Computer-Assisted Audit Tools & Techniques (CAATTs)

Ed Williams, CFE, CIA
Manager – Internal Audit and Forensic Services

James Specht, CFE, CPA, CIA, CISA, MBA
Professional – Internal Audit and Forensic Services
Agenda

• What are CAATTs?
• CAATTs Benefits
• Fraud 101
  - Types of Fraud
  - Fraud Triangle
  - Fraud Detection Plan
• Rules of Electronic Evidence
• Data Mining
• Benford’s Law
• Data Mining Scenarios
• Questions
What are CAATTs?

**Computer-Assisted Audit Tools & Techniques**
- Any computer program or tool used in the audit process
  - Work papers
  - Programs
  - Test plans
  - **Data mining tools**
  - Report writers

**Other Terminology**
- CAATs
- Digital analysis
- Data interrogation
- Digital forensics
- Data analysis

*Our focus today is on the use of data mining tools and techniques to help detect fraud*
CAATTs Benefits

• Expanding digital world - audit processes need to evolve
• Allow for more thorough, productive and efficient audits and investigations
• Provide for automation of previously manual procedures
• Provide efficient monitoring capabilities
• Add credibility to the audit and investigation process
Fraud 101

• Finding potentially fraudulent situations or transactions is just the start of a much larger process
  • Must understand the basics of fraud prevention, detection and investigation
  • Why people commit fraud
  • Types of fraud that can be committed
  • How to develop fraud detection and investigation plan
  • Who should be involved
  • Collecting and preserving evidence
  • Chain of Custody protocol
  • Attorney-Client Privilege

• Computer Forensics (beyond the scope of today’s discussion)
  • Hard-drive imaging
  • File viewers
  • Un-erase tools
  • Text searches
  • Forensic Suites – both collect and analyze computer data
Fraud 101

Prevention is key!

• Much more productive and cost effective than fraud detection and investigation

Fraud Prevention Activities

• Internal controls
• Fraud awareness training
• Culture/Code of Ethics
• Hotlines
• Internal Audit
Types of Fraud

Asset misappropriations (80%)
• Theft or misuse of company assets
• Payroll schemes
• Billing and Procurement
• Expense reimbursement
• Check tampering

Corruption (13%)
• Wrongful use of influence in a business transaction to procure some benefit
• Kickbacks
• Bid-Rigging
• Bribery

Fraudulent statements (7%)
• Falsification of financial records, transactions or statements

Source: ACFE 2008 Report to the Nation
Fraud Triangle

Pressure
- Greed
- Living beyond one’s means
- High personal debt
- High medical bills
- Poor credit
- Personal financial loss
- Unexpected financial needs
- Addictions
- Extramarital affairs

Opportunity
- Lack or circumvention of internal controls
- Past failure to discipline offenders
- Management apathy
- Ignorance or incapacity to detect fraud
- Lack of an audit trail

Rationalization
- The organization owes it to me
- I am only borrowing the money
- They can afford it
- I deserve more
- Everyone else is doing it

Fraud Detection Plan

Once a fraud is reported or suspected …

- Develop a fraud hypothesis
- Obtain relevant data
- Design and execute CAATTs
- Analyze results
- Determine if the hypothesis is supported
Rules of Electronic Evidence

- Acquire evidence without altering or damaging the original
- Authenticate that recovered evidence is the same as the original
- Analyze the data without modifying it
- Maintain chain of custody
# Chain of Custody

## Evidence Information / Description

<table>
<thead>
<tr>
<th>General Description:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Make:</td>
<td>Model:</td>
</tr>
<tr>
<td>Serial #:</td>
<td>Part #:</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Date/Time</th>
<th>Released By</th>
<th>Received By</th>
<th>Reason</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date:</td>
<td>Name:</td>
<td>Name:</td>
<td></td>
</tr>
<tr>
<td>Time:</td>
<td>Signature:</td>
<td>Signature:</td>
<td></td>
</tr>
<tr>
<td>Date:</td>
<td>Name:</td>
<td>Name:</td>
<td></td>
</tr>
<tr>
<td>Time:</td>
<td>Signature:</td>
<td>Signature:</td>
<td></td>
</tr>
<tr>
<td>Date:</td>
<td>Name:</td>
<td>Name:</td>
<td></td>
</tr>
<tr>
<td>Time:</td>
<td>Signature:</td>
<td>Signature:</td>
<td></td>
</tr>
</tbody>
</table>

*Confidential Information*
Data Mining
Data Mining

Definition
• The automated extraction of hidden predictive information from a data source

Benefits
• Detects fraud sooner
• Provides for real-time analysis
• Very targeted analysis
• Continuous and automated monitoring
• Audit trail/logs*
• Ability to analyze unlimited amount of data*
• Analysis always sits on top of the data*

(*) Specialized software tools only
Data Mining Approach

Phase I
- Define analysis objectives
- Gain an understanding of the business and technical environment
- Define data requirements

Phase II
- Request and collect data
- Validate completeness and integrity of data
- Develop an understanding of the data

Phase III
- Perform the audit and analysis
- Identify exceptions and discrepancies
- Confirm and validate errors
- Determine the impact of observations

Document the process in order to reproduce the data
Data Mining Hurdles & Challenges

- Perception that you need to be an IT expert
- Hardware and software limitations
- Numbers and dates that get imported as text or other formats
- Desired data broken between multiple fields
- Data integrity (use control totals)
- Attempting to accomplish too much in single queries
- Very large data sets
- Keeping your data/results organized
- False positives/jumping to conclusions
Data Mining
Lessons Learned

• Don’t underestimate the initial time commitment
• Anticipate multiple attempts to get it just right
• Clearly define the objectives of your analysis
  • controls assessment/audit
  • trending and analysis
  • proactive fraud detection; reactive fraud investigation
• Develop and use a formal data-request procedure
• Coordinate (partner) data extraction efforts with authorized and capable IT or database personnel
• Be aware of lead times
Data Mining Lessons Learned

- Determine data transfer media
- Obtain a thorough understanding of the transactional processes
- Obtain a thorough understanding of the data you are analyzing
- Understand how data is stored (databases, data warehouses, etc.)
- Maintain integrity of the source data – “read only”
- Ensure the integrity of the data collected – check it
- Determine if the data extraction procedures are going to affect normal processing
Data Mining Tools

**Spreadsheet software**
- MS Excel
- Lotus
- Apple Numbers

**Database software**
- MS Access
- Lotus
- SQL
- File Maker Pro
- Paradox

**Specialized Software**
- ACL
- IDEA
- DATAS
- Active Data

- No one tool can do it all
- Be proficient with several different tools
Data Mining Techniques

**Aging**
Produces aged summaries of data based on established cutoff dates.

**Append/Merge**
Combines two files with identical fields into a single file. An example would be to merge two years’ worth of accounts payable history into one file.

**Calculated Field/Functions**
Creates a calculated field using data within the file. For example, the net payroll pay to an employee could be recalculated using the gross pay field and deducting any withholding/taxes.
Data Mining Techniques

**Cross-Tabulate**
Allows you to analyze character fields by setting them in rows and columns. By cross-tabulating character fields, you can produce various summaries, explore areas of interest and accumulate numeric fields.

**Benford's Law**
Benford's Law gives auditors the expected frequencies of the digits in tabulated data. If a data set doesn't follow these patterns, this may be a cause for auditor concern and review.

**Duplicates**
Identifies duplicate items within a specified field in a file. For example, identify duplicate billings of invoices within the sales file.
Data Mining Techniques

Export
Creates a file in another software format (e.g., Excel, Word) for testing. An example would be to export customer address information to Word for "Mail Merging" to customer confirmation letters.

Extract/Filter
Extracts specified items from one file and copies them to another file, normally using an "if" or "where" statement. Examples include extracting all balances over a predefined limit.

Gaps
Identifies gaps within a specified field in a file. For example, identify any gaps in check sequence.
Data Mining Techniques

**Index/Sort**
Sorts a file in ascending or descending order. An example would be sorting a file by Social Security number to see if any blank or "999999999" numbers exist.

**Join/Relate**
Combines specified fields from two different files into a single file using key fields. This function is used to create relational databases on key fields. It can also be done in an unmatched fashion to identify differences between data files.

**Regression**
Regression analysis uses statistical means to calculate a dependent variable balance (such as net sales) based on various independent variables (e.g., product purchases, inventory levels, number of customers, etc.).
Data Mining Techniques

**Statistics**
Calculates various statistics on a selected numeric field. These may be total positive items, negative items, average balance, etc.

**Stratify**
Counts the number and dollar value of records of a population falling within specified intervals. Stratifications also provide a useful view into the largest, smallest and average dollar transactions.

**Summarize**
Accumulates numerical values based on a specified key field. An example would be summarizing travel and entertainment expense amounts by employee to identify unusually high payment amounts.
Benford’s Law
Benford’s Law

- Named for the late Dr. Frank Benford, a physicist at the General Electric Company. In 1938, he noticed that pages of logarithms corresponding to numbers starting with the numeral 1 were much dirtier and more worn than other pages.

- Data profiling technique - law of digit frequency

- In large naturally occurring data sets, the initial digits of a numeric field will follow a very predictable pattern with “1” being the most common digit and “9” appearing the least.
Benford’s Law

• Useful in helping identify anomalies or red flags in data

• Tax authorities in the United States and other countries use it to analyze tax returns

• Aids in fraud detection because fraudulent transactions/data are not to conform to Benford's Law and therefore will be detectable. The fraud perpetrator isn’t writing checks, creating journal entries, or creating fraudulent invoices with Benford’s Law in mind.

• Does not apply to manipulated, invented number sets or sets with imposed ceiling and/or floors

• Does not apply to totally random number sets

• Does not apply to small number sets
Benford’s Law

Numbers appearing on the front pages of newspapers

3,141 county populations in the 1990 U.S. Census

Dow Jones Industrial Average from 90-93

All follow Benford's law within 2%

(From "The First-Digit Phenomenon" by T. P. Hill, American Scientist, July-August 1998)
Data Mining Scenarios
Data Mining Scenarios

General Ledger
• Analyze, select and confirm shareholders’ accounts and equity
• Calculate financial ratios (and changes) for sales/assets, debt/equity, etc.
• Compare summaries by major account in any order (low-high, high-low)
• Compute weighted and average interest rates across any periods
• Calculate and sort percentage variances in accounts between periods
Data Mining Scenarios

General Ledger

- Prepare trial balances by tax groups and calculate various tax provisions
- Print custom balance sheets, P&L statements, cash-flow analyses, etc.
- Select any journal entry for test and audit purposes, from any period
- Speed closing with quick account reconciliations and adjustment transactions
- Verify feeder systems data in a consolidated ledger
- Confirm consistent use of ledger accounts and codes
Data Mining Scenarios

Accounts Payable
• Calculate average days to pay invoices
• Correlate vouchers or invoices posted vs. purchase order amounts
• Create activity summary for suppliers with duplicate products
• Determine frequency of non-purchase order check requests
• Evaluate purchasing contract compliance
• Extract invoices posted with duplicate purchase order numbers
Data Mining Scenarios

Accounts Payable

- Extract total posted invoices for the year for accurate vendor rebates
- Find duplicate invoice payments
- Find freight and tax overcharges
- Generate cash requirements by bank, period, product, vendor, etc.
- Identify cash discounts not taken
- Examine P-Card transactions, activity and trends
Data Mining Scenarios

Accounts Payable
- Identify cash discounts not taken
- Identify distributions to accounts not in suppliers’ account ledgers
- Isolate vendor unit price variances by product, over time
- List missing accounts payable check numbers
- Reconcile check register to disbursements by vendor invoice
- Review 1099 compliance
- Review recurring monthly expenses and compare to posted/paid invoices
Data Mining Scenarios

Cash Disbursements

- Extract cash disbursements by vendor/supplier for audit verification
- Generate vendor cash activity summary for support in rebate negotiations
- Identify check requests that exceed set or selected dollar amounts
Data Mining Scenarios

Accounts Receivable

• Calculate lost revenue from unpaid carrying charges by client, etc.
• Determine carrying costs by comparing days in collection by customer
• Extract variances in matching remittances to open receivables
• Generate invoice summaries by customer, invoice, amounts, etc.
• Identify credits taken beyond discount terms of payment days
• Identify duplicate invoices, credits, or receipts in any order
Data Mining Scenarios

Accounts Receivable

- Identify high-value credit notes, balances and invoices
- Report and age total receivables in any format
- Report gaps in the sequence of invoices generated
- Show variances between delivery documents and invoices
- Report on creditor balances aged in any format
- Identify accounts with oldest activity for sales follow-up
- Report gaps in a sequence of invoices generated
- Isolate average sales amounts by product, sales representative, region
- Summarize client days to payment by client, sales representative, region, etc.
Data Mining Scenarios

**Stock and Inventory Control**
- Identify obsolete inventory by sorted turnover analysis
- Isolate and analyze high-value transactions by value, group, etc.
- Report on products in order of profitability (low-high, high-low)
- Segregate variances from standard pricing in dollar amount order
- Statistically analyze usage and ordering to improve turnover
- Summarize and stratify turnover by stock item in any order
Data Mining Scenarios

Stock and Inventory Control

• Analyze the difference between standard costs and actual costs
• Check product reordering volume by item, warehouse and vendor
• Compare the value of physical counts to general ledger value
• Extract products with zero quantities or zero prices
• Identify items with yearly volume under on-hand quantity
Data Mining Scenarios

Stock and Inventory Control
• Identify starting and ending period balances by class or group
• Match stock receipts with vendor ledger and report variances
• Reconcile selected physical counts to computed amounts
• Report on stock and high-value balances using any selection criteria
• Select items from perpetual stock for test reconciliation
• Summarize products by group, location, type, etc.
Data Mining Scenarios

Salaries and Payroll

• Analyze productivity – hours of patient care, etc.
• Compare and summarize costs for special pay, overtime, premium, etc.
• Compare entries against authorization records
• Compare time-card rates and pay to payroll, indicate variances
• Compare wage and salary rates to approved ranges
• Extract all payroll checks where the gross amount exceeds set amount
Data Mining Scenarios

Salaries and Payroll
- Identify changes in exemptions, gross pay, hourly rates, salary amounts, etc.
- Identify duplicate or missing payroll checks by check, bank, etc.
- Identify false/invalid/duplicate Social Security numbers
- Identify overtime abuses
- Identify payments to terminated or non-existent employees
- Match payroll records to accounts payable records
Data Mining Scenarios

Salaries and Payroll

• Recap distributions to jobs, projects, etc.
• Report entries against authorization records for new or terminated employees
• Summarize and print payroll by selection criteria for audit review
• Summarize payroll distributions for reconciliation to general ledger
• Verify commission calculations
Data Mining Scenarios

Vendor Management
• Check product reordering volume by item, warehouse and vendor
• Compare speed and accuracy of delivery by product and vendor
• Match stock receipts with vendor ledger and report variances
• Audit receipt and ordering by selective reports by vendor, date, etc.
Data Mining Scenarios

Vendor Management
- Compare vendor performance by summarizing item delivery and quality
- Extract pricing and receipt quantity variations by vendor and purchase order
- Identify duplicate vendors on vendor master file
- Identify unused vendors
- Match vendor names/addresses/tax IDs to payroll records for employees
- Summarize large invoices without purchase orders, by amount, vendor, etc.
Data Mining Scenarios

Purchase Order Management
• Eliminate stale purchase orders by analyzing and reporting on partial receipts
• Identify duplicate purchase orders or receipts without purchase orders
• Isolate purchase order types (blanket, release, drop-ship) for audit tracking
• Reconcile receipts by comparing accrued payables to received items
• Report on best purchasing performance by various locations
Data Mining Scenarios

**Purchase Order Management**
- Identify charges posted outside of proper GL period
- Identify invalid or high-dollar charges
- Identify late charges by department, by month, etc.
- Identify lost charges; match supplies used to supplies billed
- Match procedure codes to appropriate billed charges
Data Mining Scenarios

Materiel Management and Inventory
• Divide inventory into classes and compare percent investment
• Identify duplicate supply items on inventory master
• Identify items with yearly volume under on-hand quantity
• Identify obsolete inventory by sorted turnover analysis
• Identify starting and ending period balances by class or group
• Isolate and analyze high-value transactions by value, group, etc.
Materiel Management and Inventory

- Physical inventory variance analysis
- Profile supply usage by month, by department, etc.
- Report on stock and high-value balances using any selection criteria
- Statistically analyze usage and ordering to improve turnover
- Summarize and stratify turnover by stock item in any order
Data Mining Scenarios
Healthcare-Specific

Patient Billing, Accounts Receivable and Managed Care

• Recalculate bad debt and contractor allowance reserves
• Age receivables on date-of-service vs. invoice date to recalculate cash flow
• Age service billing dates to assess timely collections and write-offs
• Aging analysis by varying dates and sorts
Data Mining Scenarios
Healthcare-Specific

Patient Billing, Accounts Receivable and Managed Care

- Analyze rejected payments by financial class, procedure code, cost center
  Calculate average days from discharge to bill, bill to payment, by payer or department
- Calculate days in accounts receivable
- Match patient refund payments to accounts receivable records
- Calculate reimbursement percentages by payer
Data Mining Scenarios
Healthcare-Specific

Patient Billing, Accounts Receivable and Managed Care
- Compare date-of-service to invoice date to identify opportunities to re-engineer charge processing
- Compare detailed aging of receivables to timing cycle required by Medicare to identify opportunities for improvements in charge processing
- Analyze volume/usage of all chargemaster codes
Data Mining Scenarios
Healthcare-Specific

Patient Billing, Accounts Receivable and Managed Care

- Determine appropriate level of doubtful accounts reserves
- Determine average billing amount by financial class or cost center
- Determine average number of days from invoice to payment by financial class
- Determine profit margin by DRG, physician, financial class, etc.
- Determine typical billing amount range (stratification) by financial class or cost center
Data Mining Scenarios
Healthcare-Specific

Patient Billing, Accounts Receivable and Managed Care

• Evaluate managed care payer performance
• Identify high-dollar accounts
• Identify missed stop-loss payments on managed care accounts
• Identify unbilled accounts
• Identify duplicate charges on patient bills
Data Mining Scenarios
Healthcare-Specific

Patient Billing, Accounts Receivable and Managed Care

- Identify underpaid managed care accounts
- Report on payer cash payment lag by region
- Report on prompt billing trends/internal performance
- Summarize and age receivables by financial class, cost center, procedure code
- Support managed care contract negotiations with payment history
Data Mining Scenarios
Healthcare-Specific

**Compliance**
- Identify single-day stays and subsequent re-admissions
- Identify inappropriate unbundling of lab tests
- Identify frequent use of high-risk organizations
- Match OIG-excluded providers list with vendor, employee master files
- Identify up-coding
- Analyze patient data access logs to identify possible HIPAA risks/violations
Data Mining Scenarios
Healthcare-Specific

Clinical Subsystems
• Compare patient visit data on lab/radiology systems to patient master
• Identify pricing discrepancies between master and subsystems
• Compare list of valid signed-up employees to list of people actually receiving health benefits from insurance company
• Match service dates to member eligibility dates
Data Mining Scenarios
Healthcare-Specific

Medical Claims
• Analyze timeliness of claims payments by comparing claim date, date claim received and date claim paid

Physicians
• Determine physician contract compliance
• Evaluate physician practice history by patient type, payer, etc.
• Report on incomplete physician profiling information
Data Mining Scenarios
Healthcare-Specific

Marketing
• Compare patient ordering histories to patient demographic information for accuracy
• Develop patient statistics by Zip codes or other demographic data
• Identify profitable segments of patient population for advertising focus
• Report on incomplete or miscoded patient demographic information
QUESTIONS?