

Facial Trauma Report

MSSP Consulting G2

December 12 2021

Abstract:

Introduction:

Facial Trauma, also called maxillofacial trauma, involves different facial trauma sites and injury mechanisms. In this report, our main focus is to explore a facial trauma database collected for any patient with facial trauma admitted in Maine Medical Center from 2016 to 2020. The goal of this project is to analyze the relationship between patient's injury sites and residency aiming to design specific facial trauma treatment plans for future patients with respect to their residency, from urban or rural areas; and provide efficient treatment corresponding to their facial trauma injury sites. For better access to facial trauma prediction, we are going to apply the Generalized Linear Mixed-Effect Regression Model which is best-fit for binary outcomes.

Data Cleaning and Processing:

For ICD codes: The major differences between the two coding systems include the number of characters involved. ICD-9 has up to five characters while ICD-10 has up to seven. ICD-10 adds laterality to the coding system, which ICD-9 lacks. ICD-10 offers much more specificity, including episodes of care, body area, etc. Also, the ICD-10 coding structure utilizes a dummy placeholder, which ICD-9 does not. Because the clients plan to analyze the injury area differences in facial trauma patients in urban and rural settings. Transformation of ICD codes is very important.

For cause of injury(Injury mechanism): From the data screening provided by the client, we found that there are 10 levels for injury mech in the dataset and some of them are same at some content, for example, there are several mechanisms for falls but with different falling height. We need to categorize them in order to tidy the dataset for further use. For categorization of ICD10 codes: Now we have the ICD10 codes, but what we want to study is the relationship between areas of facial trauma instead of the codes. So we need to categorize the codes to specific areas of injury: as the client instructs, we categorize the codes into three levels: Midface, Mandible and Superior face for further use.

When we want to do ICD Code transformation Among 318 patients, we found that 158 of them are recorded with ICD 9, while 160 take ICD 10. When the client tried to convert ICD 9 code into ICD 10 code. He encountered two major problems: A) Not Convertible: 75 out of 432 ICD 9 code can not be converted to ICD 10 code. B) Multiple Converted: 143 of the ICD 9 codes can be converted into multiple ICD 10 codes. In order to solve these two major problems. We plan to solve B first. After solving B, we can drop the codes mentioned in A and only leave the useful observations for further analysis.

To solve problem A. For the tidy dataset so far, we now have 211 patients with corresponding injury areas. For the rest of the observations, our client David conducted a more detailed investigation and he gave a table for the corresponding convertible codes for us to do the conversion. After finalizing the conversion, we are now having 292 observations for the further analysis. To solve problem B, we first let all the missing values be blank. And use for loops for each ICD9 code column, when the ICD9 column has the code that can be converted to useful ICD10 codes, then we assign the converted ICD10 codes to the corresponding ICD10 code column. Categorization of codes (ICD10) Among 318 patients, 311 of them have explicit injury mechanism

records, 7 of them are blank or not clarified clearly. For 311 patients who have records, we extracted the key words of descriptions, and divided them into 6 categories:

Converted injury mechanism categories	Original injury mechanism
MVC	Motorcycle, MVC
Falls	Fall under 1m, Fall 1m-6m, Fall over 6m
Gun	Hand gun, Shotgun
Others	Other, Assault, Biting, Pedestrian, 0
Bicycle	Bicycle
Other_Blunt	Blunt Mechanism

Data Description:

Variable	Explanation	Description
Study.ID..	ID for Patients	Total 292 observations left
Age	Age information for patients	From 17 to 90
Gender	Gender information for patients	Male, Female
Race	Ethnicity information for patients	Asian, White, African American and Other Races
Urban.Rural	Living area information for patients	Urban and Rural
Superior	Dummy variable to check if patient was hurt on Superior site	Binary
Mandible	Dummy variable to check if patient was hurt on Mandible site	Binary
Midface	Dummy variable to check if patient was hurt on Midface site	Binary
Otherff	Dummy variable to check if patient was hurt on other face trauma site	Binary
Injury.Mech.category	Variable to check what mechanisms caused facial trauma	6 mechs as showed above

In order to show the information of the data set and the relationship between various variables more intuitively, let's move to our EDA part:

EDA:

After data cleaning and processing, we started our exploratory data analysis. As shown in Figure 1, 2, 3, and 4, there are more data samples collected from midface injury patients than the other three facial trauma sites. Among all four facial trauma sites, falls and motor vehicle accidents are the two major types of injury mechanisms for both residency areas. It is hard not to notice that people in rural areas have more facial trauma caused by MVCs than people with facial trauma in urban areas. In addition, we have a small sample size for the injury mechanism of "Gun" which makes sense because gunshot accidents are rare in our daily life and gunshot accidents can cause immediate death that do not even have a chance to be admitted into hospital for treatment and gunshots are majorly targets on midface and superior.

Figure 5 is a bar plot which compares the different population sizes of people with facial trauma and bars are filled with color coded facial injury mechanism categories. From the bar plot, we can see that the database consists of more male patients than female patients. Falls and motor vehicle accidents are the two major types of injury mechanisms for both female and male patients.

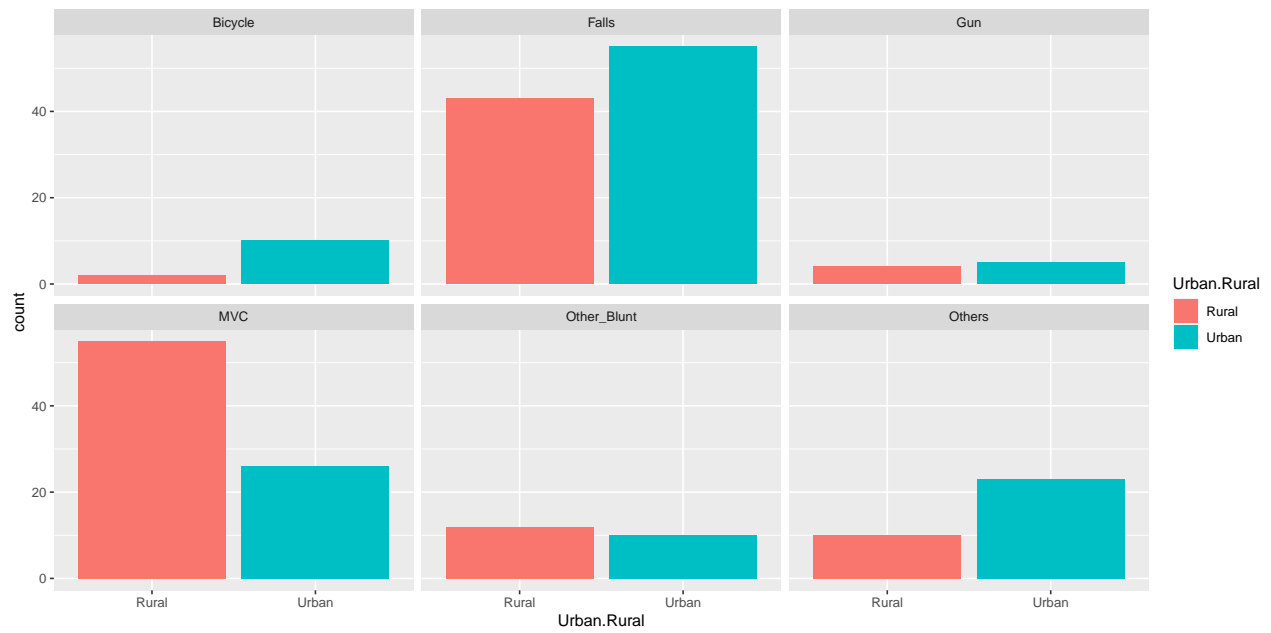


Figure 1: Midface area injured patient number vs. demographic information

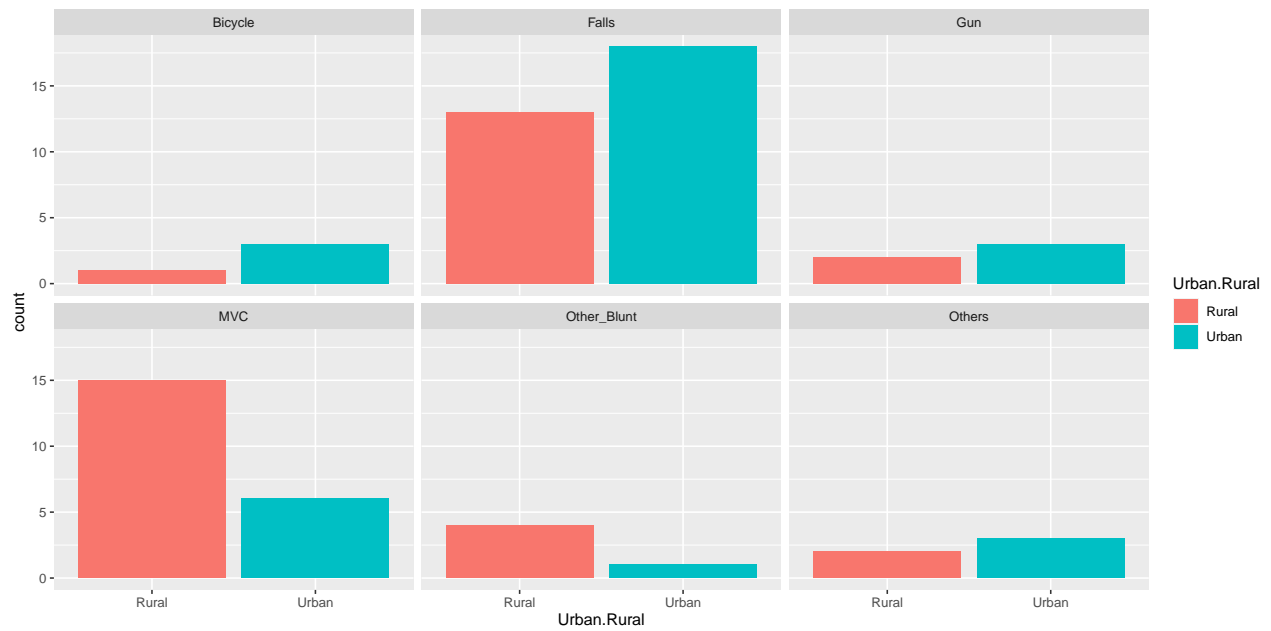


Figure 2: Superior area injured patient number vs. demographic information

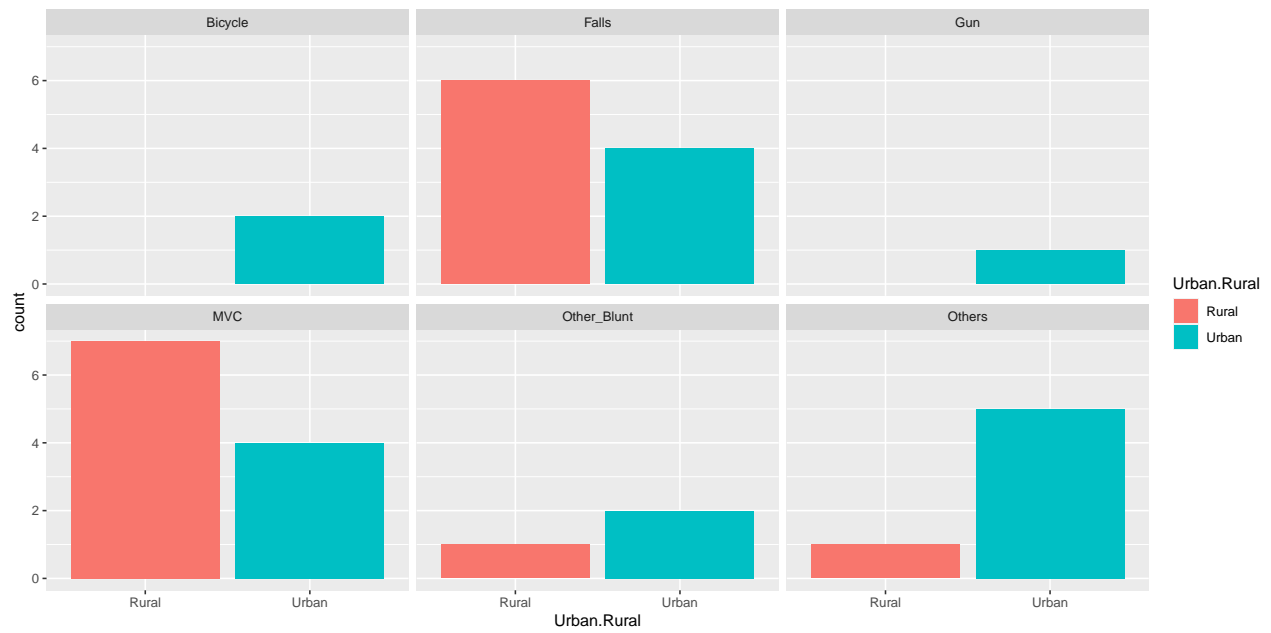


Figure 3: Mandible area injured patient number vs. demographic information

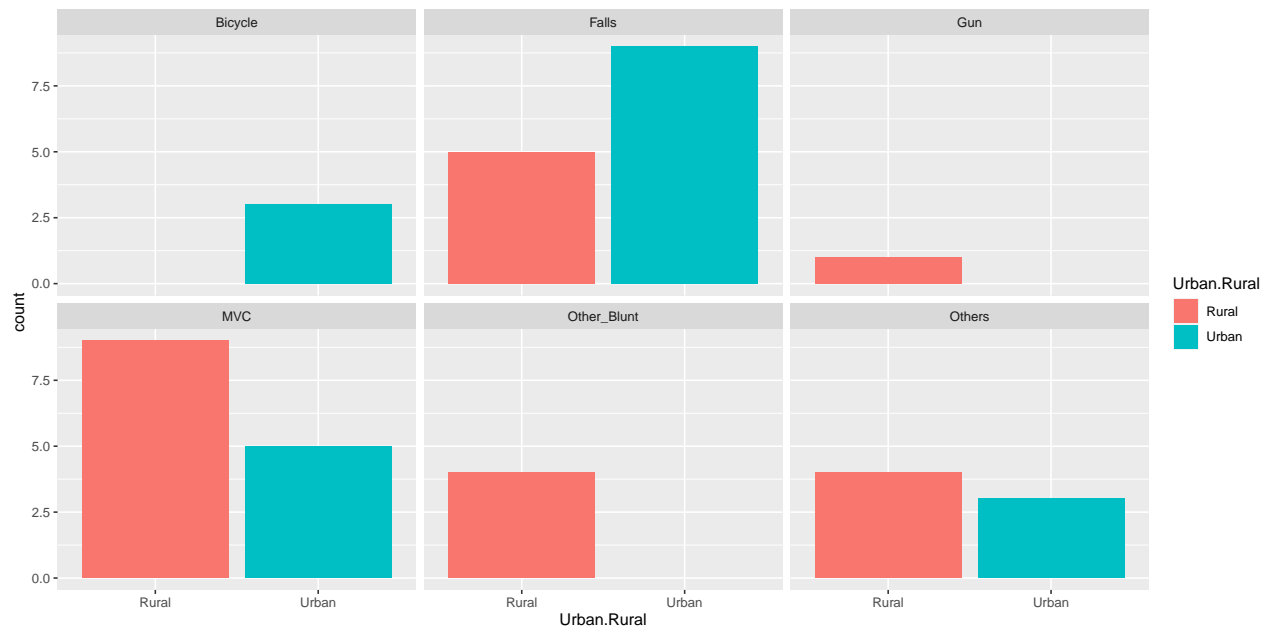


Figure 4: Other facial trauma sites injured patient number vs. demographic information

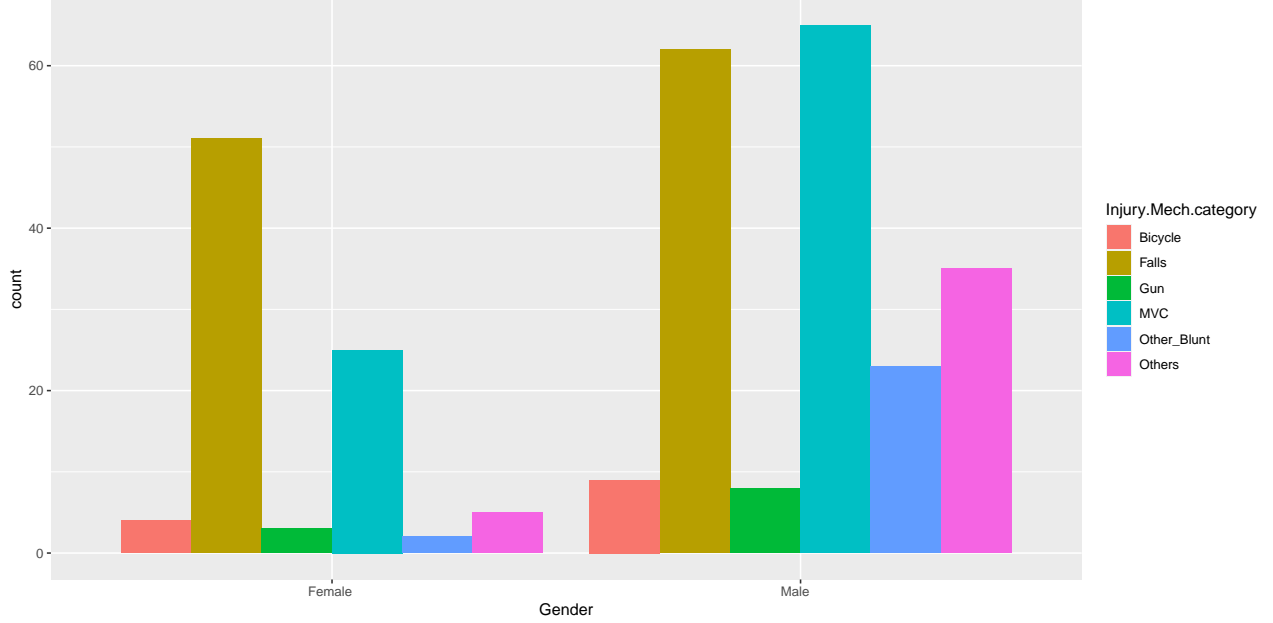


Figure 5: Distribution for injury mech for different gender

Model Fitting:

Based on finding from our EDAs, we use the Generalized Linear Mixed-Effects Regression Model to estimate the mixed effects of getting facial trauma in one of four trauma sites using “Age” and “Gender” as the predictors, add “Urban.Rural” for the random slope effect, and use “Injury.Mech.category” as the random intercept. The function of the model is shown below:

```
stan_glmmer("Trauma site" ~ as.numeric(Age) + Gender + Injury.Mech.category + (1+Urban.Rural|Injury.Mech.category),
data = cleaned_set, family = binomial(link="logit"), refresh=0)
```

By arranging all four fitted models into one plot, we can see the same model generates different results in different trauma sites. The mixed-effects have greater impacts on superior and midface trauma sites. One reason that can cause this discrepancy is that midface and superior have bigger sample sizes than mandible and other. More information can be found in the four fixed effect table of the models. We also need validation to further assess the effectiveness of this model.

Generalized Linear Mixed-Effects Regression Models are validated by posterior predictive checks and binned residual plots. The posterior predictive checks plot displays the comparison distribution of the observed outcome of trauma site(y) and simulated outcome of trauma site(y_rep) in which both distributions match well in the pattern, but simulated y_rep show more predicted zero which means zero inflation occurred except for the midface. On the other hand, binned residual plots show that some points lie outside the confidence limits for the mandible site and there is a pattern in the plots for the superior facial trauma site. Therefore, the generalized linear mixed-effects model is not the best-fit in assessing the prediction of relationship between all four trauma sites and patients residencies. Other models should be taken into consideration.

Table 3: Fixed effects of Mandible

	x
(Intercept)	-0.7863484
as.numeric(Age)	-0.0193353
GenderMale	-0.5042826
Injury.Mech.categoryFalls	-0.1202088
Injury.Mech.categoryGun	-0.6954799

	x
Injury.Mech.categoryMVC	-0.2015606
Injury.Mech.categoryOther_Blunt	-0.2229458
Injury.Mech.categoryOthers	0.0821334

Table 4: Fixed effects of superior

	x
(Intercept)	-0.8256959
as.numeric(Age)	-0.0116207
GenderMale	0.6216181
Injury.Mech.categoryFalls	0.1838416
Injury.Mech.categoryGun	0.6942662
Injury.Mech.categoryMVC	-0.4185096
Injury.Mech.categoryOther_Blunt	-0.7744858
Injury.Mech.categoryOthers	-1.2905018

Table 5: Fixed effects of midface

	x
(Intercept)	1.6559658
as.numeric(Age)	0.0320454
GenderMale	-0.3130533
Injury.Mech.categoryFalls	-1.4296686
Injury.Mech.categoryGun	-1.1207503
Injury.Mech.categoryMVC	-0.2485157
Injury.Mech.categoryOther_Blunt	-0.3606297
Injury.Mech.categoryOthers	-0.9688617

Table 6: Fixed Effects of other

	x
(Intercept)	-1.5076333
as.numeric(Age)	-0.0050647
GenderMale	0.4035537
Injury.Mech.categoryFalls	-0.4936797
Injury.Mech.categoryGun	-1.2055677
Injury.Mech.categoryMVC	-0.3561549
Injury.Mech.categoryOther_Blunt	-0.2792595
Injury.Mech.categoryOthers	-0.1184003

Result:

From the model fitting, we can do some preliminary inferences as below:

Firstly, a patient's age has no influence on facial trauma sites, which means, people with different ages have the same possibility of getting hurt on different facial trauma sites.

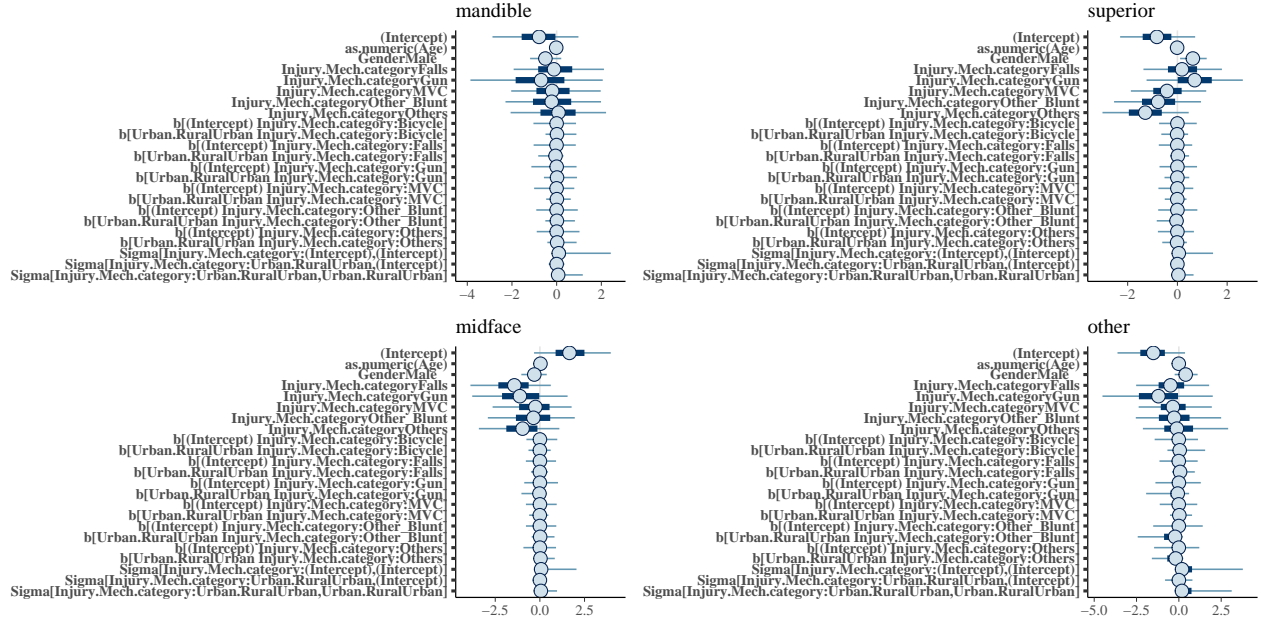


Figure 6: Plots of model fitting for four trauma sites

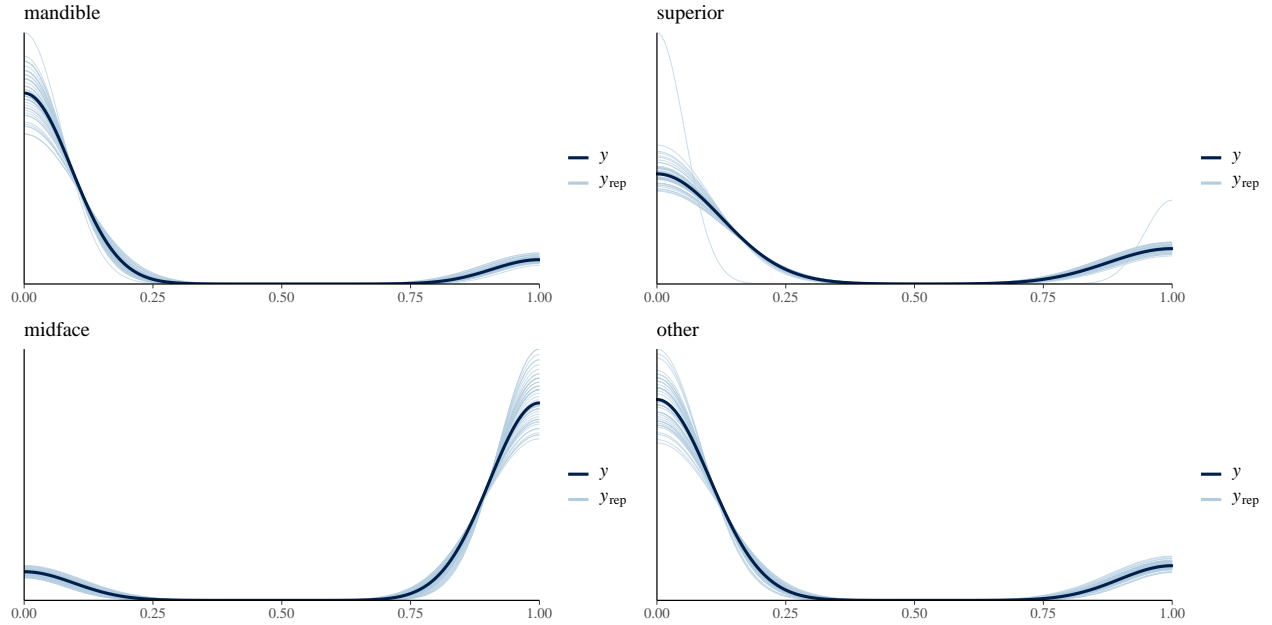


Figure 7: Posterior Predictive Checks Plots for Four Trauma Sites

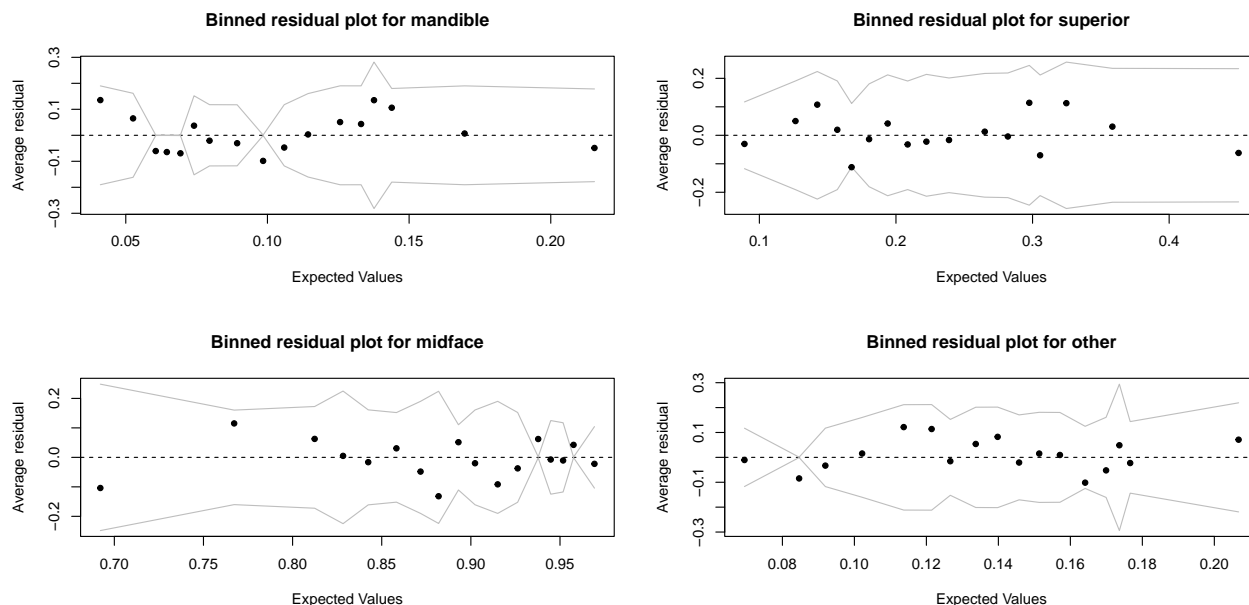


Figure 8: Binned Residual Plots for Model Fitting

Secondly, for patients with the same injury mechanism and from the same residency, men have 16% more possibility to get hurt at superior and 10% at other facial sites compared with women. Similarly, women are easier to get hurt at mandible and midface, the differences are 12.5% and 7.5%.

Thirdly, injury mechanisms also show their impacts on facial trauma sites. For patients with the same gender and from the same residency, they are 17% more likely to get hurt at superior face if they are hurt by a gun. For patients hurt by falls, they are 37% less likely to get hurt at midface and 12% at other types of facial trauma sites. For patients hurt because of motorcycles and other blunts, they have 15% and 21% less possibility to get hurt at superior sites, and show slightly less possibility at mandible and midface. For patients with other injury mechanisms in our sample, they show 33% and 30% less possibility of getting hurt at superior and midface.

Fourthly, for patients with the same gender and same injury mechanism, if the accidents occur in urban areas, patients have approximately 7% more likely to get hurt at mandible face, and 5% or so less likely at superior face.

Discussion:

Phenomena and future Steps:

Firstly, the data we used was collected from one hospital in one state. We understand that this data collection approach will allow hospitals to tailor treatments for facial injury patients from different residential areas to local characteristics in Maine. However, if our client wants to extend this method of improving diagnostic efficiency to the entire East Coast and even the entire United States, we suggest that we can select representative hospitals in each state to collect data and conduct more overall data analysis and modeling.

Secondly, after we cleaned the data set we got, we found that there were only 3-4 factors we can use for building models for checking the relationship between patients' living areal information and the facial trauma sites they got hurt. And we also noticed the information about the patient's place of residence. Our customer divides it into urban and rural areas. We understand that in this way, we can more efficiently derive the relationships that our customers want. But this will also limit our model construction and overly generalize the patient's residence information. Because the diversity of residence will also have a certain impact on the injured area of the patient's face. If our clients want to allow the hospital to formulate more accurate

treatment methods for patients with facial trauma from different living areas, we suggest that we can add some residential information.

Last but not least, we also found some imbalances in the amount of data caused by objective factors. For example, the number of patients with injuries to the middle face and upper forehead is much more than in the other two weeks. And because of the regional characteristics of Maine, the number of patients injured by guns and bicycles is far less than the number of patients caused by other reasons. These are imbalances that are hard to avoid. The positive aspect is that the imbalance of these data volumes can reflect local characteristics, and the negative aspect is that it will cause some obstacles to model selection. Our suggestion is that, if possible, we hope that more state data can be collected in the future to make our research results more convincing. Because of the limited time, we did not try more statistical models, but based on the current data situation. Knowledge of machine learning should be helpful.

References:

1. <https://icdcodelookup.com/icd-10/codes>