

Multidimensional Audit Data Selection (MADS)

Dave Dauksas (PricewaterhouseCoopers LLP)

Won No (Rutgers Business School)



Rutgers and AICPA Data Analytics Research Initiative

- ❖ AICPA, CPA Canada, eight big accounting firms, and Rutgers Business School
- ❖ Focus on the integration of analytics into the audit process at a foundational level, in an effort to enhance audit quality.
- ❖ Perform the testing of theory and methodology to inform the development of professional guidance on the application of audit data analytics, which will benefit the CPA profession as a whole.
- ❖ Objectives
 - Help an auditing profession to improve audit effectiveness through the integration of data analytics and related technologies into everyday practice.
 - Accelerate active engagement by firms and universities in fundamental applied research for continuous improvement of the auditing profession.
 - Encourage collaboration in other scholarly, practical research into audit data analytics.
 - Seek to spread learning in the field via publications, colloquia, conferences, and networking opportunities.
 - Facilitate a process for open dialogue with U.S. standards setters, regulators and other interested parties.
- ❖ <http://raw.rutgers.edu/radar.html>

Rutgers and AICPA Data Analytics Research Initiative

❖ Current projects



■ Sandbox

- Seek automation in the evaluation of internal control effectiveness with audit analytic methods.
- Apply process mining (a technique that allows for the analysis of business processes based on event logs) to the log data from the ERP system to evaluate the effectiveness of internal control.

■ Visualization

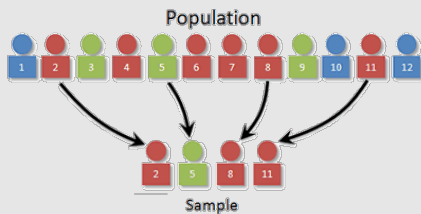
- Demonstrate/Illustrate that visualization can be used as audit evidence.
- Provide various types of visualizations generated from exploratory and confirmatory data analysis of a dataset that can be used in external audit.

■ Multidimensional Audit Data Selection (MADS)

- Provide a systematic approach for using data analytics in the audit sampling process.



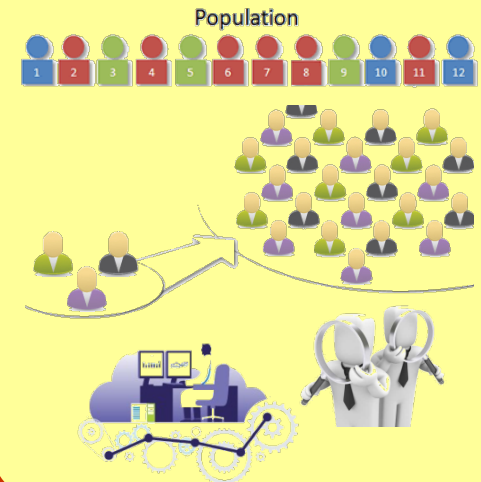
Traditional sampling approach



Advance in data
processing ability &
data analytics
techniques and tools

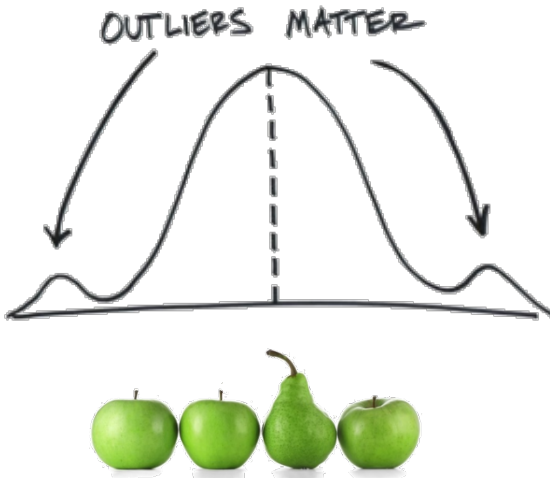


New approach



- Existing literature has introduced various statistical and machine learning techniques which process the entire population and identify unusual transactions/data which are more likely to be subject to misstatement or deviation.
- **Increases the effectiveness of the audit by decreasing the risk of erroneously concluding that the recorded account balance is not materially misstated when it is materially misstated.**
- Increase the efficiency by decreasing the risk of erroneously concluding that the recorded account balance is materially misstated when it is not materially misstated.

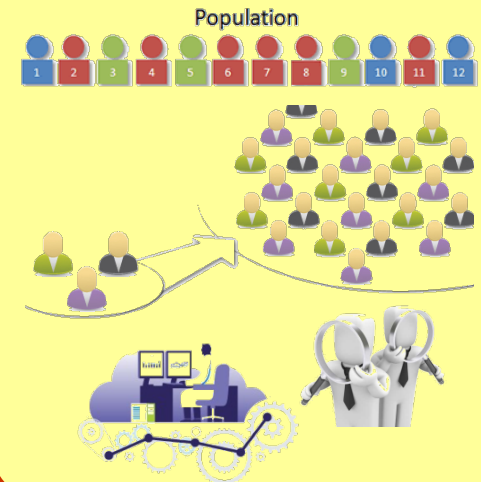
BACKGROUND



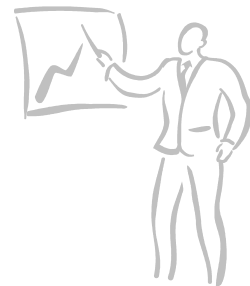
- Often generate large numbers of outliers.
- Impractical for auditors to investigate entire outliers

- Most prior studies do not address the problem of post-detection processing of outliers.
- Crucial to develop a framework that can help auditors effectively deal with large amounts of data, but also assist them to efficiently handle a massive number of outliers.

New approach



- ❖ Propose a multidimensional data prioritization methodology.
 - Introduce a Multidimensional Audit Data Selection (MADS) analytic framework to guide auditors in the identification of outliers (i.e., suspicious transactions/data that are more likely to be problematic) for a substantive test of details.
 - Empirically test whether the MADS process leads to more effective audits compared to the current audit sampling approach.





MADS ANALYTIC FRAMEWORK

Multidimensional Audit Data Selection (MADS) Analytic Framework

Overall Audit Objectives (Risk & Materiality)

- ❖ To assist auditors identifying questionable transactions/data in performing substantive test of details
 - Developed based on prior literature and professional guidelines.
 - Modified based on comments from several panel discussions of scholars and auditing professionals.
 - Consist of seven components.
- ❖ The practice of these seven components is guided by the overall objectives of audit, specifically audit risk and materiality.

Multidimensional Audit Data Selection (MADS) Analytic Framework

Objective and Criteria Identification

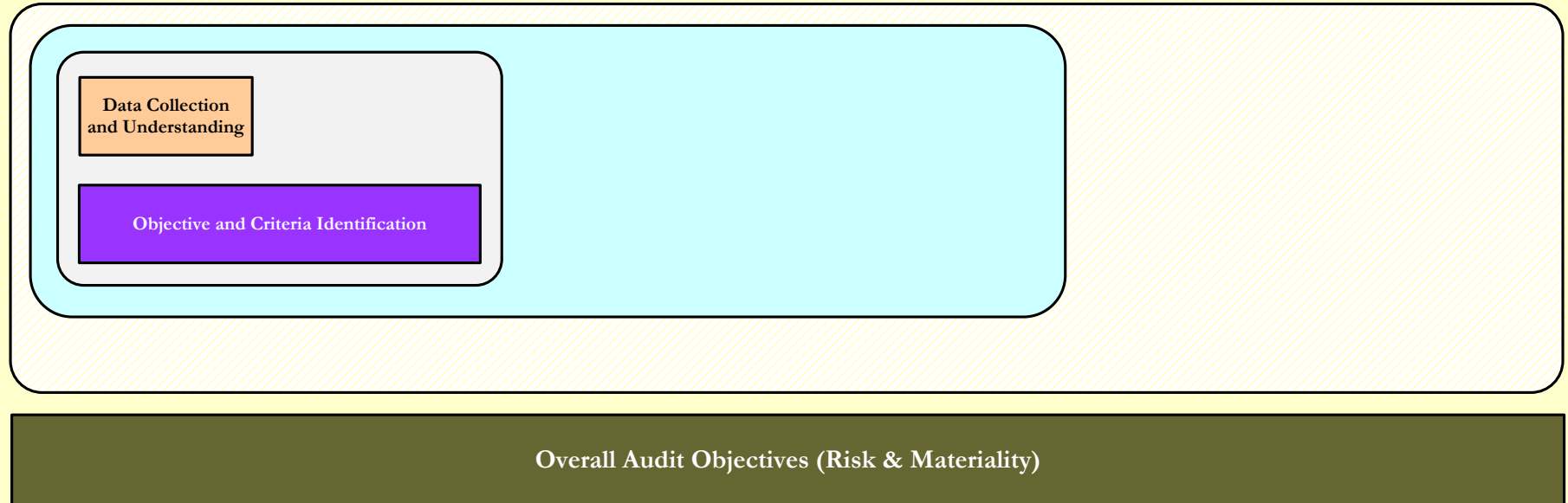
Overall Audit Objectives (Risk & Materiality)

- ❖ Identify the specific audit objective (e.g., fraud detection and identification of suspicious transactions/data identification) to be achieved by gaining as much insight as possible into the nature, extent, and timing of a substantive test of details.
- ❖ Specify success criteria (e.g., increase in the accuracy or timing of a substantive test of details) that provide a benchmark to assess whether all aspects of the process are satisfactory.



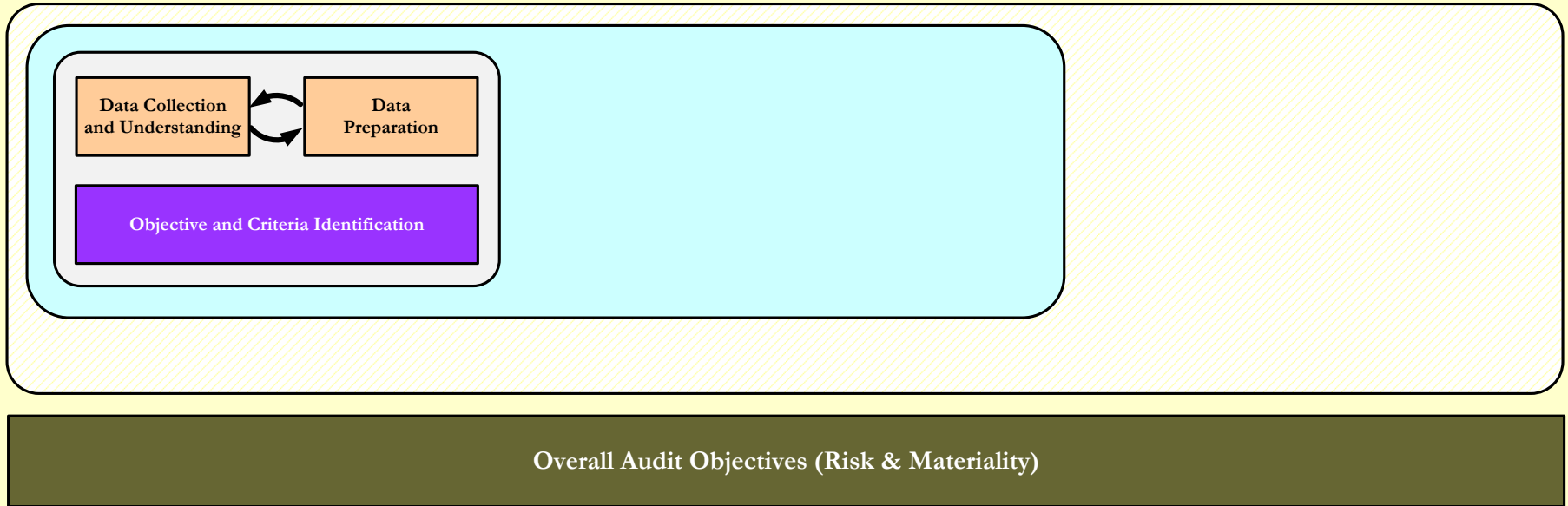
MADS ANALYTIC FRAMEWORK

Multidimensional Audit Data Selection (MADS) Analytic Framework



- ❖ Collect data and form an understanding of the data.
- ❖ Include initial data collection, data description, data exploration, and data quality verification.
- ❖ Determine whether the existing data is enough to meet their needs, various data sources that need to be merged, which attributes are relevant and irrelevant, how to handle missing data, etc.
- ❖ Often require to revise the objectives and success criteria, and thus need multiple iterations to identify practical objectives for the available data sources and criteria.

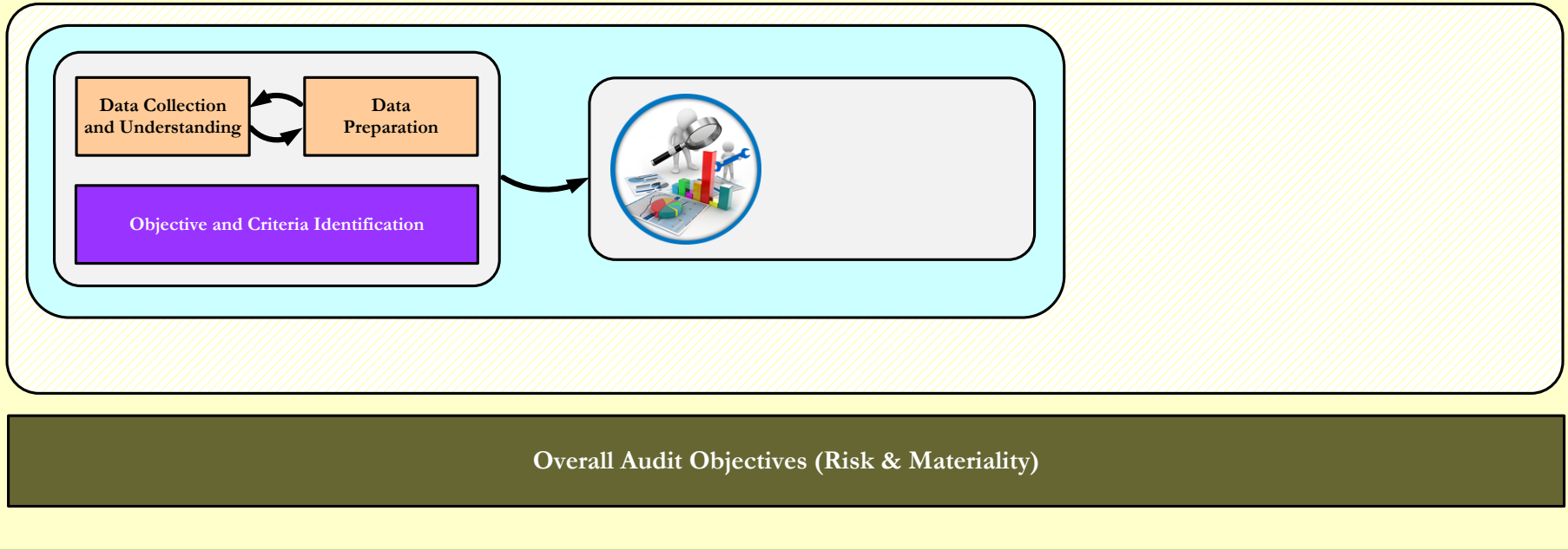
Multidimensional Audit Data Selection (MADS) Analytic Framework



- ❖ Time-consuming, but essential for any successful data analysis
- ❖ Convert original data into a format suitable for further processing and analysis.
- ❖ Include cleaning messy, inconsistent or unstandardized data, removing or inferring missing values, combining data from multiple sources, and checking the accuracy of the data.
- ❖ May require to revert to the prior stage to have a better understanding about currently available data as well as data needs, and thus need multiple iterations to prepare adequate data.

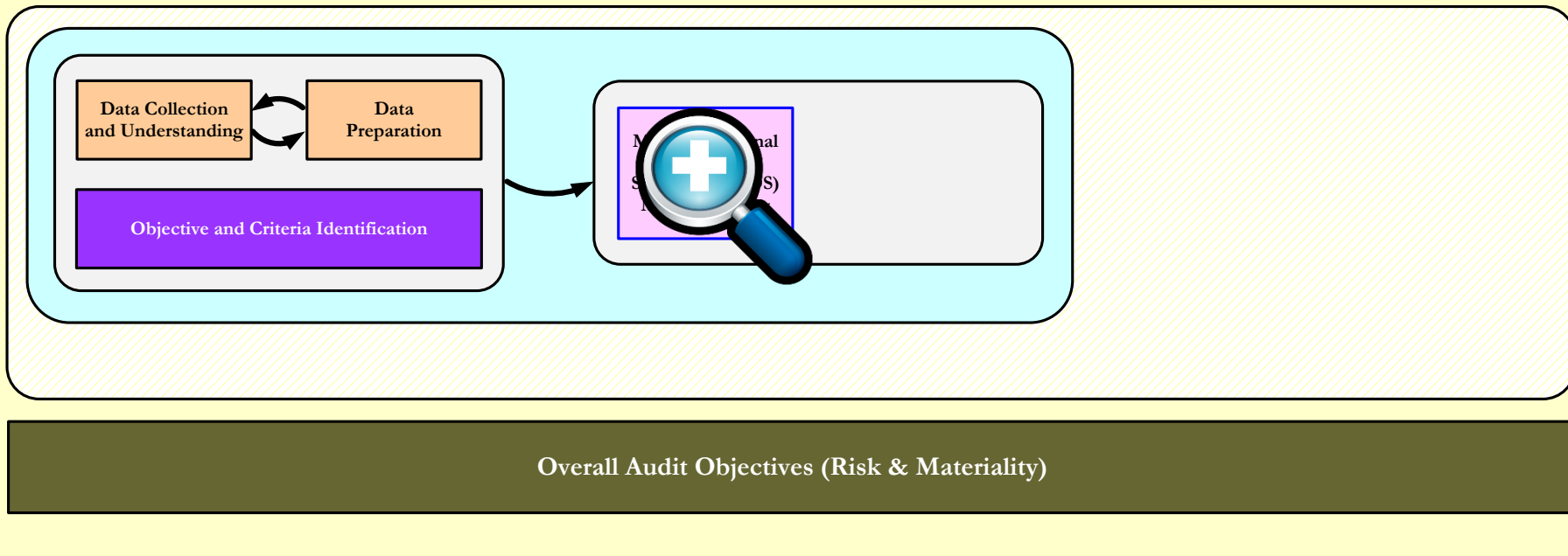
MADS ANALYTIC FRAMEWORK

Multidimensional Audit Data Selection (MADS) Analytic Framework



- ❖ Auditors tend to suffer from information overload when encountering large numbers of suspicious transactions/data (i.e., outliers) that require analysis, often including verification.
- ❖ Applying audit procedures to each of the outliers may lead to audit inefficiency.

Multidimensional Audit Data Selection (MADS) Analytic Framework



- ❖ Auditors tend to suffer from information overload when encountering large numbers of suspicious transactions/data (i.e., outliers) that require analysis, often including verification.
- ❖ Applying audit procedures to each of the outliers may lead to audit inefficiency.
- ❖ Aim to reduce the potential audit inefficiency by prioritizing the outlier candidates.
- ❖ Require to prepares several models and then modify them to obtain the optimal results.

MADS ANALYTIC FRAMEWORK

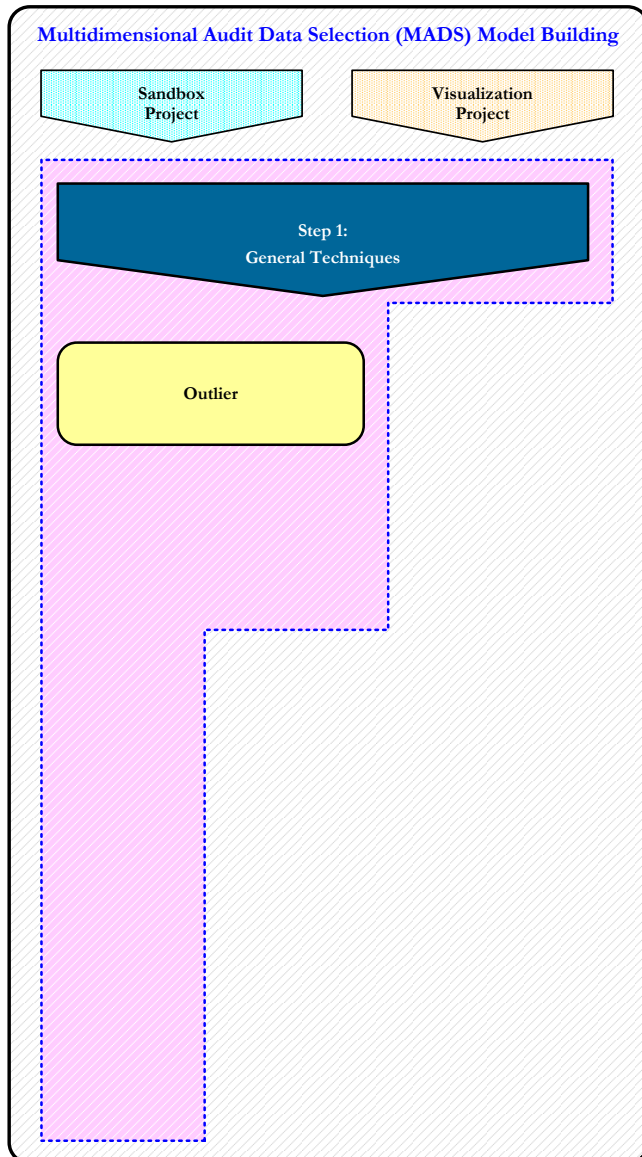
Multidimensional Audit Data Selection (MADS) Model Building

Step 1:
General Techniques

- ❖ Apply outlier detection techniques to evaluate the entire population instead of examining just a chosen sample, resulting in enhanced effectiveness of the audit.
- ❖ Select the most appropriate techniques based on the objective identified in the objective and criteria identification stage, available data, and the specific MADS model requirements (e.g., data size, data type, and model assumptions).
- ❖ Use a general guideline designed to assist auditors on how to choose appropriate techniques (or algorithms) to detect outliers given the different types of tasks and data on hand.

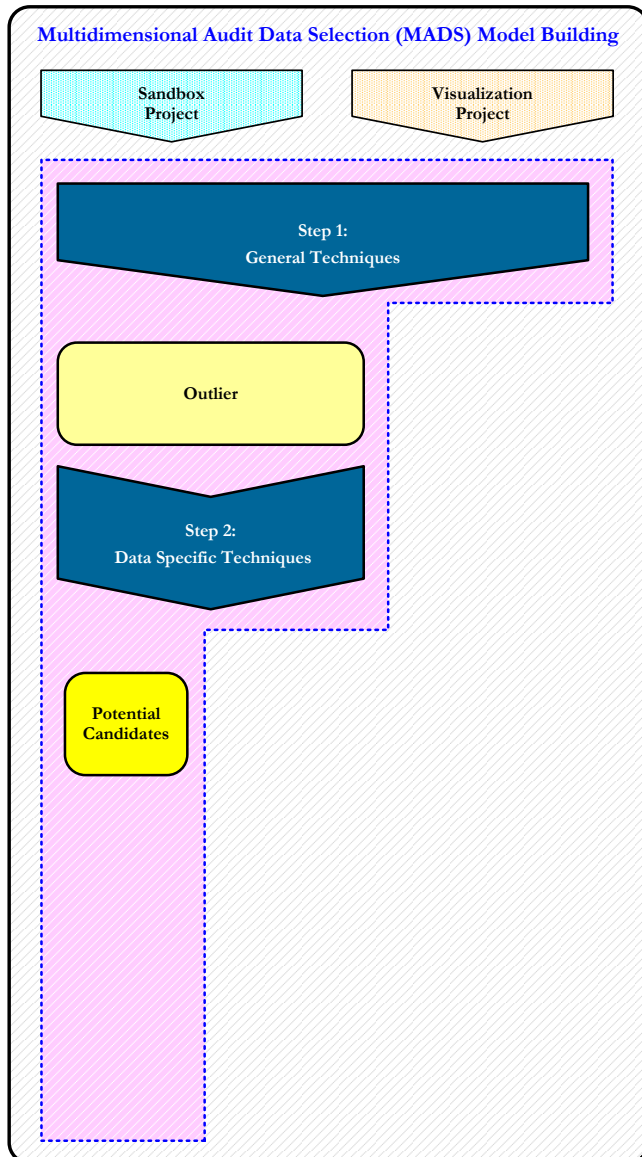
Task		Numerical data	Category data	Text data	Time series data
Supervised	Classification-based	Active learning; Novel class detection; Neural network; Bayesian network; SVM; Rule-based; K-NN	Bayesian network-based; Subspace methods	SVM	Hidden Markov Models; Rule-based;
	Statistical-based (Regression)	Linear regression;	Linear regression; Logistic regression	Logistic regression	Regression-based forecasting; deviation-based; PCA;
Unsupervised	Clustering-based	Proximity-based clustering (Cluster-based, Distance-based, Density-based)	Proximity-based clustering	Proximity-based clustering; K-means; Latent semantic indexing (LSI)	Proximity-based clustering; Transformation-based / distance-based sharp outlier method

MADS ANALYTIC FRAMEWORK



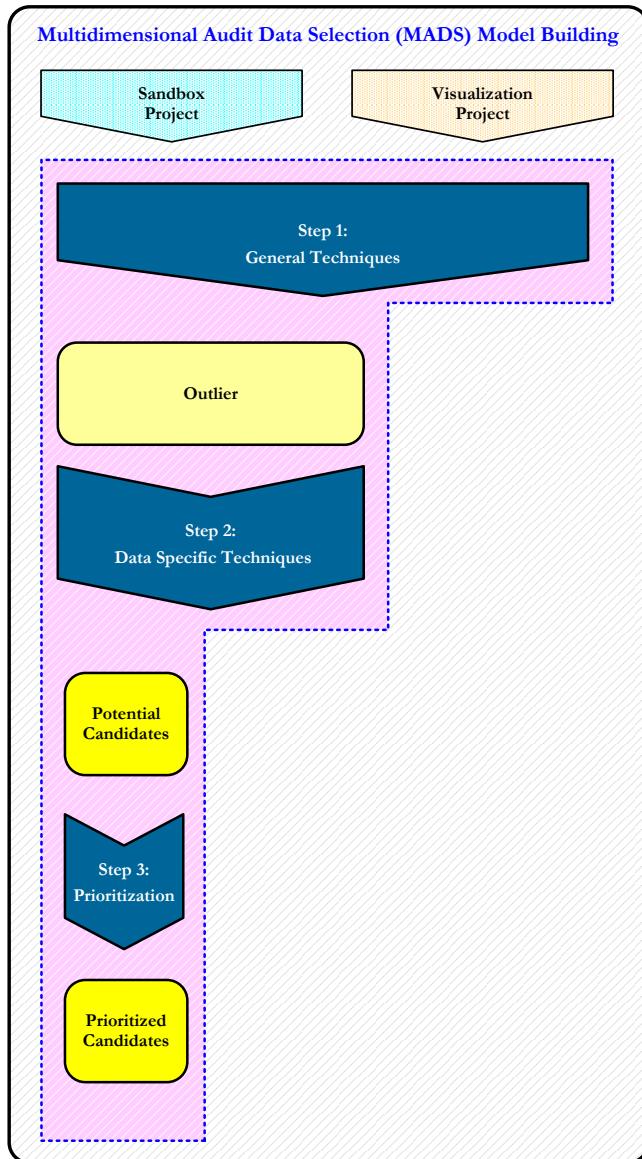
- ❖ Apply outlier detection techniques to evaluate the entire population instead of examining just a chosen sample, resulting in enhanced effectiveness of the audit.
- ❖ Select the most appropriate techniques based on the objective identified in the objective and criteria identification stage, available data, and the specific MADS model requirements (e.g., data size, data type, and model assumptions).
- ❖ Use a general guideline designed to assist auditors on how to choose appropriate techniques (or algorithms) to detect outliers given the different types of tasks and data on hand.
- ❖ The sandbox project can provide more information regarding the techniques to apply at this stage.
- ❖ Visualization techniques in visualization project can utilized to identify outliers.

MADS ANALYTIC FRAMEWORK



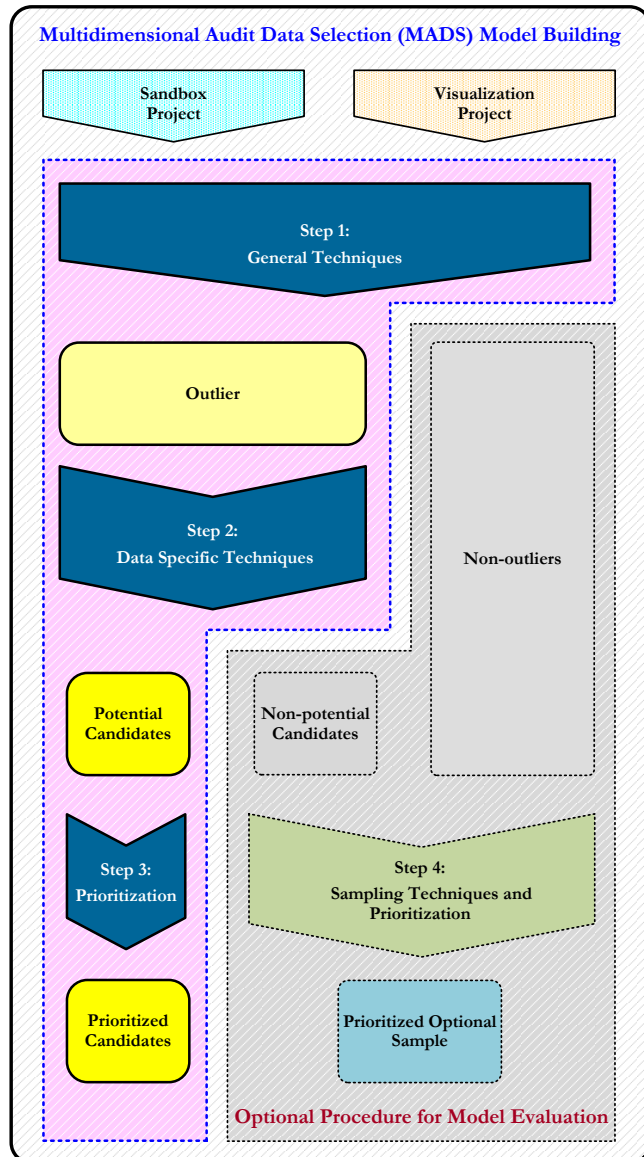
- ❖ Apply another set of data analytic techniques to the outliers in case step 1 results in a large number of outliers.
- ❖ Need to develop specific measures, a group of anomaly indicators, to identify outlier candidates that are more likely to be problematic.
- ❖ The criteria should be based on transaction types, industry, materiality of irregularity, auditors' prior experiences, and especially, the criteria should also tie in with the identified objectives of the MADS process.

MADS ANALYTIC FRAMEWORK



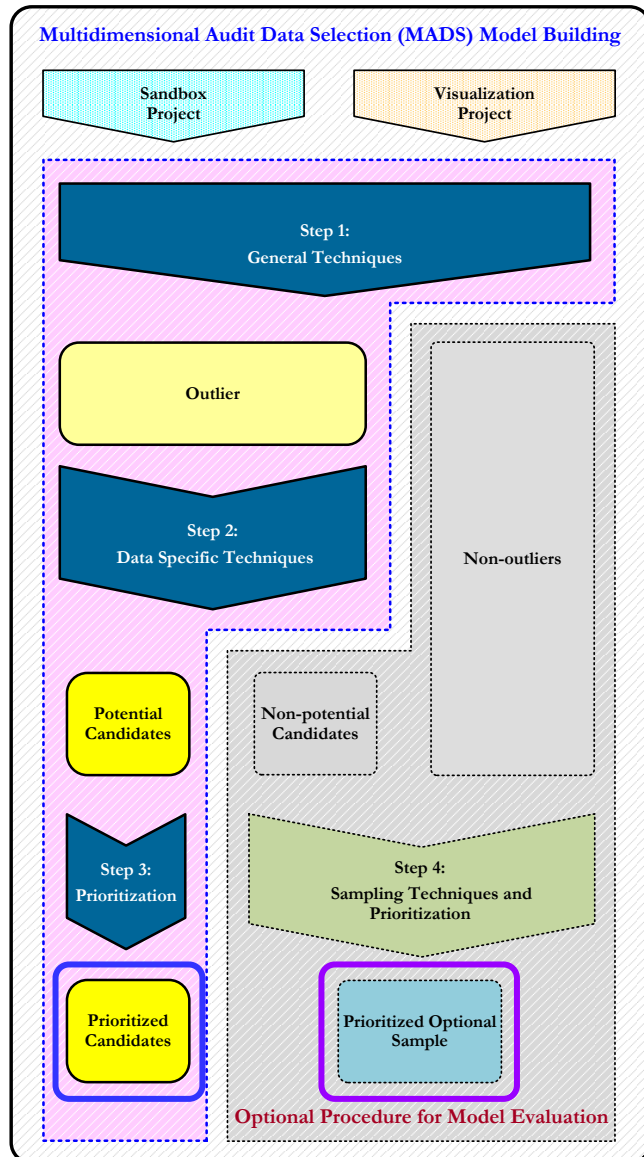
- ❖ Rank based on a prioritization method.
- ❖ Use suspicious scores based on weights.
- ❖ Weights can come from probability estimation, Bayes' Rule, or auditors' knowledge.

MADS ANALYTIC FRAMEWORK



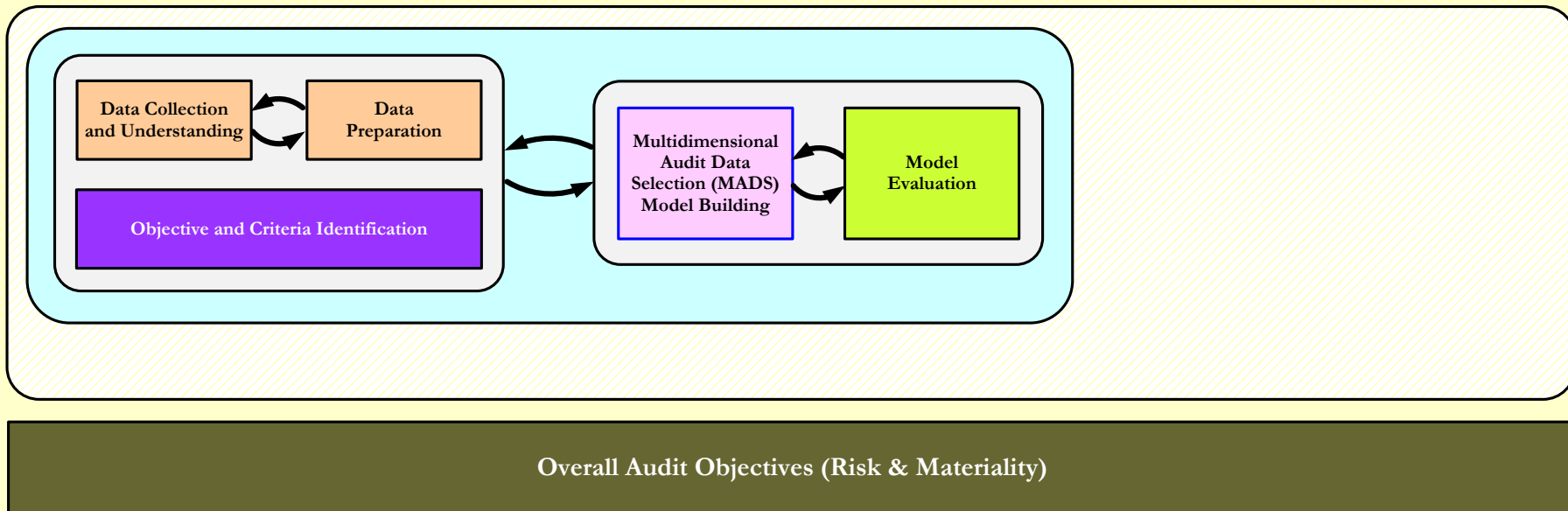
- ❖ May perform an optional procedure to confirm the effectiveness of the model.
- ❖ Employ the traditional sampling approach on the non-outliers and non-potential candidates resulting from steps 1 and 2.
- ❖ Apply the prioritization procedure to identify a prioritized optional sample.
- ❖ Performs substantive tests on the sample, which will be determined effective when the residual risk in the non-outliers and non-potential candidates is sufficiently low such that a follow up on these items becomes unnecessary.
- ❖ Only require when the auditors need to assess the effectiveness of the final MADS model.

MADS ANALYTIC FRAMEWORK



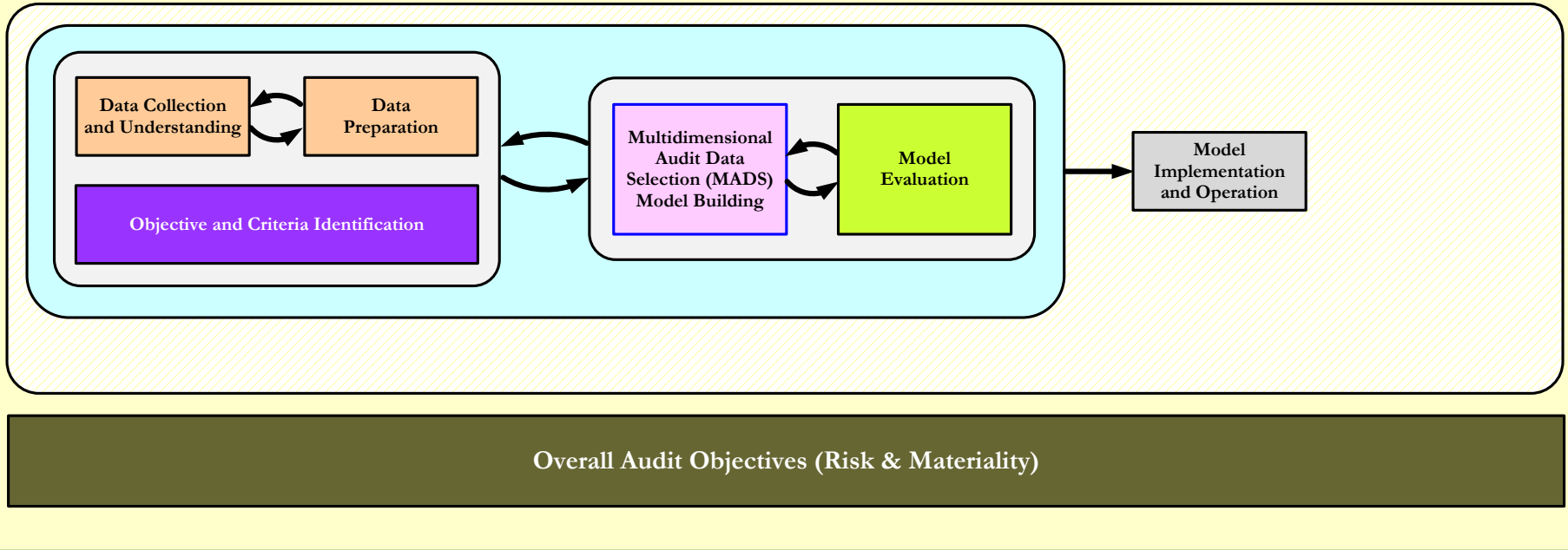
- ❖ Can examine the entire prioritized candidates, but most of the time, investigate a part of the prioritized candidates.
- ❖ Crucial question
 - How much audit data should follow up on.
 - Decide based on the determined risk of the specific account, the acceptable level of risk, and constraints such as the cost and time required to examine the data.
- ❖ If the optional procedure is performed,
 - Obtain the list by using random sampling technique or alternative sampling techniques (e.g., ranked set sampling and risk-based sampling) that take the characteristics of interest into account to improve estimation of a population parameter.

Multidimensional Audit Data Selection (MADS) Analytic Framework



- ❖ Assess the results rigorously to gain confidence that they are relevant and reliable.
- ❖ Need to iterate the procedures several times to produce a model that meets the acceptable level of accuracy.
- ❖ May need to revisit the data understanding and data preparation stages for adjustments required by its model of choice.

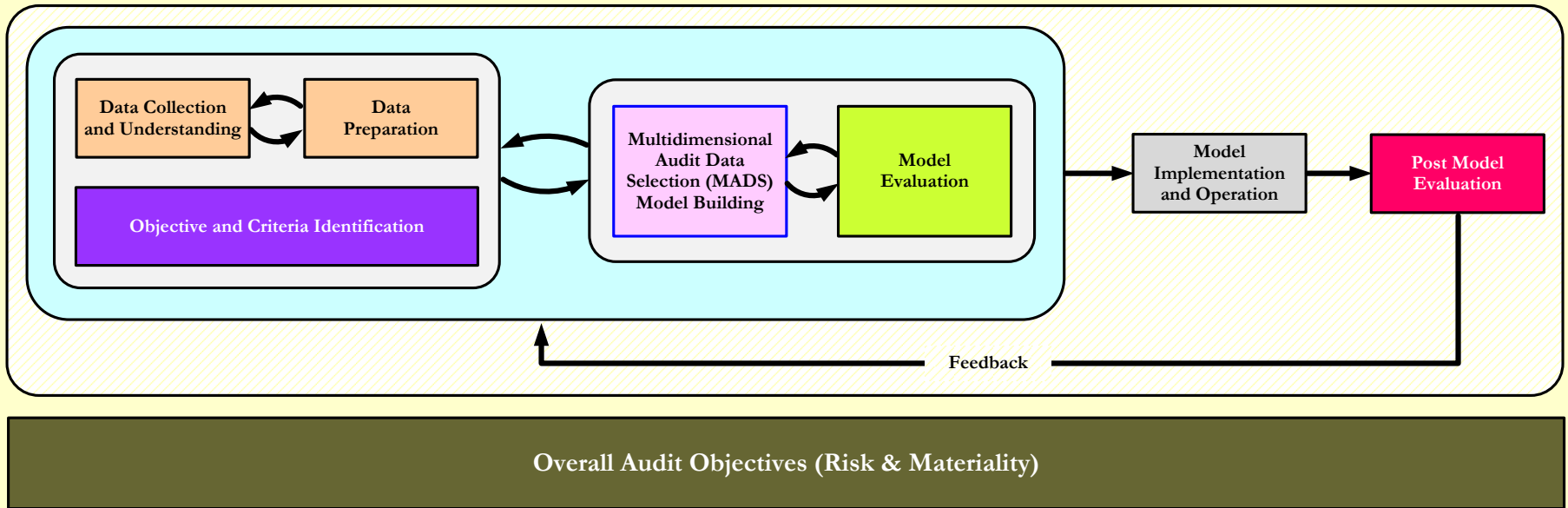
Multidimensional Audit Data Selection (MADS) Analytic Framework



- ❖ Implement and operate the model into a real substantive test of details.
- ❖ Monitor the operation of the model.

MADS ANALYTIC FRAMEWORK

Multidimensional Audit Data Selection (MADS) Analytic Framework



- ❖ Evaluate the model periodically to ensure its effectiveness and to make continuous improvements.
- ❖ Feed the results of evaluation back to the prior phases to refine or replace the existing MADS model.

- ❖ Conduct a survey of the current sampling practices in performing substantive tests of details.
- ❖ Obtain real data.
- ❖ Carry out an empirical test of whether or not the MADS process results in a more effective auditing process compared to the current sampling processes.
- ❖ Report the findings of the empirical test to facilitate a process for the public accounting profession to have open dialog with standard setters, regulators and other interested parties.



QUESTIONS & SUGGESTIONS

