CHAPTER ONE

1 Introduction to Fraud Data Analytics

The world's best auditor using the world's best audit program cannot detect fraud unless their sample includes a fraudulent transaction. This is why fraud data analytics (FDA) is so critical to the auditing profession.

How we use fraud data analytics largely depends on the purpose of the audit project. If the fraud data analytics is used in a whistle blower allegation, then the fraud data analytics plan is designed to refute or corroborate the allegation. If the fraud data analytics plan is used in a control audit, then the fraud data analytics would search for internal control compliance or internal control avoidance. If the fraud data analytics is used for fraud testing, then the fraud data analytics is used to search for a specific fraud scenario that is hidden in your database. This book is written for fraud auditors who want to integrate fraud testing into their audit program. The concepts are the same for fraud investigation and internal control avoidance—what changes is the scope and context of the audit project.

Interestingly, two of the most common questions heard in the profession are, "Which fraud data analytic routines should I use in my audit?" and, "What are the three fraud data analytics tests I should use in payroll or disbursements?" In one sense, there really is no way to answer these questions

because they assume the fraud auditor knows what fraud scenario someone might be committing. In reality, we search for patterns commonly associated with a fraud scenario or we search for all the logical fraud scenario permutations associated with the applicable business system. In truth, real fraud data analytics is exhausting work.

I have always referred to fraud data analytics as code breaking. It is the auditor's job to search the database using a comprehensive approach consistent with the audit scope. So, the common question of which fraud data analytics routines should I use can only be answered when you have defined your audit objective and audit scope. A key element of the book is the concept that while the fraud auditor might not know what fraud scenario a perpetrator is committing, the fraud auditor can identify and search for all the fraud scenario permutations. Therefore, the perpetrator will not escape the long arm of the fraud data analytics plan.

Once again, the question arises as to which fraud data analytic routines I should use in my next audit. Using the fraud risk assessment approach, the fraud data analytics plan could focus on those fraud risks with a high residual rating. The auditor could select those fraud risks that are often associated with the particular industry or with fraud scenarios previously uncovered within the organization—or the auditor might simply limit the scope to three fraud scenarios. Within this text, we plan to explain the methodology for building your fraud data analytics plan; readers will need to determine how comprehensive to make their plan.

WHAT IS FRAUD DATA ANALYTICS?

Fraud data analytics is the process of using data mining to analyze data for red flags that correlate to a specific fraud scenario. The process starts with a fraud data analytics plan and concludes with the audit examination of documents, internal controls, and interviews to determine if the transaction has red flags of a specific fraud scenario or if the transaction simply contains data errors.

Fraud data analytics is not about identifying fraud but rather, identifying red flags in transactions that require an auditor to examine and formulate a decision. The distinction between identifying transactions and examining the transaction is important to understand. Fraud data analytics is about creating a sample; the audit program is about gathering evidence to support a conclusion regarding the transaction. The final questions in the fraud audit process:

Is there credible evidence that a fraud scenario is occurring? Should we perform an investigation?

It is critical to understand that fraud data analytics is driven by the fraud scenario versus the mining of data errors. Based on the scenario, it might be one red flag or a combination of red flags. Yes, some red flags are so overpowering that the likelihood of fraud is higher. Yes, some red flags simply correlate to errors. The process still needs the auditor to examine the documents and formulate a conclusion regarding the need for a fraud investigation. It is important to understand the end product of data analytics is a sample of transactions that have a higher probability of containing one fraudulent transaction versus a random sample of transactions used to test control effectiveness. One could argue that fraud data analytics has an element of Las Vegas. Gamblers try to improve their odds of winning. Auditors try to improve their odds of detecting fraud. Figure 1.1 illustrates the concept of improving your odds by reducing the size of the population for sample selection.

Within most literature, a vendor with no street address is a red flag fraud. But a red flag of what? Is a blank street address field indicative of a shell company? How many vendors have no address in the accounts payable file because all payments are EFT? If a vendor receives payment through the EFT process, then is the absence of a street address in your database a red flag? Should a street address be considered a red flag of a shell company? Is the street address linked to a mailbox service company? What are the indicators of a mailbox service company? Do real companies use mailbox service companies? Fraud examiners understand that locating and identifying fraudulent transactions is a matter of sorting out all these questions. A properly developed fraud data-mining plan is the tool for sorting out the locating question.

To start your journey of building your fraud data analytics plan, we will need to explain a few concepts that will be used through the book.

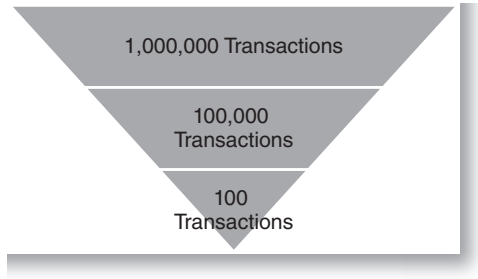

FIGURE 1.1 Improving Your Odds of Selecting One Fraudulent Transaction

What Is Fraud Auditing?

Fraud auditing is a methodology to respond to the risk of fraud in core business systems. It is a combination of risk assessment, data mining, and audit procedures designed to locate and identify fraud scenarios. It is based on the theory of fraud that recognizes that fraud is committed with intent to conceal the truth. It incorporates into the audit process the concept of red flags linked to the fraud scenario concealment strategy associated with data, documents, internal controls, and behavior.

It may be integrated into audit of internal controls or the entire audit may focus on detecting fraud. It may also be performed because of an allegation or the desire to detect fraudulent activity in core business systems. For our discussion purposes, this book will focus on the detection of fraud when there is no specific allegation of fraud.

Fraud auditing is the application of audit procedures designed to increase the chances of detecting fraud in core business systems. The four steps of the fraud audit process are:

1. *Fraud risk identification*. The process starts with identifying the inherent fraud schemes and customizing the inherent fraud scheme into a fraud scenario. Fraud scenarios in this context will be discussed in Chapter 2.
2. *Fraud risk assessment*. In the traditional audit methodology the fraud risk assessment is the process of linking of internal controls to the fraud scenario to determine the extent of residual risk. In this book, fraud data analytics is used as an assessment tool through the use of data-mining search routines to determine if transactions exist that are consistent with the fraud scenario data profile.
3. *Fraud audit procedure*. The audit procedure focuses on gathering audit evidence that is outside the point of the fraud opportunity (person committing the fraud scenario). The general standard is to gather evidence that is externally created and externally stored from the fraud opportunity point.
4. *Fraud conclusion*. The conclusion is an either/or outcome, either requiring the transaction to be referred to investigation or leading to the determination that no relevant red flags exist. Chapters 6 through 15 contain relevant discussion of fraud data analytics in the core business systems.

What Is a Fraud Scenario?

A fraud scenario is a statement as to how an inherent scheme will occur in a business system. The concept of an inherent fraud scheme and the fraud risk

structure is discussed in Chapter 2. A properly written fraud scenario becomes the basis for developing the fraud data analytics plan for each fraud scenario within the audit scope. Each fraud scenario needs to identify the person committing the scenario, type of entity, and the fraudulent action to develop a fraud data analytics plan. The auditing standards also suggest identifying the impact the fraud scenario has on the company.

While all fraud scenarios have the same components, we can group the fraud scenarios into five categories. The groupings are important to help develop our audit scope. The groupings also create context for the fraud scenario. Is the fraud scenario common to all businesses or is the fraud scenario unique to our industry or our company? There are five categories of fraud scenarios:

1. *The common fraud scenario*. Every business system has the same listing of common fraud scenarios. I do not need to understand your business process, conduct interviews of management, or prepare a flow chart to identify the common fraud scenarios.
2. *The company-specific fraud scenario*. The company-specific fraud scenario in a business cycle because of business practices, design of a business system, and control environment issues. I do need to understand your business process, conduct interviews of management, or prepare a flow chart to identify the common fraud scenarios.
3. *The industry-specific fraud scenario*. The industry specific fraud scenarios are similar to the common fraud scenario, except the fraud scenario only relates to an industry. To illustrate the concept, mortgage fraud is an issue for the banking industry. This category of fraud scenarios requires the fraud auditor to be knowledgeable regarding their industry. However, using the methodology in Chapter 2, a nonindustry person could create a credible list of fraud scenarios.
4. *The unauthorized fraud scenario*. The unauthorized fraud scenario occurs when an individual, either internal or external to the company, commits an act by overriding company access procedures.
5. *The internal control inhibitor fraud scenario*. The concept of internal control inhibitor is to identify those acts or practices that inhibit the internal control procedures from operating as designed by management. The common internal control inhibitors are collusion and management override.

Chapter 2 will explain the concept of the fraud risk structure and how to write a fraud scenario that drives the entire fraud audit program. Chapter 2

will also cover the concept of fraud nomenclature. In the professional literature, we use various fraud words interchangeably, which I believe creates confusion within the profession. Words like *fraud risk statement*, *fraud risk*, and *inherent fraud schemes*, *fraud scenario*, *fraud schemes*, and *inherent fraud risk* are used to describe how fraud occurs for the purpose of building a fraud risk assessment or fraud audit program. Within this book, I will use the phrase *fraud scenario* as the words that drive our fraud data analytic plan.

What Is Fraud Concealment?

Fraud concealment is the general or specific conditions that hide the true nature of a fraudulent transaction. A general condition is the sheer size of database, whereas a specific condition is something that the perpetrator does knowingly or unknowingly to cause the business transaction to be processed in the business system and hide the true nature of the business transaction.

To illustrate the concept, all vendors need an address or a bank account to receive payment. On a simple basis, the perpetrator uses his or her home address in the master file. On a more sophisticated level, the perpetrator uses an address for which the linkage to the perpetrator is not visible within the data—for example, a post office box in a city, state, or country that is different from where the perpetrator resides. The fraud data analytics plan must be calibrated to the level of fraud sophistication that correlates to the specific condition of the person committing the fraud scenario. In Chapter 3, the sophistication model will describe the concepts of low, medium, and high fraud concealment strategies. The calibration concept of low, medium, and high defines whether the fraud scenario can be detected through the master file or the transaction file. It also is a key concept of defining the audit scope.

It is important to distinguish between a fraud scenario and the associated concealment strategies. Simply stated, the fraud scenario is the fraudulent act and concealment is how the fraudulent act is hidden. From an investigation process, concealment is referred to as the intent factor. From a fraud audit process, the concealment is referred to as the fraud concealment sophistication factor.

What Is a Red Flag?

A red flag is an observable condition within the audit process that links to the concealment strategy that is associated with a specific fraud scenario. A red flag exists in data, documents, internal controls, behavior, and public records.

Fraud data analytics is the search for red flags that exist in data that links to documents, public records, persons, and eventually to a fraud scenario.

The red flag is the inverse of the concealment strategy. The concealment strategy is associated with the person committing the fraud scenario and the red flag is how the fraud auditor observes the fraud scenario.

The red flag theory becomes the basis of developing the fraud data profile, which is the starting point of developing the fraud data analytics plan. The red flags directly link to the fraud concealment strategy. The guidelines for using the red flag theory are discussed in Chapter 3.

What Is a False Positive?

A false positive is a transaction that matches the red flags identified in the fraud data profile but the transaction is not a fraudulent transaction. It is neither bad nor good. It simply is what it is. What is important is that the fraud data analytics plan has identified a strategy for addressing false positives. Fundamentally, the plan has two strategies: Attempt to reduce the number of false positives through the fraud data analytics plan or allow the fraud auditor to resolve the false positive through audit procedure. There may be no correct answer to the question; however, ignoring the question is a major mistake in building your plan.

What Is a False Negative?

A false negative is a transaction that does not match the red flags in the fraud data profile but the transaction is a fraudulent transaction. From a fraud data analytics perspective, false negatives occur due to not understanding the sophistication of concealment as it related to building your fraud data analytics plan. Other common reasons for a false negative are: data integrity issues, poorly designed data interrogation procedures, the lack of data, and the list goes on.

While false positives create unnecessary audit work for the fraud auditor, false negatives are the real critical issue facing the audit profession because the fraud scenario was not detected.

The false positive conundrum: Refine the fraud data analytics or resolve the false positive through audit work.

There is no real correct answer to the question. The fraud data analytics should attempt to provide the fraud auditor with transactions that have a higher probability of a person committing a fraud scenario. The fraud data interrogation routines should be designed to find a specific fraud scenario. That is the purpose of fraud data analytics. However, by the nature of data and fraud, false positives will occur. Deal with it. The real question is how to minimize the number of false positives consistent with the fraud data analytics strategy selected for the fraud audit.

Remember, fraud data analytics is designed to identify transactions that are consistent with a fraud data profile that links to a specific fraud scenario. There needs to be a methodology in designing the data interrogation routines. The methodology needs to be based on a set of rules and an understanding of the impact the strategy will have on the number of false positives and the success of fraud scenario identification.

The reality of fraud data analytics is the process will have false positives; said another way, there are transactions that will have all the attributes of a fraud scenario, but turn out to be valid business transactions. That is the reality of the red flag theory. Unfortunately, the reality of fraud data analytics is that there will also be false negatives based on the strategy selected. This is why before the data interrogation process starts, there must be a defined plan that documents the auditor judgment. Senior audit management must understand what the plan is designed to accomplish and why the plan is designed to fail. Yes, based on the correlation of audit strategy and sophistication of fraud concealment, you can design a plan to fail to detect a fraud scenario. At this point in the book, do not read this as a bad or good; Chapter 3 will explain how to calibrate your data interrogation routines consistent with the sophistication of concealment.

To provide a real-life example, in one project involving a large vendor database, our fraud data analytics identified 200 vendors meeting the profile of a shell company. At the conclusion, we referred five vendors for fraud investigation. In one sense, the project was a success; in another sense, we had 195 false positives.

If I could provide one suggestion based on my personal experience, the person using the software and the fraud auditor need to be in the same room at the same time. As reports are created, someone needs to look at the report and refine the report based on the reality of the data in your database. Fraud data analytics is a defined process and with a set of rules. However, the process is not like the equation $1 + 1 = 2$. It is an evolving process of inclusion and exclusion based on a methodology and fraud audit experience. So, do not worry about the

false positive, which simply creates unnecessary audit work. Worry about the false negative.

FRAUD DATA ANALYTICS METHODOLOGY

I commonly hear auditors talk about the need to play with the data. This is one approach to fraud detection. The problem with the approach is that it relies on the experience of the auditor rather than on a defined methodology. I am not discounting audit experience, I would suggest that auditor experience is enhanced with a methodology designed to search for fraud scenarios. In fact, the data interpretation strategy explained in Chapter 3 is a combination of professional experience and methodology.

The fraud data analytics methodology is a circular approach to analyzing data to select transactions for audit examination (Figure 1.2).

- *Fraud scenario*. The starting point for building a fraud data analytics plan is to understand how the fraud risk structure links to the audit scope. The process of identifying the fraud scenarios within the fraud risk structure and how to write the fraud scenario is discussed in Chapter 2.
- *Strategy*. The strategy used to write data interrogation routines needs to be linked to the level of sophistication of concealment. For purposes of this book there are four general strategies, which are explained in Chapter 3.

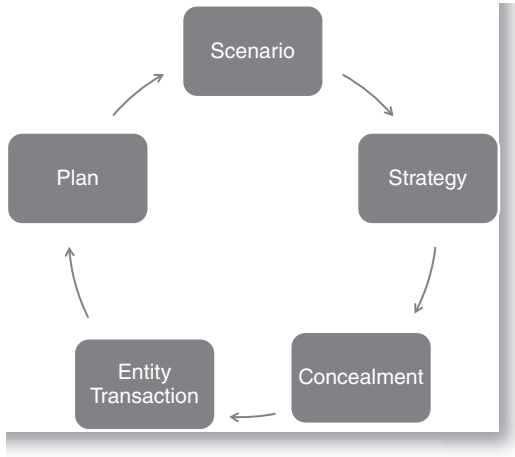

FIGURE 1.2 Circular View of Data Profile

- *Sophistication of concealment impacts the success of locating fraudulent transactions*. A common data interrogation strategy for searching for shell companies is to match the addresses of employees to the address of vendors. While a great data analytics step, the procedure is not effective when the perpetrator is smart enough to use an address other than a home address. So, at this level of concealment, we need to change our strategy. A complete discussion of fraud concealment impact on fraud data analytics is in Chapter 3.
- *Building the fraud data profile is the process of identifying the red flags that correlates to entity and transaction*. All fraud scenarios have a data profile that links to the entity structure (i.e., name, address, etc.) and the transaction file (i.e., vendor invoice). The specific red flags will be discussed in Chapters 6 through 15.
- *The plan starts with linking the fraud scenario to the fraud data profile*. Then it uses the software to build the data interrogation routines to identify the red flags and overcome the concealment strategies.
- *In reality, the search process is seldom one-dimensional*. It is a circular process of analyzing data and continually refining the search process as we learn more about the data and the existence of a fraud scenario in the core business system.

Assumptions in Fraud Data Analytics

1. *The certainty principle*. The degree of certainty concerning the finding of fraud will depend on the level of concealment sophistication and the on/off access to books and records. When the fraud is an on-the-book scheme and has a low level of sophistication, the auditor will be able to obtain a high degree of certainty that a fraud scenario has occurred. Consequently, with an off-the-book fraud scenario and high level of sophistication, the auditor will not achieve the same degree of certainty that a fraud scenario has occurred. Therefore, the auditor must recognize the degree of certainty differences when developing the fraud audit program.

   The difficulty in ascertaining the degree of certainty directly influences the quality and quantity of evidence needed. If an auditor assumes a low level of certainty with regard to a fraud scenario occurring, then the auditor may not incorporate the gathering of credible evidence at all. However, if an auditor is well versed in fraud scenario theory and, therefore, establishes some degree of certainty that a scenario has occurred, the audit plan needs to incorporate the obtaining of the appropriate amount and quality of evidence to justify that degree of certainty.

Specifically, as part of the fraud audit plan, it should first be determined what elements of proof will be necessary to recommend an investigation. Then a decision is needed to determine if the chosen elements are attainable in the context of a fraud audit based on the specific scenario, concealment sophistication, and access to books and records.

2. *The linkage factor*. The term *link* is used extensively throughout the entire book as it aptly highlights the relationship between the various fraud audit program components and objectives. For example, the fraud audit program is built by linking the data mining, audit testing procedures, and audit evidence considerations to a given fraud scenario found in the risk assessment. At its core, the concept of linkage is a simple one; however, with the traditional audit program as a frame of reference, many auditors have difficulty grasping the idea that fraud audit procedures should be designed, and therefore, linked to a specific fraud scenario. The entire book is based on the linkage factor. All fraud data analytic routines must be linked to a fraud scenario or all fraud scenarios must be linked to a fraud data analytics routine.
3. *Cumulative principle*. Seldom is one red flag sufficient to identify a fraud scenario within a database. It is the totality of the red flags that are indicative of a fraud scenario. The process should incorporate a summary report of the tests to score each entity or transaction. When we search for fictitious employee, commonly referred to as a ghost employee, a duplicate bank test will identify false positives because two or more employees are family members. However, when one of the employees is a budget owner and the second employee has a different last name, address, no voluntary deductions, postal box address, and no contact telephone number, it is the totality of the red flags versus anyone red flag. This is an important concept to incorporate into the fraud data analytics plan.
4. *Basis for selection for testing*. Fraud data analytics is all about selecting transactions for fraud audit testing. The basis for selection must be defined and understood by the entire team.

THE FRAUD SCENARIO APPROACH

The approach is simple. In essence, you develop an audit program for each fraud scenario. The starting point is to identify all the fraud scenarios within your audit scope. Within the audit project this is the process of developing your fraud risk assessment. The final step in the fraud risk assessment is the concept of

residual risk. The dilemma facing the profession is how the concept of residual risk should impact the decision of when to search for fraud in core business systems. The question cannot be ignored, but there is no perfect answer to the question. It is what I call the likelihood conundrum.

The Likelihood Conundrum: Internal Control Assessment or Fraud Data Analytics

Does the auditor rely on internal controls or does the auditor perform fraud data analytics? There is no simple answer to the question; I suspect one answer could be derived from the professional standards that the auditor follows in the conduct of an audit. In my years of teaching audit professionals the concept of fraud auditing, I have seen the struggle on the auditors' faces. The reason for the struggle is that we have been told that a proper set of internal controls should provide reasonable assurance in preventing fraud scenarios from occurring. There are many reasons why an internal control will fail to prevent a fraud scenario from occurring. The easiest fraud concept to understand why internal controls fail to prevent fraud is the concept of internal control inhibitors. We cannot ignore collusion and management override in regard to fraud.

We need to understand that fraud can occur and comply with our internal controls. I suspect this is an area of great disagreement in the profession between the internal control auditors and the fraud auditors. Even if you believe that internal controls and separation of duties will prevent fraud, what is the harm in looking for fraud? So, we give management a confirmation that fraud scenarios are not occurring in the business system. We do the same confirmation with internal controls: Because we see the evidence of an internal control we assume that the control is working. If the auditor is serious about finding fraud in an audit, then the auditor must start looking for fraud. For me, the likelihood conundrum is much ado about nothing. Management, stockholders, and boards of directors all think we are performing tests to uncover fraud.

How the Fraud Scenario Links to the Fraud Data Analytics Plan

With each scenario, the auditor will need to determine which scenarios are applicable to fraud data analytics and which fraud scenarios are not applicable to fraud data analytics. For example: A product substitution scheme can occur when the receiver accepts an inferior product but indicates the product conforms to the product requirements. This fraud scenario does not lend itself

to fraud data analytics because the clue is not in the data. However, a vendor that consistently submits invoices exceeding the purchase order within the payment tolerances can be identified. Once the list of scenarios relevant to the plan are identified the next step is to understand how the three critical elements of the scenario impact the plan.

The elements of scenarios that are relevant to creating an effective fraud data analytics plan are: the person who commits the scenario, the type of entity, and the type of action we are looking for.

To illustrate the concept, as a starting point we will consider the "who" as either the budget owner, accounts payable function, or a senior manager. A common test is to search for vendors created in the master file at off-periods. If the scenario is focusing solely on the budget owner, is the off-period test relevant to the scope of the project? Now let's change the person committing the scenario to someone in the accounts payable function. Now the off-period test is relevant to the audit scope.

The second aspect of a scenario is the type of entity. Are we searching for a false vendor or a real vendor? If the vendor is real, then searching for vendors with P.O. boxes is not relevant because real vendors tend to use P.O. boxes, whereas if we are searching for real vendors operating under multiple names, then a duplicate test on the address field is relevant.

The third aspect of a scenario is the fraudulent action. If the vendor is real and the fraud scenario is overbilling based on unit price inflation, then searching for a sequential pattern of invoices is not relevant. The test should focus on changes in unit price or comparisons of unit prices for similar items among common vendors.

The fourth element of a fraud scenario is the impact statement. While critical to the fraud scenario statement, the impact statement is not typically associated with the data analytics plan but is critical to the investigation process. The following two scenarios illustrate the concept:

1. Senior manager acting alone or in collusion with a direct report/causes a shell company to be set up on the vendor master file/causes the issuance of a purchase order and approves a false invoice for services not received/**causing the diversion of company funds.**
2. Senior manager acting alone or in collusion with a direct report/causes a shell company to be set up on the vendor master file/causes the issuance of a purchase order and approves a false invoice for services not received/**depositing the funds in an off-the-book bank account for the purpose of paying bribes.**

A close examination of the two fraud scenarios reveals that the fraud data analytics plan is exactly the same for both scenarios. In both scenarios, the fraud data analytics is searching for a shell company and a pattern of false invoices.

From a fraud investigation plan, the first scenario is an asset misappropriation scenario while the second scenario is associated with a corruption scheme mostly connected to an FCPA violation.

SKILLS NECESSARY FOR FRAUD DATA ANALYTICS

Building a fraud data analytics plan requires a defined skill set. The absence of one skill set will diminish the effectiveness of the plan. The audit team needs to ensure all the right skills are contained within the team:

- *Knowledge of fraud*. Since fraud data analytics is the process of searching for fraudulent transactions, the auditor must have a full understanding of the fraud concepts.
- *Fraud scenarios*. This skill relates to how to write a fraud statement that correlates to developing a fraud data analytics statement. For an analogy, the scenario approach should be considered the system design aspect of the project and creating the routines is the program aspect of the project, or the scenario creates the questions and the fraud data analytical plan creates the answers.
- *Information technology knowledge*. Data reside in large, complex database systems. The ability to communicate with the IT function to locate and extract the data is the starting point of the data interrogation phase of the plan.
- *Audit software knowledge*. Coding software, whether writing scripts or using software functions, is necessary to write the data interrogation routines. The ability of the auditor to clean data, reformat data, combine data, and create reports is an absolutely necessary skill.
- *Audit knowledge*. Fraud data analytics is just one aspect of conducting an audit. Understanding fraud risk assessment, building audit scopes, designing audit steps, and formulating conclusions based on audit evidence rules is what fraud data analytics is all about. Second, designing fraud test procedures for the selected items is just as important as the fraud data analytics.
- *Understand data from a real-world perspective*. In each data column there is information. We need to understand how to use that information.

To illustrate the concept, using something as easy as an address field in a vendor database, the information in the field may correlate to a payment address, a physical address, a public mailbox service address, a nonpublic mailbox service address, mail forwarding services, or a bookkeeping service company. Yes, you must understand the data in a data field from a business perspective to develop a data interrogation routine. A vendor invoice number may have several patterns, depending on the industry and size of the business. The patterns are: no invoice number, date format, sequential ascending project number with a progress billing number, numeric or alpha format, and a sequential number linked to a customer number. So, how does the pattern link to the fraud scenario or the fraud concealment?

SUMMARY

As a conclusion to Chapter 1 and throughout the remainder of the book, I would like to offer some of the lessons learned throughout my fraud audit career. First, note the important points to understand about fraud data analytics, and then note some of the common mistakes one can make in fraud data analytics. I hope you find the points useful as you conduct your next fraud data analytics project.

Axioms of Fraud Data Analytics

- The world's best audit program and the world's best auditor cannot detect fraud unless their sample includes a fraudulent transaction.
- I do not know what a perpetrator will do, but I do know everything the perpetrator can do.
- While we do not know how a perpetrator will commit a fraud or how he will conceal the fraud, we can determine the logical permutations.
- The better you can describe the fraud scenario, the more likely you will be able to find it.
- False positives will occur. You try to resolve false positives either through your fraud data analytics or through an auditor performing audit procedures.
- In fraud data analytics, fraud likelihood is based on data versus the effectiveness of internal controls.
- We search for transactions that mirror the red flag theory of the fraud scenario.

- The better we understand data, the better we can use data to search for a fraudulent transaction.
- Errors and fraud have a lot in common.
- Red flags correlate to both errors and fraud.
- Data are not perfect.
- Databases contain data errors, caused either by mistake or with intent.
- We can only search data when the data reside in our databases.
- Fraud data analytics is both a science and an art.

Common Mistakes in Fraud Data Analytics

- No plan. Please do not jump in without a plan.
- Starting the fraud data analytics process without a clearly defined fraud scope.
- Creating reports that do not link to a specific fraud scenario.
- Searching for data exceptions versus the red flags of a fraud scenario.
- Assuming that a data integrity issue is an indicator of fraud.
- Failure to understand the integrity of the data being examined.
- Failure to understand the type of data that reside in a data field.
- No effective plan for false positives.
- Not worrying about false negatives.
- The fraud data analytics strategy is not calibrated for the level of fraud concealment sophistication.
- No planned audit procedure for the fraud data analytics report.

Chapters 2 to 5 are intended to provide a methodology for building your fraud data analytics plan. The remaining chapters are intended to describe the common fraud scenarios in a core business system and how to build your fraud data analytics plan to locate the fraud scenario in core business systems.