



# Predicting the Outcome of Malpractice Lawsuits

BUDT733: Fall 2005

Andrea Chod

David Hofberg

John Jepsen

Christine Nguyen

Eric Zirotsky



# Objectives

- To predict the outcome of a malpractice suit that did result in a payout - settlement or judgment.
- To predict the relative size of a malpractice suit payout.



# Data Set

- **Data Source:** National Practitioner Data Bank Healthcare Integrity and Protection Data: National Practitioner Data Bank [http://www.npdb-hipdb.com/pubs/stats/Public\\_Use\\_Data\\_File.pdf](http://www.npdb-hipdb.com/pubs/stats/Public_Use_Data_File.pdf)
- **Sample Size:** ~ 3500 Cases
- **Data:**

Report Type	Region	License Field	Practitioner Age	Practitioner Grad	Allegation Group	Payment cat	Judgement/Se	Patient Age	Patient Ge	Nbr of Mal
101	South Centra	Nurse/Assistar	40	1990	10	low	S	70	F	1
102	Midwest	Pharmacy	40	1980	30	low	S	50	M	1
102	South Centra	Pharmacy	30	1980	30	low	S	50	F	1
102	South Altant	Pharmacy	20	2000	30	low	S	60	F	1
101	Midwest	Physician	40	1980	90	low	J	20	F	3
102	West	Pharmacy	30	1990	30	low	J	70	F	1



# Data Reduction

- Logic/Knowledge
  - Orig year: We are analyzing the most recent data, which is 2004-2005, therefore there is not much value to this variable
  - Wrk country, home state, home country: We kept the work state variable because it is the same concept as the other three variables but more applicable
  - Alleg 1 and 2: These variables give more specifics about the type of allegation, for our purpose the allegation variable is specific enough
  - Basis for action: The format and meaning of this variable changed from the way it was captured in the beginning of the database and now, therefore, it is not as meaningful as the other variables which have been consistent throughout
- Incomplete



# Predictors

- **Report Type:** Type of report filed – insurance or non-insurance
- **Region:** Paired 50 states into 5 regions
- **License Field:** 9 fields of practice (e.g. psychiatry, dental, etc.)
- **Practitioner Age:** Age at time of event
- **Practitioner Graduation Year**
- **Allegation Group:** Type of malpractice involved (e.g. diagnosis, anesthesia, surgery, etc.)
- **Payment:** Amount of specific payout
- **Judgment/Settlement**
- **Patient Age:** Age at time of event
- **Patient Gender**
- **Number of Prior Malpractice Payments**



# What are the Response Variables?

- Judgment/Settlement: Useful for matters of professional reputation.
- Pay Range: New data field created. Distinguishes whether a case payment is over (high) or under (low) \$100 thousand (roughly the average payment of all cases). Units changed to US\$ million.



# Initial Thoughts from Data Analysis

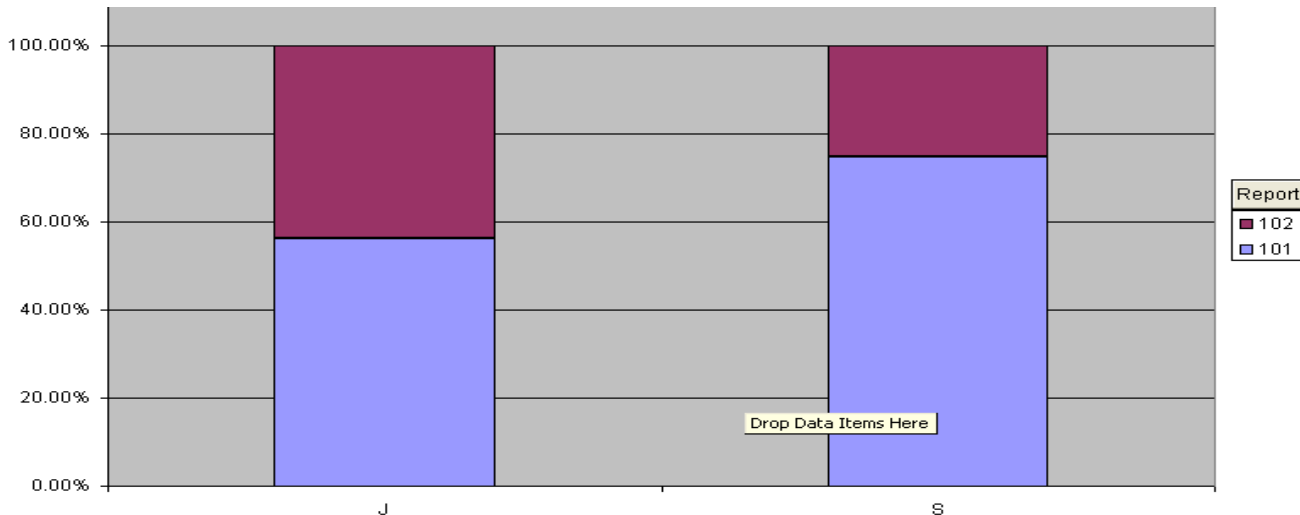
- Allegation Types
  - Out of all settlement cases diagnosis-related cases are the most common
  - Out of all judgment cases (go to court) surgery-related cases are the most common
  - Equipment-related cases have a higher probability of resulting in a low payment, while obstetrics-related cases have a higher probability of resulting in a high payment

Count of Allegation Group	Allegation Group											
Judgement	1	10	20	30	40	50	60	70	80	90	100	Grand Total
J	21.14%	0.81%	40.65%	4.88%	1.63%	3.25%	23.58%	1.63%	0.00%	2.44%	0.00%	100.00%
S	32.12%	3.10%	25.24%	6.24%	0.20%	2.72%	23.48%	4.19%	0.70%	1.79%	0.20%	100.00%
Grand Total	31.74%	3.02%	25.78%	6.19%	0.25%	2.74%	23.49%	4.10%	0.68%	1.81%	0.20%	100.00%

Count of Allegation Group	Allegation Group											
Pay_Range	1	10	20	30	40	50	60	70	80	90	100	Grand Total
High	64%	61%	59%	43%	11%	66%	41%	50%	8%	27%	29%	54%
Low	36%	39%	41%	57%	89%	34%	59%	50%	92%	73%	71%	46%
Grand Total	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

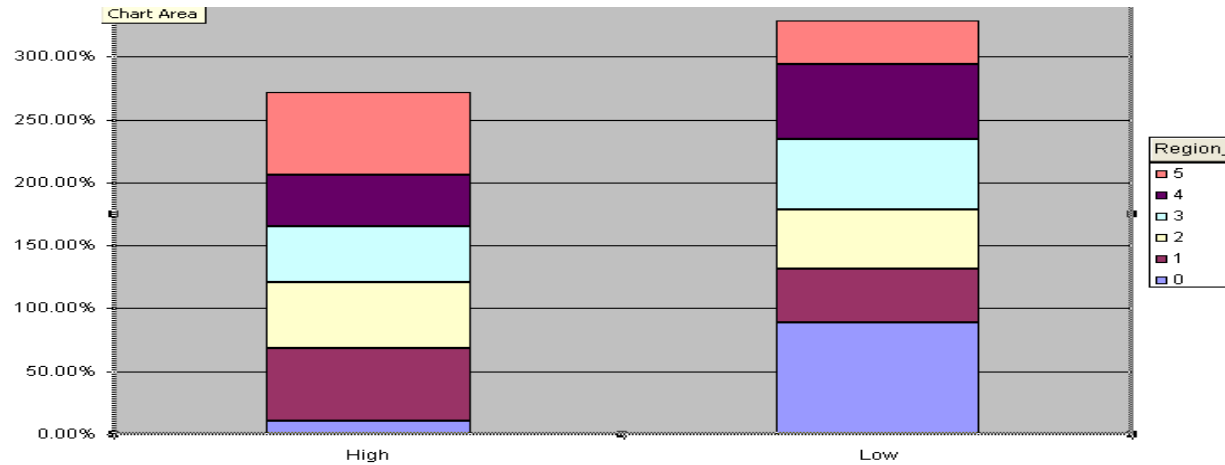
- Report Type
  - More likely to result in settlement if filed by insurance or in court if non-insurance



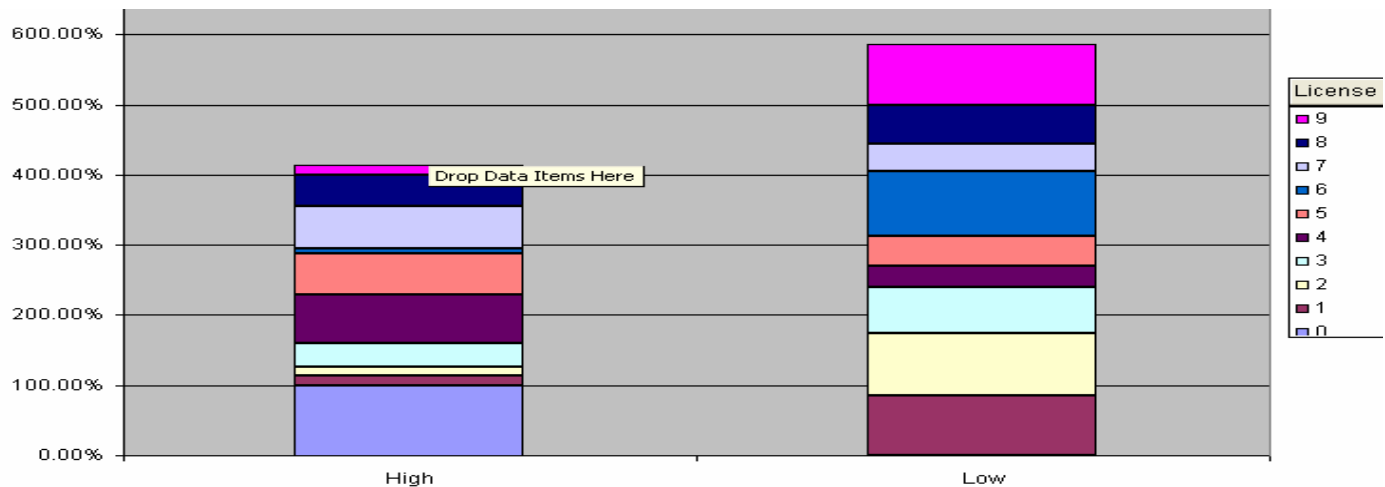


# Initial Thoughts from Data Analysis

- Region
  - Region 0 is more likely to result in a low payment
  - The Midwest & Northeast regions are more likely to result in high payment



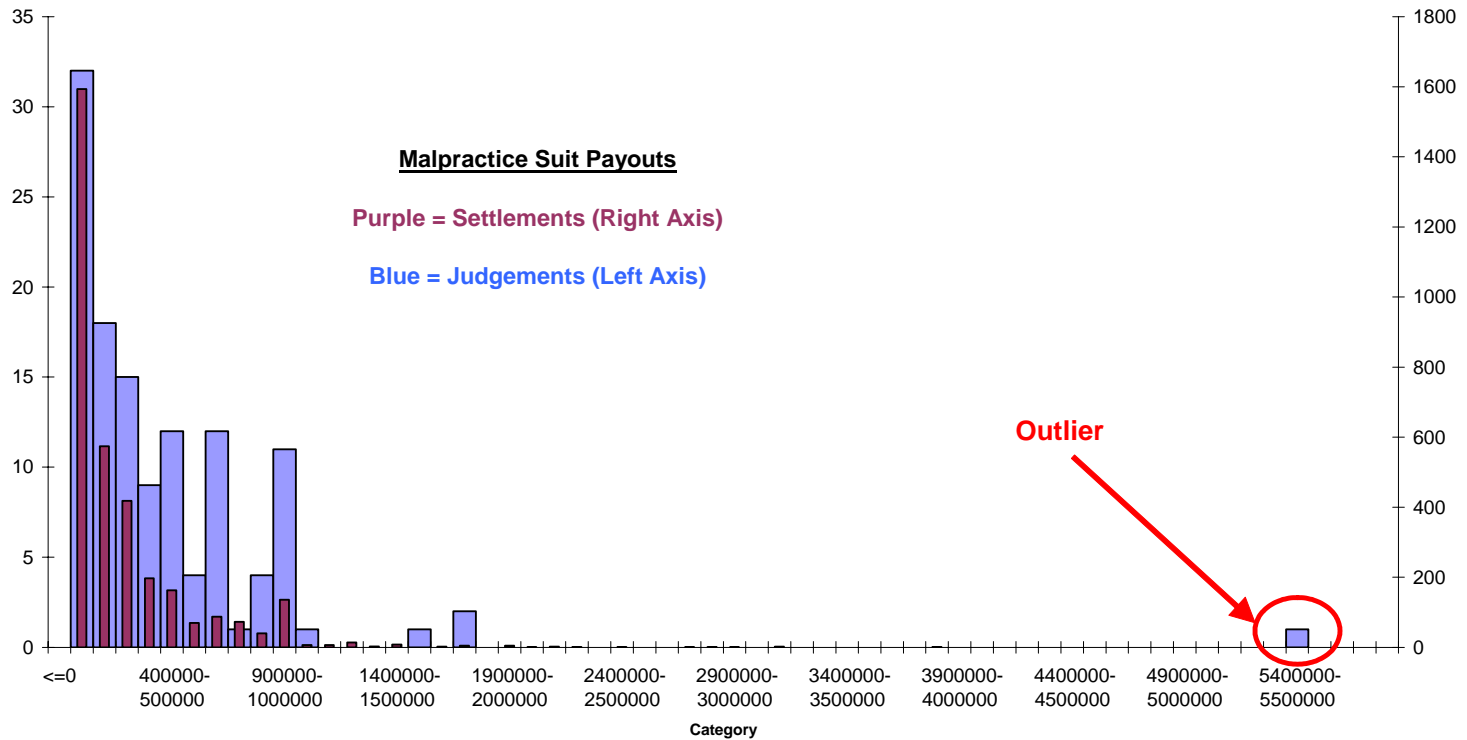
- License Field
  - Nutrition is nearly 100% likely to result in a high payment
  - Therapy, Pharmacy & Recreation will most likely result in a low payment





# Judgment/Settlement Histogram of Payouts

Histogram for Payment (\$)





# Analysis Leads to a More Compelling Question

- Our original question lead us towards a more compelling hypothesis
- Regardless of a settlement or judgment, can we predict whether a guilty verdict will result in a high or low pay out
- Reason:
  - High costs of malpractice suits can be offset if there is a high chance of paying out a large sum.
    - If we can predict that there will be a large payout, it may be better to settle early and pay less.



# Pay Range Validation Set Confusion Matrices

## Logistic Regression

### Training Data scoring - Summary Report

Cut off Prob.Val. for Success (Updatable) **0.5**

Classification Confusion Matrix		
	Predicted Class	
Actual Class	Low	High
Low	418	568
High	193	944

Error Report			
Class	# Cases	# Errors	% Error
Low	986	568	57.61
High	1137	193	16.97
<b>Overall</b>	<b>2123</b>	<b>761</b>	<b>35.85</b>

## Discriminant Analysis

### Training Data scoring - Summary Report

Cut off Prob.Val. for Success (Updatable) **0.5**

Classification Confusion Matrix		
	Predicted Class	
Actual Class	Low	High
Low	385	601
High	154	983

Error Report			
Class	# Cases	# Errors	% Error
Low	986	601	60.95
High	1137	154	13.54
<b>Overall</b>	<b>2123</b>	<b>755</b>	<b>35.56</b>

### Validation Data scoring - Summary Report (Using Best Pruned Tree)

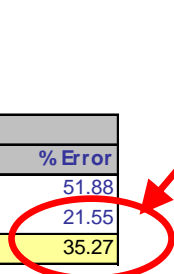
Cut off Prob.Val. for Success (Updatable) **0.5**

Classification Confusion Matrix		
	Predicted Class	
Actual Class	Low	High
Low	308	332
High	167	608

Error Report			
Class	# Cases	# Errors	% Error
Low	640	332	51.88
High	775	167	21.55
<b>Overall</b>	<b>1415</b>	<b>499</b>	<b>35.27</b>

## Classification Tree

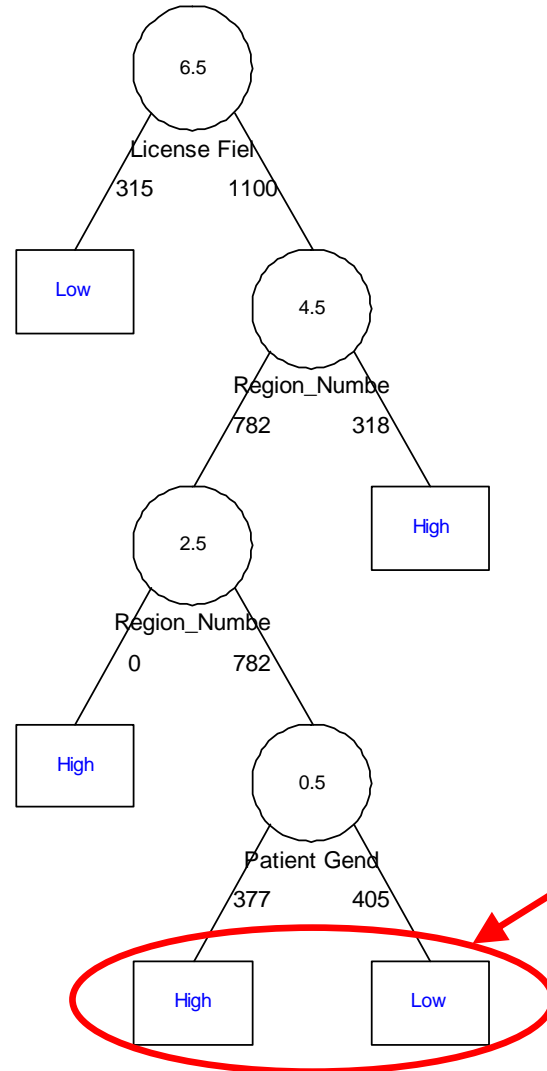
- Best Model for Prediction
- Robust To Outlier



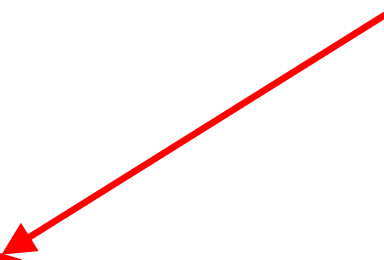


# Pay Range Classification

## Best Pruned Tree



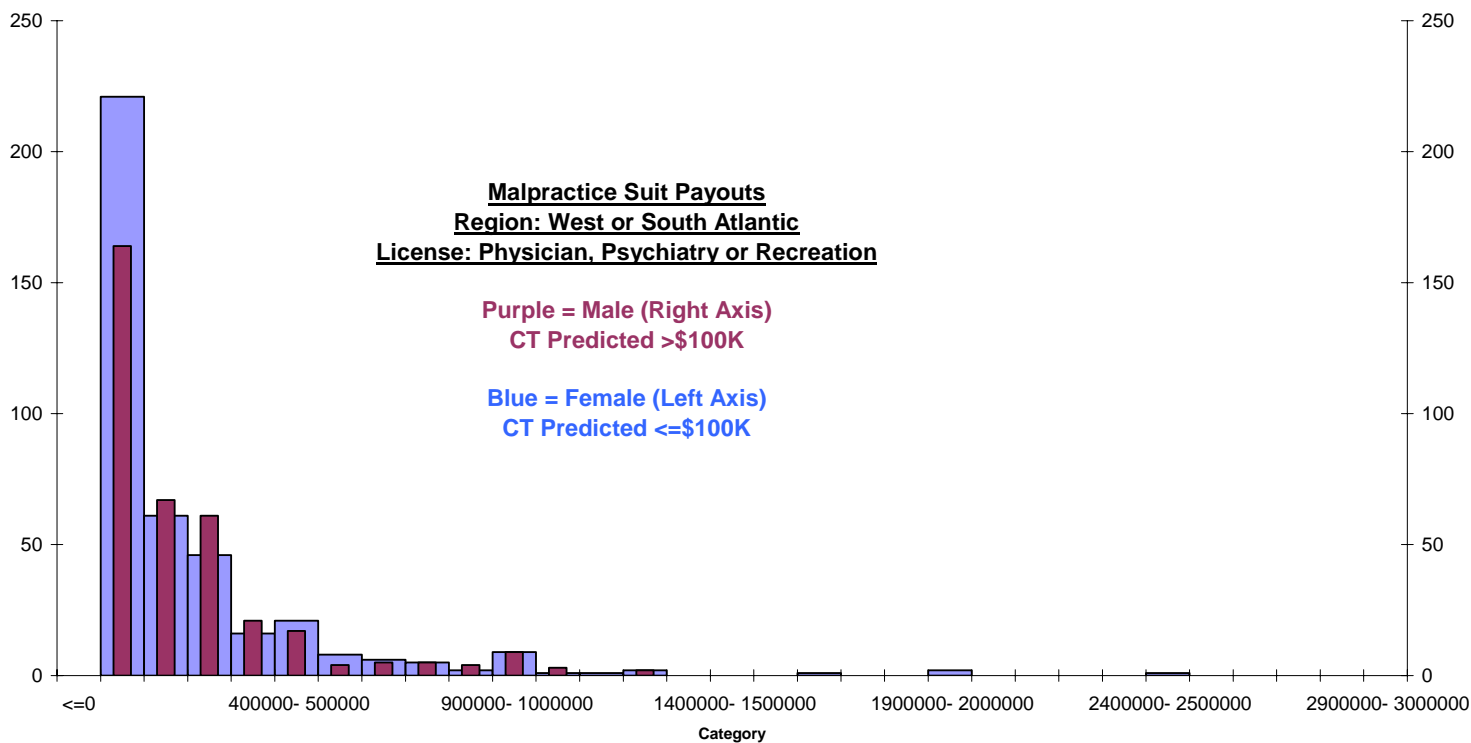
**INTERESTING!**





# Pay Range Histogram of Payouts

Histogram for Payment (\$)





# Conclusions

- The utility of prediction of judgment or settlement is limited as the range of payments judgment (\$750 - \$5 million) or settlement (\$30-\$4 million) contain roughly the same values. However, judgment/settlement is an important consideration for professional reputation
- Major predictors
  - License Field, Geographic Region, Patient Gender
- Data relatively weak to meet our objectives



# Future Studies

- Does the year of the cause play a role in the decision? As time goes on do more or less plaintiff's win?
- Do lower monetary value suits settle in favor of the plaintiff over higher dollar value suits (i.e. suit reasonableness)?
- What about including data about cases that were won by the defendant? (Judgments made - 0\$ payout)