Development of rapid visual screening form for Nepal based on the data collected from - its 2015 earthquake

D Rupakheti 1*, P Apichayakul 2

1Department of Civil Engineering, Naresuan University, Muang, Phitsanulok, Thailand
2Department of Electrical and Computer Engineering, Naresuan University, Muang, Phitsanulok, Thailand.

*Email: dharmkar13@gmail.com

Abstract. An earthquake on the 25th April 2015 in Nepal caused more than 9,000 deaths and 22,000 causalities. The main reason for such huge casualties is the lack of earthquake awareness and poor construction practices. With a large number of vulnerable houses, Nepal faces a huge risk from future earthquakes, due to the recent construction in urban areas of poorly designed and constructed buildings. Such unplanned and haphazard growth will be potentially dangerous from natural events like earthquakes. A vulnerability map gives the exact location of sites where people, natural environments and or properties are at risk due to a potentially catastrophic event. This will allow them to decide on mitigating measures to prevent or reduce loss of life, injury and environmental consequences before a disaster occurs. While Rapid Visual screening is a classical method of preliminary vulnerability study, it is one of the most economical, reliable, simple and efficient methods to determine the vulnerability level of buildings. The most known rapid visual screening methods have been developed in countries of high seismic risk such as the USA, Japan, Indonesia, New Zealand, India and Canada and are briefly described in this paper. The main objective of this paper with the help of ordinal regression is to calculate variables of damage grade model in a rapid visual screening form. This is useful to identify vulnerable and non-vulnerable buildings very quickly, and help make a plan for the implementation of a disaster risk mitigation program. Based on SPSS Statistics Software, an ordinal regression method was used to model the relationship between outcome variables. The analysis was performed by employing a database after the 25th April Gorkha earthquake of 2015. Preparing RVS special features of buildings in Nepal, have been given due consideration, and were evaluated for adherence of age, plan configuration, position, land surface condition, plinth area, building height, floor count, foundation types, ground floor types, roof types, condition, are highly significant parameters in analysing the vulnerability of them during an earthquake.

1. Introduction
Nepal is positioned in the Himalayan concave chain, 870 km long in a West-East direction and 130-260 km long in a North-South direction. The Himalayan belt is one of the most seismically active regions in the world because of the faulting between the sub ducting Indian Plate and the overriding Eurasian Plate to the north. The Indian plate converges with the Eurasian plate at a rate of approximately 45 mm/year towards the north–northeast [1] [2]. The 11th most earthquake-prone country in the world is Nepal. In 1255 one-third of the population of the Kathmandu Valley were killed in an earthquake which was the first one ever recorded in the country’s history, and have experienced further major quakes, every few generations since then. The last great earthquake with a magnitude of 8.4, occurred in 1934 resulting in more than 10,000 deaths in the Kathmandu Valley.
which caused most of the infrastructure and major heritage sites to be rebuilt. There have since been earthquakes causing severe human and physical loss in 1980, 1988 and 2011 [3].

The Gorkha earthquake with a magnitude of 7.8 occurred at 11.56 am local time, on April 25th, 2015, with the epicentre 77 km northwest of Kathmandu [4] [5]. This earthquake was one of most devastating disasters in the history of modern Nepal. At this time there were no comprehensive studies or guidelines on the reconstruction of buildings and infrastructure capable of dealing with the necessary challenges [6] [7]. The Gorkha earthquake was followed by a 7.3-magnitude quake on May 12, 2015, killing in total nearly 9000 people and injuring 22,400 [8][9], and also damaged or destroyed around 800,000 houses causing widespread devastation [10].

This study aims to calculate variables of a damage grade estimation model which will help to determine seismic vulnerability scenarios after the initial impact of the earthquake. One of the major important facts is that earthquakes are inevitable in Nepal, due to the fact that they occur frequently, and massive ones occur approximately every 75 years. The occurrence of earthquakes cannot be reduced; however, their impact can be by having defined plans and facilities in place to cope with the disasters. This research highlights the need for surveying building parameters after an earthquake, such as; type, age, height, roof type, floor types, number of floors, type of foundation, condition of the land, plan configuration, to supply the necessary data for future design and construction of earthquake proof buildings.

2. Ordinal Regression Model
An ordinal logistic regression is an extension of binary logistic regression, which may provide a useful model for the analysis of this experiment. The most widely recognized binary regression is logistic regression or a logit model. Ordinary logistic regression is used to model response variables for which the response outcome for each subject is a “success” or a “failure”. The logistic regression model has linear form for the logit of this probability,

$$\logit[\pi(x)] = \log \frac{\pi(x)}{1-\pi(x)} = \alpha + \beta x$$

(1)

Where x is a single explanatory variable and \(\pi(x)\) denotes the “success” probability of a binary response variable at value of the explanatory variable [11].

At the point, when response categories are ordered, logistic regression can utilize the requesting. This instance of calculated logistic regression has logits with cumulative probabilities such an extent that:

$$\logit[\Pr(Y \leq j)] = \log \frac{\Pr(Y \leq j)}{1-\Pr(Y \leq j)}$$

(2)

\(j = 1, 2, \ldots, J - 1\)

Where Y is the response categories with counts that have a multinominal distribution and J is the last ordered category.

In Ordinal Logistic Regression, the dependent variable is the ordered response category variable and the independent variable may be categorical, interval or a ratio scale variable. When the response categories have a natural ordering, model specification should take that into account so that the extra information is utilized in the model. This ordering is incorporated directly in the way the logits is specified. Sometimes the responses may be some continuous variable, which is difficult to measure, so that its range is divided into \(i^{th}\) ordinal categories with associated probabilities \(\pi_1, \pi_2, \ldots, \pi_J\). There are two commonly used models for this situation. [12], [13]
Cumulative Logit Model

The cumulative odds for the $j^{th}$ response category are given by:

$$
\frac{\pi_1 + \pi_2 + \ldots + \pi_j}{\pi_{j+1} + \pi_{j+2} + \ldots + \pi_J}
$$

(3)

where $J$ is the total number of response categories.

The cumulative logit model is:

$$
\log\left(\frac{\pi_1 + \pi_2 + \ldots + \pi_j}{\pi_{j+1} + \pi_{j+2} + \ldots + \pi_J}\right) = \beta_0 j + \beta_1 j x_1 + \ldots + \beta_p j x_p
$$

(4)

Where, variable with $j$ category, $x^j = (x_1, x_2, \ldots, x_p)$ is explanatory variable vector.

with $j = 1, 2, \ldots, J-1$ and $\beta_0$ are threshold model and $\beta$ is regression coefficient vector. The used parameter estimation method in this model is maximum likelihood method. This method assumes that observations are independent to each other.

Proportional Odds Model (POM)

The Proportional Odds Model for Ordinal Logistic Regression provides a useful extension of the binary logistic model to situations where the response variable takes on values in a set of ordered categories. The model might be entitled by a progression of logistic regressions for dependent binary variables, with normal regression parameters reflecting the proportional odds assumption. Key to the acceptable application of the model is the assessment of the proportionality assumption (Brant, 1990).

In the cumulative logit model, each of the parameters rely upon the category $j$. It is based on the assumption that the effect of the covariates $x_1, x_2, \ldots, x_p$ are the explanatory variable vectors, on the logarithmic scale. Thus, in this model, only the intercept term $\beta_0 j$ depends on the category $j$ so that the model is:

$$
\log\left(\frac{\pi_1 + \pi_2 + \ldots + \pi_j}{\pi_{j+1} + \pi_{j+2} + \ldots + \pi_J}\right) = \beta_0 j + \beta_1 x_1 + \ldots + \beta_p x_p
$$

(5)

The relevance of the model can be tested independently for each variable by comparing it to a cumulative odds model with the appropriate parameter not depending on $j$.

3. Methodology

The data used in this study is secondary data that was obtained from the central bureau of statistics, and the National planning commission, Nepal. The data on the number of buildings damaged during the 2015 earthquake was collected from the household registration of the reconstruction programme, which helped in proposing the appropriate actions to strengthen them. The response variable used in this study was a request for the level of damage to superstructures, while the independent variables were Age ($X_1$), the Height ($X_2$), Land Surface Condition ($X_3$), Foundation Type ($X_4$), Roof Type ($X_5$), Ground Floor Type ($X_6$), Position ($X_7$), Plan Configuration ($X_8$), Floor Count ($X_9$), and Plinth Area of Building ($X_{10}$). The first test performed in this study was to test the assumption of multicollinearity, which uses the correlation matrix. Furthermore, the Ordinal logistic Regression model was analysed, and used to sort out the variables that will be used in a Plum Ordinal Logistic Regression model, and generate a parameter estimator. After that, a coefficient, and likelihood ratio test were performed simultaneously on the
Ordinal Logistic Regression model, followed by a coefficient test on it, by using the Wald method. After determining the best regression model, further testing is performed to confirm its suitability.

3.1 Define Variables
The dependent variable is Damage Grade of the superstructure, which is categorized as Grade 1, Grade 2, Grade 3, Grade 4 and Grade 5.

| Categories | Description      | Code No |
|------------|------------------|---------|
| Grade 1    | Damage Grade     | 1       |
| Grade 2    | Damage Grade     | 2       |
| Grade 3    | Damage Grade     | 3       |
| Grade 4    | Damage Grade     | 4       |
| Grade 5    | Damage Grade     | 5       |

One or more independent variables that are continuous, ordinal or categorical (including dichotomous variables). Though, ordinal independent variables must be treated as being either continuous or categorical. The independent variables that used in this study are shown in Table 2.

| Variable designation | Description          | Value labels                      | Codes/Value |
|----------------------|----------------------|-----------------------------------|-------------|
| X1                   | Age of building      | Continuous                        | None        |
| X2                   | Height of building   | Continuous                        | None        |
| X3                   | Land surface Condition | Nominal                         | 1= Flat, 2= Moderate, 3= Steep slope |
| X4                   | Foundation Type      | Nominal                           | 1=Mud Mortar-Stone/Brick, 2= Cement-Stone/Brick, 3= RC, 4= Bamboo/Timber, 5= Other |
| X5                   | Roof Type            | Nominal                           | 1= Bamboo/Timber-Heavy, 2= Bamboo/Timber-light, 3= RCC/RB/RBC |
| X6                   | Ground Floor Type    | Nominal                           | 1= Mud, 2= Brick/Stone, 3= RC, 4= Timber, 5= Other |
| X7                   | Position             | Nominal                           | 1= Attached-1, 2= Attached-2, 3= Attached-3 side, 4= Not attached |
| X8                   | Plan Configuration   | Nominal                           | 0= Other Plan, 1= Square Plan |
| X9                   | Floor Count          | Nominal                           | 1-3= Low Rise (1-3), 4-6= Mid Rise (4-6), 7-9= High Rise (7-9) |
| X10                  | Plinth Area of Building | Continuous                    | None        |

4. Results and Discussion
Ordinal regression analysis is used in this study with the help of SPSS software. The Statistical Package for the Social Sciences (SPSS) is a software package used for the analysis of statistical data. An ordinal logistic regression must also meet two assumptions that relates to how your data fits the model in order to provide a valid result: (a) there should be no multicollinearity and (b) you have proportional odds, which is a fundamental assumption of this type of model. A Multicollinearity occurs when you have two or more independent variables that are highly correlated with each other. This leads to problems with understanding which variable contributes to the explanation of the dependent variable and technical issues in calculating an ordinal logistic regression. To find out if you might have a problem with multicollinearity, you need to consult the “Tolerance” and “VIF” values in
the Coefficients table. Proportional Odds assumption is tested in SPSS Statistics with a full likelihood ratio test comparing the fit of the proportional odds model to a model with varying location parameters.

4.1 Multicollinearity

Multicollinearity occurs when two or more independent variables that are highly correlated with each other. This leads to problems with understanding which variable contributes to the explanation of the dependent variable and technical issues in calculating an ordinal logistic regression. To find out if it might have a problem with multicollinearity, need to consult the "Tolerance" and "VIF" values in the Coefficients table. The tolerance statistic is the calculation of the variance of each of the independent variables in the model not explained by all of the other independent variables in the model. A higher tolerance value recommends low levels of collinearity. Menard (2010) suggests that a tolerance of less than .2 is alarming. If we consult the Tolerance value of less than 0.1 – which equates to a VIF value of 10 or greater – it might have a collinearity problem. In Table 3, all the Tolerance values are greater than 0.1 and VIF values are much less than 10, but only the count floor (Low-rise and Mid-rise) are less than 0.1 and VIF value is greater than 10. So, here we have a problem with collinearity in this particular data set. What we find here is that the height of building and floor count are co-related with each other. To fixed this problem, the floor count is removed from this model.

| Model                        | Collinearity Statistics |
|------------------------------|-------------------------|
|                             | Tolerance | VIF     |
| age building                | .822      | 1.217   |
| height_ft_pre_eq            | .292      | 3.425   |
| plinth_area_sq_ft           | .917      | 1.091   |
| Coded plan configuration    | .932      | 1.073   |
| Flat                        | .448      | 2.235   |
| Moderate Slope              | .454      | 2.202   |
| Mud Mortar- Stone/Brick     | .654      | 1.530   |
| Cement- Stone/Brick         | .873      | 1.146   |
| Bamboo/Timber-Heavy roof   | .899      | 1.113   |
| Bamboo/Timber-Light roof    | .901      | 1.110   |
| Mud                         | .727      | 1.376   |
| Brick/Stone                 | .923      | 1.084   |
| Timber                      | .843      | 1.186   |
| Attached-1 side             | .736      | 1.358   |
| Attached-2 side             | .585      | 1.708   |
| Attached-3 side             | .860      | 1.163   |
| Low Rise 1-3                | .026      | 39.193  |
| Mid Rise 4-6                | .029      | 34.421  |

a. Dependent Variable: Damagegrade
code
4.2 Full likelihood ratios test

One of the assumptions underlying ordinal logistic regression is that the relationship between each pair of group outcomes is the same. This is normally referred to as the test of parallel lines for the reason that the null hypothesis states the slope coefficients in the model are the same across response categories (and lines of the same slope are parallel), therefore, if we fail to reject the null hypothesis, we conclude that the assumption holds. Table 4 shows the test of parallel lines for the general model with the chi square value = 49.488, and the p value = 0.534 which is greater than the 5% level of significance, fails to reject the null hypothesis. Therefore, there is not enough evidence to reject the null hypothesis, thus, the proportional odds assumption appears to have held for the general model.

### Table 4. Test of parallel lines

| Model          | -2 Log Likelihood | Chi-Square | df | Sig. |
|----------------|-------------------|------------|----|------|
| Null Hypothesis| 1387.176          |            |    |      |
| General        | 1337.689<sup>b</sup> | 49.488<sup>c</sup> | 51 | .534 |

The null hypothesis states that the location parameters (slope coefficients) are the same across response categories.

- a. Link function: Logit.
- b. The log-likelihood value cannot be further increased after maximum number of step-halving.
- c. The Chi-Square statistic is computed based on the log-likelihood value of the last iteration of the general model. Validity of the test is uncertain.

4.3 Results

Ordinal logistic regression is a type of logistic regression analysis, that when categorized shows the response variable as more than two and having a natural order or rank, therefore we can rank the values, however the real distance between categories is unidentified. Under Ordinal Logistic Regression Analysis, we can use Model Fitting Information, Goodness-of-Fit, Pseudo R-Square, Parameter Estimates and Test of parallel lines.

The Logit link function is used in the analysis, because it evenly distributes categories which are reasonable choices when the changes in the cumulative probabilities are gradual, and logit involves all ranks of the response and dichotomizes the response scale.

### Table 5. Model fitting information

| Model          | -2 Log Likelihood | Chi-Square | df | Sig. |
|----------------|-------------------|------------|----|------|
| Intercept Only | 1460.671          |            |    |      |
| Final          | 1387.176          | 73.495     | 17 | .000 |

Link function: Logit.

The Model Fitting Information Table 5, which gives the -2-log likelihood for the intercept only and final models can be used in comparisons of nested models. The statistically significant chi-square statistic (p<0.05) indicates that the Final model gives a significant enhancement over the baseline intercept only model. This tells us that the model gives better predictions than if we just guessed based
on the marginal probabilities for the outcome categories. Therefore, the Full model (with factors that affect Damage grade as a predictor) is significantly better than the "damage grade" model.

**Table 6. Goodness-of-fit**

|        | Chi-Square | df | Sig. |
|--------|------------|----|------|
| Pearson | 2038.379   | 1947 | 0.073 |
| Deviance | 1384.168   | 1947 | 1.000 |

Link function: Logit.

From the above Table 6, the results for our analysis suggest that the model does fit very well (p>0.05) (i.e. fail to reject the null hypothesis depending on the observed data). Also, the model fits adequately.

**Table 7. Pseudo R-square**

|                      |            |
|----------------------|------------|
| Cox and Snell        | 0.137      |
| Nagelkerke           | 0.144      |
| McFadden             | 0.050      |

Link function: Logit.

What creates a “good” R² value depends upon the nature of the outcome and the descriptive variables. Here, the pseudo R² values (e.g. Nagelkerke = 14.4%) indicates that there is relatively small proportion of the variation in damage grade between buildings. This is just as we would expect because there are numerous factors that affect buildings.

**Table 8. Parameter estimates**

|        | Estimate | Std. Error | Wald | df | Sig. | 95% Confidence Interval |
|--------|----------|------------|------|----|------|-------------------------|
|        |          |            |      |    |      | Lower Bound | Upper Bound |
| Threshold | [Y = 1]  | -.907      | 1.907 | 0.226 | 1  | .634 | -4.645 | 2.831 |
|        | [Y = 2]  | .518       | 1.907 | 0.074 | 1  | .786 | -3.219 | 4.255 |
|        | [Y = 3]  | 1.882      | 1.909 | 0.972 | 1  | .324 | -1.860 | 5.624 |
|        | [Y = 4]  | 2.758      | 1.914 | 2.077 | 1  | .150 | -.993  | 6.508 |
|        | X1       | .063       | .013  | 24.386 | 1  | .000 | .038  | .089  |
|        | X2       | -.002      | .007  | .045  | 1  | .831 | -.016  | .013  |
|        | X10      | .001       | .000  | 5.677 | 1  | .017 | .000  | .001  |
|        | [X3=1]   | -.342      | .265  | 1.662 | 1  | .197 | -.862  | .178  |
|        | [X3=2]   | -.059      | .312  | .036  | 1  | .850 | -.671  | .553  |
|        | [X3=3]   | 0º         | .312  | .036  | 0  | .850 | -.671  | .553  |
|        | [X4=1]   | 1.847      | .783  | 5.564 | 1  | .018 | .312  | 3.382 |
|        | [X4=2]   | .115       | .288  | .160  | 1  | .689 | -.449  | .679  |
|        | [X4=3]   | 0º         | .288  | .160  | 0  | .689 | -.449  | .679  |
|        | [X5=1]   | 1.185      | .811  | 2.132 | 1  | .144 | -.406  | 2.775 |
There is also a strong association between factors that affect Building, even when p-values are less than alpha level. We can see significant for Age, plinth Area of building, Foundation, and Roof type. In the Parameter Estimates Table 8, we see the coefficients, their standard errors, the Wald test and associated p-values (Sig.), the 95% confidence interval of the coefficients and odds ratios. If the p-values less than alpha level, they are statistically significant; otherwise not. The thresholds are shown at the top of the parameter estimates output, and indicate where the latent variable is cut to make the five groups that we observe in our data. The threshold coefficients represent the intercepts, specifically the point (in terms of a logit) where Damage grade might be predicted into the five categories.

In this coefficient the negative sign indicates that some variables on height, land surface conditions, ground floor types, positions and plan configurations are the factors that have negative effects on Buildings in table 8. The location of the labelled estimates are the coefficients for the predictor variables, which appears to have a relationship between the damage grade and the factors that affect building. From the observed significance levels in table 8, we can see that all explanatory variables are factors that affect the damage grade in buildings. Based on the small observed significance level, we can reject the null hypothesis that is zero. The location of the labelled estimates is the ones we are interested in, because they are the coefficients for the predictor variables. The Wald statistic is the square of the proportion of the coefficient to its standard error. The significance of the Wald statistic in the column with heading sig (< 0.05) indicates the importance of the predictor variables in the model (we reject the Null hypothesis Ho: =0) and high values of the Wald statistic illustrations that the corresponding predictor variable is significant.

5. Conclusion
The main objective of this study is to calculate the variables of a damage grade estimation model in a rapid visual screening form which will help to determine seismic vulnerability scenarios after the
initial impact of an earthquake. While testing the multicollinearity there was a problem with a parameter, so after removing it, the model was fit for further process.

In this analysis we looked at regression models that can be applied when our outcome is represented by an ordinal variable. When the proportional odds assumption is justified, ordinal regression models can be a powerful means of summarizing relationships that utilizes all the information present in the ordinal outcome. Furthermore, the findings indicate that the damage grade of a building is strongly associated with age, foundation type, roof type and the plinth area. The independent variables that have no significant association with the damage grade of a building is its height, land surface condition, type of ground floor, position and plan configuration.

Acknowledgments
The authors are thankful to the anonymous reviewers for very valuable review on an earlier draft of this paper which helped them significantly revise the work. Central bureau of statistics, and the National planning commission, Nepal, are acknowledged for the data they collected. Dr. Phisut Apichayakul who helped in data analysis and preparation of the manuscript.

References

[1] J. Paul et al., “The motion and active deformation of India,” Geophys. Res. Lett., vol. 28, no. 4, pp. 647–650, 2001.
[2] L. T. White and G. S. Lister, “The collision of India with Asia,” J. Geodyn., vol. 56–57, pp. 7–17, 2012.
[3] Global Facility for Disaster Reduction and Recovery et al., “Post Disaster Needs Assessment,” pp. 1–20, 2015.
[4] G. Chiaro, T. Kiyota, R. M. Pokhrel, K. Goda, T. Katagiri, and K. Sharma, “Reconnaissance report on geotechnical and structural damage caused by the 2015 Gorkha Earthquake, Nepal,” Soils Found., vol. 55, no. 5, pp. 1030–1043, 2015.
[5] K. Goda et al., “The 2015 Gorkha Nepal Earthquake: Insights from Earthquake Damage Survey,” Front. Built Environ., vol. 1. pp. 1–15, 2015.
[6] K. Sharma and A. KC, “Challenges of reconstruction after 2015 Nepal earthquake and a comparison with Haiti and Chile reconstruction,” Ski. Knowl. Shar. Nepal. Community. NESA Symp., no. May, pp. 2015–2017, 2017.
[7] K. Sharma, K. C. Apil, M. Subedi, and B. Pokharel, “Challenges for reconstruction after M w 7.8 gorkha earthquake: A study on a devastated area of Nepal,” Geomatics, Nat. Hazards Risk, vol. 9, no. 1, pp. 760–790, 2018.
[8] K. Sharma, L. Deng, and C. C. Noguez, “Field investigation on the performance of building structures during the April 25, 2015, Gorkha earthquake in Nepal,” Eng. Struct., vol. 121, pp. 61–74, 2016.
[9] K. Sharma, M. Subedi, R. R. Parajuli, and B. Pokharel, “Effects of surface geology and topography on the damage severity during the 2015 Nepal Gorkha earthquake,” Lowl. Technol. Int., vol. 18, no. 4, pp. 269–282, 2017.
[10] K. Sharma, L. Deng, and D. Khadka, “Reconnaissance of liquefaction case studies in 2015 Gorkha (Nepal) earthquake and assessment of liquefaction susceptibility,” Int. J. Geotech. Eng., vol. 13, no. 4, pp. 326–338, Jul. 2019.
[11] S. Edition, An Introduction to Categorical Data Analysis Second Edition.
[12] G. M. Ombui, M. Geoffrey, and a W. Gichuhi, “Using Ordinal Regression Modeling To Evaluate the Satisfaction of Jomo,” Jomo Kenyatta Univ. Agric. abnd Technol., vol. 13, no. 1, pp. 164–176, 2011.
[13] A. O. Adejumo and A. A. Adetunji, “Application of Ordinal Logistic Regression in the Study of Students’ Performance,” Math. Theory Model., vol. 3, no. 11, pp. 10–19, 2013.