Multidimensional Audit
Data Selection (MADS)

PCAOB – CARLab Meeting
Advances in data processing ability & data analytic techniques allows auditors to evaluate the entire population instead of examining just a chosen sample.

• **BUT, often generate large numbers of outliers.**
• **Impractical for auditors to investigate entire outliers**

• **Crucial to develop a method that can help auditors effectively deal with large amounts of data, but also assist them to efficiently handle a massive number of outliers.**
Multidimensional Audit Data Selection (MADS) Analytic Framework

- 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 six components.

- The practice of these six components is guided by the overall objectives of audit, specifically audit risk and materiality.
MADS Model Build Process

Whole Transaction Data (Entire Population)

Step 1: Filters for Significant Potential Risk Factors

Step 1 Outputs

Step 2: Data Analytic Techniques

Apply a set of filters to examine significant risks (i.e., What Could Go Wrong) (e.g., duplicate payment)

- Additional Filters
- Visualization Techniques (e.g., scatter plots)
- Professional Judgement (e.g., knowledge and experiences)
- Outlier Detection Techniques (e.g., classification & clustering).
Use professional judgement based on the importance of each step 1 filter and step 2 filter.

• Additional Filters
• Visualization Techniques (e.g., scatter plots)
• Professional Judgement (e.g., knowledge and experiences)
• Outlier Detection Techniques (e.g., classification & clustering).

Apply a set of filters to examine significant risks (i.e., What Could Go Wrong) (e.g., duplicate payment)

• Use professional judgement based on the importance of each step 1 filter and step 2 filter.
• Use the step 1 and/or step 2 results.
• Use a reasonable factor (e.g., dollar amount).
Multidimensional Audit Data Selection (MADS) Analytic Framework

Overall Audit Objectives (Risk & Materiality)

- Carry out empirical tests of whether or not the MADS process results in a more effective auditing process compared to the current sampling processes.
  - Revenue Cycle (Order-to-Cash)
  - Expenditure Cycle (Purchase-to-Pay)
  - Payroll Cycle
  - General Ledger
Expenditure (Procure-to-Pay) cycle

- **From Hub of Analytics Education**
  (http://www.hubae.org)
- **Bibitor LLC** is a retail liquor chain company that sells wine and spirits.
  - 1 year dataset (6/21/2016 - 6/20/2017)
  - 2,291,725 records and 5,234 invoices

### MADS Model Build Process

**Whole Transaction Data**
(Entire Population)

**Step 1:**
Filters for Significant Potential Risk Factors

**Step 1 Outputs**

**Step 2:**
Data Analytic Techniques

**Step 2 Outputs**

**Step 3:**
Prioritization

**Prioritized Notable Items**
## MADS Model Build Process

### Whole Transaction Data

**Entire Population**

- **Step 1:** Filters for Significant Potential Risk Factors

<table>
<thead>
<tr>
<th>ID</th>
<th>Potential Test (or Filters)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PUR-03</td>
<td>Identify purchases that are not properly approved (i.e., authorization limits - $250,000) by the authorizer (i.e., Chief Operating Officer).</td>
</tr>
<tr>
<td>PUR-06</td>
<td>Identify purchases that are received after payment.</td>
</tr>
<tr>
<td>PUR-09</td>
<td>Identify unusual purchases by producing exception reports of order amount/quantity that is too high (e.g., higher than percentile 95 value or greater than $5M/500 Units).</td>
</tr>
<tr>
<td>PUR-17</td>
<td>Identify purchases made to vendors who are not on the approved vendor list.</td>
</tr>
<tr>
<td>INV-02</td>
<td>Identify invoices where the order amount is different from the invoice amount.</td>
</tr>
<tr>
<td>INV-13</td>
<td>Identify multiple invoices at or just under approval cut-off levels (i.e., $250,000).</td>
</tr>
<tr>
<td>PAY-08</td>
<td>Identify payments that are made to invoices without purchase orders.</td>
</tr>
<tr>
<td>DUP-02</td>
<td>Identify duplicate invoices and/or amounts.</td>
</tr>
</tbody>
</table>
MADS Model Build Process

Whole Transaction Data (Entire Population)

**Step 1:**
Filters for Significant Potential Risk Factors

---

Whole Transaction Data

2,291,725 purchase transactions records
5,234 invoices

---

After applying step 1 filters, **384 invoices (7%)** has been identified for further investigation.
MADS Model Build Process

Whole Transaction Data (Entire Population)

Step 1: Filters for Significant Potential Risk Factors

Step 1 Outputs

Step 2: Data Analytic Techniques

Whole Transaction Data

2,291,725 purchase transactions records
5,234 invoices

Step 1: 8 Filters

384 invoices (7%)
MADS Model Build Process

Whole Transaction Data (Entire Population)

Step 1: Filters for Significant Potential Risk Factors

Step 1 Outputs

Step 2: Data Analytic Techniques

Whole Transaction Data

2,291,725 purchase transactions records
5,234 invoices

<table>
<thead>
<tr>
<th>ID</th>
<th>What Could Go Wrong?</th>
<th>S_ID</th>
<th>Potential Test (or Filters)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PUR-09</td>
<td>Purchases are made with unusual order amount and quantity.</td>
<td>SUB-01</td>
<td>Actual price is larger than approved price.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SUB-02</td>
<td>Invoice amount is significantly larger than order amount (&gt; $100,000)*.</td>
</tr>
<tr>
<td>INV-02</td>
<td>Order amount does not match with invoice amount.</td>
<td>SUB-01</td>
<td>Actual price is larger than approved price.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SUB-02</td>
<td>Invoice amount is significantly larger than order amount (&gt; $100,000)*.</td>
</tr>
<tr>
<td>INV-13</td>
<td>Purchases are made just under approval cut-off amount.</td>
<td>SUB-01</td>
<td>Actual price is larger than approved price.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SUB-03</td>
<td>Identify purchases made to vendors who are not in the approved vendor list.</td>
</tr>
<tr>
<td>PAY-08</td>
<td>Payments are made to invalid purchase orders.</td>
<td>SUB-04</td>
<td>Identify unusual payment without purchase orders (&gt; $5,000)**.</td>
</tr>
</tbody>
</table>

* 1% of Performance Materiality
** Based on Judgement
MADS Model Build Process

Whole Transaction Data (Entire Population)

Step 1: Filters for Significant Potential Risk Factors

Step 1 Outputs

Step 2: Data Analytic Techniques

After applying step 2 filters, 58 invoices (out of 384 step 1 outputs) has been detected.
MADS Model Build Process

Whole Transaction Data
(Entire Population)

Step 1: Filters for Significant Potential Risk Factors

Step 1 Outputs

Step 2: Data Analytic Techniques

Step 2 Outputs

Step 3: Prioritization

Whole Transaction Data
2,291,725 purchase transactions records
5,234 invoices

Step 1: 8 Filters
384 invoices (7%)

Step 2: 4 Filters
58 invoices
MADS Model Build Process

Whole Transaction Data (Entire Population)

Step 1: Filters for Significant Potential Risk Factors

Step 1 Outputs

Step 2: Data Analytic Techniques

Step 2 Outputs

Step 3: Prioritization

Whole Transaction Data

2,291,725 purchase transactions records
5,234 invoices

Prioritization

- Use the step 1 and step 2 results.

- Example
  - Assume that invoice #273 (one of 58 notable items) has three violations in step 1 and one violation in step 2, and the dollar amount is $265,000.
  
  - Violation score will be calculated as:
    \[
    \frac{\text{Step 1 Violations} + \text{Step 2 Violations}}{\text{Number of Step 1 and 2 Filter Applied}} = \frac{3 + 1}{8 + 2} = 0.4
    \]

  - Suspicion score will be calculated as:
    \[
    \text{Amount} \times \text{Violation Score} = 265,000 \times 0.4 = 106,000.
    \]

- Using the suspicion score, step 2 outputs (i.e., 58 notable items) are prioritized.
MADS Model Build Process

Whole Transaction Data (Entire Population)

Step 1: Filters for Significant Potential Risk Factors

Step 1 Outputs

Step 2: Data Analytic Techniques

Step 2 Outputs

Step 3: Prioritization

Prioritized Notable Items

Whole Transaction Data
2,291,725 purchase transactions records
5,234 invoices

Step 1: 8 Filters
384 invoices (7%)

Step 2: 4 Filters
58 invoices

58 Prioritized Notable Items
EVALUATIONS

❖ Statistical Sampling vs. Non-statistical Sampling vs. MADS

❖ Two Aspects

▪ Effectiveness - More suspicious items (i.e., errors) in the sample
▪ Efficiency - Less sample size (?)

❖ Three Potential Evaluation Methods

▪ Method 1: Benchmark (based on the assumption that we have already identified filters which can discover all errors in the full population)
▪ Method 2: Random Transaction Changes
▪ Method 3: Realistic Error Seedings by Experienced Auditors (Preferred)
Statistical Sampling vs. Non-statistical Sampling vs. MADS

Two Aspects

- **Effectiveness** - More suspicious items (i.e., errors) in the sample
- **Efficiency** - Less sample size (?)

Three Potential Evaluation Methods

- **Method 1: Benchmark** (based on the assumption that we have already identified filters which can discover all errors in the full population)
- **Method 2: Random Transaction Changes**
- **Method 3: Realistic Error Seedings by Experienced Auditors (Preferred)**

<table>
<thead>
<tr>
<th></th>
<th>Method 1 – Benchmark</th>
<th>Method 2 – Random Transaction Changes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pro</strong></td>
<td>• Use original data set (no manipulation)</td>
<td>• Provide error population and error items</td>
</tr>
<tr>
<td><strong>Con</strong></td>
<td>• Based on the assumption that identified filters can detect all errors</td>
<td>• Change the values of original data at random.</td>
</tr>
<tr>
<td></td>
<td>• MADS is inherently more effective than traditional sampling approaches since MADS filters are a subset of identified filters.</td>
<td>• Random value changes may not represent realistic errors.</td>
</tr>
</tbody>
</table>
Benchmark
- Apply all 27 filters (8 step 1 filters + 19 additional filters).
- Assume that items filtered by 27 filters are all errors in the entire population.
- 539 (10.3%) items are identified and regarded as benchmark.

Monetary Unit Sampling (MUS)
- Use CaseWare IDEA.
- Identify 67 items.

Non-statistical Sampling
- Based on judgement, stratify items into four groups.
- All 23 large items (>= $1M) are included.
- 44 items are randomly selected from the arbitrarily allocated three groups (i.e., 50%, 30% and 20%).

MADS
- All 58 notable items are selected.

<table>
<thead>
<tr>
<th>Population</th>
<th>$306,093,663</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tolerable Misstatement (75% of OM) (OM: 5% of Total Revenues)</td>
<td>$15,494,054</td>
</tr>
<tr>
<td>Expected Misstatement (1%)</td>
<td>$3,060,937</td>
</tr>
<tr>
<td>Risk of Incorrect Acceptance</td>
<td>10%</td>
</tr>
<tr>
<td>Sample Size</td>
<td>67</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Amount</th>
<th>Allocation</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;= $1M</td>
<td>23</td>
</tr>
<tr>
<td>&gt;= $250,000</td>
<td>50%</td>
</tr>
<tr>
<td>&gt;= $100,000</td>
<td>30%</td>
</tr>
<tr>
<td>&gt;= 0</td>
<td>20%</td>
</tr>
<tr>
<td>Sample Size</td>
<td>67</td>
</tr>
<tr>
<td>Sampling Risk (10%)</td>
<td>MUS</td>
</tr>
<tr>
<td>--------------------</td>
<td>-----</td>
</tr>
<tr>
<td>Sample Size</td>
<td>67</td>
</tr>
<tr>
<td>No. of Error Items</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>(2.4%)</td>
</tr>
<tr>
<td>No. of Violations Detected</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>(2.6%)</td>
</tr>
</tbody>
</table>

- More Effective
- More Efficient
## EVALUATIONS
### METHOD 1 – BENCHMARK

<table>
<thead>
<tr>
<th>Sampling Risk (10%)</th>
<th>MUS</th>
<th>Non-statistical</th>
<th>MADS</th>
<th>Benchmark</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Size</td>
<td>67</td>
<td>67</td>
<td>58</td>
<td>5,234</td>
</tr>
<tr>
<td>No. of Error Items</td>
<td>13</td>
<td>23</td>
<td>58</td>
<td>539</td>
</tr>
<tr>
<td></td>
<td>(2.4%)</td>
<td>(4.3%)</td>
<td>(10.7%)</td>
<td>(100%)</td>
</tr>
<tr>
<td>No. of Violations Detected</td>
<td>20</td>
<td>35</td>
<td>125</td>
<td>751</td>
</tr>
<tr>
<td></td>
<td>(2.6%)</td>
<td>(4.6%)</td>
<td>(16.6%)</td>
<td>(100%)</td>
</tr>
</tbody>
</table>

**Traditional Sampling**
- Traditional Sample
- Perform Substantive Testing
  - Inspection
  - Confirmation
  - Physical Examination
  - Inquiries
  - Tests for Significant Risks

**MADS Approach**
- Prioritized Notable Items
- Perform Substantive Testing
  - Inspection
  - Confirmation
  - Physical Examination
  - Inquiries
Values are randomly changed at the purchase transaction level.

- Total number of value changes
  - 0.01% of total purchase transactions (2,291,725): 225 transactions
- The values of amount, price, receiving date, and pay date are randomly changed.

Whole Transaction Data
2,291,725 purchase transactions records
5,234 invoices

Step 1: 8 Filters
481 invoices (9.2%)

Step 2: 4 Filters
99 invoices

99 Prioritized Notable Items
### EVALUATIONS
**METHOD 2 – RANDOM ERROR SEEDINGS**

<table>
<thead>
<tr>
<th>Sampling Risk (10%)</th>
<th>MUS</th>
<th>Non-statistical</th>
<th>MADS *</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Size</td>
<td>67</td>
<td>67</td>
<td>67</td>
</tr>
<tr>
<td>No. of Suspicious Items</td>
<td>4 (3%)</td>
<td>5 (3.8%)</td>
<td>24 (18%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>No. of Errors</th>
<th>133 Items</th>
</tr>
</thead>
</table>

* For comparison, choose top 67 items from 99 notable items.

- ✔️ More Effective
- ✔️ More Efficient
Different Data Analytics Techniques in Step 2 (e.g., Clustering)

Different Prioritization Criteria (e.g., Professional Judgement)

Different Evaluation Methods (e.g., Manual Realistic Error Seedings)

Different Data Sets