ambo University: a case of natural and computational sciences college, Ambo University, *International Journal of Modern Chemistry and Applied Science, vol.2,no.3, pp.153-163,2015.*

[10] P. Sur, Y. Chen and E.J. Candès , The Likelihood Ratio Test in High-Dimensional Logistic Regression Is Asymptotically a Rescaled Chi-Square, retrieved at 10/10/2019 at: https://statweb.stanford.edu/~candes/papers/LRT.pdf, 2017.

[11] Peng, Chao-Ying Joann; Lee, Kuk Lida; & Ingersoll, Gary M. (2002). An introduction to logistic regression analysis and reporting. *Journal of Educational Research*, vol. 96, no.1, pp. 3-13, 2002.

[12] Spicer, J. (2004) *Making sense of multivariate data analysis*, Sage Publications, California.

[13] O. Franklin and O. Emmanuel, Logistic Regression: A Paradigm for Dichotomous Response Data, *International Journal Of Engineering And Science (IJES)*, vol.3, no.6,pp.1-5,2014.