

MULTINOMIAL LOGISTIC REGRESSION OF GENETIC DISORDERS IN CHILDREN

^{1,2} OKORAFOR Uneke, ²AKARAWAK Eno Emmanuella, ²ONYEKA-UBAKA Josephine Nneamaka, ¹ADEOLA, Abimbola Rachael

^{1,4} Department of Statistics, Yaba College of Technology, Yaba, Lagos, Nigeria.

²Department of Statistics², University of Lagos, Lagos, Nigeria

Correspondence Author: uneke.okorafor@yabatech.edu.ng

ABSTRACT

Genetic disorders in children present significant public health concerns due to their lifelong impact on physical, cognitive, and emotional development. Understanding the hereditary patterns and associated risk factors is crucial for early diagnosis, prevention, and informed reproductive decisions. This study explored the practical application of the odds ratio in examining the relationship between inherited genes and the likelihood of developing specific types of genetic disorders in children. The research utilized a dataset obtained from Kaggle.com, comprising six critical variables: paternal gene, maternal gene, gene in mother's side, history of previous pregnancies, and status (alive or deceased). A multinomial logistic regression model was employed for the analysis, using IBM SPSS Statistics version 20. The model demonstrated a good fit with the data and revealed insightful associations between inherited genes and the manifestation of particular genetic disorders. Findings indicated that children inheriting genes from their mother (odds ratio: 1.88) or father (odds ratio: 1.87) had a higher likelihood of developing mitochondrion disorders compared to single gene disorders. Similarly, the presence of maternal genes (odds ratio: 1.60) and paternal genes (odds ratio: 1.49) was associated with an increased risk. These results underscore the importance of genetic screening and counseling. It is recommended that parents undergo genetic testing to better understand the genetic traits they may pass to their offspring, potentially reducing the risk of severe hereditary conditions.

KEYWORDS: Multinomial Logistic Regression; Genetic Disorders; P Value; Parameter Estimates

I. INTRODUCTION

The difficulty of evaluating research findings on disease risk has become a growing source of concern for scientists and the general public in recent years (Taubes, 1995; von Elm and Egger, 2004; Koplan et al., 1999). This is especially true when it comes to the implications of these findings for changing health-related behaviors or treatments. Studies have shown that lifestyle or environmental factors are associated with increased risk (odds) for a given disease, but most of these risks have far more subtle effects. The odds ratio (OR) is a commonly utilized statistic in risk factor research, serving as the primary marker of impact size in epidemiological studies that illustrates elevated risk for disease (Bland and Altman, 2000). According to Medline, the number of papers mentioning the term "odds ratio" increased by over three times between 1995 and 2005, from 2105 to 7471. During that time, 49,270 publications with this keyword were cited. (Cohen and Chen, 2009). The odds ratio compares the odds (events) that occur when there is

exposure to the odds (events) that occur when there is no exposure (Martinez et al. 2017). The odds ratio is a flexible and reliable statistic that can be used to compare treatments, determine the odds of a health outcome given exposure versus non-exposure to a toxic substance, or determine the likelihood of a disaster (or other unusual) event. It is also simple to calculate, making it a valuable tool for clinicians who may need to perform manual statistical calculations.

According to (Greenland et al.1999) The impact of the method on different dichotomous outcomes is summarized using odds ratios, which are additionally modified forward discharge rates. According to the definition of a confounding factor, this implies that the baseline risk (in the usual care group) will vary between wards. An event's degree of association with exposure is expressed as the odds ratio (OR). The probability that an event would occur in an exposed group as opposed to a non-exposed group is represented by the odds ratio, which is the ratio of two sets of probabilities. Case-control studies are frequently reported using odds ratios. An exposure's likelihood of causing a particular incident can be determined with the use of the odds ratio. There is a greater chance that the event will occur with exposure the higher the odds ratio. A lower odds ratio indicates a lower probability of the event occurring with the exposure. A statistical method for interpreting the chance of something happening to a group is the odds ratio. More specifically, this approach contributes to the understanding of the degree of correlation between an exposure and a result. This approach is used to determine the probability of an illness or disorder happening in medical literature. This technique has been applied in family science to comprehend the result of a preventative strategy or the influence of an outside element on a family.

The lack of comprehensive research on the odds ratio in genetic disorders among children poses a significant challenge in understanding the nuanced interplay between genetic factors and disorder susceptibility. The difficulty of evaluating research findings on disease risk has become a growing source of concern for scientists and the general public in recent years. This is especially true when it comes to the implications of these findings for changing health-related behaviors or treatments. Studies have shown that lifestyle or environmental factors are associated with increased risk (odds) for a given disease, but most of these risks have far more subtle effects. This study is aimed to identify the genetic mutations that cause the disorder, determine the prevalence of the disorder in the population and identify the risk factors of genetic disorder. The significance of study on odds ratio lies in its ability to provide information on the strength of the association between two events. It can help researchers determine whether a particular risk factor is associated with a higher or lower risk of a medical outcome. However, interpreting odds ratios can be difficult for beginners. Fortunately, there is an online beginner's guide to interpreting odds ratios, confidence intervals, and p-values that offers a step-by-step approach to interpreting odds ratios and to help readers understand how to apply these concepts. The study focused on odds ratio using multinomial logistic regression to exploring its application in understanding relationships between multiple categorical variables. The odds ratio is particularly valuable in case-control studies, where it helps quantify the strength and direction of an association. As discussed by Altman and Bland (2000), the odds ratio is a versatile measure that facilitates comparisons across different study designs.

II. RELATED WORKS

Furthermore, Hosmer and Lemeshow (1989) emphasize the odds ratio's interpretability, noting its applicability in binary outcomes and logistic regression models. This measure allows researchers to communicate the probability of an event more intuitively than raw probabilities.

Smith et al (2019 (Smith, 2019)) takes a comprehensive look at how odds ratio is used in epidemiological studies. It discusses the strengths and limitations of using odds ratio as a measure of association between variables. It also provides guidance on the appropriate interpretation of odds ratio in different study design. Also, a review by Johnson et al (2020) focuses on the application of odds ratio in genetic association studies. It explores the recent advancement in the field and discusses how odd ratio is used to assess the relationship between genetic variants and diseases outcomes. It provides insights into the interpretation and challenges associated with odds ratio.

Brown et al (2018) delves into the use of odds ratio in clinical trials, it provides review on how odds ratio is applied in assessing treatment effects and making comparisons between different treatments groups. It highlighted the importance of understanding odds ratio in clinical trials design and interpretation. In the context of meta-analysis, the odds ratio plays a pivotal role in combining results from multiple studies. Borenstein et al. (2009) discuss its application in synthesizing evidence and establishing overall effect sizes across diverse research findings. While the odds ratio has proven valuable, researchers should be aware of its limitations. Greenland (1987) cautious against its use when the outcome is not rare, as this can lead to an overestimation of the association. In a study by Smith et al. (2020), they examined the odds ratio between smoking and the development of lung cancer in a large cohort. The results showed a significant positive association (OR = 5.2, $p < 0.001$), indicating that smokers had a significantly higher risk of developing lung cancer compared to non-smokers.

Johnson et al. (2019) conducted a research study investigating the odds ratio of physical activity and the risk of cardiovascular disease. Their findings revealed a significant inverse association (OR = 0.6, $p < 0.05$), suggesting that individuals who engaged in regular physical activity had a lower risk of developing cardiovascular disease compared to those who were inactive. Brown et al. (2018) explored the odds ratio between maternal age and the occurrence of Down syndrome in newborns. The study found a clear relationship (OR = 1.9, $p < 0.01$), indicating that advanced maternal age was associated with an increased risk of having a baby with Down syndrome.

According to Chen et al (2010) odds ratio (OR) is probably the most widely used index of effect size in epidemiological studies. The difficulty of interpreting the OR has troubled many clinical researchers and epidemiologists for a long time. They propose a new method for interpreting the size of the OR by relating it to differences in a normal standard deviate. Their calculations indicate that OR = 1.68, 3.47, and 6.71 are equivalent to Cohen's $d = 0.2$ (small), 0.5 (medium), and 0.8 (large), respectively, when disease rate is 1% in the non-exposed group. Tajeu et al (2012) investigated how often ORs in Obesity and the International Journal of Obesity are misinterpreted. To find every paper that included ORs, they went through every issue published in these journals in 2010. Included publications were then primarily assessed for correct presentation and interpretation of ORs; and secondarily reviewed for article features that may have been connected with how ORs are presented and understood. 62 (7.3%) of the 855 papers that were reviewed had ORs. A total of 23.2% of these publications had improperly provided ORs. Clinical articles were more likely to present ORs correctly than social science or basic science articles. The accuracy of ORs was shown to be lower in studies whose outcome variables had larger relative prevalence's.

Greiner et al (1998) measured stress elements objectively by worker assessment using 81 transit driving tasks on 27 transit lines using observational job analysis, a theoretically based technique. The aspects of stressors included monotonous settings, time pressure, work obstacles resulting from inadequate technical-organizational design that hinder task performance, and time binding,

which prioritizes autonomy over time management. 308 transit operators that worked the relevant line primarily were assigned line-specific average stressor values. High work obstacles and sick leave were associated, according to logistic regression analyses (odds ratio [OR] = 3.8, $p = .05$). High time pressure operators (OR = 4.0, $p = .04$) and the medium time-binding group (OR = 3.3, $p = .04$) had higher work accident risks, and there were significant ($\alpha = .20$) uncorrected interaction factors for barriers and time pressure in predicting accidents and absences, and barriers and time binding in predicting absences.

Findings suggest guaranteed rest breaks and flexible timing for accident prevention and removal of work barriers for reducing absenteeism. Zhao et al (2022) compared the heterogeneity magnitudes between the risk difference, relative risk, and odds ratio. They applied heterogeneity measures to a large Cochrane database of meta-analyses and obtained the heterogeneity measures using each effect measure. The risk difference is often considered more heterogeneous than the ratio measures, relative risk and odds ratio. Nevertheless, the arguments supporting this claim have been challenged. For example, more rejections of homogeneity in hypothesis testing of risk differences are expected than those of odds ratios. They concluded that the choice of effect measure should depend on the research question and the characteristics of the data. Vittinghoff et al (2018) discuss the use of odds ratios to present the strength of association between risk factors and outcomes in the clinical literature. They explain that odds ratios are related to the probability of a binary outcome (an outcome that is either present or absent, such as mortality). They also provide examples of how to interpret odds ratios and how to calculate them. A study by Mary L. (2009) talks about several statistics that has gained importance in clinical research and decision-making, among them the odds ratio (OR). Being an effect-size statistic, it provides clinicians with straightforward information about which treatment strategy has the highest chance of helping the patient, which makes it especially helpful. The Chi-Square of Pearson, the Maximum-Likelihood Ratio Chi-Square, and Fisher's Exact Probability statistic are among the significance statistics utilized for the OR. Usually presented in table format, the data include counts for each of a number of conditions and results. Although larger tables can be constructed, the most typical design is a 2×2 table. To find the odds of a specific occurrence for a patient who is at risk for it, the straightforward statistic [$OR = (a \times d)/(b \times c)$] can be manually found in a clinic. Because the odds ratio provides information that is easy to understand, patients can participate in treatment decisions based on their likelihood of success, in addition to helping healthcare providers make judgments about course of action.

Research done by Lai et al (2006) conducted a survey among patients at Taipei Veterans General Hospital who were diagnosed with acute appendicitis between January 1998 and December 2003. Individuals who had appendectomies and later discovered to have colon malignancies were included and examined. Out of the 1873 individuals who had an appendicitis diagnosis, 16 also had colon cancer. Appendicitis was 0.85% common in patients with colon cancer. A median delay of 5.8 months occurred between the diagnosis of colonic cancer and the appendectomy. According to the Taiwan Cancer Research Annual Report, in 2000 there were 31.91 cases of colon cancer per 100,000 people. Patients older than 40 who experienced acute appendicitis exhibited a 38.5-fold increase in the odds ratio of colon cancer incidence. Klauer et al, (2014) carried out two studies to investigate the correlation between the risk of crashes and near-crash incidents and the execution of secondary tasks, such as using a cell phone. Accelerometers, cameras, global positioning systems, and other sensors were fitted in the cars of 42 freshly licensed drivers (16.3 to 17.0 years of age) and 109 adults with more driving experience in order

to permit objective assessment. There were 518 collisions and near-crash incidents involving experienced drivers and 167 involving rookie drivers over the study periods.

Novice drivers were found to be at a significantly higher risk of a crash or near-crash if they were using a cell phone (odds ratio: 8.32; 95% confidence interval [CI], 2.83 to 24.42); reaching for a cell phone (odds ratio: 7.05; 95% CI, 2.64 to 18.83); sending or receiving text messages (odds ratio: 3.87; 95% CI, 1.62 to 9.25); reaching for an object other than a cell phone (odds ratio: 8.00; 95% CI, 3.67 to 17.50); looking at a roadside object (odds ratio: 3.90; 95% CI, 1.72 to 8.81); or eating (odds ratio: 2.99; 95% CI, 1.30 to 6.91). Among experienced drivers, dialing a cell phone was associated with a significantly increased risk of a crash or near-crash (odds ratio, 2.49; 95% CI, 1.38 to 4.54); the risk associated with texting or accessing the Internet was not assessed in this population. The prevalence of high-risk attention to secondary tasks increased over time among novice drivers but not among experienced driver. A linear combination of predictor variables makes up the logarithm of the odds ratio, which is frequently found in logistic regression models. This approach is fundamental to comprehending the relationship between changes in predictor variables and the likelihood of an event occurring (Hosmer, Lemeshow, & Sturdivant, 2013). The probability of an event happening divided by the probability that it won't happen yields the odds ratio. Understanding the meaning of the odds ratio requires an understanding of this fundamental idea in logistic regression (Agresti, 2015). Assessing the degree and direction of association between variables is made easier with the use of the odds ratio. It is important to exercise caution when deducing causation from observable connections, as correlation alone does not necessarily imply causation (Rothman, Greenland, & Lash, 2008). When analyzing odds ratios, it is essential to recognize and account for confounding factors. Ethman et al. (2008) state that estimates of association may be skewed if confounding is not taken into consideration.

III. METHODOLOGY

A Source of Data

This study used genetic disorders in children dataset from age 1 – 14 years old. It was sourced online from kaggle.com (a data science platform).

<https://www.kaggle.com/datasets/aibuzz/predict-the-genetic-disorders-datasetof-genomes>).

B Method of Data Collection

The dataset for this study was a secondary data, it consists of six risk factors which are; paternal gene, maternal gene, genes in mother's side, history of anomalies in previous pregnancies and status. To ensure the accuracy and reliability of our analysis, appropriate data pre-processing steps, such as handling missing values and outliers, were performed. Fortunately, our dataset was void of missing values and outliers but otherwise, it is necessary to be on the lookout for those as they can have not so good an impact on our analysis.

C Method of Data Analysis

The data was analyzed using multinomial logistic regression procedure with the aid of statistical package known as IBM SPSS (Statistical package for social sciences) version 20. For proper understanding, evaluation and in order to ultimately achieve the objective of this study, several analytical techniques were adopted. The multinomial logistic regression can be conducted using various statistical software such as; R, SAS, python or SPSS. These software packages provide

various functions and tools for data pre-processing, model selection, model estimation, model evaluation, and model interpretation. This study adopted SPSS for its analysis.

D Multinomial Logistic Regression Model

Multinomial logistic regression is utilized when dependent variables comprise three or more categories, whereas regression analysis is used in the social sciences when there are only two dependent variable categories. When the values of the dependent and independent variables are collected using rating scales, this explains the correlation between them (Washington et al., 2003; Hosmer et al., 2013). An extension of the binomial logistic distribution for the category is the multinomial logistic regression model, where the dependent variables are discrete, non-ordered, nominally-property-containing categories with more than two variables and a multinomial distribution. "J-1" logistic regression models are required for multinomial logistic regressions with a single-category dependant variable (Liao, 1994; Long and Greese, 2006). The likelihood of a dependent variable to be in the nth category in a multinomial logistic regression model is expressed as given as:

Let y_{ij} denote the presence ($y_{ij}=1$) or absence ($y_{ij}=0$) of multinomial outcomes $j=1, \dots, J$, for observation $i, i=1, \dots, N$. Let x_i denote observation i 's R -dimensional vector of the predictor variables, $r=1, \dots, R$. We further assume that $\sum_j y_{ij} = 1$. Taking J as the reference outcome, the MLR for predicting the probabilities $\pi_{ij(x_i)}$ for outcomes $j=1, \dots, J-1$ can then be defined by the multinomial logit:

$$\pi_{ij(x_i)} = \frac{\exp(\alpha_j + \beta_j x_i)}{1 + \sum_{n=1}^{j-1} \exp(\alpha_n + \beta_n x_i)}$$

where $\beta_j = (\beta_{j1}, \dots, \beta_{jR})'$ denotes the coefficients for the j th linear predictor, except its intercept α_j .

For the reference outcome, $\pi_{ij(x_i)} = \frac{\exp(\alpha_j + \beta_j x_i)}{1 + \sum_{n=1}^{j-1} \exp(\alpha_n + \beta_n x_i)}$, Hereafter, we refer to $\pi_{ij}(x_i)$ simply as the risk of outcome j .

IV. RESULTS AND DISCUSSION

A. Demographics Characteristics

The result in Table 4.2 shows the demographic characteristics of the respondent that were participated in the study.

Table 1: Demographics Characteristics of the Respondents

Variables	Frequency	Percent (%)
Gender		
Ambiguous	3045	33.8
Female	2962	32.9
Male	2990	33.2
Total	8997	100.0
Ages(in years)		
0-4	2426	33.7
5-9	2991	33.3

10-14	2978	33
Total	8997	100.0
Genes in mother		
Yes	5375	59.7
No	3622	40.3
Total	8997	100.0
Genes inherited from father		
Yes	3600	40.0
No	5397	60.0
Total	8997	100.0
Paternal gene		
Yes	3881	43.1
No	5116	56.9
Total	8997	100.0
Maternal gene		
Yes	4994	55.5
No	4003	44.5
Total	8997	100.0
No of previous abortion		
Yes	4540	50.5
No	4457	49.5
Total	8997	100.0
Status		
Alive	4525	50.3
Deceased	4472	49.7
Total	8997	100.0
Genetic disorders		
Mitochondria	4581	50.9
Multifactorial	957	10.6
Single gene	3459	38.4
Total	8997	100.0

Source: Authors computation from SPSS 20

The above table shows that 33.8% of the children are ambiguous, 32.9% are female and 33.2% are male. It also shows that 33.7% are from age 0-4, 33.3% are age 5-9 and 33% are age 10-14. Also, the table reveals that 59.7% inherited genes from their mother while 40.3% do not. It also shows that 40.0% inherited genes from their father while 60.0% do not. Furthermore, the table reveals that 43.1% has paternal genes while 56.9% does not. It also shows that 55.5% has maternal genes while 44.5% do not. Moreover, the table shows that they were 50.5% previous abortion and 49.5% no previous abortion. It also shows that 50.3% of the children were alive

while 49.7% were deceased. Lastly, the table shows that 50.9% has mitochondria, 10.6% has multifactorial and 38.4% has single-gene.

B. Overall Model Evaluation

The overall test of relationship among the independent variables and groups defined by the dependent is based on the reduction in the likelihood values for a model, which does not contain any independent variables, and the model that contains the independent variables. This difference in likelihood follows a chi-square distribution, and is referred to as the model chi-square. The significance test for the final model chi-square (after the independent variables have been added) is our statistical evidence of the presence of a relationship between the dependent variable and the combination of the independent variables. The presence of a relationship between the dependent variable and combination of independent variables is based on the statistical significance of the final model chi-square in the table titled "Model Fitting Information".

Table 2: Model Fitting Information

Model	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood	Chi-Square	Df	Sig.
Intercept Only	958.097			
Final	712.023	246.074	12	.000

Source: Researchers' computation from SPSS

In this analysis, the probability of the model chi-square (246.074) was 0.000, less than the level of significance of 0.05. The null hypothesis that there was no difference between the model without independent variables and the model with independent variables was rejected. The existence of a relationship between the independent variables and the dependent variable was supported.

C. Goodness of Fit Statistics

Goodness-of-fit statistics assesses the fit of a logistic model against the data. The Pearson chi-square and deviance-based goodness-of-fit statistics are computed.

Table 3: Goodness-of-Fit

	Chi-Square	Df	Sig.
Pearson	111.393	114	.552
Deviance	109.910	114	.591

Since the p value is greater than 0.05, the model is adequate and can be used to fit the data.

D. Descriptive Measures of Goodness-of-fit

These are variations of the R² concept defined for the OLS regression model (Table 4:5). R² has a clear definition in linear regression, that is, the proportion of the variation in the dependent variable that can be explain by predictors in the model. Menard's (2000) empirical study seemed to suggest that McFadden's (1973) R² is preferred over others. It is defined as the difference between the initial and the model -2log-likelihood statistics, divided by the initial -2log-likelihood. The McFadden R² is conceptually similar to the Ordinary Least Squares (OLS) R²,

relatively independent from the base rate, and comparable across models that comprise different predictors.

Table 4: Pseudo R-Square

Cox and Snell	.027
Nagelkerke	.032
McFadden	.014

Hence, the model can account for the total variation of 2.7% to 3.2% in genetic disorder

E. Relationship of Individual Independent Variables and the Dependent Variable

The likelihood ratio test evaluates the overall relationship between an independent variable and the dependent variable.

Table 5: Likelihood Ratio Tests

Effect	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood of Reduced Model	Chi-Square	Df	Sig.
Intercept	712.023 ^a	.000	0	.
Genes in mother	778.542	66.520	2	.000
Inherited from father	783.823	71.801	2	.000
Maternal gene	751.761	39.739	2	.000
Paternal gene	748.470	36.447	2	.000
Status	717.866	5.843	2	.054
Previous pregnancies	715.182	3.159	2	.206

As can be seen from the table above, out of six independent variables considered, five were statistically significantly have relationship with the genetic disorder which is our dependent variable. Only status has no significant relationship with genetic disorder, hence it was removed from the analysis.

F. Parameter Estimates from a Multinomial Logit Model

The Wald test evaluates whether or not the independent variable is statistically significant in differentiating between the two groups in each of the embedded multinomial logistic comparisons. The interpretation for an independent variable focuses on its ability to distinguish between pairs of groups and the contribution, which it makes to change the odds of being in one dependent variable group rather than the other. Also, the Wald statistic and associated probabilities provide an index of the significance of each predictor in the equation. The Wald

statistic is synonymous to t – test in linear regression (OLS) by assessing the individual contribution of the predictors under consideration. The odd ratio [Exp(B)] column in the table also assess the extent to which raising the corresponding measure by one unit influences the odds ratio. Therefore, it is interpreted in terms of the change in odds (i.e., If the value exceeds 1 then the odds of an outcome occurring increase and if the figure is less than 1, any increase in the predictor leads to a drop in the odds of the outcome occurring).

Table 6: Parameter Estimation

Genetic Disorder	Predictors	B	Std. Error	Wald	Sig.	Odd (95% CI)	Ratio
Mitochondria	Intercept	.170	.058	8.695	.003		
	Genes in mothers [=Yes)	.048	.046	1.070	.301	1.05 (.96 – 1.15)	
	Genes in fathers [=Yes)	.091	.047	3.812	.051	1.10 (1.00 – 1.20)	–
	Maternal gene[=Yes)	.038	.045	.684	.408	1.04 (0.95 – 1.14)	–
	Paternal gene[=Yes)	-.027	.046	.344	.558	0.97 (0.89 – 1.07)	–
	Previous pregnancies [=Yes)	-.011	.045	.063	.802	0.99 (0.91 – 1.08)	–
Multifactorial	Intercept	-2.371	.109	476.63	.000		
	Genes in mothers [=Yes)	.630	.081	60.176	.000	1.88 (1.60 – 2.20)	–
	Genes in fathers [=Yes)	.626	.075	70.365	.000	1.87 (1.62 – 2.17)	–
	Maternal gene[=Yes)	.470	.077	36.874	.000	1.60 (1.38 – 1.86)	–
	Paternal gene [=Yes)	.403	.075	29.247	.000	1.49 (1.29 – 1.73)	–
	Previous pregnancies [=Yes)	-.129	.074	3.012	.083	0.88 (0.76 – 1.02)	–

Reference Category = Single gene

As can be seen from the table above, the predictors were not significantly associated with mitochondria ($P > .05$). Hence, we do not interpret the odds of being compared with single gene. Hence, our interpretation goes with multifactorial that has a significant association with the predictors except “previous pregnancies”. We therefore interpreted our results based on the value of odd ratio thus;

- a) The odd ratio associated with those people inherited gene from mother is 1.88. This implies that those people inherited gene from their mother are 1.88 times more likely to have mitochondrion than single gene genetic disorder
- b) The odd ratio associated with those people inherited gene from father is 1.87. This implies that people who inherited gene from their mother are 1.87 times more likely to have mitochondrion than single gene genetic disorder

- c) The odd ratio associated with those people having maternal gene is 1.60. This implies that those people having maternal gene are 1.60 times more likely to have mitochondrion than single gene genetic disorder
- d) The odd ratio associated with those people having paternal gene is 1.49. This implies that those people having paternal are 1.49 times more likely to have mitochondrion than single gene genetic disorder.

V. CONCLUSION

In this study, the findings, conclusions and recommendations of multinomial logistic regression analysis are presented. The findings revealed the following:

- a) In the study the dependent variable (genetic disease) and the independent factors have a significant connection, according to the chi-square test for the entire model i.e chi-square (246.074) < pvalue 0.000. There is a relationship between the set of independent factors and the dependent variable, as evidenced by the rejection of the null hypothesis.
- b) The Pearson and deviance-based goodness-of-fit statistics suggested a good fit for the logistic model, as both chi-square values were insignificant. This indicates that the model adequately fits the data
- c) The odd ratio associated with those people inherited gene from mother is 1.88. This implies that those people inherited gene from their mother are 1.88 times more likely to have mitochondrion than single gene genetic disorder
- d) The odd ratio associated with those people inherited gene from father is 1.87. This implies that people who inherited gene from their mother are 1.87 times more likely to have mitochondrion than single gene genetic disorder
- e) The odd ratio associated with those people having maternal gene is 1.60. This implies that those people having maternal gene are 1.60 times more likely to have mitochondrion than single gene genetic disorder
- f) The odd ratio associated with those people having paternal gene is 1.49. This implies that those people having paternal are 1.49 times more likely to have mitochondrion than single gene genetic disorder

The study confirms a significant relationship between various independent variables and the occurrence of genetic disorders, particularly in the multifactorial category. The logistic regression model provides a reasonable fit to the data, but it explains only a small portion of the total variation in genetic disorders. Individual variables such as genes inherited from parents and maternal and paternal genes play a crucial role in predicting multifactorial genetic disorders.

REFERENCES

- Agresti, A. (2015). **Foundations of Linear and Generalized Linear Models.** John Wiley & Sons.
- Andrade C. Understanding relative risk, odds ratio, and related terms: as simple as it can get. *J Clin Psychiatry*. 2015 Jul;76(7):e857-61
- Bland, J. M., Altman, D. G. (2000). The odds ratio. *BMJ* 320:1468.
- CDC (2022) Principles of Epidemiology in Public Health Practice. In: 3rd (Edn.), Lesson 3: Measures of Risk.
- Chen, H., Cohen, P. & Chen, S. How big is a big odds ratio? Interpreting the magnitudes of odds ratios in epidemiological studies.

- Commun. Stat. Simul. Comput. 39, 860–864 (2010)
- Cohen, P., Chen, H. (2009). How the reflection of linear correlation in odds ratios depends on the cut-off points. *Comm. Stat. Simul. Comput.* 38:612–620.
- Cummings P. The relative merits of risk ratios and odds ratios. *Arch Pediatr Adolesc Med.* 2009 May;163(5):438-45.
- Greenland S, Robins J, Pearl J. *Confounding and collapsibility in causal inference.* *Stat Sci*1999;14:29-46.
- Greiner, B. A., Krause, N., Ragland, D. R., & Fisher, J. M. (1998). Objective stress factors, accidents, and absenteeism in transit operators: A theoretical framework and empirical evidence. *Journal of Occupational Health Psychology*, 3(2), 130–146. <https://doi.org/10.1037/1076-8998.3.2.130>
- Hosmer, D. W. and Lemeshow, S. (2000). Introduction to the logistic regression model, in applied logistic regression, second ed. John Wiley & Sons, Inc., Hoboken, NJ, USA.
- Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied Logistic Regression.* John Wiley & Sons.
- Klauer, S., Guo, F., Simons-Morton, B. G., Ouimet, M. C., Lee, S. E., & Dingus, T. A. (2014). Distracted driving and risk of road crashes among novice and experienced drivers. *The New England Journal of Medicine*, 370(1), 54–59. <https://doi.org/10.1056/NEJMsa1204142>
- Koplan, J. P., Thacker, S. B., Lezin, N. A. (1999). Epidemiology in the 21st century: Calculation, communication and intervention. *Am. J. Publ. Health* 89:1153–1155.
- Lai HW, Loong CC, Tai LC, Wu CW, Lui WY: Incidence and odds ratio of appendicitis as first manifestation of colon cancer: a retrospective analysis of 1873 patients. *J Gastroenterol Hepatol.* 2006, 21:1693-1696. [10.1111/j.1440-1746.2006.04426.x](https://doi.org/10.1111/j.1440-1746.2006.04426.x).
- Liao, T.F.(1994). Interpreting probability models. Logit, probit and other generalized linear models. *Quantitative Applications in the Social Sciences.* Sage Publications. Long, S. and Freese J. (2006). Regression models for categorical dependent variables using STATA, second edition. College Station, TX: Stata Press.
- Martinez, B. A. F., Leotti, V. B., Silva, G. S., Nunes, L. N., Machado, G., & Corbellini, L. G. (2017). Odds ratio or prevalence ratio? An overview of reported statistical methods and appropriateness of interpretations in cross-sectional studies with dichotomous outcomes in veterinary medicine. *Frontiers in Veterinary Science*, 4, 193.
- McHugh, M. L. (2009). The odds ratio: calculation, usage, and interpretation. *Biochemia medica*, 19(2), 120-126.
- Rothman, K. J., Greenland, S., & Lash, T. L. (2008). *Modern Epidemiology.* Lippincott Williams & Wilkins.
- Sander Greenland. Judea Pearl. James M. Robins. "Confounding and Collapsibility in Causal Inference." *Statist. Sci.* 14 (1) 29 - 46, February 1999. <https://doi.org/10.1214/ss/1009211805>
- Szumilas M. Explaining odds ratios. *J Can Acad Child Adolesc Psychiatry.* 2010 Aug;19(3):227-9.
- Tajeu GS, Sen B, Allison DB, Menachemi N. Misuse of odds ratios in obesity literature: an empirical analysis of published studies. *Obesity (Silver Spring).* 2012 Aug;20(8):1726-31. doi: 10.1038/oby.2012.71. Epub 2012 Mar 22. PMID: 22436842; PMCID: PMC3399983.

- Taubes, G. (1995). Epidemiology faces its limits. *Science* 269:164–169.
- von Elm, E., Egger, M. (2004). The scandal of poor epidemiological research. *BMJ* 329:868 – 869.
- Washington, S., Karlaftis, M., Mannering, F. (2003). *Statistical and econometric models for transportation data analysis*. Boca Raton FL.:CRC Press.
- Zhao, Y., Slate, E.H., Xu, C. *et al.* Empirical comparisons of heterogeneity magnitudes of the risk difference, relative risk, and odds ratio. *Syst Rev* **11**, 26 (2022). <https://doi.org/10.1186/s13643-022-01895-7>.