RUTGERS
THE STATE UNIVERSITY
OF NEW JERSEY

An application in Fraud Risk Assessment

-- Healthcare Fraud Detection

❖ Motivation and Contribution

- According to the Office of Management and Budget, in 2013, about 9.5%, or around $47.8 billion of the US's Medicare expenditure was lost due to fraud.
- Current studies are essentially technical in character; few of them discuss healthcare fraud detection from accounting/auditing point of view.
- This chapter intends to demonstrate how auditors can take advantage of advanced EDA techniques to assess healthcare fraud risk by following the proposed conceptual EDA application process.

Healthcare Fraudulent Behavior

Service provider's fraud 69% Research Effort	• Billing services that are not actually performed; • Unbundling; • Upcoding; • Perform medically unnecessary services and make them legal.
Insurance subscribers' fraud 31% Research Effort	• Falsifying records of employment/eligibility for obtaining a lower premium rate; • Filing claims for medical services which are not actually received; • Using other persons' coverage or insurance card to illegally claim the insurance benefits.
Insurance carriers' fraud	• Falsifying reimbursements; • Falsifying benefit/service statements.
Collusive fraud	• Fraud involving more than one party, eg. patient and physician

Jing Li. Kuel-Ying Huang, Jionghua Jin, Jianjun Shi (2007)

Methodology—Healthcare Data

- Current research mainly use outpatient claim data from private insurance company or governmental health departments outside U.S.
- Data used in this study purchased from the center for Medicare and Medicaid services (http://www.cms.gov/).
- Includes all the Medicare inpatient claims in 2010.
- There are in total 12,453,186 records and 1627 fields in the dataset.

All Attributes	1627
Less:	
Blank Attributes	-1181
Filled Attributes with more than 50% missing values	-134
Attributes with Single Values	-55
Remaining Attributes	**257**

Summary of Attributes Information

Methodology— Standard Audit Procedure

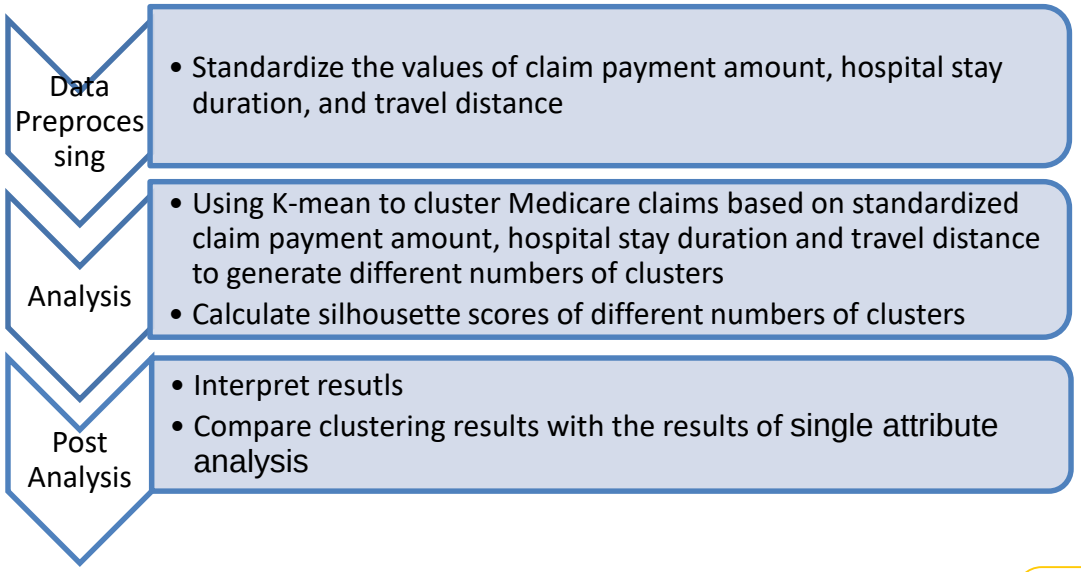
- Internal auditors' major concern of health care fraud is the payment
- Medicare claims with the same diagnosis, those having extreme large payment amounts are considered as high-risk instances.
- Threshold: Mean+3*Standard Deviation
- Payment can also be used with other risk indicators to prioritize the suspicious cases that need further investigation
  - Service providers obtaining extreme large payment amount from Medicare
  - Service providers filing large number of Medicare claims

Methodology—EDA Process

❖ Attribute Selection

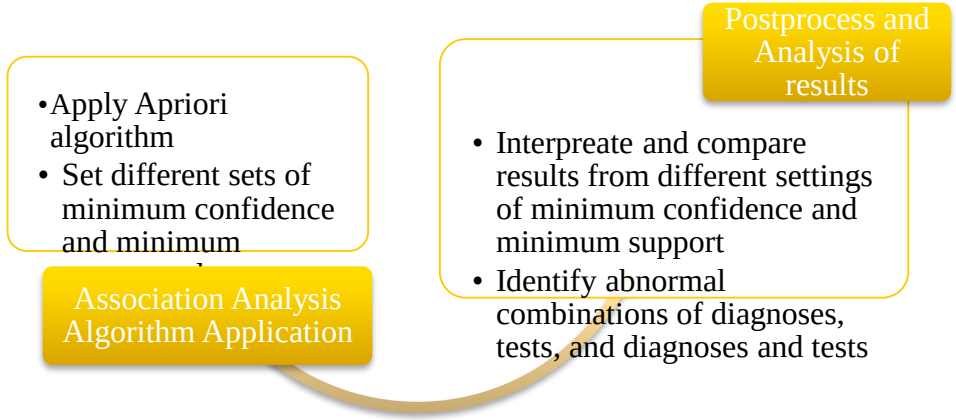
Attribute Name	Description
CLAIM_NO	Claim number
DESY_SORT_KEY	Beneficiary identifier
CLM_PMT_AMT	Claim payment amount
PRVDR_NUM	Provider number
CLM_UTLZTN_DAY_CNT	Claim utilization day count
Distance (derived)	Distance between beneficiary's residence county and provider's state
CLM_DGNS_CD{1-10}	Claim diagnosis code
CLM_PRCDR_CD{1-6}	Claim procedure code

Methodology—EDA Process



Cluster Analysis Process

Association Analysis Process



Results -- Standard Audit Procedures

- 180,644 high-risk payments are identified
- Analysis on service providers are performed to prioritize high-risk Medicare claims

Mean	Standard deviation	Minimum	Maximum	Count
1500.02	2534.86	1	39200.00	8302

Descriptive Statistics of Service Providers' Frequency Distribution

Mean	Standard deviation	Minimum	Maximum	Median
10515.5	8101.31	-375	159499.28	8302

- vo

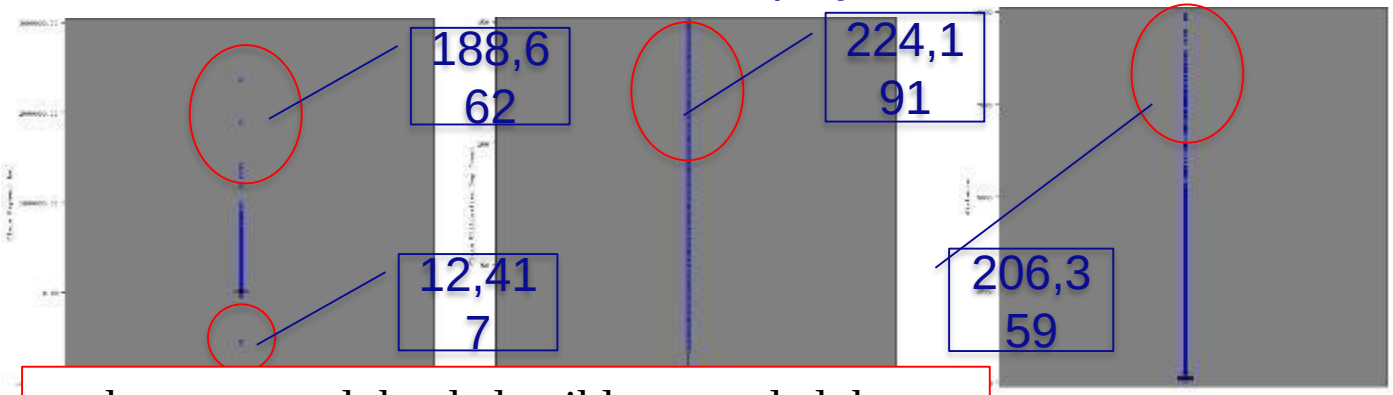
analyses, which relate to 56,267 high-risk payments.

Descriptive statistics of Service Providers' Payment Summary

EDA Results -- Generate and Testing Potential Explanations

EDA Results -- Identify Salient Features

EDA Results -- Display Distributions



- short stay and the deductible exceeded the amount Medicare pays
- long stay and the coinsurance amount exceeds the amount Medicare pays

Travel Distance

28 beneficiaries

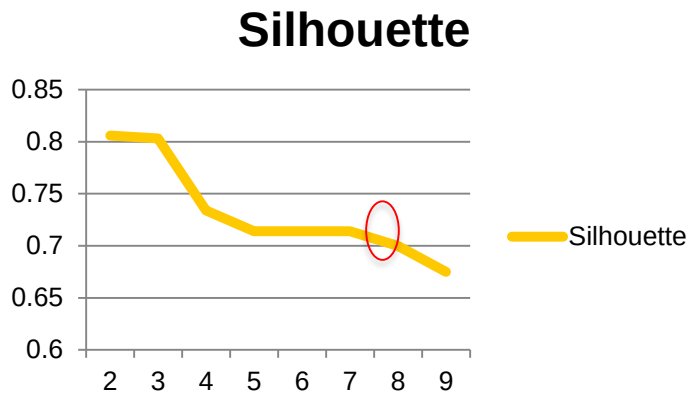
			Minimum	Maximum
			1	103
Payment summary	17480.15	24400.12	-534561.55	3384174.56
Hospital Stay Summary	9.7433405	13.9969875	0	668
Travel Distance Summary	218.6821081	398.0258267	0	110925.55

Descriptive Statistics of Beneficiary Related Distributions

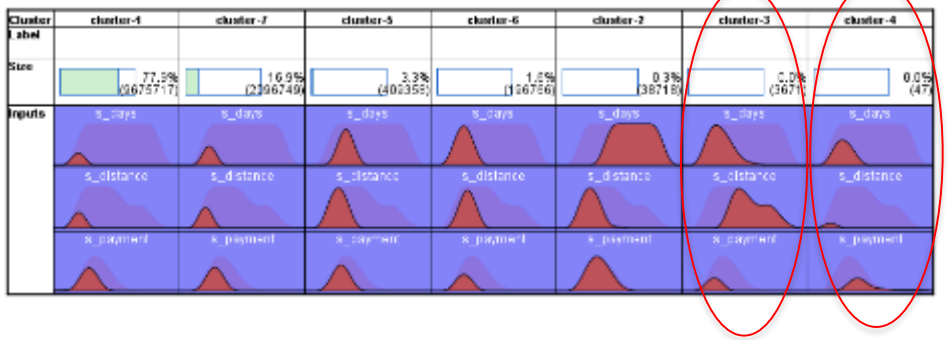
Identify Suspicious Cases – Traditional EDA Techniques

- 7 out of 12,417 claims with negative payment amount associated with zero deductible amount and coinsurance amount.
- 25 out of 28 beneficiaries who were paid for more than 365 days are not actually stayed in hospital for such long.
- Among the 138 claims relating to these 25 beneficiaries, 6 potential duplicate claims are identified.

Identify Suspicious Cases – Cluster Analysis



Number of Clusters and Resulting Silhouette



Cluster Analysis Results of 7 Clusters

Identify Suspicious Cases – Association Analysis

212 Claims

	S_{min}=1%	S_{min}=0.5%	S_{min}=0.25%
C>95%	2	11	75
90%<C<95%	1	11	39
85%<C<90%	2	4	24
Total	5	26	138

12,298 Claims

29,530 Claims

Generated Association Rules

- Inappropriate combinations can imply extensive misuse of certain procedures.
- Confirmed high confidence rules can be used to identify abnormal cases from the dataset.

Explore the Causes of Exceptional Cases, Confirm Relationships, and Generate New Audit Objectives and Report Finding

- Additional supporting information or experts' domain knowledge are required to explore and confirm the causes of exceptional cases.
- New Audit Objectives:
  - Negative payment amount was presented when deductible amount or coinsurance amount exceeded the amount Medicare pays.
  - One beneficiary was paid for at most 365 days' hospital stay per year.
  - The number of days paid by Medicare did not exceed the actual number of days the beneficiary stayed in hospital. .
  - Large payment amounts and long distance travels were associated with long hospital stays.
  - All the verified association rules can be considered as new audit objectives.

Conclusion

- This chapter demonstrates how EDA process can be applied to healthcare data to assess fraud risk.
- Traditional EDA methods as well as two advanced EDA techniques, cluster analysis and association analysis, are applied.
- **Major Finding**
  - Traditional EDA techniques discover 25 beneficiaries who were inappropriately paid for more than 365 days hospital stayed in 2010
  - Cluster analysis identifies 3,671 Medicare claims having long travel distances, short hospital stay periods, and small payment amounts; and 47 claims with large payment amounts and short hospital stay periods.
  - Association analysis creates up to 75 strong rules to describe relationships among diagnoses and procedures, which can discover at least 212 exceptional Medicare claims from the data.