Detecting Suspicious Claims:
The Case of individual Claim Processing

Richard A. Derrig, Ph.D.
President. OPAL Consulting LLC
Visiting Professor, Temple University

CAS RPM Seminar PMGMT-1
March 16, 2010
Chicago, IL

Agenda

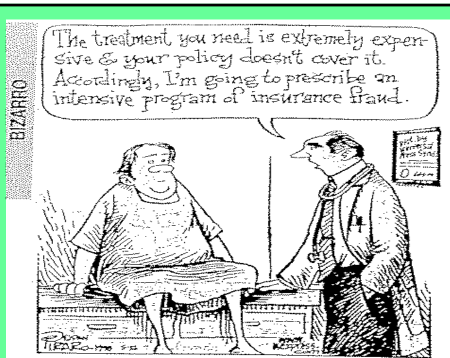
- What Is a Suspicious Claim?
- Hard v Soft Fraud
- Explanatory Models for Auto BI
- The Mass. Detail Claim Database
- Predictive Models for Claims
- Questions and Comments

Fraud Definition

PRINCIPLES

- Clear and willful act
- Proscribed by law
- Obtaining money or value
- Under false pretenses

Abuse: Fails one or more Principles



Derrig Top Five Fraud Ideas

- 1. “FRAUD” is ambiguous, ill-defined.
- 2. “FRAUD” should be reserved for criminal behavior (Hard Fraud). “Abuse” (Soft Fraud)
- 3. “FRAUD” ambiguity muddles the discussion and responsibility. Criminal Justice v Claim Management. Both are necessary (CIFI)
- 4. Criminal Fraud is several orders of magnitude less than popular estimates.
- 5. Fraud and Systematic Abuse can and should be mitigated by computer-assisted claim trained adjusters and special investigators dealing with “suspicious” claims.

HOW MUCH CLAIM FRAUD?

ABUSE DEFINITION

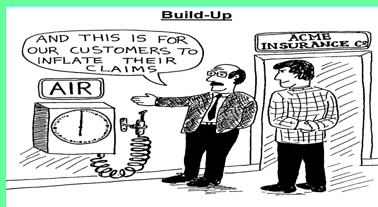
PRINCIPLES

- Not (Criminal) Fraud
- Unwanted, Unintended, Unnecessary Claims
- Disputable Damages
- Civil Matter
- Company's Problem
- Regulator's Problem

Non-Criminal Fraud

- General Deterrence – Ins. System, Ins Dept + DIA, Medical & Bar Associations, Other Government Oversight, Fraud Bureaus (CIFI)
- Specific Deterrence – Company SIU, Auditor, Data, Predictive Modeling for Suspicious Claims (and Underwriting).

BUILD-UP



Detail Claim Database DCD Background

DCD Background

In 1993, the Commissioner of Insurance mandated carriers to report to the DCD specific information on all Closed Bodily Injury (BI), Uninsured Motorist (U1), Underinsured Motorist (U2), Medical Payments and Personal Injury Protection (PIP) claims on Massachusetts private passenger and commercial policies including claims closed with no payment with claim handling activity

DCD Background

The DCD:

- **contains a broad array of data on injuries, injury and treatment patterns, and the professionals involved in automobile insurance claims**
- **has become an important tool in claim review, cost containment, and the battle against insurance fraud**
- **is available to insurance company staff via online database searches and reports**

Data Elements

Company: Insured's auto insurance carrier.

- **Premium Town, Claimant/Insured Address**.
- **Claimant DOB, SSN, Coverage**.
- Injury Type (32): Minor, Strain/Sprain, Major
- Outpatient Providers (2) Individual & Org
- Attorney Individual & Organization
- Medical Bills; Medicals "Paid"
- Investigation: IME, Med Audit, SIU Outcomes

DCD PROVIDER FILE

- Dynamic Audited File of Currently Active Medical Providers and Attorneys.
- File contains individual providers, organizations, and individuals linked to organizations.
- File streamlines reporting by requiring only six digit codes instead of all name and address data.
- File standardizes reporting - same file is used at each company which reduces errors.

DCD Provider File

- Providers on Auto Insurance Claims on Mass Auto Policies
- Automated Access by Companies
- March 2010: 126,928 Entries
- Medical 94,783; Attorney 32,145

DCD 2008 Closed Claims

138,000 Claims for $911 Million Paid
BI 34% , PIP 63% , UM+UIM 3%
Strain & Sprain 79%, $575 Million
Major & Fatal 15%, $319 Million
BI: Med 28%, Chiro 39%, PT 24%
PIP: Med 36%, Chiro 27%, PT 16%
BI: ATTY 88%, Avg $13,454, NA $5,460
PIP: ATTY 47%, Avg $ 4,150, NA $2,157

REFERENCES

- Derrig, R. A., (2002), Insurance Fraud, *Journal of Risk and Insurance*, 69:3, 271-289.
- Derrig, R. A., L. K. Krauss (1994), First Steps to Fight Workers Compensation Fraud, *Journal of Insurance Regulation*, 12:3, 390-415
- Francis, L. and Derrig, R. A., (2008) Distinguishing the Forest from the TREES, A Comparison of Tree-Based Data Mining Methods, *VARIANCE*, 2:2, 184-208
- Johnston, D.J., Derrig, R A., Sprinkel, E.A.,(2006) Auto Insurance Fraud: Measurments and Efforts to Combat It, *Risk Management and Insurance Review, 9:2,109-130.*
- Weisberg H.I., Derrig, R.A. (1992), Massachusetts Automobile Bodily Injury Tort Reform, *Journal of Insurance Regulation,* 10:2, 384-440.

Distinguishing the Forest from the Trees A Comparison of Tree-Based Data Mining Methods, *VARIANCE*, 2:2, 184-208

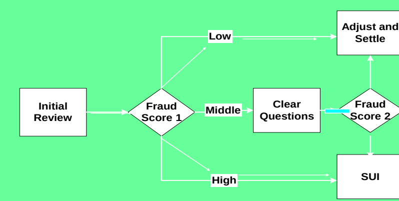
Richard Derrig, PhD,
Opal Consulting
www.derrig.com
Louise Francis, FCAS, MAAA
Francis Analytics and Actuarial Data Mining, Inc.
www.data-mines.com

Desirable Features of a Data Mining Method:

- Any nonlinear relationship can be approximated
- A method that works when the form of the nonlinearity is unknown
- The effect of interactions can be easily determined and incorporated into the model
- The method generalizes well on out-of sample data

Sequential Handling of Suspicious Claims

- Decision Flow Model



The Fraud Surrogates used as Dependent Variables

- Independent Medical Exam (IME) requested; IME successful
- Special Investigation Unit (SIU) referral; SIU successful
- Data: Detailed Auto Injury Claim Database for Massachusetts
- Accident Years (1995-1997)

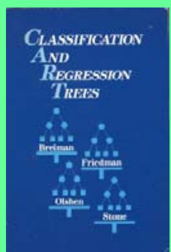
Explanatory Variables

- Claim file variables
  - Provider bill, Provider type
  - Injury
- Derived from claim file variables
  - Attorneys per zip code
  - Docs per zip code
- Using external data
  - Average Household income
    Households per zip

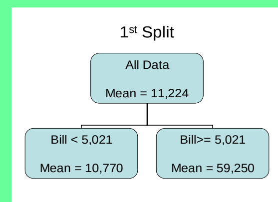
Decision Trees

- In decision theory (for example risk management), a **decision tree** is a graph of decisions and their possible consequences, (including resource costs and risks) used to create a plan to reach a goal. Decision trees are constructed in order to help with making decisions. A decision tree is a special form of tree structure.
  - **www.wikipedia.org**

The Classic Reference on Trees
Brieman, Friedman Olshen and Stone, 1993



CART Example of Parent and Children Nodes
Total Paid as a Function of Provider 2 Bill



Decision Trees Cont.

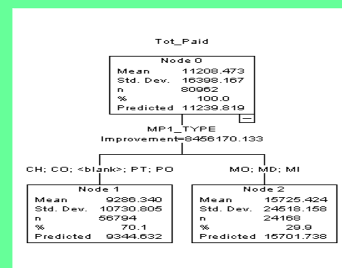
- After splitting data on first node, then
  - Go to each child node
  - Perform same process at each node, i.e.
  - Examine variables one at a time for best split
  - Select best variable to split on
  - Can split on different variables at the different child nodes

Classification Trees: Categorical Dependent

Find the split that maximizes the difference in the probability of being in the target class (IME Requested)

- Find split that minimizes *impurity*, or number of records not in the dominant class for the node (Too Many No IME)

Recursive Partitioning: Categorical Variables



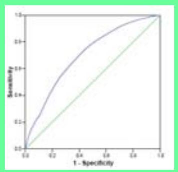
Different Kinds of Decision Trees

- Single Trees (CART, CHAID)
- Ensemble Trees, a more recent development (TREENET, RANDOM FOREST)
  - **A composite or weighted average of many trees (perhaps 100 or more)**
  - **There are many methods to fit the trees and prevent overfitting**
    - **Boosting: Iminer Ensemble and Treenet**
    - **Bagging: Random Forest**

Specificity/Sensitivity

- Sensitivity:
  - The proportion of true positives that are identified by model.
- Specificity:
  - The proportion of True negatives correctly identified by the model

TREENET ROC Curve – IME
AUROC = 0.701



Results for IME Requested

Area Under the ROC Curve – IME Decision				
	CART Tree	**S-PLUS Tree**	**Iminer Tree**	**TREENET**
AUROC	0.669	0.688	0.629	0.701
Lower Bound	0.661	0.680	0.620	0.693
Upper Bound	0.678	0.696	0.637	0.708
	Iminer Ensemble	**Random Forest**	**Iminer Naïve Bayes**	**Logistic**
AUROC	0.649	703	0.676	0.677
Lower Bound	0.641	695	0.669	0.669
Upper Bound	0.657	711	0.684	0.685

Results for IME Favorable

Area Under the ROC Curve – IME Favorable				
	CART Tree	S-PLUS Tree	Iminer Tree	TREENET
AUROC	0.651	0.664	0.591	0.683
Lower Bound	0.641	0.653	0.578	0.673
Upper Bound	0.662	0.675	0.603	0.693
	Iminer Ensemble	Random Forest	Iminer Naïve Bayes	Logistic
AUROC	0.654	0.692	0.670	0.677
Lower Bound	0.643	0.681	0.660	0.667
Upper Bound	0.665	0.702	0.681	0.687

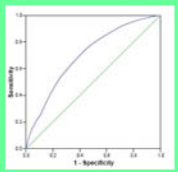
Results for SIU Referral

Area Under the ROC Curve – SIU Decision				
	CART Tree	S-PLUS Tree	Iminer Tree	TREENET
AUROC	0.607	0.616	0.565	0.643
Lower Bound	0.598	0.607	0.555	0.634
Upper Bound	0.617	0.626	0.575	0.652
	Iminer Ensemble	Random Forest	Iminer Naïve Bayes	Logistic
AUROC	0.539	0.677	0.615	0.612
Lower Bound	0.530	0.668	0.605	0.603
Upper Bound	0.548	0.686	0.625	0.621

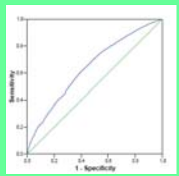
Results for SIU Favorable

Area Under the ROC Curve – SIU Favorable				
	CART Tree	S-PLUS Tree	Iminer Tree	TREENET
AUROC	0.598	0.616	0.547	0.678
Lower Bound	0.584	0.607	0.555	0.667
Upper Bound	0.612	0.626	0.575	0.689
	Iminer Ensemble	Random Forest	Iminer Naïve Bayes	Logistic
AUROC	0.575	0.645	0.607	0.610
Lower Bound	0.530	0.631	0.593	0.596
Upper Bound	0.548	0.658	0.625	0.623

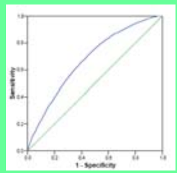
TREENET ROC Curve – IME
AUROC = 0.701



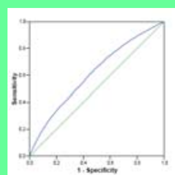
TREENET ROC Curve – SIU
AUROC = 0.677



Logistic ROC Curve – IME
AUROC = 0.643



Logistic ROC Curve – SIU
AUROC = 0.612



Ranking of Methods/Software – IME Requested

Method/Software	AUROC	Lower Bound	Upper Bound
Random Forest	0.7030	0.6954	0.7107
Treenet	0.7010	0.6935	0.7085
MARS	0.6974	0.6897	0.7051
SPLUS Neural	0.6961	0.6885	0.7038
S-PLUS Tree	0.6881	0.6802	0.6961
Logistic	0.6771	0.6695	0.6848
Naïve Bayes	0.6763	0.6685	0.6841
SPSS Exhaustive CHAID	0.6730	0.6660	0.6820
CART Tree	0.6694	0.6613	0.6775
Iminer Neural	0.6681	0.6604	0.6759
Iminer Ensemble	0.6491	0.6408	0.6573
Iminer Tree	0.6286	0.6199	0.6372

Ranking of Methods/Software – SIU Requested

Method/Software	AUROC	Lower Bound	Upper Bound
Random Forest	0.6772	0.6681	0.6863
Treenet	0.6428	0.6339	0.6518
SPSS Exh CHAID	0.6360	0.6270	0.6460
MARS	0.6280	0.6184	0.6375
Iminer Neural	0.6230	0.6136	0.6325
S-PLUS Tree	0.6163	0.6065	0.6261
Iminer Naïve Bayes	0.6151	0.6054	0.6247
Logistic	0.6121	0.6028	0.6213
SPLUS Neural	0.6111	0.6011	0.6211
CART Tree	0.6073	0.5980	0.6167
Iminer Tree	0.5649	0.5552	0.5745
Iminer Ensemble	0.5395	0.5305	0.5484