
Statistical Models of Economic Burden: A Case Study in Medicine

Submitted 30/08/19, 1st revision 19/09/19, 2nd revision 27/10/19, accepted 20/11/19

Lakman I.A.¹, Maksimenko Z.V.², Shangareeva R.Kh.³, Gindullin R.V.⁴

Abstract:

Purpose: The main aim of this article is to use statistical methods for the estimation of the economic burden and the survival rate of deeply premature babies.

Design/Methodology/Approach: The results of a survey of 2.222 children with a birth weight of 501-1500 grams and a gestational age of 23-37 weeks were used as input data. Cox's proportional hazards model was used as a survival tool.

Findings: The results of Cox survival regression model showed a series of statistically significant predictors of survivability ($p < 0.05$) for three age cohorts: neonatal, postnatal and pediatric (until 2 years). One of the statistically significantly predictors of survivability of premature infants with very low birth weight (VLBW) and extremely low birth weight (ELBW) in every age cohort is the volume of primary resuscitation measure and the length of stay in the neonatal pathology unit (NPU).

Practical Implications: The results permitted to assess the amount of nursing care measures, the duration of care in a neonatal pathology unit, the rehabilitation of children with VLBW and ELBW in the long run. The assessment will ultimately help to estimate the overall economic burden associated with maintaining health and quality of life of premature babies.

Originality/Value: The scientific contribution of the study consists in the use of an integrated approach to the problem of estimating the economic burden of nursing very premature babies, taking into account their survival and subsequent disability risks in the neonatal, postnatal, and pediatric periods.

Keywords: Premature newborns, very low birth weight, extremely low birth weight, survival analysis, risk.

JEL Classification Codes: C01, C21, F43, F62, F63, O47.

Paper type: Research article: Statistical Methods of Economic Burden.

¹Associate Professor of the Department of "Computing Mathematics and Cybernetics", Ufa State Aviation Technical University, lackmania@mail.ru

²Associate Professor of the Department of "Mathematical Methods in Economics", Bashkir State University, zubazz@mail.ru

³Professor of the Department of "Pediatric Surgery", Bashkir State Medical University, shanrosa@mail.ru

⁴Associate Professor of the Department of "Mathematical Methods in Economics", Bashkir State University, ramizgindullin@gmail.com

Acknowledgment:

The study is carried out with financial support of Russian Foundation for Basic Research (RFBR) within the scientific project №18-010-00872 «Development of methodology, mathematical and software tools for intellectual evaluation and planning of expenses for medical treatment, rehabilitation, health and life quality support of premature children born with very low and extremely low body weight».

1. Introduction

Currently, there are no common standards and protocols to assess the gross economic burden as nursing and rehabilitation of children born with VLBW and ELBW. In this regard, there is a need to develop a common methodology for assessing and long-term planning of the costs involved. When planning the costs, survival potential assessment should be included, as well as the disability probability rate of very premature babies.

The work by Johnson *et al.* (2014) is devoted to the problem of estimating the economic burden in Canada, considering the prediction of the survival rate of premature and deeply premature infants. The paper deals with the assessment of the economic burden of nursing and rehabilitation within 10 years of a child's life. Moreover, in the calculations, premature babies are divided into 3 categories in accordance with the gestational age of: less than 28 weeks, 28-32 weeks, 33-36 weeks.

An assessment of the economic burden, taking into account both direct and indirect costs associated with nursing very premature babies within the period up to 1 year in Nigeria, was considered in the paper by Tongo *et al.* (2009). Gilbert's study (2006) in USA estimated the economic costs associated with preterm labor, including the cost of initial hospitalization, the cost of any chronic diseases related to prematurity, and social costs, including the loss of paid work by a family member looking after an infant or child, plus potential future earnings loss. Muraskas and Parsi (2008) investigated the rationale for allocating funds for nursing premature infants on improving the quality of antenatal care and reducing the birth rate of underweight children.

Lewit *et al.* (1995) proved that, considering the additional costs of nursing and adapting of preterm newborns in 1988 up to the age of 15, the total cost is about \$6 million higher than in the case when children were born full-term. Schmitt *et al.* (2005), based on data about the births of more than 500.000 children in the state of California in the year 2000, also conclude that the neonatal costs of nursing of preterm infants are disproportionate to the costs on newborns born on time.

In Russia, the issues of estimating the economic burden of preterm birth and nursing of premature infants have hardly been studied. Costs of treatment and nursing of the children born with VLBW and ELBW, due to a high tariff, are covered only within the neonatal period (up to 28 days) from the funds allocated for high-tech medical care. The actual costs of treatment, according to estimates by various medical

institutions involved in care and rehabilitation of premature babies, are much higher than the established tariff (Bakhitova *et al.*, 2018).

The existing approach to cost estimate does not consider differentiation according either to body mass, or to the period of essential intensive care. The risks of disability, the expected quality of life in the future and the anticipated costs in the long term are not taken into account either. In addition to the nursing and rehabilitation expenses within the first year of life, the authors suggest evaluating direct (for example, the costs of treatment of chronic diseases associated with prematurity) and indirect (social costs associated with possible disability and employment incapacity of the child in the future) costs. Since the assessment of the total economic burden should be determined at different periods in the life of children with VLBW and ELBW (neonatal, postnatal and pediatric periods), the mortality risks of children in the neonatal, postnatal and pediatric periods and disability risks in their future should be assessed at the first analysis stage.

2. Methodology

There are different statistical methods that allow estimation of the change of economic burden of health care, depending on premature babies' survivability in different age cohorts (neonatal, postnatal and pediatric). For example, in Abolfotouh *et al.* (2018) and Santhakumaran *et al.* (2018) logistic regression is used to estimate the survival rate of deeply premature babies up to 6 and 20 years respectively, the main predictors being gestational age, birth weight, sex, and Apgar scores. The use of such a tool is not always justified, since survival by period is subject to the class of exponential distributions, and in this case it is advisable to use survival analysis regressions, such as the Cox models. The Cox proportional hazards models are used to assess the survival of hypotrophic babies (Inoue, Ochiai and Yasuoka, 2017).

As an alternative method to assess the risk of developing disability in premature babies, is proposed to use tools for solving binary classification problems such as binary multivariate regression with the distribution function matching/selection. In the work of Li *et al.* (2018) and Silva *et al.* (2018) the multivariate logistic regression is used to estimate the likelihood of disability of neuro-psychic development of deeply premature babies.

Based on the analysis of available research data devoted to the problem of survival modeling and disability of very premature babies, as well as consultations with neonatologists, it was decided to investigate the survival rate by period: neonatal (up to one month), postnatal (up to one year) and pediatric (up to two years), and to study disability of premature children under 24 months. The limitation of the study of survival rates and disability prediction of premature babies within a two-year period is due to the fact that representative data sampling sufficient for reliable conclusions were available only for this period of time. The empirical base of the study was the data from the examination and treatment of 277 children born with VLBW and ELBW

who were hospitalized in 2013-2018 into the Children's City Clinical Hospital No. 17 and the Republican Clinical Perinatal Center in Ufa.

The methodology of the statistical research on the analysis of survival rate can be arranged in the following scheme. At the first stage, the target variables in terms of the longevity of an infant born with VLBW or ELBW, up to 28 days (in days) and from one month to one year (in months), were formed. Also censored variables were formed, limited to zero associated with a life expectancy variable, taking the value of 1 in the event of the child's death and 0 in case of their survival by this time.

At the second stage, we checked the significant differences in the survival of very premature babies in groups formed on the basis of the distinction between alternative features (for example, the use of IVF technology), based on the nonparametric Gekhan-Wilcoxon test (Gehan, 1965). The test checked the null hypothesis about indistinction in the survival of the groups formed according to the principle of diversity of features. This allowed us to define statistically significant factors that determine the survival rate of children with VLBW and ELBW in the neonatal and postnatal periods.

In accordance with the Hekhan-Wilcoxon criterion, each object Z_{0i} of a group of survivors is compared with each object Y_{1j} of a group of people who died (according to their life length) and the statistics are calculated (1):

$$U_{ij} = U\{Z_{0i}, Y_{1j}\} = \begin{cases} 1, & \text{if } T_{0i} > T_{1j} \text{ and } X_{0i} \geq T_{1j}, \\ 0, & \text{if } T_{0i} = T_{1j} \text{ or the object with the shortest life} \\ & \text{length was censored,} \\ -1, & \text{if } T_{0i} < T_{1j} \text{ or } T_{0i} \leq X_{1j}. \end{cases} \quad (1)$$

The statistics of the test $W = \sum_{i=1}^{n_0} \sum_{j=1}^{n_1} U_{ij}$ is calculated, which is compared to the standard normal distribution to make a decision in favor of the zero and alternative hypothesis.

At the third stage, a survival model was evaluated, which allowed determining the death risk multipliers of infants with VLBW and ELBW in the period of up to 28 days and in the period of up to one year based on previously identified factors using the Cox proportional hazards modeling (Cox, 1975).

$$\lambda(t|x) = \lambda_0(t) \exp(x'\beta), \quad (2)$$

where $\lambda_0(t)$ is the basic death risk function in the considered periods of time, equal to the risk function under the condition that there are no influence factors ($x = 0$), $\exp(x)$ is a function linking the risk of dying by a certain period with the corresponding predictors. Incrementing the factor $x_j, j = 1, \dots, k$ by 1 means an increase in the linear

combination of $x'\beta$ by the corresponding coefficient β_j and the growth of the survival function at each point t by $e^{(\beta_j t)}$ times.

Evaluation of model coefficients was done with maximum likelihood estimation procedures proposed by Breslow (1974) and Efron (1977). Most suitable likelihood equation were selected based on the Akaike and Schwarz minimum information criteria.

To assess the quality of the constructed Cox proportional hazards models, the quality indicators of the model were adapted to the actual data based on the coefficient of determination R^2 . The coefficient shows the “explanatory power” of the model obtained. To assess the consistency of the influence of explanatory factors on the survival function, the coefficient of concordance was used, whose proximity to unity indicates the consistency of the predicted data with the actual ones. To assess the statistical significance of the Cox proportional hazards model, a likelihood ratio criterion (LR test) and the Wald test were used, testing the null hypothesis of the absence of the factors’ impact on the resulting survival variable.

Methods to determine the probability of disability development can be laid into the following scheme. At the first stage, the binary regression is estimated, with the selection of the distribution function (logistic, normal or Cauchy) in accordance with the equation:

$$P(y_i) = F(x'\beta) \quad (3)$$

Where y_i is a binary variable that takes on the value of 1 if there is a disability of a premature baby and 0 in the contrary case, $x'\beta$ is a linear combination of the disability predictors considered in the model, F is the distribution function (normal, logistic or Cauchy) based on the Akaike and Schwarz minimum information criteria.

At the second stage, the quality of the model is assessed. Hosmer-Lemeshow criterion is used to check the consistency between calculated and factual data. Null hypothesis of this criterion assumes that the prediction model is well calibrated.

At the third stage, the marginal effects of the influence of the considered statistically significant factors on the disability probability in very premature babies are calculated, based on which the interpretation of the simulation results is made.

At the fourth stage, the quality metrics of the obtained model for predicting the disability of premature infants are estimated based on an analysis of the conjugacy matrix and calculation of the area under the ROC-curve (AUC), which is a graph of the ratio of true positive predictions to false positive predictions at different cut-off thresholds for the probability of a positive case.

3. Study Results

As a result of the analysis carried out in accordance with the proposed stages, the statistical significance of the influence of independent factors on the survival rate of premature infants in the neonatal, postnatal and pediatric periods (up to 2 years) was verified via the Gekhan-Wilcoxon test. As a result, birth weight, an infant's sex, fetal pathology, Apgar score, the volume of primary resuscitation measures, the duration of traditional mechanical ventilation, the duration of stay in a NPU, the presence of pneumonia at birth were found to be statistically significant at $p < 0.1$.

At the second stage, estimates of the Cox survival models were obtained on the basis of the maximum likelihood method in the Breslow estimates for a survival period of up to 30 days, up to one year and up to 2 years, which are presented in Tables 1-3, respectively. The tables show the coefficients of the Cox model, the multipliers calculated from them, standard errors and the corresponding p-level deviation of the null hypothesis. In the analysis, only statistically significant survival factors were left, for which $p < 0.1$.

Table 1. Estimates of survival model coefficients for the neonatal survival period (up to 30 days)

Influence factor	Coefficient β_j	Multiplicator e^{β_j}	Statistical error	p-level
Birth weight	-0,00608***	0,9939***	0,0015	0,0000
Fetal pathology	2,46003***	11,7051***	0,7515	0,0011
Apgar score	-0,83378***	0,2108***	0,4344	0,0000

Note: *, **, *** - the significance of coefficients at the levels of 10%, 5%, 1%.

The model obtained for infant survival in the neonatal period contains only three statistically significant influence factors. In general, the model is statistically significant (Kendall's coefficient of concordance, showing consistency of features, 0.901, $R^2 - 0.39$, model general significance statistics $LR = 41.63$ ($p = 0.0000$)). As can be seen, an increase in birth weight in relation to the average weight for the sample of premature babies in question reduces the probability of a child's death in the period of up to 30 days (the corresponding multiplier $e^{\beta_j} < 1$). On the contrary, the presence of fetus pathology increases the likelihood of a child not surviving one month by 11.7 times compared with infants who have no pathology. A twofold increase in Apgar score compared to the average value for premature babies reduces the chances of not surviving 1 month by 0.21 times.

Table 2. Estimates of survival model coefficients for a postnatal survival period (up to 1 year)

Influence factor	Coefficient β_j	Multiplicator e^{β_j}	Statistical error	p-level
Fetal pathology	1,0403**	2,8301**	0,4629	0,0246
Not fully provided primary resuscitation	1,8255***	6,2059***	0,4783	0,0001

Length of mechanical ventilation	0,0239**	1,0242**	0,0096	0,0131
Duration of stay in a NPU	-0,1662***	0,8469***	0,0311	0,0000
Pneumonia at birth	2,1766**	8,8164**	1,0311	0,0348

Note: *, **, *** - the significance of coefficients at the levels of 10%, 5%, 1%.

The model obtained for the survival of infants in the postnatal period already contains five statistically significant influence factors. Overall, the model is also statistically significant (the concordance coefficient is 0.957, R^2 is 0.67, the statistics of the overall significance of the model are $LR = 111.4$ ($p = 0.0000$)). As for the survival in the neonatal period, the factor of the fetus pathology affects survival in the post neonatal period, and the multiplier of the mortality risk up to 1 year is 2.8. Also, a significant factor that has a remote effect is pneumonia at birth: the risk of not surviving until the age of one for these children is 8.8 times greater than for the children who do not have it at birth. The factors associated with nursing of very premature babies also significantly affect postnatal survival. With deficient primary resuscitation, the risk of not surviving a year for the children with VLBW or ELBW increases 6.2 times, the insufficient duration of traditional mechanical ventilation increases the risk of not surviving one year by 1.02 times. At the same time, the stay of children in a NPU 2 times longer than the average time (in days) reduces the risk of death in the postnatal period by 0.85 times.

Table 3. Estimates of survival model coefficients for pediatric survival period (up to the age of 2)

Influence factor	Coefficient β_j	Multiplicator e^{β_j}	Statistical error	p -level
Fetus pathology	0,9858**	2,6800**	0,4645	0,0338
Male sex of an infant	0,7928*	2,2095*	0,4286	0,0643
Deficient resuscitation	1,4552***	4,283***	0,4484	0,0012
Duration of stay in a NPU	-0,1793***	0,8358***	0,0328	0,0000
Pneumonia at birth	2,1409**	8,5077**	1,0244	0,0366

Note: *, **, *** - the significance of coefficients at the levels of 10%, 5%, 1%.

The model obtained for the survival of infants during the pediatric period (up to the age of 2) already contains five statistically significant influencing factors. Overall, the model is also statistically significant (the concordance coefficient is 0.95, R^2 is 0.67, the statistics of the overall significance of the model are $LR = 108.6$ ($p = 0.0000$)). As for the survival in the neonatal and postnatal periods, the factor of the fetus pathology continues to affect the survival rate during the pediatric period, and the multiplier of the death risk under the age of 2 is 2.68. As well as for the survival in the postnatal

period, the most significant factor that has a remote effect in the period of up to the age of 2 is pneumonia at birth: the risk of not surviving until the age of 2 for these children is 8.5 times greater than for the children, not suffering from it at birth.

As before, the volume of resuscitation measures is a significant factor, in case of resuscitation measures are deficient or incomplete, the risk not to survive until the age of two for deeply premature babies increases by 4.28 times. The stay of children in a NPU is twice as long as the average time (in days) reduces the death risk in the pediatric period by 0.84 times. Unlike the survival models built for the neonatal and postnatal periods, the male sex factor becomes significant for the mortality risk up to the age of two, and this risk will increase 2.2 times in comparison with female children.

When modeling the occurrence of disability of premature infants with VLBW and ELBW in the pediatric period, all four specifications of binary choice models were used: a Logit model, a Probit model, a Gompit model, and the Cauchy model. The same factors as in survival models were considered as predictors of disability. As a result of the analysis of the Akaike and Schwartz information criteria for all the model specifications in consideration, the Probit model was selected, since its criteria values were minimal. As a result of the sequential selection of only statistically significant factors ($p < 0.05$), a Probit model was obtained, which included the following factors: a free member, infections of the urogenital system in the mother, the length of stay in a NPU and the absence of cerebral ischemia at birth. Table 4 shows the values of the Probit-model coefficients estimated by the maximum likelihood method, the corresponding marginal effect, standard errors, the p-level deviation of the null hypothesis that the corresponding coefficients are equal to zero.

Table 4. Estimates of the coefficients of the Probit model of the disability onset in very preterm infants

Influence factor	Coefficient β_j	Marginal effect dF/dx	Statistical error	p-level
Duration of stay in a NPU	-0,0530***	-0,007***	0,0158	0,0008
Absence of cerebral ischemia at birth	-1,7897***	-0,337***	0,5032	0,0004
Absence of infections of the urogenital system in the mother	-1,5391**	-0,125**	0,7654	0,0443
Free member	-1,2216***	0,8358***	0,0328	0,0000

Note: *, **, *** - the significance of coefficients at the levels of 10%, 5%, 1%.

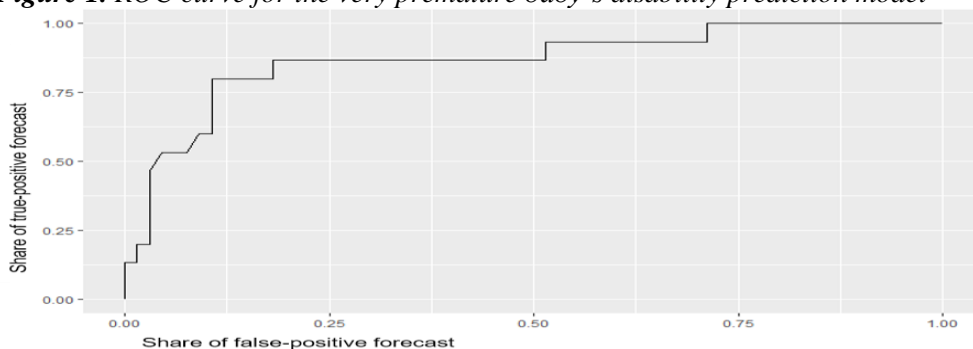
The conclusion about the validity of the constructed binary regression model was made based on the analysis of the McFadden determination coefficient ($R^2MF = 0.66$),

indicating a good predictive capacity of the resulting model. The conducted likelihood ratio test ($LR = 27.861$, $p = 0.001$) rejected the null hypothesis of non-effect of the predictors under consideration on the disability variable. The Hosmer-Lemeshow test ($HL = 1.21$, $p = 0.521$) confirmed the null hypothesis of consistency of the calculated and actual data.

At the second stage of developing a binary regression model, marginal effects were analyzed to better interpret the results obtained. As a result, it was shown that a twofold increase in the length of stay in a NPU compared to an average value reduces the likelihood of a child's disability in the future by 0.7%, the absence of urinary system infections in the mother reduces the risk of disability by 12.5%, and the absence of cerebral ischemia at birth reduces the risk of determining disability of premature babies in the pediatric period by 33.7%.

At the third stage, the metric of the obtained binary regression model validity was estimated based on the ROC analysis (Figure 1). As a result, the model predicting the likelihood of disability of premature babies in the pediatric period is a reliable classifier, since the area indicator under the ROC curve is close to 1 ($AUC = 0.8687$).

Figure 1. ROC curve for the very premature baby's disability prediction model



4. Conclusion

As a result of the analysis required for calculating the total economic burden and the survival rate of deeply premature babies, the effect of various factors on the death risk in the neonatal, postnatal, and pediatric periods was revealed. In general, the results obtained are consistent with the studies on the issue of the survival rate of children with VLBW and ELBW. These results could also be used to make a detailed calculation of the odds ratio for the survival of very premature babies at different ages. This calculation simplifies and refines the process of the evaluation of economic burden for different age cohorts of premature babies. Additionally, the risk factors for developing disability in premature babies were characterized. This allows to compute probabilities for premature babies with potential disabilities and to increase the accuracy of economic burden predictions.

The results of the study are extremely important for the evaluation of preventive measures in the postnatal and prenatal periods of newborns nursing. The obtained design models of survival and disability risks enable us to assess the amount of nursing care measures, the duration of care in a neonatal pathology unit, the rehabilitation of children with VLBW and ELBW in the long run, which ultimately will help estimate the overall economic burden associated with maintaining health and quality of life of premature babies.

References:

- Abolfotouh, M.A., Al Saif, S., Altwaijri, W.A., Al Rowaily, M.A. 2018. Prospective study of early and late outcomes of extremely low birthweight in Central Saudi Arabia. *BMC Pediatrics*, 18(1), Paper number 280.
- Bakhitova, R.Kh., Maksimenko, Z.V., Lakman, I.A. 2018. An approach to evaluation and planning of small premature infants nursing costs. *Economics and Management: Research and Practice Journal*, 5(143), 144-148.
- Breslow, N. 1974. Covariance Analysis of censored survival data. *Biometrics*, 30, 89-99.
- Cox, D.R. 1975. Partial Likelihood. *Biometrika*, 62, 269-276.
- Efron, B. 1977. The Efficiency of Cox's Likelihood Function for Censored Data. *Journal of American Statistical Association*, 72, 557-565.
- Gehan, E.A. 1965. A generalized Wilcoxon test for comparing arbitrarily singly censored samples. *Biometrika*, 52, 203-223.
- Gilbert, W.M. 2006. The cost of preterm birth: The low cost versus high value of tocolysis. *BJOG: An International Journal of Obstetrics and Gynaecology*, 113(3), 4-9.
- Han, R.H., McKinnon, A., Creve-Coeur, T.S., Baksh, B.S., Mathur, A.M., Smyser, C.D., Strahle, J.M., Olsen, M.A., Limbrick, D.D. 2018. Predictors of mortality for preterm infants with intraventricular hemorrhage: a population-based study. *Child's Nervous System*, 34(11), 2203-2213.
- Inoue, H., Ochiai, M., Yasuoka, K. 2017. Early Mortality and Morbidity in Infants with Birth Weight of 500 Grams or Less in Japan. *Journal of Pediatrics*, 190, 112-117.
- Johnston, K.M., Gooch, K., Korol, E., Vo, P., Eyawo, O., Bradt, P., Levy, A. 2014. The economic burden of prematurity in Canada. *BMC Pediatrics*, 14(1), No. 93.
- Lewit, E.M., Schuurmann Baker, L., Corman, H., Shiono, P.H. 1995. The Direct Cost of Low Birth Weight. *The Future of Children Low Birth Weight*, 5(1), 35-56, DOI: 10.2307/1602506.
- Li, Y., Wei, Q.F., Meng, D.H., Pan, X.N., Mo, Y., Yao, L.P., Jing, L.F., Zhao, D., Shen, K.Y., Xu, J. 2018. Treatment of the conditions and associated factors in Guangxi, China. *Pediatrics and Neonatology*, 59(3), 263-266.
- Muraskas, J., Parsi, K. 2008. The Cost of Saving the Tiniest Lives: NICUs versus Prevention. *Virtual Mentor*, 10(10), 655-658. DOI:10.1001/virtualmentor.2008.10.10.pfor1-0810.
- Santhakumaran, S., Statnikov, Y., Gray, D., Battersby, C., Ashby, D., Modi, N. 2018. Survival of very preterm infants admitted to neonatal care in England 2008-2014: time trends and regional variation. *Archives of disease in childhood. Fetal and neonatal edition*, 103(3), F208-F215.
- Schmitt, S.K., Sneed, L., Phibbs, C.S. 2006. Costs of Newborn Care in California: A Population-Based Study. *Pediatrics*, 117(1), 154-160. DOI: 10.1542/peds.2005-0484.

- Silva, L.V., De Araújo, L.B., De Oliveira Azevedo, V.M.G. 2018. Assessment of the neuropsychomotor development in the first year of life. *The Remata Brasileira de Terapia Intensiva*, 30(2), 174-180.
- Stephens, A.S., Lain, S.J., Roberts, C.L., Bowen, J.R., Nassar, N. 2016. Association of Gestational Age and Severe Neonatal Morbidity with Mortality in Early Childhood. *Pediatric and Perinatal Epidemiology*, 30(6), 583-593.
- Tongo, O.O., Orimadegun, A.E., Ajayi, S.O., Akinyinka, O.O. 2009. The economic burden of preterm/very low birth weight care in Nigeria. *Journal of Tropical Pediatrics*, 55(4), 262-264.