International Journal of Community Medicine and Public Health
Rosaiah K et al. Int J Community Med Public Health. 2020 Aug;7(8):3014-3019
http://www.ijcmph.com
pISSN 2394-6032 | eISSN 2394-6040

DOI: http://dx.doi.org/10.18203/2394-6040.ijcmph20203082

Original Research Article

An application of multiple logistic regression for identifying lipid profile changes towards assessing maternal and fetal outcomes

K. Rosaiah[1], Naga Saritha Kolli[2]*, N. S. Sanjeeva Rao[2]

[1]Department of Statistics, Acharya Nagarjuna University, Guntur, Andhra Pradesh, India
[2]Department of Community Medicine, NRI Medical College, Chinakakani, Guntur, Andhra Pradesh, India

Received: 20 June 2020
Revised: 08 July 2020
Accepted: 09 July 2020

***Correspondence:**
Dr. Naga Saritha Kolli,
E-mail: sariu.chandu@gmail.com

ABSTRACT

Background: The millennium development goals encourage governments to address and reduce various developmental issues, two of the important ones being maternal and child health. The one of the important causes of maternal mortality in India is pregnancy induced hypertension (PIH) and present study is to identify the relationship between disturbed lipid profile and preeclampsia its effect on fetal and maternal outcome.
Methods: This was a descriptive cross-sectional study done on data of maternal care and outcomes from the NRI General Hospital, Guntur district in the year 2013. Multiple logistic regression analysis is applied and results are adjusted to covariates maternal age and gravida.
Results: Systolic blood pressure, diastolic blood pressure of normal group (n=50) and PIH group (n=60) are 116.08±7.77, 76.08±4.93, and 165.66±16.8 105.5±14.07 respectively. Birth weights of infants in normotensives and PIH group are 2.85±0.33 and 1.93±0.659 respectively. Percentages of fetal and maternal complications in PIH group are 88.33% and 25%. Still births are present in 31.66% of PIH cases. Mean and SD of gestational age in weeks in normal and PIH groups are 37.92±1.94 and 34.36±3.44 respectively.
Conclusions: The model showed significant association between the selected independent variable, covariates and outcomes. The study demonstrates that multiple logistic regression may be applied to medical data in developing predictor models which are useful in clinical settings.

Keywords: Multiple logistic regression, Dichotomous outcome, Fetal complications, Hypertensive disorders

INTRODUCTION

Linear models assume that response variables and error terms are normally distributed. Least squares estimation procedure minimizes error around the line of best fit. But this method is not applicable in situations where the dependent variable (dv) is dichotomous and a mix of dichotomous and continuous independent variables (iv) are present. Generalized linear modeling introduced and developed by Nelder, Wedderburn, and Mccullough is the suitable technique for modeling non normal response and error terms. Logistic regression is the application of Generalized linear modeling technique. Underlying distribution is binary, predictors being both continuous and categorical. The link used is logit link.

Logistic regression predicts only two variables, for example probability that the person belongs to group with PIH (P=1) or group without PIH (P=0) on the basis of risk factors considered in the study. The objective here is to correctly predict the outcome.

Let P (x) represents the probability of an event that depends on n independent variables and (or) covariates.

$$P(x) = \frac{e^{\beta_0+\beta_1X_1+\beta_2X_2+....+\beta_nX_n}}{1+e^{\beta_0+\beta_1X_1+\beta_2X_2+....+\beta_nX_n}}$$

$$\text{logit}[P(x)] = \ln\left(\frac{P(x)}{1-P(x)}\right)$$

$$\text{Logit } (P(x)) = \beta_0+\beta_1X_1+\beta_2X_2+\beta_3X3+.....+\beta_nXn$$

In the above equation β_0 is intercept. If all covariates are set to zero, β_0 is the baseline event rate. $\beta_1,\beta_2, \beta_3,.....,\beta_n$ are regression coefficients(slopes).

Odds ratio= $e^{Z\beta i}=(e^{\beta i})^Z$, odds ratio of β_i represents the odds ratio for a change of size Z for that variable X_i, i=1,2,3,...,n.

$$\text{odds} = \frac{P(x)}{1-P(x)} = \frac{\text{probability of presence of characteristic}}{\text{probability of absence of characteristic}}$$

Logistic regression become one of the most frequently used procedure in the obstetrics and gynecology research.[1]Over the last two decades significant increase in the use of multivariate logistic regression was documented by Kenneth et al, Levy et al, Chin and Khan et al.[1-4] Application of standard logistic regression model to medical data was illustrated by Bender et al.[5] Applicability of logistic regression was enhanced with the advent of maximum likelihood estimation according to Paul et al.[6] Hypertensive disorders in pregnancy such as PIH, preeclampsia and eclampsia are considered a global burden. Though well documented information about incidence of hypertensive disorders is not available in India, national incidence of preeclampsia is reported to be 8-10% of pregnancies as per Gohil et al.[7] WHO facts sheet (2013) reveals that severe forms of preeclampsia and eclampsia are more common ranging from 4% to 18% in developing countries. A woman in a developing country is seven times likely to develop preeclampsia than a woman in a developed country.[8]

Islam et al documented that increased triglycerides levels, delayed triglycerides clearance and high blood pressure are proved to be the significant reasons for development of preeclampsia and eclampsia.[9] Preterm delivery and reduction in anthropometric measurements is associated with hypertensive disorders in pregnancy was proved by Oniyirwika et al.[10] Applied multiple logistic regression analysis was used by Yadav et al and identified some of the serum profile parameters as significant predictors of preeclampsia.[11] The ABCD study identified disturbed lipid profile during early pregnancy which is negatively associated with pregnancy outcomes.[12]

In this paper we develop multiple logistic regression models for maternal and fetal outcomes in pregnancy induced hypertension (PIH) with altered lipid profiles.

METHODS

To apply and develop a multiple logistic regression model, secondary data was obtained from the department of Obstetrics and Gynecology, NRI Medical College and General Hospital from January to June 2013. Information collected was age, height, weight, gestational age, systolic blood pressure (SBP), diastolic blood pressure (DBP), lipid profile done in first trimester (serum total cholesterol (SeTC), total triglycerides (TG), high density lipoprotein (HDL) and low-density lipoprotein (LDL), maternal and fetal complications. Lipid profiles were available because of an ongoing study in the department.

Sample size was 110 pregnant women who include 60 PIH cases (including gestational hypertensive cases, Preeclampsia cases and Eclampsia cases) and 50 normotensive pregnant women. Sample size was calculated by following the guidelines of Peduzzi et al Scott et al.[13,14] N= (10 k)/p where k is the co-variates in the study and p is the smallest proportion of positive or negative cases in the population. In the present study k=2 (covariates are age and gravida) and prevalence of hypertensive disorders is 18% in developing countries. Resulting calculations yield 111 and nearest rounded digit 110 was decided as the sample size.

Data was entered into Microsoft Excel and exported to Medcalc 14.10.2, an easy to use statistical software to develop multiple logistic regression models. The above software develops the following multiple logistic regression equation:

$$\text{Logit } (p) = \beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_3+.....+\beta_nXn$$

Overall fit of the estimated logistic regression equation was tested with the Chi square statistic which measures the difference in null model (-2*Ln (Lo)) and full model (-2*Ln (L)). P value less than 0.05 indicates that at least one of the independent variables is contributing to the prediction of the outcome variable.

The relative amount by which odds of the outcome increases or decreases when the value of the independent variable is increased by one unit is explained with the help of Odds ratio for the independent variable Xi (i=1,2,3,...,n). 95% confidence intervals were also calculated for the odds ratio.

Goodness of fit for the fitted regression model is tested with Hosmer and Lemeshow test. Results of the test yields Chi square value. A small Chi square value with large p value closer to 1 indicates a good logistic regression model fit.

Area under the curve (AUC) resulted from receiver operating characteristic (ROC) curve analysis gives power of the model's predicted values to discriminate between positive and negative cases. AUC values vary

from 0.5 (discriminating power not better than chance) to 1.0 (perfect discriminating power).

Z test of difference between two independent means and proportions are used to test the difference in averages and proportions of different parameters under study respectively.

RESULTS

Different continuous variables under study in the two groups are summarized in Table 1. There is a statistically significant difference between averages of all the parameters under study between normotensive and PIH groups except age at pregnancy.

Table 1: Basic data characteristics of subjects under study.

Variables	Normotensive (n=50)	PIH cases (n=60)	P value
Age in years	23.12±3.39	24.48±5.19	0.098
BMI (Kg/m^2)	24.22±1.63	27.54±2.71	<0.0001
SBP (mmHg)	116.08±7.77	165.66±16.8	<0.0001
DBP (mmHg)	76.08±4.93	105.5±14.07	<0.0001
SeTC (mg/dl)	179.04±14.82	241.31±33.48	<0.0001
TG (mg/dl)	136.94±15.94	213.3±54.04	<0.0001
HDL (mg/dl)	48.61±4.72	40.51±4.08	<0.0001
LDL (mg/dl)	115.35±11.74	157.9±33.96	<0.0001
Gestati-onal age (in weeks)	37.92±1.94	34.36±3.44	<0.0001
Birth weight (in kg)	2.85±0.33	1.93±0.659	<0.0001

Table 2: Pregnancy outcomes in both the groups under study.

Pregnancy outcome	Normotensive (n=50)	PIH cases (n=60)	P value
Fetal complications	13 (26)	53 (88.33)	<0.0001
Maternal complications	0	15 (25)	NA
Pre term deliveries	2 (4)	25 (41.66)	<0.0001
Low birth weight	2 (4)	45 (75)	
Still births, Neonatal deaths and intra uterine deaths	0	19 (31.66)	NA

Pregnancy outcomes between the groups are also summarized in Table 2. There is a statistically significant difference between proportions of complications between the two groups.

Four multiple logistic regression models were developed from the basic data under the following headings i.e. presence or absence of PIH, preterm delivery (PD), low birth weight (LBW), fetal deaths (FD)

Presence or absence of PIH

Presence of PIH is coded as "1" and absence as "0". Continuous iv's are SeTC, TG, HDL and LDL. Age (age <30=0 and age ≥30=1) and gravida (gravida <3=0 and gravida ≥3=1) are categorical covariates. The following is the fitted regression equation for the relation between PIH and above-mentioned iv's and covariates.

Logit (PIH)=-15.89+0.8923SeTC+1.2308TG+0.95HDL+ 1.03LDL-1.62 age -1.16 gravida.

Overall model fit looking at difference between null and full model yields a Chi square statistic of 93.78 which is statistically significant with p=0.000001 showing that at least one of the iv's are contributing to PIH.

The regression coefficients corresponding Odds ratio (OR), confidence interval (C.I) for Odds ratio and P-values for each iv and covariates are given in Table 3. Data from Table 3 shows that all the 4 iv's concerning serum lipid profile are significantly contributing to PIH and are independent risk factors for development of PIH.

Hosmer Lemeshow statistic test yields a Chi square value 2.035 with p value 0.898 showing a goodfit of the multiple logistic regression model.

The value of AUC 0.895 explains models significant discriminating power between PIH positive cases and negative cases.

Preterm delivery

Analysis of occurrence of preterm delivery in the presence of continuous iv's such as SeTC, TG, HDL, LDL, SBP and DBP and categorical covariates such as Age and Gravida was done. The following is the fitted regression equation for the relation between PD and above-mentioned iv's and covariates.

Logit (PD)=-16.55+0.43SeTC+0.658TG-0.023HDL+ 0.835LDL+2.578SBP+1.973DBP+0.877age-0.096 gravida.

Overall model fit looking at difference between null and full model yields a chi-square statistic of 42.108 which is statistically significant with p=0.00001 showing that at least one of the iv's are contributing to preterm delivery.

The regression coefficients corresponding OR, C.I for OR and p values for each iv and covariates are given in Table 4. Table 4 shows that the 4 serum lipid profile parameters, SBP and DBP are significantly contributing to preterm delivery and are independent risk factors for development of preterm delivery.

Hosmer Lemeshow statistic test yields a Chi square value 4.589 with P- value 0.695 showing a goodfit of the multiple logistic regression model for preterm delivery. The value of AUC 0.824 explains models significant discriminating power between preterm and full term delivery cases on the basis of given independent and covariates data.

Table 3: Regression coefficients of serum lipid profile.

Variables	β	P value	Adjusted OR	95% C.I of adjusted OR
SeTC	0.8923	0.032	2.441	(1.57, 3.79)
TG	1.2308	0.0001	3.42	(2.216, 5.269)
HDL	0.95	0.02	2.58	(1.674, 3.978)
LDL	1.03	0.001	2.8	(1.815, 4.315)
Age	-1.62	0.43	0.197	(0.128, 0.304)
Gravida	-1.16	0.55	0.313	(0.30, 0.848)

P values, OR and, 95% C.I of OR.

Table 4: Regression coefficients of different variables on PD.

Variables	β	P value	Adjusted OR	95% C.I of adjusted OR
SeTC	0.433	0.04	1.54	(0.33,7.04)
TG	0.6588	0.043	1.93	(0.42,8.828)
HDL	-0.023	0.567	0.977	(0.213, 4.473)
LDL	0.8352	0.04	2.305	(0.504,10.548)
SBP	2.578	0.00001	13.17	(2.87,60.279)
DBP	1.9734	0.0001	7.19	(1.57,32.92)
Age	0.877	0.04	2.4	(0.128,0.304)
Gravida	-0.096	0.88	0.908	(0.30,0.848)

P values, OR and, 95% C.I of OR.

Table 5: Regression coefficients of different variables impact on low birth weight.

Variables	β	P value	Adjusted OR	95% C.I of adjusted OR
SeTC	0.2973	0.078	1.322	(0.281,6.203)
TG	0.033	0.42	1.03	(0.219,4.83)
SBP	0.959	0.029	2.6	(0.554, 12.19)
DBP	0.561	0.05	1.75	(0.37,8.207)
Age	2.42	0.00001	11.24	(2.39,52.72)
Gravida	0.72	0.03	2.05	(0.436,9.61)

P values, OR and, 95% C.I of OR.

Table 6: Regression coefficients of different variables impact on fetal death.

Variables	β	P value	Adjusted OR	95% C.I of adjusted OR
SeTC	0.032	0.43	1.03	(0.13,8.09)
TG	0.058	0.38	1.05	(0.135,8.25)
SBP	1.298	0.0001	3.66	(0.465,28.36)
DBP	0.862	0.042	2.367	(0.301, 18.59)
Age	1.053	0.004	2.86	(0.364,22.46)
Gravida	0.438	0.049	1.54	(0.196,12.109)

P values, OR and, 95% C.I of OR.

Low birth weight

The fitted regression equation for the relation between LBW (<2.5 kg) and above-mentioned iv's and covariates is:

Logit (LBW)=-12.369+0.2793SeTC+0.033TG+0.959 SBP+0.561 DBP+2.42 age+0.72 gravida

Overall model fit yields a chi-square statistic of 56.73 which is statistically significant with p=0.00001. The regression coefficients corresponding OR, confidence interval (C.I) for OR and p values for each iv and covariates are given in table 5. Data shows that TG, SBP and DBP, age and gravida are significantly contributing to LBW and are independent risk factors for this outcome i.e. low birth weight babies.

Hosmer Lemeshow statistic test yields a Chi square value 9.844 with p value 0.2761 showing a moderate fit of the multiple logistic regression model for low birth weight with the above iv s and covariates. The value of AUC 0.744 explains models good discriminating power between low birth weight and normal birth weight cases.

Fetal death

For the relation between occurrence of fetal death and above mentioned iv's and covariates the fitted regression equation is:

Logit (FD)=-18.473+0.032SeTC+0.58TG+1.298 SBP +0.862 DBP +1.053 age+0.438gravida

Overall model fit looking at difference between null and full model yields a Chi-square statistic of 23.72 which is statistically significant with p value 0.0002. The regression coefficients corresponding OR, confidence interval (CI) for OR and p values for each iv and covariates are given in table 6 which shows that SBP and DBP, and age at pregnancy are significantly contributing to fetal death and are independent risk factors for fetal deaths.

Hosmer Lemeshow statistic test yields a Chi square value 12.55 with p value 0.3780 showing a moderate fit of the multiple logistic regression model for fetal deaths. The value of AUC 0.719 explains the model's good discriminating power between fetal deaths and fetal survival.

DISCUSSION

The present study looks at the relationship between serum lipid profile, PIH and the consequences in pregnancy outcome. Differences in serum lipid profile values, SBP and DBP, gestational age and birth weight between women with the presence of PIH and normotensive groups are significantly different. Results of the present study given in table-1 are correlated with the other studies.[7,15]

25% of cases in PIH group had maternal complications. 88.3% of the PIH group had fetal complications while 26% in the normal group had fetal complications. Difference in proportions of preterm deliveries, low birth weight and fetal deaths (including still births, neonatal deaths and intra uterine deaths) are significantly higher in the PIH group.

Looking at the occurrence of PIH, given high lipid profile values, the developed multiple logistic regression model showed a significant relationship. The OR for individual lipid profile parameters showed highly significant relationship between all of them with particular emphasis on triglycerides. Age and gravida are not contributing to the occurrence of PIH. By looking at C.I for OR of TG in Table 3, pregnant women with high triglycerides had a risk of developing PIH five times greater than in those with normal triglycerides

It is observed that preterm deliveries are associated with SeTC, TG and SBP and DBP, and age at pregnancy. Increase in SBP and DBP directly related to preterm deliveries. OR of SBP and DBP given in Table 5 are correlated with the findings of Villar et al and confirm that the risk of preterm births is three fold in women with hypertensive disorders than normal pregnant women.[16] HDL and LDL were not found to show any significant contribution towards preterm delivery. Gravida is not significantly related to preterm deliveries.

Relationship between SeTC, TG and low birth weight is not statistically significant. Whereas increased levels of SBP and DBP, age and gravida are significantly resulting in low birth weight babies. Occurrence of fetal death is associated with increased values of SBP, DBP and age at pregnancy. TC and TG are not significantly related to the occurrence of fetal death.

Overall fit of the four models was statistically significant. Multiple logistic regression models developed and Odds ratios in the present study are correlating with the results of earlier studies indicating that dyslipidaemia (with particular reference to triglicerides), presence of PIH and maternal age greater than 30 years are resulting in adverse pregnancy outcomes.[9-12]

Hosmer and Lemeshow test values and AUC values of all the four developed multiple logistic regression models showed that models are good fit to the given iv's data and the discriminating power is close to one.

Limitations of present study were the assumption of multicollinearity was not checked with any of the procedures, the process was left to the software. The sample size 110 was on the edge of minimum number of cases to consider i.e 100. But the results were on par with previous studies and the relationships developed in the

four multiple logistic regression models are biologically plausible.

CONCLUSION

Maternal mortality and poor fetal outcome in India is significantly high due to PIH leading to preeclampsia and eclampsia. It is necessary to identify and estimate reliable markers like serum lipid profiles (Cholesterol, TG, HDL and LDL) which can predict pre-eclampsia in pregnant women and also poor fetal outcomes like preterm births, low birth weights and fetal deaths. Taking into account the binary outcomes with both continuous and categorical predictors a logistic regression linear modeling technique is applied to develop models of good fit. The model showed significant association between the above independent variable, covariates and outcomes. The study demonstrates that Multiple logistic regression may be applied to medical data in developing predictor models which are useful in clinical settings.

ACKNOWLEDGEMENTS

The authors wish to thank the management of NRI Medical College and the Department of Obstetrics and Gynecology for their co-operation.

Funding: No funding sources
Conflict of interest: None declared
Ethical approval: Not required

REFERENCES

1. Ottenbachera KJ, Ottenbacherb HR, Toothc L, Ostird GV. A review of two journals found that articles using multivariable logistic regression frequently did not report commonly recommended assumptions. J Clin Epidemiol. 2004;57:1147-52.
2. Levy PS, Stolte K. Statistical methods in public health and epidemiology: a look at the recent past and projections for the next decade. Stat Methods Med Res. 2000;9:41-55.
3. Chin S. The rise and fall of logistic regression. Aust Epidemiol. 2001;8:7-10.
4. Khan KS, Chien PF, Dwarakanath LS. Logistic regression models in obstetrics and gynecology. literature. Obest Gynecol, 1999;93:10014-20.
5. R Bender, U Grouven. Ordinal Logistic Regression in medical research. J Royal Coll Physicians London. 1997;31(5):546-51.
6. Schmitz PIM. Developments in logistic regression methodology from 1970-1986. Chapter 1:13-31. Logistic Regression in Medical Decision Making and Epidemiology. 1986
7. Gohil JT, Patel PK, Gupta P. Estimation of Lipid Profile in Subjects of Preeclampsia. J Obstetr Gynaecol of India. 2011;61(4):399-403.
8. WHO Facts sheet on maternal mortality. 2013. Available at: http://www.who.int/mediacentre/factsheets/fs348/en/. Accessed on 3 January 2020.
9. Islam NAF, Chowdhury MAR, Kibria GM, Akhter S. Study of Serum Lipid Profile in Pre-Eclampsia and Eclampsia. Faridpur Med Coll J. 2010;5 (2).
10. Onyiruika AN, Onakewhor JU, Okolo AA. Effects of Hypertensive Disorders in Pregnancy on Preterm Delivery and Anthropometric Indices in the Resultant Newborn Infants. Ann Biomed Sci. 2004;3(1&2):12-22.
11. Singh U, Yadav S, Mehrotra, Natu SM, Kumari K, Yadav YS. Serum lipid profile in early pregnancy as a predictor of Preeclampsia. Int J Med Rev. 2013:1(2):56-62.
12. Vrijkotte TGM, Krukziener N, Hutten BA, Vollebregt KC, van Eijsden M, Twickler MB. Maternal Lipid Profile During Early Pregnancy and Pregnancy Complications and Outcomes: The ABCD Study. J Clin Endocrinol Metabol. 2012;97(11):3917-25.
13. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol. 1996;49(12):1373-9.
14. Scott LJ, Freese J. Regression models for categorical dependent variables using Stata. Stata press, 2006.
15. Evrüke IC, Demir SC, Ürünsak IF, Özgünen FT, Kadayıfçı O. Comparison of lipid profiles in normal and hypertensive pregnant women. Ann Saudi Med 2004;24(5):382-5.
16. Villar J, Carroli G, Wojdyla D, Abalos E, Giordeno D, Ba'aqeel, et al. Preeclampsia, gestational hypertension and intrauterine growth restriction, related or independent conditions?. Am J Obstetr Gynaecol. 2006;194(4):921-31.

Cite this article as: Rosaiah K, Kolli NS, Rao NSS. An application of multiple logistic regression for identifying lipid profile changes towards assessing maternal and fetal outcomes. Int J Community Med Public Health 2020;7:3014-9.