**Development of a Multivariate Model for Predicting Injuries Based on Attack Type and Other Independent Factors**

**1. Introduction**

Predicting the number of injuries resulting from attacks is essential for resource allocation and emergency response in the context of gauging reconstructive surgery need. We developed a multivariate regression model that estimates daily injury counts based on attack characteristics, such as attack type, population density, and other relevant independent variables. Given that multiple factors contribute to injury severity and frequency, a multivariate approach was chosen to capture these complex relationships over the univariate model which only evaluate the correlation between two isolated variables and may fail to capture the complex interactions between multiple contributing factors.

**3. Preliminary Modeling**

*3.1 Number of Attacks x Attack Distance from Al-Shifa Hospital*

As an initial step, we constructed a preliminary model using two independent variables: Number of attacks per day (sourced from Ministry of Health daily reports) and Distance from Al-Shifa Hospital, the main medical complex in Gaza (calculated using the Euclidean distance between each attack location in the ACLED database and the hospital’s coordinates).

This finding may be explained by population density differences—Al-Shifa is located in Gaza City, a densely populated urban area where attacks are likely to result in more casualties. In contrast, attacks occurring further from the hospital may happen in rural areas with fewer civilians, leading to lower reported injury counts. Additionally, underreporting may occur in remote areas lacking medical infrastructure, where victims have limited access to hospitals.

To assess alternative modeling approaches, we tested a Poisson regression model and a Negative Binomial regression model. Poisson Regression exhibited signs of overfitting (pseudo R² = 1.000), suggesting the model did not generalize well, while Negative Binomial Regression produced similar results to the least-squares model while better accounting for overdispersion in the data (A. 1).

*3.2 Attack-Type Categorization*

A major challenge was defining attack type classifications in a way that balanced statistical robustness and meaningful categorization. The ACLED dataset pre-classifies events using “event type” (e.g., explosions, battles) and “sub-event type” (e.g., shelling, grenade attacks). However, we found that manually aggregating similar attack types into broader categories (e.g., air attacks, ground attacks) improved the statistical significance of the model while also serving as distinct, meaningful classifications. Thus, the final attack type classification consisted of: Airdrone attacks, Shelling/artillery attacks, Ground attacks, Civil unrest, and Other attack types (A.2).

**4. Descriptive Statistics - Multivariate Modeling**

*4.1 Data Processing and Categorization*

To improve model performance, we categorized population density into three groups based on OCHA situational reports: High density (> 20,000 people/km²), Medium density (5,000–20,000 people/km²), Low density (< 5,000 people/km²). Infrastructure type was also classified into five categories using satellite imagery: Barren, Camp, Tent settlements, Suburban, Urban environments.

4.2 Selection of Multivariate Model

To develop a robust predictive framework for estimating injury counts, we evaluated four multivariate models: Least-Squares Regression, a Generalized Poisson Linear Model, a Generalized Negative Binomial Model, and a Least Angle Regression Model. Given the observed overdispersion in injury count data, which suggests greater variability than a Poisson distribution would typically accommodate, we considered the Generalized Negative Binomial Model as the final predictive framework.

*4.2 Selection of Independent Variables*

The selection of independent variables was guided by prior research, situational reports, and structured attack databases such as the Armed Conflict Location & Event Data Project (ACLED). We hypothesized that that the following factors could significantly influence injury counts: Attack type (e.g., airstrikes, shelling, armed clashes), Number of attacks per day, Population density of the affected region, Proximity to hospitals, Attack location and infrastructure, and Time of day. This model incorporated the following independent variables: Attack type (5 categorical variables), Mean population density, and Infrastructure type (5 categorical variables). For each attack listed in the ACLED database, each was categorized into one of the five following standardized groupings: Airdrone attacks, Shelling/artillery attacks, Ground attacks, Civil unrest, and Other attack types. Likewise, each attack was also assigned to an Infrastructure type from the five groupings: urban, suburban, rural, tent, or camp.

*4.3 Interaction Factor Analysis*

The existence of interaction factors was a possibility that was addressed. A significant interaction effect between two independent variables would indicate the presence of a dampening effect of one factor on the other—essentially, the significant effect of the former independent variable on the dependent variable is exacerbated by the latter. In the case where interaction effect is significant within the independent variables selected in our final model, this may lead to the overemphasis on the significance of a certain factor in affecting injury count, which would result in a biased estimation.

We hypothesized that infrastructure categorical variables when compared with mean population density would yield the most likely and most significant interaction effects, given the intuitive relationship between them. An interaction factor analysis was performed on the negative binomial distribution model on each infrastructure type categorical variable as compared to mean population density variable (A.3).

5. Sensitivity Analysis

We estimated the sensitivity of each model using 95% confidence intervals for each injury prediction (A.3).

**5. Discussion**

*5.1 Rationale Behind Variable Selection*

Due to the lack of standardized protocol for reporting civilian injury or death during wartime, oftentimes reporting of injuries by independent organizations remains biased and underreported [https://doi.org/10.1186/s13049-024-01299-7]. Given the variability in conflict reporting and the complexity of factors contributing to injury outcomes, our selection of independent variables was guided by both statistical considerations and real-world applicability. Our primary goal was to construct a model that could effectively predict injury counts while incorporating interpretable and actionable variables relevant to humanitarian response and medical resource planning.

Attack type was chosen as a core variable due to its direct influence on the scale and severity of injuries. Airborne attacks, such as airdrone strikes and shelling, tend to cause mass casualties due to their wide-area effects, while ground attacks and civil unrest events typically involve more localized engagements. Including attack type allowed us to differentiate between the distinct mechanisms of injury and the expected severity of casualties associated with different forms of violence.

*5.1 Comparisons Between Multivariate Models*

Our findings demonstrate that a multivariate approach is essential for accurately predicting injury counts resulting from attacks, as multiple interacting factors contribute to the variability in casualty outcomes. Each model was assessed based on statistical fit metrics and predictive performance. The Least-Squares Regression model yielded an R-squared value of 0.303, indicating a moderate linear relationship between independent variables and injury count, though its adjusted R-squared (0.267) suggested limited generalizability. The Generalized Poisson model demonstrated a strong fit to the data, with a pseudo R-squared of 1.000 and a Pearson chi-squared statistic of 5.38e+04, suggesting potential overfitting. The Generalized Negative Binomial Model mitigated this issue, achieving a pseudo R-squared of 0.5533 while accounting for overdispersion with a Pearson chi-squared statistic of 135.

The Negative Binomial regression model provided the most reliable results, effectively addressing the overdispersion in the data and significantly outperforming both the Least-Squares Regression model (R² = 0.303) and the Poisson Regression model (Pseudo R² = 1.000), which exhibited overfitting. The predictive power of the final model suggests that attack type, population density, and infrastructure characteristics play critical roles in determining injury severity.

A key finding from our model was the strong association between attack type and injury severity. Airborne attacks, such as airdrone strikes and shelling, were the most significant predictors of higher injury counts, likely due to their potential for large-scale destruction and indiscriminate targeting in densely populated areas. Ground attacks and incidents categorized as civil unrest exhibited weaker but still relevant associations, indicating that while direct engagement events result in casualties, they tend to produce fewer injuries per incident compared to aerial bombardment. The strong positive correlation between population density and injury counts further supports the hypothesis that attacks in urban areas result in significantly more casualties, consistent with prior research on conflict impact in densely populated regions. Similarly, infrastructure type was found to contribute to injury outcomes, with urban environments showing higher casualties than barren or tented areas, likely due to both population concentration and structural vulnerabilities.

*5.2 Limitations*

While the Negative Binomial model performed best, several limitations should be considered. One primary limitation is the granularity of available population density data. While we categorized population density into three broad groups, a more precise dataset reflecting dynamic population movements in conflict zones could improve model accuracy. Additionally, our model relies on reported injury counts, which may underestimate the true impact of attacks, particularly in areas where medical access is limited. Underreporting is a persistent issue in conflict zones, particularly in areas where transportation to medical facilities is restricted or where a substantial number of injured individuals are trapped under rubble. Furthermore, the model does not account for potential confounding variables or other potentially significant factors, such as injury severity of demographical information, all of which may influence injury numbers.

5.4 Future Directions

Future research should explore modeling approaches that incorporate additional contextual factors to enhance predictive accuracy. While our model incorporated factors that are considered to be substantially impactful on predicting injury rates based on past observations and current analysis, predicting injury rates depends on a nuanced and complex number of variables that is difficult to quantify. For example, our analysis of injury rates did not consider the effect of other potentially significant factors such as

An analysis past predicting the crude number of injuries into effect of the postulated factors on other assessments such as injury patterns (e.g. correlation between attack variables and extremities affected). In the context of providing reconstructive surgery care to Palestinians in a war-torn region, simply predicting the number of injuries and using a crude, constant estimate that assumes a constant percentage of injuries as requiring reconstructive surgical need may not reflect the complex and dynamic nature of injuries accrued across a few years.

Future work should also focus on optimizing the quality and quantity of data obtained from online, open source datasets and available situation reports published throughout the conflict.

The findings of this study have significant implications for humanitarian response efforts and medical resource planning in conflict zones in the context of reconstructive surgical need.

**Appendix**

Table A.1 Injury prediction with preliminary model

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Least-squares Regression** | | | | **Generalized Poisson linear model** | | | | **Generalized Negative Binomial Model (manuscript model)** | | | |
| Descriptive Statistics | R-squared | | Adj. R-squared | | Pseudo R-squ. | | Pearson chi2 | | Pseudo R-squ. | | Pearson chi2 | |
| 0.100 | | 0.091 | | 1.000 | | 1.58e05 | | 0.5533 | | 135 | |
| Independent Variable | Coef. | Std. error | | P-value | Coef. | Std. error | | P-value | Coef. | Std. error | | P-value |
| const | 246.5270 | 402.407 | | 0.541 | 5.5159 | 0.032 | | 0.000 | 5.7802 | 0.588 | | 0.000 |
| number\_of\_attacks | 21.4763 | 5.597 | | 0.000 | 0.0458 | 0.000 | | 0.000 | 0.0453 | 0.008 | | 0.000 |
| Distance\_km | -46.4522 | 21.011 | | 0.028 | -0.0979 | 0.002 | | 0.000 | -0.1154 | 0.031 | | 0.000 |
| Predicted # of injuries (95% CI) | 81971.08 (-142495.65, 306437.82) | | | |  | | | |  | | | |
| Predicted # of surgeries (95% Ci) | 19263.20 (-33486.48, 72012.89) | | | |  | | | |  | | | |

Table A.3 Interaction Effect Descriptive Statistics – Negative Binomial Model

|  |  |  |  |
| --- | --- | --- | --- |
|  | Coef | Std error | p-value |
| Population density vs. camp interaction | -0.0743 | 0.118 | 0.531 |
| Population density vs. barren interaction | -0.0108 | 0.046 | 0.816 |
| Population density vs. suburban interaction | -1.9057 | 0.689 | 0.006 |
| Population density vs. urban interaction | -0.3775 | 0.189 | 0.046 |
| Population density tent interaction | -1.1815 | 0.401 | 0.003 |

Table A.3 Injury prediction using other multivariate models

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Least-squares Regression** | | | | **Generalized Poisson linear model** | | | | **Generalized Negative Binomial Model** | | | | | **Least Angle Regression Model (manuscript model)** | | | | |
| Descriptive Statistics | R-squared | | Adj. R-squared | | Pseudo R-squ. | | Pearson chi2 | | Pseudo R-squ. | | Pearson chi2 | | |  | |  | | |
| 0.303 | | 0.267 | | 1.000 | | 5.38e+04 | | 0.5533 | | 135 | | |  | |  | | |
| Independent Variable | Coef. | Std. error | | P-value | Coef. | Std. error | | P-value | Coef. | Std. error | | P-value |  | |  | |  |
| const | -1340.7519 | 468.033 | | 0.005 | 1.3448 | 0.049 | | 0.000 | 2.2980 | 0.763 | | 0.003 |  | |  | |  |
| airdrone\_attack\_count | 49.2180 | 13.814 | | 0.000 | 0.0896 | 0.001 | | 0.000 | 0.0662 | 0.023 | | 0.003 |  | |  | |  |
| shelling\_attack\_count | -4.0323 | 15.066 | | 0.789 | -0.0220 | 0.002 | | 0.000 | -0.0006 | 0.025 | | 0.979 |  | |  | |  |
| ground\_attack\_count | 21.6320 | 22.318 | | 0.334 | 0.0552 | 0.002 | | 0.000 | 0.0762 | 0.036 | | 0.036 |  | | | | |
| civil\_unrest\_count | -33.0944 | 33.428 | | 0.323 | -0.0763 | 0.004 | | 0.000 | -0.0364 | 0.055 | | 0.504 |  | | | | |
| other\_attack\_count | -49.1636 | 79.296 | | 0.536 | -0.1974 | 0.010 | | 0.000 | -0.1768 | 0.129 | | 0.172 |  | | | | |
| mean\_population\_density | 923.2218 | 318.167 | | 0.004 | 2.4451 | 0.032 | | 0.000 | 1.8395 | 0.519 | | 0.000 |  | | | | |
| barren\_count | -3.5402 | 14.484 | | 0.807 | 0.0061 | 0.002 | | 0.000 | -0.0147 | 0.024 | | 0.534 |  | | | | |
| camp\_count | -31.3199 | 18.939 | | 0.100 | -0.0735 | 0.002 | | 0.000 | -0.0296 | 0.031 | | 0.338 |  | | | | |
| suburban\_count | 134.4944 | 75.814 | | 0.078 | 0.2089 | 0.005 | | 0.000 | 0.2039 | 0.123 | | 0.098 |  | | | | |
| tent\_count | -43.8778 | 28.489 | | 0.125 | -0.1737 | 0.004 | | 0.000 | -0.1682 | 0.047 | | 0.000 |  | | | | |
| urban\_count | -71.1969 | 30.080 | | 0.019 | -0.1188 | 0.002 | | 0.000 | -0.0629 | 0.049 | | 0.199 |  | | | | |
| Predicted # of injuries (95% CI) | 81971.41 (-276346.31, 440289.14) | | | | 81923.43 (66918.63, 100355.73) | | | | 78626.93 (3869.67, 1801730.94) | | | |  | | | | |
| Predicted # of surgeries (95% Ci) | 19263.28 (-64941.38, 103467.95) | | | | 19252.00 (15725.88, 23583.60) | | | | 18477.33 (909.37, 423406.77) | | | |  | | | | |