SELECTING THE BEST FRAUD SAMPLE USING VARIOUS ANALYTIC TECHNIQUES

Given the difficulty in selecting a sample from which to detect fraud, a variety of analytics will be applied to the data so that a fraud score can be calculated for each sample unit. Using this fraud score by transaction, the examiner can rank and summarize by key subsets (i.e., vendor, employee, etc.) in order to arrive at the most likely transactions for selection. The ultimate goal is to focus on fraud risk that you can now mathematically calculate at the transactional level.

You will learn how to:

- Apply a variety of analytical techniques to be completed in simple software (Excel) and more advanced programs (ACL, IDEA, and Arbutus).
- Develop a combined transactional fraud score based on digital, textual, geo-mapping, and other fraud report scores.
- Graphically analyze and otherwise use this score for selecting a final sample for analysis.

RICHARD LANZA, CFE, CPA, CGMA
CEO
Cash Recovery Partners LLC
Lake Hopatcong, NJ

Rich Lanza helps companies identify their hidden financial assets, mostly using technology and referring them to specialists. He has more than two decades of experience in audit technology and recovery auditing, and has become a leading authority in these areas. In addition to making videos for AuditSoftwareVideos.com, Rich now spends much of his time building customized continuous monitoring solutions for companies that are self-sustaining through cost recoveries. He is the author of 13 publications and training courses in audit software, and he has authored more than 75 articles for major audit and accounting publications.

Looking for the Largest Fish

Many companies, after completing a risk assessment, want to immediately catch *every* fish—big or small. However, just as it's impossible for a batter to hit five baseballs pitched at the same time, a company with a limited fraud detection budget won't be able to correct every problem at the same time. You need to determine not just the likelihood of the frauds but their significance as you make your priority list. You can do this qualitatively, but, because we need to "follow the money," the quantitative data analysis usually is most beneficial.

Prioritizing with the Qualitative Approach

Before we discuss the meat of this column—detecting the most significant areas for fraud review—let's step back and ask some questions about the organization, or even more specifically, key departments:

- Who has the most spending authority within the company? More authority leads to higher spending and, therefore, more opportunity for fraud.
- Which departments provide the most revenue to the company? The larger the revenue base, the larger the risk that fraudulent sales could be posted and not detected.
- Has the organization recently been part of a merger or acquisition? Obviously, such events increase the size of the organization. But they also result in stretching of employees (I'm thinking, specifically, of IT people, among others) who can become quite unhappy.
- Has the organization experienced rapid growth, increasing business demands, and exposure? Transaction volume can rise at these times and processes become even more complex, which can be fertile ground for an undetected fraudster.

NOTES

- Has your organization recently had a system conversion? With new technologies or system implementations, the priority becomes simply ensuring the transactions are processed, with little time left for verifying their accuracy.

Answering these questions can identify departments that generally have the largest amounts of company processing. The answers reveal more about spending than revenue because companies tend to be more focused on detecting misappropriation schemes that impact their cash flow.

Prioritizing with the Quantitative Approach

After the qualitative assessment, identify a few departments for review, but validate and expand this assessment based on a dive into the company databases. The main areas of focus will be revenue, cost of goods sold, and sales and general administrative, and so you should obtain these data files (for the last two or three years):

- General ledger trial balance—general ledger file providing the summary balance for every general ledger account
- Accounts payable invoice header—accounts payable invoices with one invoice listed per row in the data
- Accounts payable invoice detail distribution—Working from the invoice header file above, this file provides the general ledger account distribution for every invoice so each row in this data file is a portion of the invoice amount allocated to a general ledger account.
- Payroll register header—Compile a list of all employee payments, with one employee payment in

NOTES

each row of the table. Preferably, a department code is provided for each employee.

- Sales history invoice header—Sales invoices with one invoice listed per row in the data

After obtaining these data files, filter and summarize to home in on those departments and the employees (as the data allows) with the most unusual trends. You can complete this analysis in Microsoft Excel (Excel 2007 might be best because it allows for a little more than one million rows of data) through the use of the AutoFilter option and Pivot Tables.

Perform the following analysis to prioritize specific opportunities:

- You can align the company's trial balance expense categories to external benchmarks for the industry and review any unusually high trends more closely. Visit www.bizstats.com for a free tool for industry benchmarking trial balance data. Also, visit www.bizminer.com and www.profitcents.com, among others, for paid subscription services.
- Compare material spending categories, based on a summary by general ledger account, to the list of fraud categories in the ACFE's Fraud Classification System and other savings opportunities identified at www.findmillions.net. My website lists contingency fee recovery firms that perform retrospective and prospective savings analysis; they collect shared fees for any realized savings.
- Find any increasing trends over time that could be in a general ledger account or a vendor, which could, for example, warrant a vendor audit to ensure contractual and other pricing compliance.
- Create a scatter graph of volume to values for general ledger accounts or vendor balances. A

NOTES

scattergraph will allow you to quickly see, for example, vendors with high values and low volumes, which might signal high-dollar fraudulent payments that were made to quickly bleed cash out of the organization.

- Summarize vendor history by general ledger account and vendor, which will quickly identify any sole-source vendors that might be the best candidates to competitively bid. It might also detect general ledger accounts that are serviced by too many vendors and therefore could use a centralization of spending to improve purchasing power in a given expense area.
- Perform simple duplicate payment tests based on the invoice header file by invoice number and vendor. A simple test using a Microsoft Access query might yield recoveries or spot an opportunity for a recovery firm to perform a more detailed analysis. It might also highlight a vendor scheme to purposely send a repetitive duplicate invoice to a company.

So, while a qualitative approach can lead to savings opportunities, it's best to enhance that approach with quantitative analysis. Through some simple data analysis of company spending, revenue, and other general ledger accounts balances, most of which can be performed in Microsoft Excel, opportunities can be identified for fraud detection and other bottom-line savings.

NOTES

NOTES

Auditing Vendor Accounts for Fraud …
Or at Least Some Cash Recovery

Above, we focused on prioritizing potential fraud opportunities with employees and vendors using qualitative and quantitative approaches. Here, we'll go one level deeper by using automated tools in uncovering associated fraud within vendor accounts. (Even if fraud isn't detected, the procedures should identify some cash recovery for the organization.)

When considering automated tests for this area, ask the following questions:

- Does the vendor provide actual goods/services to the company or is it a fraudulent billing scheme?
- Has the vendor returned mistaken overpayments to the company?
- Is the vendor charging the negotiated pricing and is it consistent with the marketplace?
- Are vendor inventories becoming obsolete?

Please note that this list focuses only on questions that can be answered using computerized tools because some questions are too subjective, such as "Are the vendor's goods and services of good quality?" and "Does the organization actually need the provided service?" Here we'll try to answer the questions in the list using data analysis.

Does the Vendor Provide Actual Goods/Services to the Company or Is It a Fraudulent Billing Scheme?

The following reports can flush out any vendors that aren't providing anything to the company and are *shells* in essence. (The list isn't complete, of course, but does provide some of the most likely places where a fraudulent vendor can turn up.)

- Check blank address information. Filtering or sorting a vendor master file by address, phone number, bank account details, and/or tax ID can reveal shell vendors.
- Identify vendors added during the period under review. It's completely normal for companies to add frequent new vendors. However, it's probably more unusual than normal to see a new vendor's material valued invoices shortly after the vendor is signed. Therefore, scrutinize those new vendors with material new purchases.
- Review payments with little or no sequence between invoice numbers. Complex algorithms can detect such occurrences in phony invoices, but you could get the same result by sorting the data file by vendor and invoice number and then manually scanning the records.
- Match the vendor master file to the employee master file on various key fields. It's amazing how many employees still forget to use different addresses or PO Box from their own when fraudulently paying themselves. Match the two master files by telephone number, address, tax ID numbers, or bank account details.
- Match vendor address information against external data sources. Matching vendor information to external databases such as Dun & Bradstreet or LexisNexis can be powerful, but it's difficult to complete tests in an automated fashion because of slight differences in the address information. Manually compare recent vendors to external data sources to determine, among other traits, if the address is residential or business and if the total payments to the vendor represent its entire reported annual revenues, which is likely for fraudulent vendors but unlikely for actual businesses.

NOTES

NOTES

Has the Vendor Returned Mistaken Overpayments to the Company?

Running the following reports might identify actual fraudulent transactions on the company's books or vendors that are ripe for an audit because they have the highest probability for holding on to credits due to the company.

- Duplicate payments based on the invoice number, invoice date, vendor number, and invoice amount. You can find duplicate invoices with any combination of these data fields by using the Microsoft Access™ Duplicates Wizard or ActiveData for Excel™ (both available at a discount in the ACFE's bookstore). The duplicates could be fraudulent attempts by employees or vendors to increase payments to the vendor or they could be innocent errors. However, they could all be eligible for cash recovery.
- Extract round dollar payments and summarize by vendor. Most software uses the MOD() function to identify round dollar items paid to vendors. For example, =MOD(A1,1000) would determine if cell A1 in Microsoft Excel™ is perfectly divisible by 1,000. If the result of this function is 0.00, then the amount in A1 is perfectly divisible by 1,000. Also, the 1,000 can be changed to 10000000 to determine all invoices that are perfectly divisible by 1,000,000.
- Final payments to any vendor that exceed the 12-month average payments to that vendor by a specified percentage (such as 20%). Scrutinize all vendor payments that are larger than usual but particularly concentrate on categories that are more likely to be associated with fraud, such as consulting, cleaning services, maintenance, etc.

- Summarize debit memos by vendor, issuer, and type. In the accounts payable trial balance, which represents all open accounts payable invoices, debit memos are negative items that are unusual because they're associated with company/vendor adjustments. Increasing trends and/or unapplied debits could be employee attempts to cover unauthorized payments or could represent forgotten credits from vendors that represent cash recovery to the company.
- Review vendor statements to company activity. If you can obtain vendor statements that have all open and paid invoices and adjustments on the vendor's books, you can compare the statements to company activity. You'll quickly notice any overpayments of invoices or adjustments not realized by the company, and they could highlight issues of a vendor's management of the company's account.

Is the Vendor Charging the Negotiated Pricing and Is It Consistent with the Marketplace?

You'll probably have to look outside your company for marketplace trends to assess the reasonableness of pricing. You might have to purchase external databases that track product pricing or team up with strategic sourcing consultants with experience in the area. The cost usually is offset by identifying any overpriced items and improving future vendor negotiations. Remember that any kickbacks a vendor gives to employees are generally based on increased profit margins offered to that vendor. If you can minimize such margins, the kickbacks and other perks offered to employees also will be minimized.

- Identify increasing vendor relationships. Align vendor activity summaries for the previous three

NOTES

years and you can identify increasing or other unusual trends.

- Locate sole-sourced vendors. Summarize accounts payable activity by general ledger account and vendor, and you can quickly identify any general ledger accounts with one vendor. Those might be legitimate but also could represent inappropriate bidding procedures and lack of competition.
- Match negotiated price tables to company purchase orders. Pricing on purchase orders normally can be adjusted to be higher or lower than the contract price with the vendor (stored in a separate data table). Compare the two to highlight unique error situations and long-term inappropriate trends.
- Calculate the ratio of the largest purchase to next largest purchase by vendor. Any large ratio difference might identify a fraudulently issued largest purchase.
- Calculate the annualized unit price changes in purchase orders for the same product in the same year. Increasing trends within the past year compared to the two previous years can quickly locate specific products wrapped up in a kickback scheme.

Are Vendor Inventories Becoming Obsolete?

While much of the focus up to this point has been on identifying overpayments from the perspective of pricing and invoice payments, a vendor can also be overpaid through quantities ordered. This will become apparent when looking at inventory through the lens of normal company usage and obsolescence.

- Age inventory by the date of last part issuance. Quickly find inventory that has been over-purchased and is sitting on shelves by looking at the last time parts or products were issued or sold.

NOTES

- For each part or product, calculate the number of months that it's been in the inventory and extract those with a high number of months. A summary of the inventory usage/sales files by part number over the previous year can be aligned back to the current inventory. Then a simple division of the on-hand inventory units to the total sales over the previous year can determine the number of years remaining in inventory usage. Any exceptionally high "years left" represent a potential for intentionally over-purchasing inventory.
- *Trend in* obsolete inventory over two or more periods. Inventory older than a few years will most likely be considered obsolete. Summarize by vendor any inventory falling into obsolescence and, if possible, trend to vendors in the prior period with obsolete inventory.

Depending on the answers to these questions, you can decide if fraud patterns exist and thus conduct a more extensive review of company documents and company employees. You might also want to go to the extra step of completing a vendor audit on-site to get to the bottom of the story.

NOTES

NOTES

1.0 Vendor Scoring—Identifying the Best Vendors to Audit for Fraud Recovery

In any organization, there may be thousands of vendors, but it is difficult to determine which vendors are the most ripe for review from an error, fraud, or cost recovery perspective. Usually, these vendors are identified as the business naturally realizes which vendors have the most issues, adjustments, and company time to reconcile their account.

However, waiting for these vendors to show their faces through process errors is a reactive versus proactive approach. In order to identify the vendors that are most problematic now, we can run a variety of reports and string the results together so as to identify those vendors showing or "scoring" the most. In essence, we can mathematically calculate risk for each vendor based on the reports we consider "risky".

Using a sample ACL project of four scripts, we can show how to calculate a score for a vendor by combining the results for three reports using a key field: the unique vendor number. You can download a zip folder containing an ACL document with scripts here: www.acl.com/2014/07/16/vendor-scoring-identifying-the-best-vendors-to-audit-for-fraud-recovery-part-1-of-2

The provided scripts will execute three reports that are simple but do have a tendency to detect vendors with a potential for fraud as explained below:

1. *Report 1 – Score = 34% – Vendors with low volume and high value transactions*: Many times a fraudster will have fewer opportunities to post a phony invoice so they will try to get as much out of the organization as possible in few attempts.

2. *Report 2 – Score = 33% – Vendors with blank information in critical fields*: Not everyone remembers to add the key fields to a vendor record when making a phony one.
3. *Report 3 – Score = 33% – Vendors with over 10 round-dollar payments (values divisible by 1,000)*: The fraudster may not be creative enough to arrive at amounts that have different digits, especially in the cents.

The key to this scoring approach is to string together these reports, which is done via the Scoring_Script. It will kick off the three report scripts (above) and then, more importantly, assign a score to each report which is joined together on a key field, the supplier number.

More specifically, in building the score the ACL-provided Scoring_Script:

- Asks for some variables around high values, low volumes, and the level of round dollars that are considered high to the company. Then it asks for an invoice and supplier table for analysis with a limited number of fields needed as listed below:
  - Invoice Table
    - Vendor Number
    - Invoice Number
    - Invoice Amount
  - Vendor Table
    - Vendor Number
    - Vendor Address
    - Vendor Postcode
    - Vendor Telephone
    - Vendor Tax ID

  ** Any fields that you do not have, you can select the BLANK_TEXT, BLANK_AMOUNT or

NOTES

BLANK_DATE fields accordingly, which we create for you when running the script.

- It runs the scripts REPORT_1, REPORT_2, and REPORT_3 in ACL with each report's results being at the Supplier or Vendor Number level. Therefore, each of the records from these scripts is a unique supplier number.
- Then, each time it runs a report, a new field is created in the report with the score being written to each unique supplier number record using the ACL code below (for Report 1 below as set to 34%):

```
DELETE FIELD REP1SCORE OK
DEFINE FIELD REP1SCORE COMPUTED .34
```

** – Please note that the scores listed above are assigned to each report, respectively in the specific report scripts (i.e., REPORT 1 has the above code listed in the report itself). This could be moved to the Scoring_Script in ACL but we have done this to make the below step of joins easier to read (see Step #4).

- Using a supplier extract table of key fields from the supplier table (*supextract*), the script joins the report result score fields (outputs from the REPORT_1, REPORT_2, and REPORT_3 script results as in step 3 above) to this *supextract* table on the supplier number.

```
OPEN "supextract"
OPEN HIGHDOLL_LOWTRAN SECONDARY
JOIN PKEY %supnum% FIELDS %supnum%
SKEY %supnum% WITH REP1SCORE
PRIMARY TO "joined" OPEN PRESORT
SECSORT
CLOSE SECONDARY
```

NOTES

NOTES

```
OPEN joined
OPEN BLANK_SUPP SECONDARY
JOIN PKEY %supnum% FIELDS %supnum%
REP1SCORE SKEY %supnum% WITH
REP2SCORE PRIMARY TO "joined1" OPEN
PRESORT SECSORT
CLOSE SECONDARY

OPEN joined1
OPEN LOT_OF_ROUND SECONDARY
JOIN PKEY %supnum% FIELDS %supnum%
REP1SCORE REP2SCORE SKEY %supnum%
WITH REP3SCORE PRIMARY TO "joined"
OPEN PRESORT SECSORT
CLOSE SECONDARY
```

- With a supplier extract table now having each of the report scores aligned, the script concludes by calculating the total value and volume for each supplier using the invoice table, which is joined back to the *joined* table above, creating the Final Score Report. It will finish by deleting all temporary join tables and calculating a total score by vendor.

Developing a vendor score is an excellent way to reduce false positives and improve sample selections of vendors. This simplistic approach only provides a binary YES or NO if the supplier is on the report; it does not provide a severity of each vendor based on their full set of vendor transactions showing on a given report. To provide an improved approach, we will now take the scoring concept explained at the vendor level and apply it to the transactions for that vendor. Transactional Scoring will take the analysis to the transaction level in order to calculate those transactions which are the most risky for review.

We arrived at this approach of transactional scoring by first running the vendor score and realizing that the summary per vendor was not accurate enough, as it would identify the entire vendor each time even if only one of the vendor's invoice transactions displayed on the report. Therefore, we determined it was better to count up the invoice transactions that showed on an invoice-level report and then divide that by the total number of invoice transactions for that vendor. Instead of one vendor transaction leading to a full score, we were now taking the number of transactions for that vendor showing on the invoice report (i.e., 100) divided by the total invoice transactions (for example 1,000) to arrive at a 10% report score to transactions ratio. We will use the same code for Vendor Scoring and explain how it can be edited in order to run at an invoice transactional level. You can download a zip folder containing an ACL document with scripts here: www.acl.com/pdfs/ACL_Vendor_Scoring_Script.zip

Transactional Score Ratio Defined and Setting the Record Number

Taking the example above, we wanted to define the transactional score ratio as follows:

The rate of (Error/Fraud/Control Failure/Recovery) report occurrence of a transaction whereby the score is set at 100% or 1 for each transaction based on the number of reports it appears on as run at this transactional level.

So, if we had 25 reports in the transactional scoring model, we would set the score of each report to 4% or .04. If the transaction appeared on each of the 25 reports, it would receive a total score of 100% (25 reports * .04 score). Over time, the score could be customized as the scores need not be equally rationed but rather customized by report to the location's data reviewed. For example, in some locations

NOTES

reviewed, Report 2 may carry twice as much weight (or set at .08) as Report 5 (as set equally at .04).

Please note in our example that we will be using the Invoice Header table where each physical invoice represents one invoice record in the file. This is normally designated by a unique transaction or voucher number assigned to that invoice. Assuming the invoice table is at the invoice transaction level, we will assign a unique record number to each record as done below:

```
DELETE FIELD RECORD_NUMBER OK
DEFINE FIELD RECORD_NUMBER
COMPUTED
LAST("000000000000"+ALLTRIM(STRING(REC
NO() 12)) 12)
```

This unique record now becomes the key field for joining the table, and the table is no longer the supplier table as in Part 1 but now is the invoice header table. In this way, you are able to take each invoice transaction, assign a unique number to it, and then join the report results for the various invoice transaction scoring reports on the unique number created above. This can be joined back to the original invoice table, so each of the 25 invoice report scores would each be a new score field joined to the original invoice table. In the Part 1 post, we completed the join on supplier number and now we are doing so on the unique record number. This would need to be adjusted in the provided ACL project script *Scoring_Script*.

Summarizing the Score and Calculating the Score Ratio

Assuming the invoice table has now been updated with all of the scores for each of the 25 reports, a total score (assumed to add to 100% or 1) would be calculated for each invoice transaction. In Part 1, we had a score by

NOTES

vendor and now it is by invoice. This is where it gets interesting.

Now that you have the score at the transactional level, you can summarize it on supplier number and, when doing so, total not only the score but also the number of transactions. This will allow you to divide the two in order to obtain a score ratio. For example, a vendor summary on the transactional score and their total transactions may lead to:

- Vendor A Summary—491.12 in transactional score out of a total of 1,000 transactions, or a 49.11% score ratio
- Vendor B Summary—491.12 in transactional score out of a total of 2,000 transactions, or a 24.56% score ratio

Beyond the supplier number summary, the score ratio can be applied to all of the *who*, *what*, *when*, and *where* perspectives individually so as to identify the ratio of error/fraud to total transactions in each perspective. You may identify a pattern between a departmental summary of the score ratio or on the month/year combination. The various perspectives in invoice data usually translate into the following:

- Who—"Enterer," Approver, Department
- What—Invoice Type, Input Source
- When—Date, Month/Year
- Where—World Location, Company Location Code

Conclusion

After developing a vendor score, an improved approach is to focus on the transactional invoice score and, further, the score ratio for each of the *who*, *what*, *when*, and *where* perspectives. This additional viewpoint analysis may spot an area that is prone to error and fraud that may not be as visible at the vendor level.

NOTES

Sampling Data in ACL—Instructional Videos

The following supplemental videos provide online instruction for different techniques in selecting sampling data using Excel and ACL Software:

- **Sampling in Excel and Other Simple Add-In Products**. Explains how a transactional score can be calculated and then combined to produce a master transactional score. (Length 1:53:04) http://bit.ly/1AKqd5Q
- **Sampling Theories—Translating into ACL Commands**. Provides an overview of the sampling process, inputs to sampling, and useful guides for future references. (Length 22:32) http://bit.ly/1KRhy7z
- **Record Sampling—Random, Cell, and Fixed Interval Sampling**. Explains the approach to completing a random sampling using the three ACL supported methods. (Length 12:33) http://bit.ly/1F3LBZc
- **Stratified Sampling—Building a Script to Help Your Sampling Efforts**. Explains the approach to completing a stratified sample in ACL and ActiveData for Excel (to show variety in approach) (Length 14:31) http://bit.ly/1F9KtFf
- **Transactional Scoring Basics and Identifying Rare Items.** Explains a simple scoring approach, which is followed by a detailed explanation of various additional ways to score data looking at number, digits, and rule-based generator engines that highlight deviations in your data automatically. (Length 17:04) http://bit.ly/1zR1wK1
- **Calculating and Combining the Score**. Explains how a transactional score can be calculated and then combined to produce a master transactional score. (Length 8:04) http://bit.ly/1cuUVdS

NOTES

Fear Not the Software

Fraud-fighting with Data Analysis Tools

FINDING THE FRAUD 'NEEDLES' IN THE HEALTH-CLAIMS HAYSTACK

By Richard B. Lanza, CFE, CPA/CITP, PMP and Si Nahra, Ph.D.

For this column, Rich Lanza is joined by coauthor, Si Nahra, Ph.D., president and cofounder of Health Decisions Inc. – ed.

Health-care fraud detection historically has focused on finding the criminals. However, fraud *prevention* is best done by addressing the underlying waste and abuse that pervades America's health-benefits market, disguises fraud, and allows it to flourish.

Health-care claims are like a giant haystack riddled with needles of waste and abuse. However, some of those needles are bent – they're fraudulent. Data mining acts like a giant magnet that pulls the needles out of the haystack so the fraud examiner can sort them, find the fraud, and reduce overall waste and abuse.

Of course, advances in computer technology and data-mining software have revolutionized fraud detection. Any interested health-plan fiduciary, manager, administrator, or internal auditor (as well as a fraud examiner) can now detect health-care fraud. Data mining narrows the field of suspects to a manageable number by categorizing them by standard criteria. This analysis (complemented by human judgment) detects potential fraud that would have gone unnoticed. Fraud examiners can then begin their examinations.

Advances in computer technology and data-mining software have revolutionized health-care fraud detection. Any interested health-plan staff member can now detect health-care fraud.

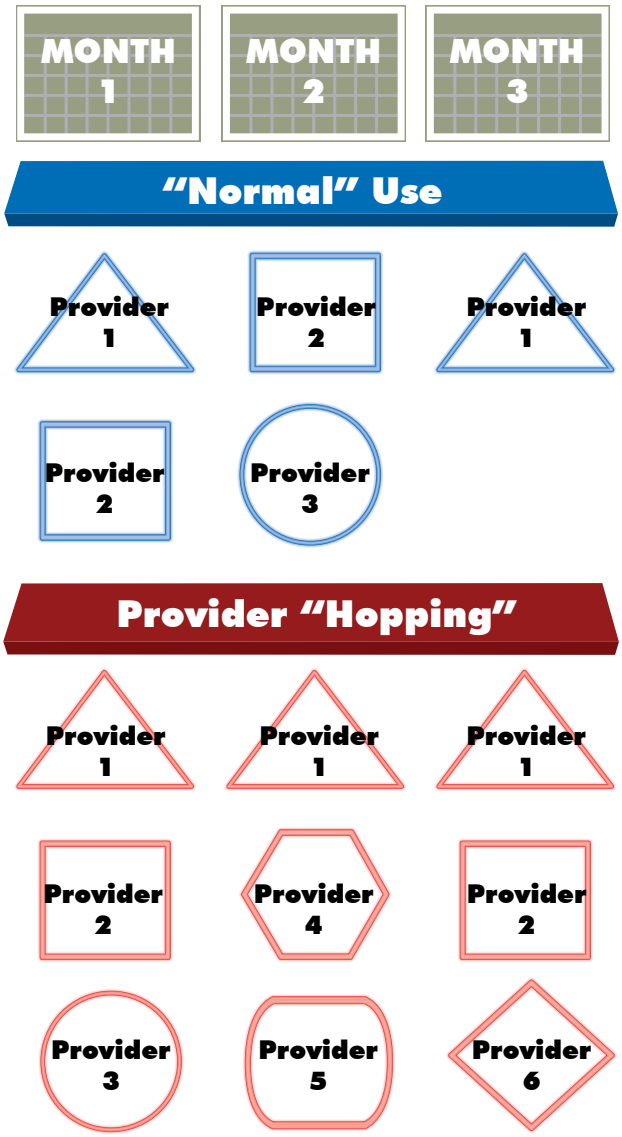
Health-care fraud takes many forms. Here we'll give some examples that illustrate how the tools of computerized data mining can be used to ferret out fraudulent behavior hiding in the midst of waste and abuse.

MAPPING UTILIZATION

Individuals might fraudulently utilize drugs covered under a medical plan to feed addictions or as kickback arrangements. Mapping networks of actual use of individuals in the haystack – showing associations with all providers they've seen over a period of time – can reveal this form of fraud. Mapping shows people who "hop" across different providers for no apparent reason. By itself, each provider visit might seem appropriate. But the total picture shows a map that differs markedly from other users in the same population and can exhibit potential fraudulent behavior that can be isolated and examined. Providers *can't* do such cross-provider histories, and payers rarely complete them because they complicate their primary task of payment processing. Data-mining analyses can reveal these patterns.

The Use Mapping illustration on page 16 shows a comparison of a "normal" person using providers and a provider hopper who might be revealed by mapping. Normal use shows a person in Month 1 going to a primary-care doctor who then refers that person to a specialist. In Month 2 the specialist directs the person to a hospital or diagnostic center. In Month 3 the person returns to the primary physician. Contrast this pattern

with the hopper's. Each month the person visits several different physicians with little rationale to the care patterns.



Use Mapping

UNIQUE PROVIDER-PROCEDURE COMBINATIONS

Providers love to maximize their income, but when do they cross the line and overbill? One way to tell is to examine each unique combination of provider and procedure. Because every health plan will have hundreds or thousands of providers each billing hundreds or thousands of procedures, the number of these unique combinations can easily reach into the millions – well beyond the range of any traditional audit review process. However, computerized data mining makes it possible to document and review each of these millions of unique combinations.

This is a particularly effective technique for Preferred Provider Organization (PPO) health plans now popular throughout the country. In these arrangements, providers gain "preferred" status and additional patients by agreeing to accept significantly discounted fees as payment in full. Gaps in the claim payment process can occur especially when an outside "re-pricer" (that is, a service vendor that receives the original bill, applies the PPO discount, and forwards the re-priced claim to the health plan for payment) is used to assess the discount. Some providers fraudulently seek out and exploit these gaps in the claim discount process by evading discounts while still reaping the added patient volumes from their "preferred" status.

The table below is a simplified version of a "missed discount" analysis that can help reveal waste and expose potential fraud. Assume each provider is a "preferred" provider.

	Provider 1	*Provider 2*	*Provider 3*	*Provider 4*
Procedure 1	*100%*	*0%*	*100%*	*10%*
Procedure 2	*100%*	*0%*	*80%*	*10%*
Procedure 3	*100%*	*0%*	*90%*	*10%*
Procedure 4	*100%*	*0%*	*80%*	*10%*
Procedure 5	*100%*	*0%*	*0%*	*10%*

Percent of Claims Discounted by Procedure and Provider

Provider 1 is clearly no problem because *all procedures were discounted*. If Provider 2 weren't preferred this result might be acceptable; however, as a preferred provider, further investigation is clearly warranted. This is a preferred provider that has not had any discounts applied. Provider 3 shows a pattern indicative of contractual or administrative issues. The fact that a majority but not all claims were discounted would entail eliminating problems with re-pricing administration before questioning the provider. The pattern for Provider 4 concerns us because it might show a situation in which some discounting was done to determine how to evade further discounts. Further analyses to document the amounts of payment would also help focus any follow-up investigation.

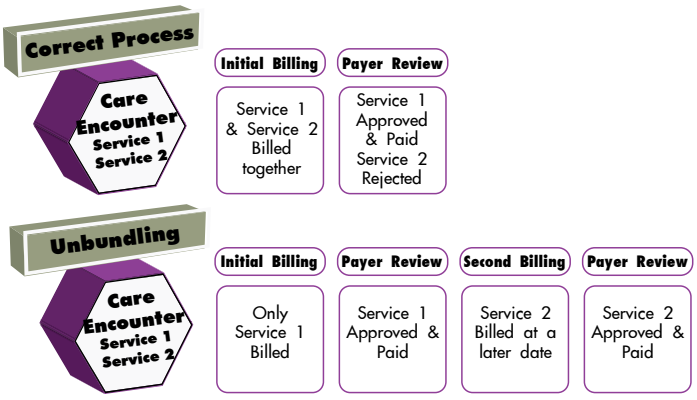
CONSTRUCTING CARE ENCOUNTERS

Providers also maximize income by billing for excessive services, which is sometimes referred to as "unbundling." To detect this practice, you must find individual care "encounters." An

encounter is an incident of care provided to the same person on the same day by the same provider. A provider that consciously spreads out billings for a single encounter over time can often evade standard detection techniques (if any are in use at all). You can detect these shady provider billing practices by compiling a claim history file covering several quarters or a year and constructing encounters from that history. The illustration below shows how unbundling can occur. In the Correct Process all services for an encounter are billed at the same time. Service 1, for example, might be a physical and Service 2 might be for supplies used during the physical. Because Service 1 covers Service 2, Service 2 is duplicative and should be denied. However, if Service 2 is held for a time and billed at a later date it can (and often is) viewed in isolation and not associated with the overall physical and paid in error. Providers who consciously and consistently engage in such billing manipulation warrant a closer examination for potential fraud.

ENROLLMENT FRAUD

With data mining, you can detect enrollment fraud perpetrated by employees or covered persons. It's not uncommon to find 5 percent to 15 percent ineligible persons enrolled in a health plan. Some of these might be innocent mistakes or administrative glitches, but others are conscious attempts to defraud. You can find this type of fraud only by mining the data to compare employment rosters to enrollment rosters, cross-reference enrollment rosters across payers, or accumulate documentation that corroborates eligibility via enrollee surveys or other follow-up.



How Unbundling Occurs

CASE STUDY: SUSPICIOUSLY HIGH COSTS

The national Health Plan X noticed that one state had unusually high costs. The company performed a variety of statistical analyses that confirmed this, but the analysts couldn't explain why it was happening.

Benefit managers dug beneath the statistics to reveal that the high costs were localized in one community but not within the whole state. Within that community there was a small group of people that had excessive levels of use associated with these providers. Plan representatives contacted these providers. Use levels quickly fell and costs in that community came into line with expected values. The company realized the possibility of fraud but didn't pursue it because they didn't have any fraud examiners on staff.

CASE STUDY: MASSIVE UNBUNDLING

Health Plan Z noticed that the volume of service billings was increasing dramatically even though the care episodes weren't. This led the company to question billing patterns and to conduct data mining that reconstructed care encounters over the prior 15-month period. The data mining documented massive amounts of unbundling. Investigation into the cause revealed that the plan administrator had stopped monitoring for unbundling and that providers had became aware of this gap. The claim administrator was terminated and procedures for monitoring excess billings were put in place.

WHAT IS DATA MINING?

There is no single definition of data mining. Nor is there any available "off-the-shelf" software for conducting data mining. The process is equal measures of technical ability and human judgment. Unlike statistical analyses that use defined calculations with known documents and data pattern descriptions, data mining is more intuitive.

Generally, data mining entails three steps:

1. **Preparation: This is largely technical and entails data acquisition, standardization, and conversion. While many techniques exist for performing these tasks, it's preferable to have them result in a new data asset that is compatible with Microsoft® products. Structured Query Language (SQL) is a proven tool for supporting data mining, and produces findings that can be communicated to any user with Microsoft® Office or similar software.**
2. **Mining: The actual conduct of data mining starts with a problem statement expressed in words by any interested party. This is then interpreted and expressed in the form of programming logic. Initial results of this program are reviewed to determine if they accurately capture the intent of the problem statement. Adjustments to the program are then made to bring it into alignment with the intent and purpose of the problem statement.**
3. **Action: Once the program is refined, results can be reviewed and interpreted. Often one of the greatest benefits of data mining is the ability to focus review activity on problem areas and quantify the overall magnitude of a perceived issue. Unlike statistical samples that review a representative selection of cases (most of which will be correct), data mining uses human judgment and analysis to review "problem" cases and decide if follow-up is warranted.**

Data mining is a powerful tool in examining health-care statistics. However, it's only as useful as the personnel available to review and act on its results. This is an ideal role for fraud examiners.

FINDING HEALTH-CARE FRAUD ONLY WITH DATA MINING

In each of these examples, you can use data mining to expose previously undetected areas of potential fraud and find waste and abuse, which really is "fraud" without the criminal intent. Just the act of looking for such inappropriate behavior can have a chilling effect on fraudsters or would-be fraudsters; they're likely to move on if they fear detection. Even when fraud isn't present, identifying and quantifying waste and abuse levels can help focus management priorities and provide the basis for quality improvement monitoring.

In the past, America's self-funded health plans and group health insurers have been able to exert a collective influence on the health-care system. They can do it again by insisting that all post-payment data be mined to expose the root causes of fraud and the waste and abuse that provide its cover.

Si Nahra, Ph.D.**, is president and co-founder of Health Decisions Inc., an IT company specializing in post-payment and enrollment data auditing services for group health benefit clients throughout the country since 1985. His e-mail address is:* ***si@healthdecisions.com.

Rich Lanza, CFE, CPA/CITP, PMP**, president of Cash Recovery Partners LLC, helps companies identify their hidden financial assets, mostly by using technology and referring them to specialists. He has two free Web sites:* ***www.findmillions.net, *and* ***www.auditsoftware.net***. *His e-mail address is:* ***rich@richlanza.com***.