**Development of a Multivariate Model for Predicting Injuries in Palestine 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 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 the complex interactions between multiple contributing factors.

**2. Preliminary Modeling**

*2.1 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) (A.1).

**3. Descriptive Statistics - Methods**

*3.1 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.

*3.2 Model Selection Criteria*

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 (A.4). Given the nature of our dataset and the challenges posed by overdispersion, model selection was based on multiple statistical criteria to ensure both predictive accuracy and goodness-of-fit.

The following metrics were used to evaluate model performance:

* Descriptive Statistics & Goodness-of-Fit Measures:
  + R-squared (R²) and Adjusted R-squared (Adj. R²): Used for the Least-Squares Regression model to assess how well independent variables explain variation in injury counts. However, R² is less reliable in count-based data due to its assumptions of normality and constant variance.
  + Pseudo R-squared: Applied to count models (Generalized Poisson and Generalized Negative Binomial) as an alternative measure of model fit. Higher values indicate better explanatory power.
  + Pearson Chi-Square: Assesses the discrepancy between observed and predicted values in count-based models, helping to evaluate model fit. Lower values generally indicate a better fit.
* Statistical Significance & Coefficients:
  + Coefficient Estimates & Standard Errors: Coefficients were examined for their interpretability and magnitude, particularly for key independent variables such as attack type, population density, and infrastructure type. The standard errors provided insight into the precision of these estimates.
  + P-values: Statistical significance of each variable was evaluated (p < 0.05 as a threshold), ensuring that included predictors contributed meaningfully to the model.
* Handling Overdispersion:
  + Overdispersion, where the variance exceeds the mean, was a key consideration in model selection. The Generalized Poisson Model and Generalized Negative Binomial Model were specifically evaluated to account for this issue, as they provide more flexibility in modeling count data with high variance.
* Predictive Stability & Confidence Intervals:
  + The width of the 95% confidence intervals for predicted injury counts and predicted required surgeries was considered. Narrower intervals suggest more precise predictions, while excessively wide intervals indicate high variability and potential instability in the model.

Based on these criteria, the Generalized Negative Binomial Model was selected as the final predictive framework, as it effectively accounted for overdispersion while maintaining a strong model fit across statistical metrics.

*3.3 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 potentially 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.

*3.4 Computational Modeling*

All statistical analyses and model implementations were conducted using Python within the VS Code environment, utilizing Jupyter Notebook for interactive computation and visualization. The models were trained on one year of injury data (October 2023 – October 2024) and subsequently used to predict injury counts through January 2025. The models were implemented using the Statsmodels library (v. 3.12.2), specifically leveraging its suite of regression and count-based modeling tools.

**4. Results/Discussion**

*4.1 Rationale Behind Variable Selection – Challenges and Final Justification of Methods and Selection*

Due to the lack of reliable, standardized protocols for reporting civilian injury or death during wartime, data collection remains inconsistent and susceptible to bias and underreporting [https://doi.org/10.1186/s13049-024-01299-7 and https://pmc.ncbi.nlm.nih.gov/articles/PMC10450422/]. 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 medical-planning applicability. Our primary goal was to construct a model that could effectively predict injury counts while incorporating measurable and interpretable variables that potentially influence these counts.

An effective medical response pertaining to modern warfare depends on a multitude of factors such as terrain, type of weaponry, time, medical service availability and nature of operations (https://dx.doi.org/10.1093/milmed/usab108). However, defining and selecting variables posed a challenge due to the unprecedented and inherently unpredictable nature of modern warfare. The Israeli-Palestinian conflict differs from traditional state-on-state warfare in that it is highly asymmetric, irregular, and urbanized [https://www.aljazeera.com/news/2023/11/6/analysis-hamass-asymmetric-warfare-against-israel-lessons-from-ukraine, https://asjp.cerist.dz/en/article/255794], resembling guerilla warfare with fast-moving frontlines, non-conventional combatants, and prominent civilian involvement. Conventional variables typically studied and quantified to predict injury and fatalities numbers not only lack in available open-source data but also struggle to fully capture the unpredictable and decentralized nature of violence in this conflict (<https://doi.org/10.3389/fpubh.2021.765261> ,<https://doi.org/10.1136/bmj.a137>). The lack of available quantifiable variables made it necessary to develop a more flexibleapproach.

Attack type was chosen as a variable due to evidence showing its potential influence on the proportion, scale and severity of injuries and casualties in past conflicts (https://doi.org/10.1186/s13031-020-0252-7, doi:10.1001/jamainternmed.2017.5723, doi: 10.1097/TA.0b013e31829a0970, https://books.google.com/books/about/Analysis\_of\_Casualty\_Rates\_Patterns\_Like.html?id=xDoUzwEACAAJ). Airborne attacks, including drone strikes and bomb shelling, often result in mass injuries and casualties due to their large-area impact, a pattern historically observed in conflicts where air superiority and strategic bombing played a dominant role (doi: [10.1097/TA.0000000000004063](https://doi.org/10.1097/TA.0000000000004063), <https://doi.org/10.1093/milmed/usab108>, <https://doi.org/10.1177/2053168020972816>). On the contrary, ground attacks and combat typically involve more localized engagements, although can nonetheless still result in widespread, severe injury (DOI: 10.1097/01.TA.0000057151.02906.27, doi: [10.1097/TA.0000000000004063](https://doi.org/10.1097/TA.0000000000004063)). An extensive literature review conducted by Aboutanos et. al. found that traumatic injuries afflicted upon civilians in recent wars and conflicts have shown a trend of explosive devices such as artillery shells to have caused the greatest proportion of injuries [https://pubmed.ncbi.nlm.nih.gov/9356079/], a trend that may be exacerbated due to the rise of new advances in weaponry such as improvised explosive devices (IEDs) (<https://doi.org/10.1093/milmed/usac366>). However, in the context of the ongoing Gazan conflict, no existing reports comprehensively analyze the relationship between attack types and injury patterns. Furthermore, the distinct nature of the conflict limits the applicability of prior conflict data on attack patterns, making it challenging to draw direct conclusions for injury count predictions. Distinguishing between attack types also proved challenging due to inconsistent classification in reports. As a result, we standardized ACLED attack event type classifications into categories that serve as meaningful classifications in predicting injury count.

Infrastructure type and population density were also included as key variables because they potentially influence both the likelihood and scale of injuries. Infrastructure type was categorized based on land use classifications (e.g., urban, rural, camp, etc.), while population density was quantified as the number of inhabitants per square kilometer within the affected area.

Infrastructure type was particularly important because different forms of infrastructure are targeted differently in conflict, with varying consequences for civilians. Attacks on residential areas often lead to high injury counts due to the dense presence of civilians, whereas strikes on industrial or critical infrastructure can trigger structural collapses, compounding injury severity and increasing the need for reconstructive surgical intervention. Additionally, the destruction of medical infrastructure, such as hospitals and clinics, exacerbates injury severity by limiting access to emergency treatment, increasing the likelihood of preventable deaths (https://pchrgaza.org/gaza-strips-health-sector-under-israeli-military-aggression-and-closure/). Urban areas, for example, tend to have denser populations and more critical infrastructure (e.g., residential buildings, hospitals, and transportation hubs), making attacks in these locations more likely to result in high injury counts. The London-based NGO Airwars has documented a strong correlation between the use of explosive weapons in urban areas and high civilian casualty and injury rates, as observed through conflict monitoring in Iraq and Syria (<https://airwars.org/wp-content/uploads/2018/05/Airwars-Death-in-the-City-web.pdf>). Conversely, rural or sparsely populated areas may experience attacks that are less deadly in absolute numbers but still devastating due to limited access to emergency medical care and evacuation routes. Surprisingly, Hoglund and colleagues found that armed conflicts are becoming increasingly more prominent in rural locations as opposed to urban regions, leading to greater casualty numbers (https://link.springer.com/chapter/10.1057/9781137550484\_4).

Additionally, infrastructure type may influence injury mechanisms, which is crucial for quantifying and assessing reconstructive surgical needs. For instance, a study by Satanovsky et al. analyzed injury patterns among infantry soldiers in two previous Gaza operations, revealing that lower extremity injuries were more prevalent in open terrain, whereas head trauma and blast injuries were more common in urban environments (<https://doi.org/10.1093/milmed/usac366>). Given that certain injuries necessitate reconstructive surgery more than others, the demand for such procedures is likely not uniform but instead skewed toward injuries associated with specific infrastructure types.

Population density was selected to account for the concentration of civilians in targeted areas, which significantly influences injury numbers. “Conflict exposure” quantifies the number of people who experience exposure to an attack event (https://acleddata.com/conflict-exposure/). High-density population settings tend to amplify the human impact of attacks, as more people are likely to be present at the point of impact (https://unocha.exposure.co/the-impact-of-conflict-in-numbersnbsp). Additionally, these environments often present challenges for emergency response efforts, as damaged roads, blocked evacuation routes, and overwhelmed hospitals can delay life-saving medical care. This is particularly relevant in conflicts where civilian displacement is ongoing, as population density fluctuates dynamically due to mass migration, leading to inconsistencies in injury and casualty estimates.

Given the challenge of obtaining standardized data on infrastructure and population density during an active conflict, we integrated multiple data sources, including satellite imagery and publicly available geospatial datasets, to approximate these variables. This information was then consolidated to align with the relevant time span for analysis. A multiple source/database integration approach is commonly leveraged by several studies in predicting injury or casualties as a result of war (https://doi.org/10.1016/ S0140-6736(24)02678-3, <https://doi.org/10.1073/pnas.2307372120>).

*4.2 Key Correlative Findings from Model*

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 injury 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, despite 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 injuries, 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.

*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 between each variable. Interaction factors that were statistically significant in predicting injury counts were also included as dependent variables in the model.

*4.4. Sensitivity Analysis*

We estimated the sensitivity of each model using +/-10% variable interval for each injury prediction (A.4). Due to the high standard error of each model as well as significant variability and unpredictability of injury count across the board, we observe a wide confidence interval for each model.

*4.5 Comparisons with Injury and Casualty Predictions/Patterns Researched with Other Global Conflicts*

The findings in this study, although revealing pertinent information relevant to the ongoing Israeli-Palestinian conflict and drawing dire attention to the general need for reconstructive surgery treatment in the region, remains unique to the specific conflict itself. Comparisons can be drawn with studies conducted on other global conflicts to contextualize our findings within a broader view of war-related injury patterns and medical needs.

Other research studies also reveal different patterns present in war. Studies on injury and fatality patterns present in other conflicts have also been previously researched. One major finding from this study is that air and shelling attacks comprised the overwhelming majority of attacks on Gaza tracked by the ACLED, likely due to increase use of targeted missile strikes or air drone attacks, showing significance in predicting injury count. However, a systematic review conducted by Khorram-Manesh et. al. on modern military conflicts found that in the most recent conflicts explosive devices and gunshot wounds served as the main cause of injuries and fatalities (<https://doi.org/10.1093/milmed/usab108>). This discrepancy in attack type patterns reveals the unique war situation present in the Israeli-Palestinian conflict, highlighting the irregular warfare present in the region. However, Guha-Sapir et. al. found that aerial bombing and shelling attacks were the predominant cause of death in women and children in the Syrian war (<https://pubmed.ncbi.nlm.nih.gov/29226821/>), alluding to the idea of similar attack and warfare patterns among certain global conflicts.

The data collection performed during the duration of this study also revealed the intrinsic lack of accurate, open-source data on the conflict. Therefore, unconventional methods to obtain enough relevant data of our final variables were performed (e.g. satellite imagery, etc.) to predict injury count. A previous study conducted by Jamaluddine et. al. incorporated data from both the Palestinian Ministry of Health (MoH) and social media obituaries to predict traumatic injury mortality in the Gaza Strip (https://doi.org/10.1016/ S0140-6736(24)02678-3), further illustrating the challenges faced in acquiring comprehensive and reliable datasets in conflict zones. It may be necessary to adopt an approach of integrating multiple data sources to construct a complete picture of injury and mortality trends.

*4.6 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, demographical information, or waiting time for medical assistance (<https://doi.org/10.1093/milmed/usab108>, <https://doi.org/10.1186/s13049-024-01299-7>), all of which may influence injury numbers or the likelihood an injury would require reconstructive care. Regardless, as stated previously, it is difficult to prescribe conventional variables used to predict casualties due to the unique nature of this conflict.

*4.7 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 are difficult to quantify. For example, our analysis of injury rates did not consider the effect of other potentially significant factors such as the availability and accessibility of medical aid, the speed of emergency response, and the influence of international intervention on conflict dynamics. Socioeconomic status, pre-existing health conditions, and patterns of forced displacement (should integrate population flow of refugees and IDPs) may also play a role in determining injury counts. Thought should also be put into methods utilized to predict injury counts in the future. The dynamic landscape pertaining to technological, social, and political change will undoubtedly affect patterns of warfare globally. The evolution of key variables in injury prediction such as medical care, weaponry, and strategic development must also be considered.

Expanding the analysis beyond injury count prediction to assess the influence of proposed factors on injury patterns (e.g., correlations between attack variables and affected extremities) would offer deeper insights. In the context of providing reconstructive surgery care to Palestinians in a war-torn region, assuming a fixed percentage of injuries requiring reconstruction may not accurately capture the complexity and evolving nature of injuries sustained over time.

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.

**5. Conclusion**

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.2 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) | | | |  | | | | |