

# Investigating the Features of Lawsuit Cases to Predict the Size of Total Damage Awards

To Jared the Client

*Mengran He*

*16 December, 2016*

## Introduction

There has been some public outcry about the total amount of damages awarded to plaintiffs in some civil lawsuit cases, as some plaintiffs are awarded large amount of money while some are awarded very little even none. This project analyzed the features of different civil lawsuit cases, for the goal of finding any predictabilities to the size of total damage awards, including whether or not they are zero. We especially focused on whether a more recent trial, a corporation as the defendant, and an increased number of plaintiffs are predictive of a higher probability that damages are awarded, and also higher damages. Plus, whether bodily injury claim will augment the effect of damages requested by plaintiff on the damages awards is also part of our interests.

## Methods

The data collection was carried out by the National Center for State Courts for the goal of investigating what features of a particular case are predictive of the probability that damages are awarded and the size of total damage awards. To measure the characteristics of each case, the data included **demanded** total amount of damages requested by plaintiff (in \$), **tridays** how many days the trial lasted, **bodinj** whether or not a bodily injury was part of the claim (1: "Yes", 0:"No"), **decorp** whether or not the defendant was a corporation (1: "Yes", 0:"No"), **degovt** whether or not the defendant was the government (1: "Yes", 0:"No"), **year** year the civil lawsuit was filed (from pre-2017 to 2001), **claimtype** type of claim (1: motor vehicle, 2: premises liability, 3: malpractice, 4: fraud, 5: rental/lease; 6: other), **totalnopl** total number of plaintiffs(1: one plaintiff, 2: two plaintiffs, 3:  $\geq 3$  plaintiffs), **totalnode** total number of defendants(1: one defendant, 2:two defendants, 3:  $\geq 3$  defendants), as well as **totdam** total amount of damages awarded to plaintiff (in \$). In addition, we added another binary variable **award** representing whether the damages were awarded. The total amount of damages awarded and total amount of damages requested are considered continuous variables as they measure the value of damages. The day the trial lasted is also continuous variable as it represents a period of time. The year the civil lawsuit was filed is initially categorical variable, but we tried to measure the recency of civil lawsuit, the year the civil lawsuit was converted as continuous variable.

In order to measure the predictability of higher probability that damages are awarded and how much the damages are awarded, we built two types of model. The first one is logistic regression model to predict whether any damages are awarded, considering the log odds of probability that the damages were awarded as dependent variable, the second one is multiple linear regression model to predict how much the damages are awarded, considering **totdam** total amount of damages awarded as dependent variable. We first made exploratory data analysis on the dataset, trying to measure any indications and relationships among variables. Then, for each type of models, we fitted several models, conducted corresponding diagnostic analysis including identifying any potential influential points/outliers, measuring residuals and linear model assumptions. Also, we determined if transformation on response and predictor variables were necessary for model fitting. After we generated our best models, we analyzed the variables' effects on response variables and compare with our initial hypotheses for predicting higher probability of whether damages are awarded and how much damages are awarded.

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	0.4353	0.4499	0.97	0.3333
log(demanded)	-0.0280	0.0276	-1.01	0.3102
tridays	0.0383	0.0180	2.12	0.0338
bodinjl	-0.8937	0.2190	-4.08	0.0000
decorp1	0.0926	0.1172	0.79	0.4294
degovt1	-0.5226	0.2339	-2.23	0.0254
year	0.1817	0.0468	3.89	0.0001
claimtype2	-0.4064	0.2051	-1.98	0.0475
claimtype3	-1.2930	0.2508	-5.16	0.0000
claimtype4	-0.8831	0.3036	-2.91	0.0036
claimtype5	-0.4924	0.3323	-1.48	0.1384
claimtype6	-0.4151	0.2292	-1.81	0.0701
totalnopl	0.1692	0.0973	1.74	0.0821
totalnode	0.0941	0.0766	1.23	0.2193

Figure 1: Summary of logistic regression model for predicting if any damages are awarded

## Results

There are 1836 civil lawsuit cases in the dataset, 64% of them received damage awards (1175 out of 1836 cases). Our exploratory data analysis suggested that as the number of plaintiffs increased or the number of defendants increased, the average amount of damages awarded and the average amount of damages requested also increased. There are 1076 out of 1836 trials lasting 2 or less days, the distribution of day the trials lasted is heavily right skewed. Furthermore, among 1175 cases where damages were awarded to plaintiffs, 411 cases included bodily injury and 764 cases didn't include body injury; 27 cases where the defendants were corporation and 548 cases where defendants were not corporation; 267 cases were motor vehicle claims, 76 cases were premises liability claims, 36 were malpractice, 81 were fraud, 67 were rental/lease, and 648 are others; 26 cases had 1 plaintiff, 181 cases had 2 plaintiffs, 68 cases had 3 or more plaintiffs; 642 cases had 1 defendant, 363 cases had 2 defendants, 170 cases had 3 or more defendants. As we can see, the number of cases which didn't include bodily injury but received damages awards is more than the number of cases which included bodily injury and received damages awards. The number of cases which didn't have corporation as defendants but received damages awards is more than the number of cases which had corporation as defendants and received damages awards. The number of cases which had 2 plaintiffs and received the damages awards is greater than the number of cases which had 3 or more plaintiffs and received the damages awards. We would compare these results with our model suggestions.

Furthermore, the total amount of damages requested by plaintiff had minimum value \$250, maximum \$100000000, average \$1030000 and median \$60290, which suggests that the mean is far greater than the median, the distribution of damages requested by plaintiff is right skewed. The potential outlier is case 965: plaintiff requested \$100025000 but no damages was awarded, trial lasted 2 days, a bodily injury was included in premises liability claim, the lawsuit was filed in 1998, and one plaintiff and more than 3 defendants were included. Similarly, the total amount of damages awarded to plaintiff had minimum \$0, maximum \$44970000, average \$213100 and median \$7795, which means the distribution of damages awarded to plaintiff is also right skewed. The potential outlier is case 266: more than three plaintiffs requested \$62000000 and they were awarded \$44968563, the trial lasted 1 day, no bodily injury was included in fraud claim, the lawsuit was filed in 1998, and there were 2 defendants. We also found out that the total amount of damages awarded to plaintiff is not linearly related to the total amount of damages requested by plaintiff. By taking logarithms transformation on both variables, their relationship appeared to be positively linear but had noises included.

After diagnostic analysis and transformation on predictor variable, the best model of predicting whether any damages are awarded is  $\log \frac{p}{1-p} = \beta_0 + \beta_1 \log(\text{demanded}) + \beta_2 \text{tridays} + \beta_3 \text{bodinj} + \beta_4 \text{decorp} + \beta_5 \text{degovt} + \beta_6 \text{year} + \beta_7 \text{claimtype} + \beta_8 \text{totalnopl} + \beta_9 \text{totalnode} + \epsilon$ , as shown in figure 1, where response variable is log odds of probability that the damages are awarded. The estimated coefficients of days the trial lasted, whether a bodily injury was included, whether the defendant was the government, the year the civil lawsuit was filed

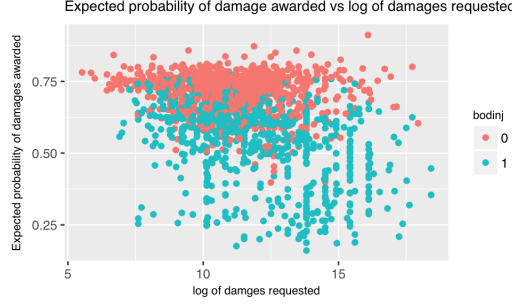


Figure 2: Expected probability of damages awarded vs log(demanded) with inclusion of bodily injury claim (1:Yes, 0:No)

and claimtype are statistically significant as their P-values are less than  $\alpha = 0.05$ . When the days the trial lasted one day longer, the odds of probability that the damages will be awarded is expected to increase by 3.9%. The odds of the probability that the damages will be awarded is expected to decrease by 59% once the bodily injury is included in the claim. The odds of the probability that the damages will be awarded is expected to decrease by 41% if the defendants was the government. The odds of probability of damages awarded is expected to increase by 19.9% if the trial is more recent. Besides, the coefficients of the number of plaintiff and whether the defendant is corporation are not statistically significant in this model. As shown in the dataset, the cases where defendants were not corporation but received damage awards were more than the cases where defendants were corporation and received damage awards. Also, the cases which had 2 plaintiffs and received damage awards were more than the cases which had 3 or more plaintiffs and received damage awards. Figure 2 suggests that the log of damages requested may have smaller effect on the probability of damages awarded if there is a bodily injury, as shown in green dots, even though log of damages requested alone doesn't significantly affect the probability of damages awarded.

Next, the linear regression model for predicting how much the damages are awarded:  $\log(\text{totdam}) = \beta_0 + \beta_1 \log(\text{demanded}) + \beta_2 \text{tridays} + \beta_3 \text{bodinj} + \beta_4 \text{decorp} + \beta_5 \text{degovt} + \beta_6 \text{year} + \beta_7 \text{claimtype} + \beta_8 \text{totalnopl} + \beta_9 \text{totalnode} + \beta_{10} \log(\text{demanded}) \times \text{bodinj} + \epsilon$ , shown in figure 3. This model is generated by log-transforming both predictor and response variable in order to fix problem of unequal variances, outliers and violation of normality assumptions. The intercept, the log of damages requested by plaintiffs, the days that trial lasted, the inclusion of bodily injury claim, the types of claim, and the interaction between the log of damages requested and whether the bodily injury was included are statistically significant. If a bodily injury was not part of the claim, the total amount of damages awarded to plaintiff is expected to increase by 6.79% given 10% increase in total amount of damages requested by plaintiff. If a bodily injury was included as part of the claim, the total amount of damages awarded is expected to increase by 4.6% given 10% increase in total amount of damages requested by plaintiff. Figure 4 supports this indication, as we can see that given the same log of damages requested by plaintiff, the expected value of log of damages awarded to plaintiff when bodily injury was included (shown in blue dots) is lower than those when bodily injury wasn't included (shown in red dots). For the days the trial lasted, for one day increase in the days the trial lasted, the total amount of damages awarded is expected to increase by 11.42%. For the type of claim, motor vehicle claim as reference, the amount of damages awarded is expected to increase by 84% when shifting from motor vehicle claim type to premises liability type, increase by 107% when shifting to malpractice type, increase by 70.8% when shifting to fraud type, increase by 69.76% when shifting to rental/lease type, increased by 71.3% when shifting to other claims. The number of plaintiffs, whether the defendant was corporation, and the year the lawsuit was filed are not statistically significant in this model.

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	2.1290	0.4361	4.88	0.0000
log(demanded)	0.6890	0.0291	23.68	0.0000
tridays	0.1081	0.0128	8.44	0.0000
decorp1	0.1374	0.0881	1.56	0.1192
degovt1	0.1971	0.2138	0.92	0.3566
year	-0.0546	0.0376	-1.45	0.1463
bodinj1	2.0391	0.5236	3.89	0.0001
claimtype2	0.6094	0.1853	3.29	0.0010
claimtype3	0.7273	0.2476	2.94	0.0034
claimtype4	0.5355	0.2483	2.16	0.0313
claimtype5	0.5292	0.2583	2.05	0.0407
claimtype6	0.5380	0.1969	2.73	0.0064
totalnopl	0.0644	0.0704	0.91	0.3607
totalnode	0.0283	0.0577	0.49	0.6233
log(demanded):bodinj1	-0.2172	0.0412	-5.28	0.0000

Figure 3: Summary of multiple linear regression model for predicting predicting how much the damages are awarded

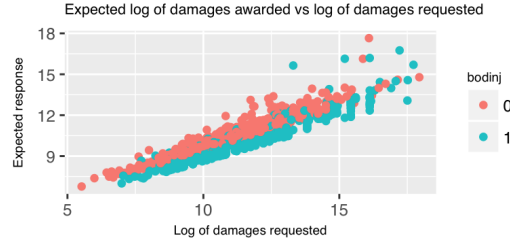


Figure 4: Expected response vs log(demanded) with inclusion of whether bodily injury claim (1:Yes, 0:No)

## Discussion

Based on the data collected by the National Center for State Courts, we found a positively linear relationship between the log of damages requested and the log of damages awarded. When analyzing the predictability of whether the damages are awarded, we concluded that a more recent trial or longer lasting trial are predictive of a higher probability that damages are awarded, and these results are consistent with our initial hypothesis. However, the effects of whether a corporation was the defendant or an increased number of plaintiffs don't have strong indication on the higher probability that damages are awarded which was assumed by the advocacy watchdog group. In addition, the inclusion of bodily injury claim reduces the effect of log of damages requested on the probability of damages awarded, which contradicts to our initial hypothesis.

When analyzing the predictability of how much the damages are awarded, we concluded that the increase of damages requested by plaintiff and longer lasting trial days and malpractice as claim type are predictive of higher damages awarded, which agree with our initial hypothesis. However, the amount of damages requested may have smaller effect on the amount of damages awarded, when bodily injury is part of the claim, which contradicts to our initial hypothesis. Whether the defendant is corporation and increased number of plaintiffs are not as predictive of higher damages awarded as hypothesized initially. This research implies that a bodily injury claim might alter the relationship between total damages and amount of damages demanded in a negative way.

The variable describing the year the civil lawsuit was file is from pre-1997 to 2001, which is 20 years ago, so the lawsuit cases might be outdated to analyze current trend in damage awards. Also, variables of total number of plaintiffs and total number of defendants only have three levels(one individual, two individuals, 3 or more individuals), so it is insufficient to identify if an increasing number of plaintiffs/defendants affect the damage awards. For future study, we should include recent lawsuit cases and change the way of recording total number of plaintiffs and defendants.