©*The Journal of Risk and Insurance*, 2002, Vol. 69, No. 3, 271-287

Insurance Fraud

Richard A. Derrig

Abstract

Insurance fraud is a major problem in the United States at the beginning of the 21st century. It has no doubt existed wherever insurance policies are written, taking different forms to suit the economic time and coverage available. From the advent of "railway spine" in the 19th century to "trip and falls" and "whiplash" in the 20th century, individuals and groups have always been willing and able to file bogus claims. The term *fraud* carries the connotation that the activity is illegal with prosecution and sanctions as the threatened outcomes. The reality of current discourse is a much more expanded notion of fraud that covers many unnecessary, unwanted, and opportunistic manipulations of the system that fall short of criminal behavior. Those may be better suited to civil adjudicators or legislative reformers. This survey describes the range of these moral hazards arising from asymmetric information, especially in claiming behavior, and the steps taken to model the process and enhance detection and deterrence of fraud in its widest sense. The fundamental problem for insurers coping with both fraud and systemic abuse is to devise a mechanism that efficiently sorts claims into categories that require the acquisition of additional information at a cost. The five articles published in this issue of the *Journal of Risk and Insurance* advance our knowledge on several fronts. Measurement, detection, and deterrence of fraud are advanced through statistical models, intelligent technologies are applied to informative databases to provide for efficient claim sorts, and strategic analysis is applied to property-liability and health insurance situations.

Introduction

When railroads were the proverbial deep pockets in the late 19th century, organized fakers slipped on banana peels, feigned paralysis, and extracted as much as $500 per fall from the railway companies, according to Dornstein, author of *Accidentally on Purpose* (1996). Personal injury cases ranged from "railway spine" of the

Richard A. Derrig is vice president of research for the Insurance Fraud Bureau of Massachusetts and senior vice president for the Automobile Insurers Bureau of Massachusetts, Boston, Massachusetts. The author acknowledges the production assistance of Eilish Browne and Julie Farrell of the Automobile Insurers Bureau and the hospitality during spring semester 2002 of the Risk Management and Insurance Department of the Wharton School, University of Pennsylvania.

1870s[1] to the similar "whiplash" auto claims of the 1950s, each reversing the bad luck of accidents into the good fortune of insurance settlements. Insurance researchers have devoted much energy to the accommodation of insurance contracts with the "moral hazard" of bogus and opportunistic claimants.[2] Outright fraud was perceived to be much more of a nuisance than the more pervasive exaggerated claim. In *Settled Out of Court* (1970), Ross relates that "(t)he adjuster typically believes that few people cut false claims from whole cloth, but that nearly everyone exaggerates his loss" (p. 45).

What has happened in the past few years to make insurance fraud such a hot topic? That is the subject of this short overview of the five major articles in this issue dedicated solely to technical aspects of insurance fraud. This article begins by recounting some insurance fraud milestones of the last 20 years.[3] A brief discussion of parsing criminal fraud from all other undesirable activities covered by *fraud* is followed by an idealized portrayal of the claims process. The twin objectives of fraud deterrence and detection are set in that process to highlight the importance of the integration of both in the Tennyson and Salsas-Forn article. Major and Riedinger attack the ubiquitous problem of systematically and strategically isolating the "interesting" claims in a database of health care claims. A discrete choice model of claim characteristics modified to estimate misclassification error in Artís et al. advances fraud measurement, specifically the problem of undiscovered fraud. Claim features set in a database can be implicitly weighted and the claim "scored" consistent with an unseen or latent variable (such as fraud) in the algorithmic PRIDIT development of Brockett et al. Finally, alternative scoring models that cover the landscape of current technologies (naïve Bayes, neural networks, decision trees, and others) are tested for intrinsic comparable advantage over logistic regression in Viaene et al. The astute reader will be able to extract the universality of these five approaches to fraud and adapt their lessons across insurance lines.

Understanding Insurance Fraud

Worldwide interest in insurance fraud, separate from moral hazard, has expanded greatly in the past ten to 20 years.[4] Growing problems with auto theft and claim fraud spawned the special investigative unit (SIU) in the United States in the 1980s (Ghezzi, 1983). Prompted by publicity about health care fraud, specifically in Medicare, Dionne (1984) began to lay the conceptual foundation for the analysis of provider fraud in Canada. Concern was evident in the United Kingdom that fraud had become a problem needing concerted attention in the travel, motor, home, and business covers

[1] Railway spine was a controversial disease first described in the 1860s by a British doctor as one with symptoms with a source in "microscopic changes to the spine that could not be seen." Naturally, railway spine became a leading cause for personal injury compensation in rail accidents in Britain and the United States (Dornstein, 1996, pp. 209-219).

[2] See Dionne et al. (1993) for workers' compensation and Cummins and Tennyson (1996) for automobile insurance. For an up-to-date general discussion of moral hazard in the insurance context, see Doherty (2001), Chapter 3.

[3] This trip though fraud publication history is necessarily short and incomplete. Apologies to those making important contributions that are omitted here.

[4] An annotated searchable bibliography on insurance fraud is available on the Insurance Fraud Bureau of Massachusetts Web site at http://www.ifb.org/IFRR/ifrr_ind.htm.

in the late 1980s (Clarke, 1989). By the end of the 1980s, an emerging consensus on the problem of fraud, but with wide variation in the responses, was noted for eight industrial nations, including the United States and Canada (Clarke, 1990). Market studies of fraud have been conducted in the United States by the Insurance Research Council (1992, 1997), by the Insurance Bureau of Canada (Fortin and Girard, 1992; Insurance Bureau of Canada, 1994), by the Insurance Councils of Australia (1994) and New Zealand (1996), and by Artís et al. (1999) for the Spanish Auto Insurance Market. Centralized associations were formed to promote solutions to the insurance fraud problem, including the International Association of Insurance Fraud Agencies (1986), the Coalition Against Insurance Fraud in the United States (1993), Comite Europeen des Assurances for the European Union (1993), and the Canadian Coalition Against Insurance Fraud (1994).

Initial systematic studies of automobile claims have been conducted in Massachusetts (Weisberg and Derrig, 1991, 1992), Florida (Florida Insurance Research Center, 1991), Canada (Caron and Dionne, 1999; Dionne and Belhadji, 1996), and nationally (Insurance Research Council, 1996), establishing the characteristics and magnitude of the auto insurance fraud problem. The Massachusetts studies have been extended to gain more insight into the players and mechanics of insurance fraud (Derrig et al., 1994). Academic researchers began to expand the understanding of insurance fraud in studies of auto insurance by Cummins and Tennyson (1992); workers' compensation insurance by Dionne et al. (1992) and Butler et al. (1996); health care insurance by Sparrow (1996); and property-liability insurance in general by Picard (1994, 1996). Economic modeling efforts followed in an attempt to promote efficient solutions in terms of contract design and auditing strategies (Bond and Crocker, 1997; Crocker and Tennyson, 1999; Picard, 2000; Watt, 1999, 2000). Practical models to sort out claims for fraud investigation began to emerge in the 1990s with database organization and selection strategies (Major and Riedinger, 1992), fuzzy clustering (Derrig and Ostaszewski, 1995), simple regression scoring models (Weisberg and Derrig, 1998; Brockett et al., 1998), and probit models (Belhadji and Dionne, 1997; Belhadji et al., 2000).[5]

Many definitions of claim fraud are in common use. In this overview, I propose that *fraud* be reserved for criminal acts, provable beyond a reasonable doubt, that violate statutes making the willful act of obtaining money or value from an insurer under false pretenses or material misrepresentations a crime (Derrig and Krauss, 1994). This strict definition of fraud will not fit with the many large dollar and claim proportion estimates of fraud so often quoted in newspaper and official industry sources.[6] For example, this definition of fraud may even be too strict to match the estimation objective of the discrete choice model applied to the Spanish auto insurance market by Artís et al. (1999). In their study, claims are labeled as fraud when the claimant admits fraud (undefined) and the insurer denies payment and/or cancels the policy. No criminality appears to be involved, so there is little in the way of a "cost" of being discovered with this sort of claiming behavior, thereby increasing the economic incentives for this

[5] It is highly likely that proprietary models were developed during this time frame but were not published.

[6] In its annual report for 2000, the CAIF (2001a) produced a section called "Pin the Tail on the Estimate." Estimates for the cost of insurance fraud ranged from a low of $18 billion by the National Insurance Crime Bureau for property-liability fraud to a high of $96 billion by Conning & Co. for all lines of private-market insurance.

TABLE 1
Types of Fraud—Auto Claimant versus Workers' Compensation (WC) Claimant

Auto	WC
1. Staged Accident	1. Deliberate Injury
2. Claimant Not Involved in Accident	2. Faked Injury (Multiple Claimants)
3. Duplicate Claims for Same Injury	3. Multiple Claims (Aliases)
4. Bills Submitted for Treatment Not Given	4. Fabricated Treatment
5. Real Injury, Unrelated to Accident	5. Non-Work-Related or Prior Injury
6. Fictitious Injury	6. Faked Injury (Single Claimant)
7. Misrepresentation of Wage Loss	7. Misrepresentation of Wage Loss
8. Other Material Misrepresentations	8. Other Material Misrepresentations

fraud type (Derrig and Zicko, 2002; Watt, 1999). Even with the low cost incentive, the Artís et al. (1999) study finds that undiscovered fraud is nearly 5 percent, comparing modeled fraud to known fraud in their data set.

If *fraud* is to represent provable criminal behavior, then what claimant actions are associated with this fraud? On a simple level, insurance claim fraud is asking for payment for an event that did not happen. To make this more concrete, we can look to auto and workers' compensation insurance for parallel actions constituting fraud.[7] Table 1 shows eight scenarios that all would agree to be fraud in the strictest sense. How often do these types of events happen?

HOW MUCH CLAIM FRAUD IS THERE?

Despite the agreed-upon importance of measuring fraud (Dionne et al., 1993; Sparrow, 1996), the vast array of estimates (CAIF, 2001a) attests to the imprecision of the measurement attempts to date. According to Sparrow (1996), the health-care industry "makes no serious attempt to measure the problem" (p. 55) and relies on the unsupported 10 percent estimate provided by the Government Accounting Office. Indeed, the differences in measuring (1) criminal or hard fraud, (2) suspected criminal fraud, (3) soft fraud or systematic abuse, and (4) suspected fraud or systematic abuse are vast. Some of those differences are generated by methodological weaknesses of the studies. Surveys routinely ask how much fraud is there, and is it getting better or worse, without providing working definitions or requiring any empirical backup to the opinions solicited.[8] Even expert opinions solicited in evaluating random samples of claim files can provide suspected fraud and abuse at most as opposed to definitive identification.[9] I suspect that orders-of-magnitude differences exist among the four categories spanning the fraud landscape.

[7] Table 1 originally appeared in Derrig and Krauss, 1994, p. 399.

[8] Of course, surveys such as the International Research Council/Insurance Services Office [IRC/ISO] 2001 survey provide an accurate picture of the perceptions of fraud, if not fraud itself.

[9] In one such sampling, four different coders provided subjective appraisals on the fraud content of Massachusetts personal injury protection (PIP) and bodily injury (BI) liability claims. While each set of coders identified 5–10 percent of the claims as suspected fraud, no claim was judged as suspected fraud by all four coders (Derrig and Ostaszewski, 1995).

TABLE 2
Fraud Measurement Example

Fraud Type	IFB Data	Size	Percent
1. Broad Suspected Fraud	Referrals	17,274	100%
2. Narrowed Suspected Fraud	Accepted Referrals	6,684	39%
3. Narrowed Suspected Fraud	Cases	3,349	39%
4. No Fraud	Closed No-Prosecution	2,084	24%
5. Probable Fraud	Refer to Prosecution	552	6%
6. Prosecuted Fraud	Prosecution Complete	293	3%
7. Fraud	(6) × (0.85)*	249	3%
8. Possible Fraud	(7) + Prosecution Pending	119	1%
9. Fraud Estimate	(7) + (8)	368	4%

Source: Insurance Fraud Bureau of Massachusetts, 1991–2000. Cases average two referrals, with about one subject per referral.
*Approximately 85 percent of prosecuted subjects are disposed as guilty or equivalent (Derrig and Zicko, 2002).

As an example to illustrate the wide possibilities of fraud estimates, I use data from a ten-year, real-time experiment in Massachusetts of identifying, investigating, and prosecuting auto and workers' compensation claim fraud. From its inception in 1991, the Insurance Fraud Bureau of Massachusetts (IFB) solicited property-liability fraud referrals from the industry, the government, and the public at large.[10] Approximately 85 percent of the 17,274 referrals to the IFB pertained to auto and workers' compensation claims. Table 2 shows the progressive weaning process from outside referral to prosecution. It demonstrates that the ratio of suspected fraud (not abuse) by industry personnel and the public to provable fraud is on the order of 25 to 1. Even if the unsupported suspected fraud estimate of 10 percent were accurate, the true level of criminal fraud would be less than one-half of 1 percent.

WHAT DO COMPANIES DO ABOUT FRAUD AND ABUSE?

When discussing fraud issues, one must often remember that insurance contracts between the company and the insured are agreements to pay for accidental damages when they occur. The business of insurance is to pay claims in a timely and efficient manner. Companies are well aware that claimants and providers may have opportunities and incentives to take advantage of accidents, even fabricate or cause them to happen, to obtain payments they might otherwise not deserve. The claim adjusting process is in theory a narrowing of the information asymmetry (the claimant knows exactly what happened; the company knows some of what happened) to the point that an appropriate payment is made or the claim is denied. Adjusters routinely investigate claims and negotiate settlements. Companies have the discretion to spend as little as possible (overhead and routine reports only) on a claim or invest in acquiring information (independent medical examinations, accident reconstruction, depositions) to

[10] While the vast majority of referrals concerned property-liability lines, occasional referrals were for other lines such as life and viaticals.

Table 3
Fraud and Abuse: The Top Ten Defenses

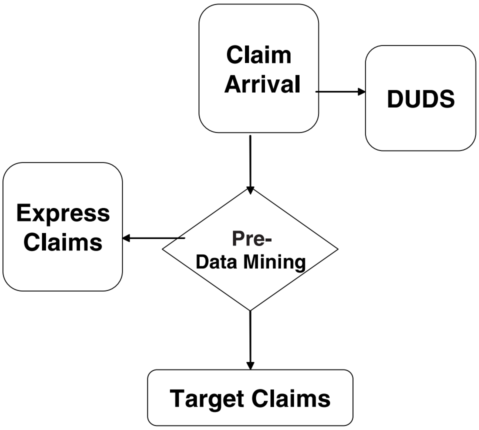
1.	Adjusters
2.	Computer Technology
3.	Criminal Investigators
4.	Data and Information
5.	Experts
6.	Judges
7.	Lawyers
8.	Legislators
9.	Prosecutors
10.	Special Investigators

resolve the asymmetry partially (negotiation) or fully (jury trial). This points to the following fundamental problem: How can you sort incoming claims efficiently into categories that require the acquisition of additional information at a cost? In the literature, this is known as costly state verification (Bond and Crocker, 1997). But these types of decisions are made thousands of times every day. The question then becomes: Are the sorting systems today effective at isolating fraud, isolating abuse, and isolating complicated claims from those that are easily paid? At least for auto bodily injury liability claims in Massachusetts, the data show that reductions are taken routinely in negotiated settlements when buildup is suspected (Weisberg et al., 1994). Claim denials occur when the information, usually gathered by an SIU, supports no payment.

As for criminal fraud, the most efficient way to deal with appropriate sanctions is centralized insurance fraud bureaus, of which there are 39 at this writing, seven of which are for workers' compensation alone (IRC/ISO, 2001, Appendix 2; see also CAIF, 2001b; Derrig and Zicko, 2002). They have the unique ability to connect those with the ability to discover fraud (SIU personnel) with those who can sanction (prosecutors, judges). The small amount of "fraud" relative to suspected fraud and abuse illustrated in Table 2 leads to two conclusions. First, company personnel must resolve the vast majority of suspicious claims, with perhaps referrals to fraud bureaus and/or regulatory authorities as an additional obligation. Second, the use of the wide definition of *fraud* inappropriately targets the law enforcement system as the solution when regulatory agencies for providers and the legislated parameters of the insurance system itself may be much more appropriate for mitigating unwanted claims and claim payments. Table 3 identifies ten components of the fraud and abuse claim handling system. It is instructive for the reader to rank them by their importance in fighting fraud versus abuse; the rankings should be quite different.[11]

[11] For example, in my opinion, prosecutors are most important (Rank 1) for deterring fraud but least important (Rank 10) for mitigating abuse. My full ranking for the Ten Defenses (in the alphabetical order of Table 3) for fraud are 5, 7, 3, 6, 8, 9, 10, 4, 1, and 2, and for abuse they are 5, 4, 9, 3, 7, 8, 1, 2, 10, and 6.

FIGURE 1



CLAIM PROCESSING

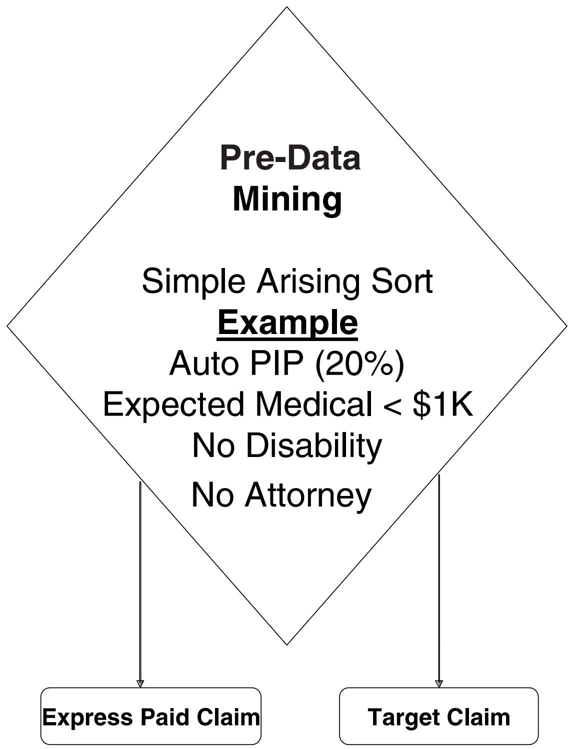
The five articles in this issue all contribute to an understanding of how the claim processing system can be enhanced by technology. The goal is not to replace adjusters and SIU personnel but to support their function by sharpening the information delivery system using more appropriate data (expanded claim features) and better technology to manipulate and deliver the data (Sparrow, 1996). Figure 1 shows the idealized initial processing of incoming claims. When the first notice of a claim arrives, a triage (pre-data mining) should sort arising claims into those that can be paid immediately, called express claims here, and those that need to be further evaluated, called target claims (Weisberg and Derrig, 1995). The remaining claims are those "duds" that never materialize into payments, about 20 percent for PIP claims in Massachusetts.

The reason that this incoming triage is separated from the main processing is twofold. First, it is easy to do. Figure 2 shows a set of decision criteria that typifies the simplicity of the triage process, requiring no heavy-duty data mining operations such as those found in Viaene et al. in this volume. Second, the presence of a large amount of the express and dud claims, which need no additional information to resolve the claim, will necessarily skew data mining fitting parameters as they "stretch" to accommodate the simple claims. The establishment of minimum criteria for investigation also comports with costly state verification theory (Bond and Crocker, 1997). It is better to reserve the more sophisticated techniques for the more difficult sorting problems.

Figure 3 illustrates conceptually how the target claims are potentially processed, including the handling of fraud and abusive claims. Data mining (DM) enhancements to the processing system are primarily valuable in delivering sorting information at the routine adjusting stage.[12] The information can range from simple routine information (prior claims) to complicated claim type profiles (Major and Riedinger, 2002).

[12] DM algorithms can also serve the more specialized sorting required within the set of claims handled by SIU personnel. Indeed, the more sophisticated algorithms may be reserved for SIU use, perhaps because of sensitive information and privacy concerns.

FIGURE 2



Once the information is in hand and the claims are sorted, those needing "investigation" (spend money) are identified. Ideally, investigation results in claims being paid, built-up claims being negotiated (down), and suspected fraud claims referred to an SIU. Civil and/or criminal proceedings result when the situation warrants, if the appropriate institutional systems (such as fraud bureaus) are effective.

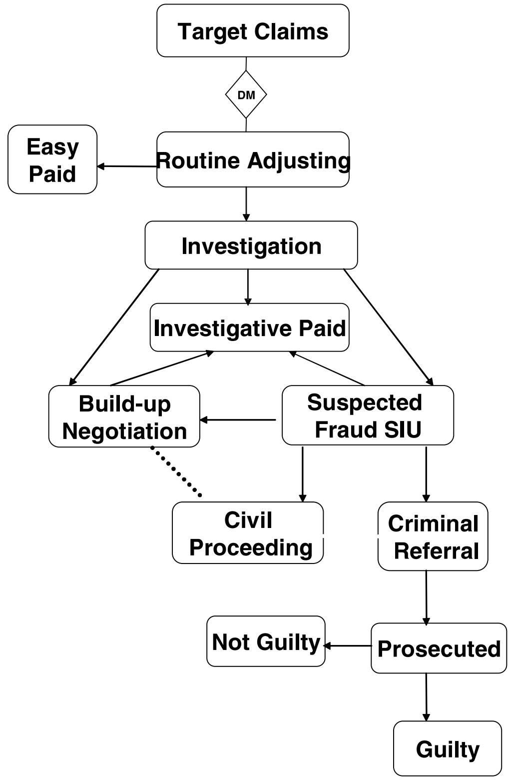
DEVELOPING THE CLAIM SORT

As discussed above, the fundamental problem in dealing with all claims is developing an effective sorting of claims into bins requiring different levels of information acquisition at a cost. At the bottom, all such claim sorting systems have the same theoretical structure, although the details can and do vary widely. Figure 4 shows the conceptual structure of the DM process.

One or more databases are accessed for each target claim. The databases are manipulated based upon some algorithm that "scores" claims in the sense that the incoming target claims are to be sorted by their "scores" into bins with associated actions. Generally speaking, suspicious and complicated claims where additional information is needed appear in one collection of bins and nonsuspicious routine claims in the remaining bins.[13] Examples of such scoring processes are given by the PRIDIT scores using data on fraud indicators (Brockett et al., 2002) and the probabilities of fraud gen-

[13] Actual claim processing systems may contain more complicated layers such as medical bill reductions of upside outliers to so-called reasonable and customary levels.

FIGURE 3

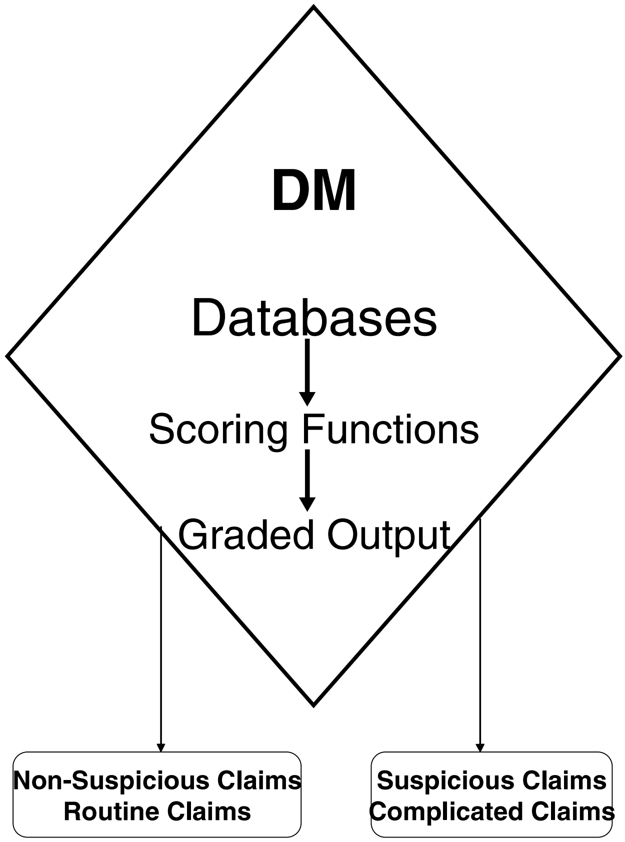


erated by the discrete choice models operating on claim characteristics (Artís et al., 2002). Typically, the scoring output will be one-dimensional graded output along a scale such as $[-1,1]$ as in PRIDIT. Occasionally, the scores may be multidimensional when more than one criterion applies,[14] in which case one can make bin assignments in multistage processes or by means of fuzzy clustering of potentially conflicting criterion (Derrig and Ostaszewski, 1995). Note that the same sorting problem can occur for claim providers, as amply demonstrated by the electronic fraud detection (EFD) system of Major and Riedinger (2002). The development flow for a data mining process built to score claims is idealized in Table 4.

One must start with real data on claim processing for the subject line of business. Economic efficiency argues for acquiring that knowledge off the production line through the analysis of random samples of claims in Step 1 of Table 4. As many relevant dimensions of the processing should be included as possible in Step 2, including both subjective and objective characteristics and assessments. Red flags pertinent to the line of insurance form a starting point (Canadian Coalition Against Insurance Fraud, 1997). Additional collected or derived variables from accessible databases will most likely help (Viaene et al., 2002). Since the data mining output is designed to assist in the (additional) information acquisition, one must pay particular attention to the

[14] One potential example would be the use of credit scores as a separate criterion not derived from the company claim data. Another example would be some scoring of providers by their propensity for involvement in questionable or exaggerated claims.

FIGURE 4



timing of the input features, with early arriving (routine) information being the most valuable for Step 3. Steps 4–6 in Table 4 are meant to emphasize the separation of the selection and value of the feature data themselves (Step 4), by whatever means are chosen, from the association of the feature data with the objective of interest via the scoring output. Expert assessment and/or historical results can provide the objective variable for Step 5 that defines the potential extent of claim sort bins needed. Step 6 covers the difficult task of matching the modeling specification to the identification problem at hand. Neural networks, fuzzy logic, genetic algorithms, and more can be used singly or in tandem to manipulate the databases (Goonatilake et al., 1995; Hastie et al., 2001). Steps 7 and 8 point out the obvious: The data mining process needs to be validated not only in a testing mode but also in a feedback production mode (Wolf et al., 1999). I turn next to the contributions of the articles in this issue to the idealized claim processing flow discussed herein.

RESEARCH THIS ISSUE

All would agree that the primary goal of the information decision flow discussed above is to both detect suspicious claims and identify the nonsuspicious claims for rapid payment. That is the detection problem. Tennyson and Salsas-Forn elaborate on the twin goals of detection and deterrence. They provide concrete tests of auditing practices and outcomes that allow for the relative valuation of the process

TABLE 4
Claim Sorting Algorithm Development Flow

STEP 1:	Sample: Systematic benchmark of a random sample of claims.
STEP 2:	Features: Isolate red flags and other sorting characteristics.
STEP 3:	Feature Selection: Separate features into objective and subjective; early, middle, and late arriving; acquisition cost levels; and other practical considerations.
STEP 4:	Cluster: Apply unsupervised algorithms (Kohonen, PRIDIT, fuzzy) to cluster claims, examine for needed homogeneity.
STEP 5:	Assessment: Externally classify claims according to objectives for sorting.
STEP 6:	Model: Supervised models relating selected features to objectives (logistic regression, naïve Bayes, neural networks)
STEP 7:	Static Testing: Model output versus expert assessment, model output versus cluster homogeneity (PRIDIT scores) on one or more samples.
STEP 8:	Dynamic Testing: Real-time operation of acceptable model, record outcomes, repeat Steps 1–7 as needed to fine-tune model and parameters. Use PRIDIT to show gain or loss of feature power and changing data patterns, tune investigative proportions to optimize detection and deterrence of fraud and abuse.

toward meeting the separate detection and deterrence goals. Their research implies that auditing outcome criteria, such as the proportion of positive outcome audits of risk classes, can be used to tune the scoring parameter for efficient allocation of directed or random audits. The positive outcomes tested in the article are subjective (raised doubt, refuted), which may be problematic for tuning. Perhaps a consistent measure of claim cost reduction or savings net of cost would be better for operational purposes.

The value of reprinting the classic Major and Riedinger health claim fraud detection article is twofold. First, the basic scientific approach to the problem of manipulating a large claim database to extract "outliers" of interest is probably more relevant today than when the article was first written. Second, potential developers should be encouraged that such a system design was actually implemented in a company (Travelers) and produced results that are valid in theory and useful in practice. Their notions of product design (EFD architecture) are similar in spirit to the development steps of Table 4 and are essential to the final design of EFD.[15] Most valuable, however, is their discussion of frontier identification (Figure 6) as a selection criterion. While this particular approach may not fit other modeling situations, it is highly illustrative of the need for careful analytic choice of decision criteria that become embedded into the model and hidden from production users.

Discrete choice models with misclassification as discussed by Artís et al. add to the many ways of estimating fraud proportions. Their contribution not only yields an estimate of "hidden fraud," but also provides a way of evaluating the importance of claim features associated with fraud. They find that the discrete choice model correctly classifies about 75 percent of claims with or without the misclassification adjustment

[15] EFD is the acronym coined by Major and Riedinger for electronic fraud detection.

(their Table 5). While percent correctly classified (PCC), or the disaggregated confusion matrix, provides an easily interpretable evaluation criterion, both the Major and Riedinger and Viaene et al. articles point to AUROC, the area under the receiver operating curve, as the better evaluation criterion, as it shows the model specificity and sensitivity to positive and negative errors at once. The addition of the misclassification variable in the Artís et al. final model and its additional claims "predicted" to be fraud illustrate a fundamental principle underlying the application of these and other scoring models. Claims with many features similar to those of fraudulent claims are (expected to be) fraudulent. We can call it the "guilt by association" criterion. It includes the nearest neighbors models in Viaene et al. and, indeed, most feature identification models.

Brockett et al. contribute to the evaluation and selection of appropriate feature data. The authors prove a valuable universal theorem—namely, that feature data with responses that are monotone in a latent or hidden binary variable (like fraud) have a unique (up-to-scale), nonlinear scoring model that separates the claims according to the latent variable and gives a relative value to each feature that contributes to that separation. This latter result allows for the testing of various features (Steps 2 and 3 in Table 4) to permit the use of only those feature variables that actually contribute to effective sorting while ignoring those with weak contributions. In addition, the embedded scoring model can be used to test (1) whether some alternatively developed scoring model "fits" the data in the sense of not being too distant from the embedded model, and (2) whether the feature data are changing over time (the embedded models should not be too far apart).

Viaene et al. deliver four major contributions to the claim sort development problem. First, they provide a large laundry list of potential Step 6 model specifications for binary classification models. Logit, decision trees, nearest neighbor, neural network, support vector machine, naïve Bayes, and tree-augmented naïve Bayes models with common parameter and smoothing variations are presented in Viaene et al.'s Table 5. Second, their results show that at least on the Massachusetts data tested, the models per se show no advantage one over the other. While testing (Steps 7 and 8 in Table 4) is always in order, a cost-conscious modeler may opt for the simpler and easy-to-interpret logit model to start the claim automated sort development. Third, the results show the superiority of the AUROC evaluation criterion (Step 7) over the PCC criterion (compare their Table 5 and Table 6 results relative to the "majority" benchmark). Fourth, the feature set augmented with non-fraud indicator data provides better results, highlighting the fact that traditional fraud indicators alone may not be adequate and that a search for augmented features is worthwhile. Readers can now see the sweep of the five articles in this volume and the relevance of each to the design of a claim sort. I now turn to the future.

FUTURE RESEARCH

The five articles in this issue contribute much to the theoretical and practical aspects of fraud detection and deterrence. As with all good research, a deeper understanding of the process raises additional, usually more difficult, or more subtle questions. The purpose of this final section of the insurance fraud overview is to pose some of those additional topics for future research. This will be an admittedly biased selection conditioned on the author's interest; others may note different priorities for follow-up projects.

Data

The single most important improvement that can be made is to assemble additional databases for production and study. Even the casual reader of these five articles will note that three of the five use data on automobile insurance generated in studies of Massachusetts automobile BI claims. The mere presence of additional data sets, such as the Spanish market claims in Artís et al., provide for alternative specifications of modeling efforts as well as tests of robustness of research results that are dependent on the data set used (Viaene et al., 2002). Centralized data sets, perhaps assembled by third-party vendors such as ISO, which can be accessed in the claim processing will play a key role in the efficiency of any system of claim sorting developed. The type of features and the quality of the reported data will determine the effectiveness of the processing or research, as the case may be. A high priority should be to assemble a data set for workers' compensation.[16]

Fraud Indicators

Determine which traditional fraud indicators carry good claim sorting information and which augmented features best supplement or replace the fraud indicators (by line).

Time and Place

Time and place of occurrence are objective features known early in the claim handling flow. Tune models to take advantage of the availability of these data and test their importance by line.

Subjective/Objective

Subjective features have the advantage of being summarized or preprocessed data elements where the processor is a domain expert. This makes the subjective feature a valuable way to incorporate subtle correlations that may be available to the human but not in a systematic way. Subjective features have the distinct disadvantage of being observer-dependent and subject to wide variation of interpretation. Objective features, however, stand alone as verifiable data elements without interpretation by the processor. Showing how to replicate subjective features by equivalent, or near-equivalent, objective features in a demonstration modeling effort would be helpful.

Investigation Models

Build a demonstration scoring model that not only identifies sorting bins but also provides probabilities of successful investigation outcomes for each claim within each bin.

Costly Misclassification

Compare the qualitative and quantitative outcomes of scoring models when misclassification costs of false positives and false negatives are equal and unequal.

---

[16] See Biddle, 2001.

PRIDIT Scores

Determine the appropriate range of "permissible" supervised models in relation to the embedded PRIDIT model for feature data that are monotone in the investigation/no investigation latent variable. Compare that model outcome with actual investigation decisions to determine value added. Compare PRIDIT determined valuation of features to otherwise determined feature values, such as significance of regression coefficients, in a demonstration model.

Expert Design

Determine the potential increase in scoring model efficiency by the use of expert tuning using advanced techniques such as boosting and bagging (Hastie et al., 2001, Chapters 8 and 10).

Detection and Deterrence

Establish objective claim cost or net savings criteria as measures for valuing detection and deterrence in the sense of Tennyson and Salsas-Forn.

Random and Targeted Audits

Determine the methods to optimize a pre-commitment policy mix for random and targeted audits.

REFERENCES

Artís, M., M. Ayuso, and M. Guillén, 1999, Modelling Different Types of Automobile Insurance Fraud Behavior in the Spanish Market, *Insurance: Mathematics and Economics*, 24(1-2): 67-81.

Artís, M., M. Ayuso, and M. Guillén, 2002, Detection of Automobile Insurance Fraud With Discrete Choice Models and Misclassified Claims, *Journal of Risk and Insurance*, 69(3): 325-340.

Belhadji, E.-B., and G. Dionne, 1997, Development of an Expert System for the Automatic Detection of Automobile Insurance Fraud, *Risk Management Chair*, Working paper 97-06, HEC-Montreal.

Belhadji, E.-B., G. Dionne, and F. Tarkhani, 2000, A Model for the Detection of Insurance Fraud, *Geneva Papers on Risk and Insurance—Issues and Practice*, October, 25: 517-539.

Biddle, J., 2001, Do High Claim-Denial Rates Discourage Claiming? Evidence From Workers Compensation Insurance, *Journal of Risk and Insurance*, 68(4): 631-658.

Bond, E. W., and K. J. Crocker, 1997, Hardball and the Soft Touch: The Economics of Optimal Insurance Contracts With Costly State Verification and Endogenous Monitoring Costs, *Journal of Public Economics*, 63: 239-264.

Brockett, P. L., X. Xia, and R. A. Derrig, 1998, Using Kohonen's Self-Organizing Feature Map to Uncover Automobile Bodily Injury Claims Fraud, *Journal of Risk and Insurance*, 65: 2.

Brockett, P. L., R. A. Derrig, L. L. Golden, A. Levine, and M. Alpert, 2002, Fraud Classification Using Principal Component Analysis of RIDITs, *Journal of Risk and Insurance*, 69(3): 341-372.

Butler, R. J., D. Durbin, and M. N. Helvacian, 1996, Increasing Claims for Soft Tissue Injuries in Workers' Compensation: Cost Shifting and Moral Hazard, *Journal of Risk and Uncertainty*, 13: 73-87.

Canadian Coalition Against Insurance Fraud, 1997, Red Flags for Detecting Insurance Fraud, Toronto.

Caron, L., and G. Dionne, 1999, Insurance Fraud Estimation: More Evidence From Quebec Automobile Insurance Industry, in: C. Laberge-Nadeau and G. Dionne, eds., *Automobile Insurance: Road Safety, New Drivers, Risks, Insurance Fraud and Regulation* (Boston, Mass.: Kluwer).

Clarke, M., 1989, Insurance Fraud, *The British journal of Criminology*, 29: 1-20.

Clarke, M., 1990, The Control of Insurance Fraud: A Comparative View, *The British Journal of Criminology*, 30: 1-23.

Coalition Against Insurance Fraud, 2001a, *Annual Report* (Washington, D.C.: Coalition Against Insurance Fraud).

Coalition Against Insurance Fraud, 2001b, *A Statistical Study of State Insurance Fraud Bureaus—A Quantitative Analysis—1995-2000* (Washington, D.C.: Coalition Against Insurance Fraud).

Comite Europeen des Assurances, 1996, The European Insurance Anti-Fraud Guide, ed., Euro Publishing System, May, 4: 1-55.

Crocker, K. J., and S. Tennyson, 1999, Costly State Falsification or Verification? Theory and Evidence from Bodily Injury Liability Claims, in: C. Laberge-Nadeau and G. Dionne, eds., *Automobile Insurance: Road Safety, New Drivers, Risks, Insurance Fraud and Regulation* (Boston, Mass.: Kluwer).

Cummins, J. D., and S. Tennyson, 1992, Controlling Automobile Insurance Costs, *Journal of Economic Perspectives*, 6: 95-115.

Cummins, J. D., and S. Tennyson, 1996, Moral Hazard in Insurance Claiming: Evidence From Automobile Insurance, *Journal of Risk and Uncertainty*, 12: 29-50.

Derrig, R. A., and L. Krauss, 1994, First Steps to Fight Workers Compensation Fraud, *Journal of Insurance Regulation*, 12: 390-415.

Derrig, R. A., and K. M. Ostaszewski, 1995, Fuzzy Techniques of Pattern Recognition in Risk and Claim Classification, *Journal of Risk and Insurance*, 62: 447-482.

Derrig, R. A., H. I. Weisberg, and X. Chen, 1994, Behavioral Factors and Lotteries Under No-Fault With a Monetary Threshold: A Study of Massachusetts Automobile Claims, *Journal of Risk and Insurance*, 61: 245-275.

Derrig, R. A., and V. Zicko, 2002, Prosecuting Insurance Fraud: A Case Study of the Massachusetts Experience in the 1990s, Working paper, Insurance Fraud Bureau of Massachusetts.

Dionne, G., 1984, The Effect of Insurance on the Possibilities of Fraud, *Geneva Papers on Risk and Insurance—Issues and Practice*, 9: 304-321.

Dionne, G., and E.-B. Belhadji, 1996, Evaluation de la Fraude à l'Assurance Automobile au Quebec, *Assurances*, 64: 365-394.

Dionne, G., A. Gibbens, and P. St.-Michel, 1992, Conséquences de la Fraude à l'Assurance, in: J.-L. Fortin and J. D. Girard, eds., *La Fraude à l'Assurance* (Montreal: University of Montreal).

Dionne, G., P. St.-Michel, and Charles Vanasse, 1993, Moral Hazard, Optimal Auditing and Workers' Compensation, in: R. Chaykowski and T. Thomason, eds., *Research in Canadian Workers' Compensation* (Kingston, Canada: IRC Press).

Doherty, N. A., 2001, *Integrated Risk Management* (New York: McGraw-Hill).

Dornstein, K., 1996, *Accidentally on Purpose* (New York: St Martin's Press).

Florida Insurance Research Center, 1991, Automobile Insurance Fraud Study, University of Florida, Tallahassee.

Fortin, J. L., and J. D. Girard, 1992, *La Fraude à l'Assurance* (Montreal: University of Montreal).

Ghezzi, S. G., 1983, A Private Network of Social Control: Insurance Investigation Units, *Social Problems*, 30: 521-530.

Goonatilake, S., and S. Khebbal, 1995, *Intelligent Hybrid Systems* (New York: Wiley and Sons).

Hastie, T., R. Tibshirani, and J. Freidman, 2001, *The Elements of Statistical Learning* (New York: Springer).

Insurance Bureau of Canada, 1994, Insurance Fraud in Canada, Report of the National Task Force on Insurance Fraud, Toronto, January.

Insurance Council of Australia, 1994, Insurance Fraud in Australia, Sydney, September.

Insurance Council of New Zealand, 1996, National Research Bureau, Insurance and Fraud, New Zealand, December.

Insurance Research Council, 1992, Fighting Fraud in the Insurance Industry, Oak Brook, Ill., October.

Insurance Research Council, 1996, Fraud and Buildup in Auto Injury Claims—Pushing the Limits of the Auto Insurance System, Wheaton, Ill., September.

Insurance Research Council, 1997, Fighting Fraud in the Insurance Industry, Wheaton, Ill., October.

Insurance Research Council and Insurance Services Office, 2001, Fighting Insurance Fraud: Survey of Insurer Anti-Fraud Efforts, Malvern, Pa.

Major, J. A., and D. R. Riedinger, 1992, EFD: A Hybrid Knowledge/Statistical-Based System for the Detection of Fraud, *International Journal of Intelligent Systems*, 7: 687-703 (reprinted this volume).

Picard, P., 1994, Auditing Claims in Insurance Market With Fraud: The Credibility Issue, Thema—Université de Cergy-Pontoise and Université de Paris X-Nanterre.

Picard, P., 1996, La Fraude à l'Assurance—Quelques elements d'analyse, *Risques*, July–September, 27: 171-177 .

Picard, P., 2000, Economic Analysis of Insurance Fraud, in: G. Dionne, ed., *Handbook of Insurance*, 315-363.

Ross, L. H., 1970, *Settled Out of Court* (Chicago, Ill: Aldine).

Sparrow, M., 1996, *License to Steal* (Boulder, Colo.: Westview Press).

Tennyson, S., and P. Salsas-Forn, 2002, Claims Auditing in Automobile Insurance: Fraud Detection and Deterrence Objectives, *Journal of Risk and Insurance*, 69(3): 289-308.

Viaene, S., R. A. Derrig, B. Baesens, and G. Dedene, 2002, A Comparison of State-of-the-Art Classification Techniques for Expert Automobile Insurance Fraud Detection, *Journal of Risk and Insurance*, 69(3): 373-421.

Watt, R., 1999, The "Zero Fraud" Indemnity Function for Insurance Contracts, Paper presented at the 26th Seminar of the European Group of Risk and Insurance Economists, September.

Watt, R., 2000, Curtailing Ex-Post Fraud in Risk Sharing Arrangements, Universidad Autónoma de Madrid, Working paper, February.

Weisberg, H. I., and R. A. Derrig, 1991, Fraud and Automobile Insurance: A Report on the Baseline Study of Bodily Injury Claims in Massachusetts, *Journal of Insurance Regulation*, 9: 497-541.

Weisberg, H. I., and R. A. Derrig, 1992, Massachusetts Automobile Bodily Injury Tort Reform, *Journal of Insurance Regulation*, 10: 384-440.

Weisberg, H. I., and R. A. Derrig, 1995, Identification and Investigation of Suspicious Claims, in: AIB Cost Containment/Fraud Filing (DOI Docket R95-12) (Boston, Mass.: Automobile Insurers Bureau of Massachusetts).

Weisberg, H. I., and R. A. Derrig, 1998, Quantitative Methods for Detecting Fraudulent Automobile Bodily Injury Claims, *Risques*, July–September: 35: 75-99.

Wolf, G. D., H. I. Weisberg, and A. J. Brand, 1999, New Technology to Reduce Opportunistic Fraud, *SIU Awareness*, June: 18-22.