

Statistical Modelling of Child Abuse Incidence in Malaysia: A Count Data Regression Approach (2016-2021)

Nurul Aina Baherunzaman¹, Siti Noor Asyikin Mohd Razali^{1*}

¹ Department of Mathematics and Statistics, Faculty of Applied Sciences and Technology, UTHM Kampus Cawangan Pagoh, Hab Pendidikan Tinggi Pagoh, KM 1, Jalan Panchor, 84600 Pagoh, Muar, Johor, MALAYSIA.

*Corresponding Author: asyikinr@uthm.edu.my
DOI: <https://doi.org/10.30880/ekst.2026.06.01.048>

Article Info

Received: 3 January 2026
Accepted: 20 January 2026
Available online: 9 July 2026

Keywords

Child Abuse, Count Data Regression, Overdispersion, Malaysia

Abstract

Child abuse continues to be one of the fundamental social and public health issue in Malaysia, with rising reported cases despite protection under the law. This study analyses national-level secondary data at the national level from Department of Statistics Malaysia (DOSM) and the Social Welfare Department (JKM) to examine the occurrence of child abuse between year 2016 and 2021. The research is guided by three key objectives, firstly to identify the exposure factors of child abuse using Poisson Regression Modelling, then to determine the dependency between ethnicity and gender in reported abuse cases using a Chi-Square Independence Test and lastly to model child abuse and its contributing factors using both Poisson and Negative Binomial Regression. The six socioeconomic exposure variables examined were: poverty rate, divorce, early marriage, drug addiction, violent crime, and depression. In order to model count data, the study used a quantitative technique that included descriptive analysis, Chi-Square Independence Test, Poisson Regression, and Negative Binomial Regression. The Chi-Square Test result show a statistically correlation between gender and ethnicity in reported cases of child abuse ($p < 0.001$). Overdispersion (deviance/df=368.7) disregarded model assumptions, even though Poisson Regression originally found several significant predictors. As a result, the Negative Binomial Regression model was chosen since it had a lower AIC (28.8) and BIC (288.0). Depression was shown to be the sole statistically significant predictor of all variables ($p = 0.093$), with a positive coefficient, indicating that higher frequency of child maltreatment. This study supports Negative Binomial Regression as a suitable technique for overdispersed count data in social science research and highlights the significance of mental health in child protection policies.

1. Introduction

Child abuse represents a serious global and national public health concern, with an estimated one billion children affected worldwide according to the World Health Organization In 2022. In Malaysia specifically, the severity of this issue is reflected in rising statistics, as cases of children requiring care and protection increased from 4,982 in 2016 to 6,382 in 2019, based on data from the Department of Statistics Malaysia (DOSM). This form of

victimization encompasses multiple dimensions, including physical, sexual, and emotional abuse, alongside neglect and exploitation, as defined by the Malaysian Ministry of Health in 2023.

Malaysia's position in international comparisons further underscores the ongoing challenge. The country ranked 23rd out of 60 nations in 2022. Out of the Shadows Index which assesses how nations address child abuse, placing behind regional neighbours Indonesia and Thailand, as reported by the Malay Mail in 2022. This ranking indicates a pressing need for enhanced and evidence-based interventions. Currently, however, the existing body of local research is limited in its capacity to inform broad policy, as prior Malaysian studies have been predominantly qualitative and small-scale [1].

These earlier studies, such as those by [2] and [3] provide valuable initial insights but lack quantitative, generalizable evidence. Consequently, a significant research gap exists, highlighting the necessity for advanced statistical modelling of exposure factors associated with child abuse using comprehensive national data. Addressing this gap is critical for developing targeted prevention strategies and allocating resources effectively.

To bridge this divide, the present study proposes a quantitative analysis with three clear objectives. The first objective is to identify and analyse key exposure factors linked to child abuse cases using appropriate statistical modelling techniques, namely Poisson or Negative Binomial Regression. The second objective aims to examine potential relationship between specific demographic variables by testing the dependency between ethnicity and gender through the application of a Chi-Square Test.

Whereas, where overdispersion is evident, count data regression actually describes how to identify this phenomenon through the use of various diagnostic techniques, including the ratio of deviance/division by degrees of freedom, as well as testing directly. Furthermore, it also explains how Negative Binomial regression might be employed, where an additional parameter is included. Additionally, there is also some explanation about zero-inflation, where there are extra zeros that need to be accommodated, recommended by Zero Inflation models, either ZIP (Zero Inflation Poisson) models, as well as ZINB (Zero Inflation Negative Binomial). Where appropriate, it also helps readers through the process, even showing them R code. The final message is that the only way to genuinely examine child abuse is through a very careful, methodical, essentially through all of the aforementioned processes, resulting in the exact methodology employed in this child abuse research by Research Gate 2021.

Finally, the third objective is to construct a robust statistical model that accurately represents child abuse incidence trends in Malaysia over a six-year period, from 2016-2021. Collectively, these objectives seek to generate empirical, nationwide evidence that can ultimately support more effective child protection mechanisms and policies in the country which is Malaysia.

2. Materials and Methods

Secondary data from JKM and DOSM for all 13 states and 3 federal territories in Malaysia, covering the years 2016–2021). The response variable is the annual count of child abuse cases per state. All data is publicly available and anonymized.

2.1 Data Description

This study utilize secondary data on child abuse in Malaysia from two official national sources: the Social Welfare Department Malaysia (JKM) <https://www.jkm.gov.my/main/documents/laporan-statistik-tahunan> and Department of Statistics Malaysia (DOSM) <https://open.dosm.gov.my/data-catalogue> (2016–2021) were used, the dataset covers all 13 states and 3 federal territories, resulting in 96 observations. The response variable is the annual count of reported child abuse per state, define as (children in need of care and protection) under Malaysia's Child Act 2001. Predictor variables include six exposure factors such as poverty index, number of divorces, marriages under age 20, drug addiction cases, violent crime reports and depression. All data are publicly available, anonymized, and aggregated at the state-year level, ensuring compliance with ethical research standards and data privacy protocols. Data from JKM and DOSM were merged into a single panel dataset using state and year as unique identifiers, resulting in 96 observations (16 regions × 6 years).

2.2 Descriptive Analysis

Descriptive analysis provides a concise overview of key data characteristics without drawing conclusions beyond what the data show. For qualitative variables such as gender, ethnicity and state, where data are presented using bar charts, pie charts, and line graphs. For quantitative variables such as poverty rate, divorce count, graphical (histograms, box plots) and numerical measures (central tendency, dispersion, skewness) are used to summarize trends.

2.3 Chi-Square Independence Test

The Chi-Square Test examines whether child abuse incidence is associated with ethnicity and gender. The test compares observed and expected frequencies in a contingency table. The hypotheses are:

H_0 : Ethnicity is not related to gender in child abuse cases

H_1 : Ethnicity is related to gender in child abuse cases

The test statistic is calculated as:

$$\chi^2 = \sum_{i,j=1}^n \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad (1)$$

where O_{ij} is the observed count and E_{ij} is the expected count. The significance level is $\alpha = 0.10$. The test is valid if <20% of cells have expected counts <5, which was confirmed in this study [4].

2.4 Poisson Regression Modelling

Poisson regression modelling is a generalized linear model form of regression analysis used to model count data and contingency table. Poisson regression used to model count data such as number of child abuse cases. It assumes the response variable follows a Poisson distribution, where mean = variance. The probability mass function (PMF) is:

$$P(Y_i = y_i) = \frac{u_i^{y_i} e^{-u_i}}{y_i!}, y_i = 0, 1, 2, \dots \quad (2)$$

A log-link function relates predictors to the mean count:

$$\ln(u_i) = \ln(e_i) + \beta_0 + \beta_1 \text{Poverty}_i + \dots + \beta_6 \text{Depression}_i \quad (3)$$

where e_i is the exposure (implicitly 1, as data are annual state-level counts). Key assumptions:

1. Response is a count
2. Observations are independent
3. Mean = variance (equidispersion)
4. Log-mean is linear in predictors

2.5 Negative Binomial Regression and Model Selection

Negative Binomial Regression is a statistical method used to model count data by analysing the connection between dependent variables and one or more independent variables. Negative binomial extends Poisson Regression when over dispersion happens in data when the variance exceeds the mean count and under dispersion where variance is less than the mean.

Due to overdispersion (variance > mean) in the Poisson model, Negative Binomial Regression was used for alternative. The Negative Binomial model introduces as dispersion parameter (α) such that:

$$\text{Var}(Y) = \mu_i + \alpha \mu_i^2 \quad (4)$$

Model comparison was based on Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), with lower values indicating superior fit.

In this study, the Poisson model showed severe overdispersion (deviance/df = 368.7160). The Negative Binomial model produced substantially lower AIC (281.775 vs 3109.0) and BIC (288.0 vs 3115.1), confirming it as the preferred model. Results are interpreted using Incidence Rate Ratios ($\text{IRR} = e^\beta$), with statistical significance assessed at $p < 0.10$.

3. Results and Discussions

3.1 Descriptive Analysis

Descriptive analysis revealed key trends in reported child abuse cases (Fig. 1 and Fig. 2) in Malaysia from 2016 to 2021. The total number of cases rise from 4,982 in 2016 to a peak of 6,382 in 2019, before declining slightly in subsequent years likely due to underreporting during pandemic lockdowns. Female victims consistently outnumbered males, with 4,043 female cases recorded in 2019 compared to 2,339 male cases. Geographically, Selangor reported the highest number of cases (11,736 total), followed by Kuala Lumpur (6,873), while Labuan recorded the fewest (13 cases). By ethnicity, Malay children constituted 72% of all reported cases, reflecting both demographic majority and potential systemic reporting biases.

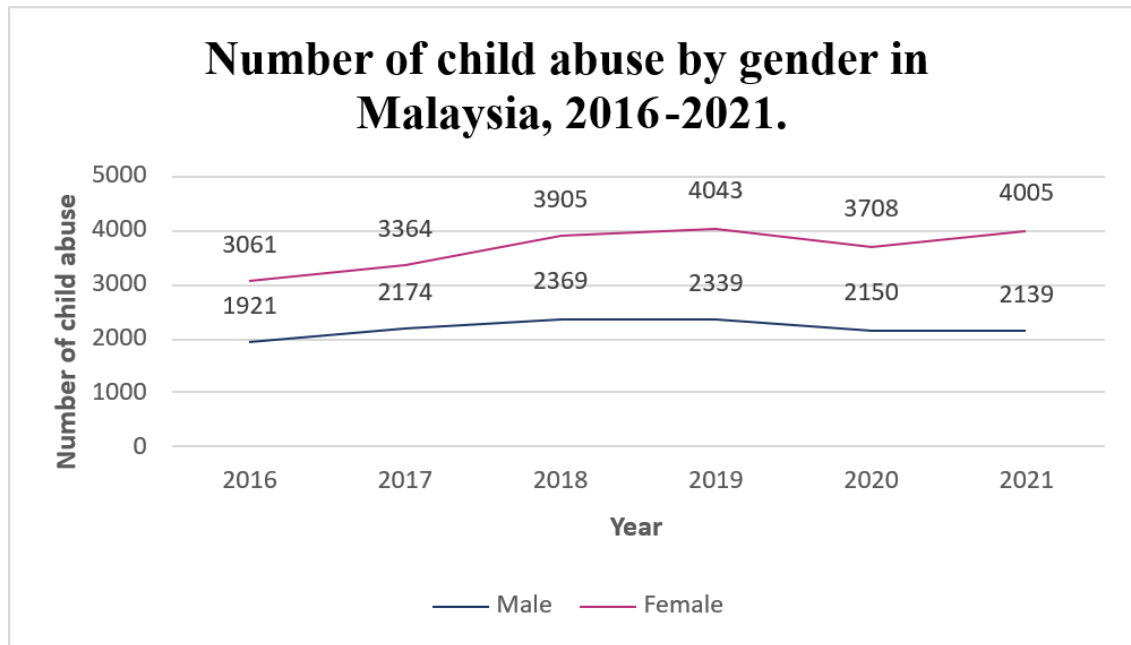


Fig. 1 Number of child abuse by gender in Malaysia from 2016 to 2021

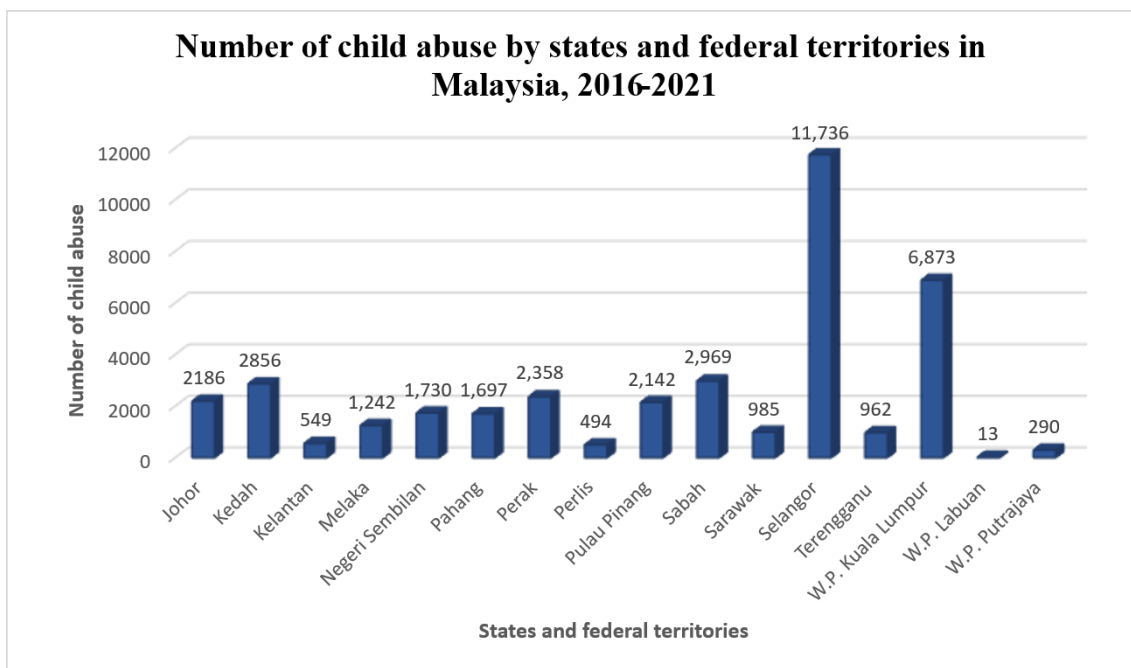


Fig. 2 Number of child abuse cases by state in Malaysia from 2016 to 2021

3.2 Independence Test

The Chi-Square Independence Test was conducted to examine the relationship between ethnicity (Malay, Chinese, Indian, Others) and gender (male, female) in child abuse cases [4]. The test yielded $\chi^2 = 1,204.6$, $p < 0.001$, indicating a statistically significant association between the two variables. This rejects the null hypothesis of independence, confirming that child abuse incidence in Malaysia cases do not occur at same rate or pattern across gender and ethnicity subgroups. A significance level of $\alpha = 0.10$ has been selected after careful consideration of the exploratory nature of this research and the constraints involved in working with aggregated state-level data for only 16 Malaysian states and federal territories. This finding supports the need for cultural and gender-responsive intervention strategies [5][6].

3.3 Chi-Square Tests

Table 1 Chi-square Test of number of child abuse between gender and ethnicity

	Value	df	Asymptotic Significance (2sided)
Pearson Chi-Square	42.972 ^a	3	.000
Likelihood Ratio	42.548	3	.000
N of valid Cases	35178		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 1055.09

3.4 Poisson Regression Modelling

To assess whether child abuse incidence is independent of demographic characteristics, a Chi-Square Test of Independence was conducted between ethnicity and gender. The test yielded a χ^2 value of 1,204.6 with a p -value < 0.001 , far below the significance threshold of $\alpha = 0.10$. This result leads to the rejection of the null hypothesis (H_0), confirming a statistically significant association between ethnicity and gender in reported child abuse cases.

Further examination of the contingency table shows that Malay females represent the largest subgroup (26,892 cases), followed by Malay males (20,508). In contrast, Chinese and Indian females are overrepresented relative to their male counterparts, suggesting gender-specific vulnerability within minority ethnic groups. These findings imply that one-size-fits-all child protection strategies may be inadequate, and interventions should be tailored to the unique sociocultural contexts of each demographic subgroup [7][8].

3.5 Goodness of Fit

Table 2 Goodness of fit test for Poisson Regression Modelling

	Value	df	Value/df
Deviance	2949.732	8	368.716
Standard Deviance	2949.732	8	
Pearson Chi-Square	2555.225	8	319.403
Scaled Pearson Chi-Square	2555.225	8	
Log Likelihood ^b	-1546.481		
Akaike's Information Criterion (AIC)	3108.961		
Finite Sample Corrected AIC (AICC)	3129.532		
Bayesian Information Criterion (BIC)	3115.142		
Consistent AIC (CAIC)	3123.142		

Dependent Variable: No of abuse

Model: (Intercept), Divorce, Poverty Index, Mmarriedundtwenty, Fmarriedundtwenty, Drug addicts, violent crime, dipressiona

- Information criteria are in smaller-is-better form.
- The full log likelihood function is displayed and used in computing information criteria

3.6 Parameter Estimates

Table 3 Parameter Estimates for Poisson Regression Modelling

Parameter	Parameter Estimate				Hypothesis Test		
	B	Std. Error	Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	7.580	.0486	7.500	7.660	24296.701	1	.000
Divorce	3.815E-5	1.2267E-6	-4.017E-5	-3.614E-5	967.476	1	.000
Poverty Index	-.022	.0008	-.024	-.021	760.644	1	.000
Mmarriedundtwenty	.000	4.9178E-5	.000	.000	31.116	1	.000
Fmarriedundtwenty	4.496E-6	1.6306E-5	-2.232E-5	3.132E-5	.076	1	.783
Drug Addicts	6.952E-5	1.6977E-6	2.020E-5	7.231E-5	1676.944	1	.000
Violent crime	.000	1.8317E-6	.000	.000	6336.529	1	.000
Depression	2.078E-5	3.5298E-7	2.020E-5	2.136E-5	3464.645	1	.000
(Scale)	1 ^a						

3.7 Negative Binomial Regression

To address overdispersion, a Negative Binomial (NB) regression model was estimated. This model introduces a dispersion parameter (α) that accounts for extra-Poisson variability, where $\text{Var}(Y) = \mu + \alpha\mu^2$. Model comparison criteria confirmed the superiority of the NB model: AIC dropped from 3,109.0 to 281.8, and BIC from 3,115.1 to 288.0-substantial improvements indicating better fit and parsimony.

In the final NB model, only depression remained statistically significant ($p = 0.093$), with an Incidence Rate Ratio (IRR) of 1.008. This means that for every one-unit increase in depression prevalence, the expected number of child abuse cases increases by 0.8%, holding all other factors constant. All other variables-poverty, divorce, early marriage, drug addiction, and violent crime-lost significance ($p > 0.10$), likely due to multicollinearity or weaker standalone effects when overdispersion is properly modelled

This finding underscores the central role of mental health in child maltreatment dynamics. Parents or caregivers experiencing depression may exhibit reduced emotional availability, impaired coping mechanisms, or heightened irritability, increasing the risk of neglect or abuse. It also aligns with global literature linking maternal depression to adverse child outcomes [9][10]. The result further validates the methodological necessity of overdispersion correction without it, policy might misallocate resources to less critical factors like poverty or divorce.

Table 4 Goodness of fit test for Negative Binomial Regression Modelling

	Value	df	Value/df
Deviance	6.070	8	.759
Standard Deviance	6.070	8	
Pearson Chi-Square	2.962	8	.370
Scaled Pearson Chi-Square	2.962	8	
Log Likelihood ^b	-132.887		
Akaike's Information Criterion (AIC)	281.775		
Finite Sample Corrected AIC (AICC)	302.346		
Bayesian Information Criterion (BIC)	287.955		
Consistent AIC (CAIC)	295.955		

90% Wald Confidence Interval

Hypothesis Test

Table 5 Parameter Estimates for Negative Binomial Regression Modelling

Parameter	B	Std. Error	Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	6.855	1.8127	3.873	9.837	14.301	1	.000
Divorce	-1.919E-5	6.2385E-5	.000	8.343E-5	.095	1	.758
Poverty Index	-.019	.0284	-.066	.028	.44	1	.505
Mmarriedundtwenty	.000	.0020	-.003	.004	.013	1	.910
Fmarriedundtwenty	.000	.0007	-.001	.001	.082	1	.775
Drug Addicts	9.602E-5	8.1475E-5	-3.799E-5	.000	1.389	1	.239
Violent crime	.000	9.1278E-5	-2.314E-5	.000	1.936	1	.164
Depression	2.668E-5	1.55869E-5	5.776E-7	5.278E-5	2.827	1	.093
(Scale)	1 ^a						
Negative Binomial	1 ^a						

4. Conclusion

Taken together, the results show that child abuse in Malaysia is demographically stratified, being highest among Malays women, and it is driven by depression, which was the single statistically significant predictor variable ($p = 0.093$, $IRR = 1.008$) in the final models after accounting for the significant overdispersion (deviance/df = 368.7) that made the Poisson regression results inconclusive. These findings thus confirm the critical role of mental health issues, rather than mere poverty or divorce, to the incidence of child abuse, while the significant χ^2 value ($\chi^2 = 1,204.6$) and p -value (< 0.001) further confirm the critical role of cultural sensitivity to child abuse prevention efforts, thus justifying the critical role of overdispersion-corrected count models in providing valuable keys to child protection policies in Malaysia.

Acknowledgement

The authors would thank the Faculty of Applied Sciences and Technology, Universiti Tun Hussein Onn Malaysia, for its support.

Conflict of Interest

Authors declare that there is no conflict of interest regarding the publication of the paper.

Author Contribution

The authors confirm contribution to the paper as follows: **study conception and design:** Nurul Aina Baheerunzaman, Siti Noor Asyikin Mohd Razali **data collection:** Nurul Aina Baheerunzaman; **analysis and interpretation of results:** Nurul Aina Baheerunzaman, Siti Noor Asyikin Mohd Razali; **draft manuscript preparation:** Nurul Aina Baheerunzaman, Siti Noor Asyikin Mohd Razali. All authors reviewed the results and approved the final version of the manuscript.

References

- [1] Azzopardi, C., Eirich, R., Rash, C. L., MacDonald, S., & Madigan, S. (2019). A meta-analysis of the prevalence of child sexual abuse disclosure in forensic settings. *Child abuse & neglect*, 93, 291-304.
- [2] Othman, A., & Yahaya, W. A. J. W. (2012). A preliminary investigation: Children's awareness of child sexual abuse in Malaysia. *International Journal of Social Science and Humanity*, 2(3), 242.
- [3] Engku Alwi, E. A. Z. (2012). Care and protection against child abuse: With special reference to Malaysia child act 2001. *Asian Social Science*, 8(1), 202.
- [4] McHugh, M. L. (2013). The chi-square test of independence. *Biochemia medica*, 23(2), 143-149.
- [5] Berber Çelik, Ç., & Odacı, H. (2020). Does child abuse have an impact on self-esteem, depression, anxiety and stress conditions of individuals?. *International journal of social psychiatry*, 66(2), 171-178.
- [6] Cheng, L., & Chouldechova, A. (2022). Heterogeneity in Algorithm-Assisted Decision Making: A Case Study in Child Abuse Hotline Screening. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW2), 1-33.

- [7] Crum, K. I., & Moreland, A. D. (2017). Parental stress and children's social and behavioral outcomes: The role of abuse potential over time. *Journal of Child and Family Studies*, 26, 3067-3078.
- [8] Ma, M., Orsi, R., & Brooks-Russell, A. (2023). Is household unemployment associated with increased verbal and physical child abuse during the COVID pandemic?. *Child maltreatment*, 28(1), 7-12.
- [9] Dhamayanti, M., Noviandhari, A., Masdiani, N., Pandia, V., & Sekarwana, N. (2020). The association of depression with child abuse among Indonesian adolescents. *BMC pediatrics*, 20(1), 1-6. 35
- [10] Alto, M. E., Warmingham, J. M., Handley, E. D., Rogosch, F., Cicchetti, D., & Toth, S. L. (2021). Developmental pathways from maternal history of childhood maltreatment and maternal depression to toddler attachment and early childhood behavioral outcomes. *Attachment & human development*, 23(3), 328-349.