# Progress Report

# Integrated Warfighter Biodefense Program (IWBP)

**Submitted By:** Quantum Leap Innovations, Inc.
3 Innovation Way, Suite 100
Newark, DE 19711-5456

**Contract Number:** N00014-10-C-0363
**Report Date:** October 21, 2011
**Reporting Period:** July 1, 2011 – September 30, 2011

**Principal Investigator:** Dr. Ganesh Vaidyanathan
Phone: (302) 894-8044 Fax: (302) 894-8001
gv@quantumleapinnovations.com

**Monitoring DoD Organization:** CDR Joseph Cohn
Program Officer, Code 341
Office of Naval Research
875 North Randolph Street
Arlington, VA 22203-1995
joseph.cohn@navy.mil



# Unclassified

# Report Documentation Page

*Form Approved*
*OMB No. 0704-0188*



| 1. REPORT DATE | 2. REPORT TYPE | 3. DATES COVERED |
|---|---|---|
| **DEC 2011** | | **01-07-2011 to 30-09-2011** |

| 4. TITLE AND SUBTITLE | 5a. CONTRACT NUMBER |
|---|---|
| **Integrated Warfighter Biodefense Program (IWBP)** | 5b. GRANT NUMBER |
| | 5c. PROGRAM ELEMENT NUMBER |
| 6. AUTHOR(S) | 5d. PROJECT NUMBER |
| | 5e. TASK NUMBER |
| | 5f. WORK UNIT NUMBER |
| 7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Quantum Leap Innovations, Inc.,3 Innovation Way, Suite 100,Newark,DE,19711** | 8. PERFORMING ORGANIZATION REPORT NUMBER |
| 9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) | 10. SPONSOR/MONITOR'S ACRONYM(S) |
| | 11. SPONSOR/MONITOR'S REPORT NUMBER(S) |

12. DISTRIBUTION/AVAILABILITY STATEMENT
**Approved for public release; distribution unlimited**

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT | 18. NUMBER OF PAGES | 19a. NAME OF RESPONSIBLE PERSON |
|---|---|---|---|---|---|
| a. REPORT | b. ABSTRACT | c. THIS PAGE | | | |
| **unclassified** | **unclassified** | **unclassified** | **Same as Report (SAR)** | **11** | |

# EXECUTIVE SUMMARY:

This report outlines Quantum Leap Innovations, Inc. (QLI) accomplishments during the three months of performance between June 30, 2011 and September 30, 2011 on ONR Contract N00014-10-C-0363 for the Integrated Warfighter Biodefense Program (IWBP). The report summarizes activities focused on continued development of the Quantum Leap Innovations Pattern Based Analytics (PBA) Platform.

# SUMMARY OF ACCOMPLISHMENTS:

**Continued Development of Quantum Leap Innovations Pattern Based Discovery:**

In the previous reporting period, we had outlined a design framework for a new search engine inspired work flow for our Discovery product to facilitate exploratory data analysis and visualization. The primary objective has been to allow the end user to easily identify patterns of most interest to them as the basis for further data exploration, visualization and analysis.

During the current reporting period, significant progress has been made in implementing the search engine work flow. The tutorial developed below provides a good summary of the progress made during this reporting period:

# PATTERN BASED DISCOVERY TUTORIAL

The Quantum Leap Pattern Based Discovery ("Discovery") product automatically discovers informative patterns against a user specified query using a search based paradigm. A ranked list of informative patterns that link to associated data subsets is generated from the search and displayed. This allows the user to easily perform targeted exploration, visualization and analysis of informative data subsets rather than all the data. The following example walks the user through a Healthcare Fraud problem using Discovery.

Healthcare Fraud Example:

It has been estimated that healthcare fraud and abuse can constitute between 3 -15% of annual healthcare expenditures in the United States. From a cost standpoint, this translates to $100-$170 billion in annual costs! Analysis of healthcare data to discover patterns that associate with different fraud types can potentially provide a proactive means for health care providers to detect fraud early on. In this example, we use a simulated data set of ~1 million patients based on an existing fraud model ("Healthcare Fraud and Abuse", Rudman et al). Six fraud types are modeled based on statistical occurrence within the nation. An additional challenge with this data set is the prevalence of MISSING data that is characteristic of healthcare data. Appendix A summarizes the data characteristics for this example.

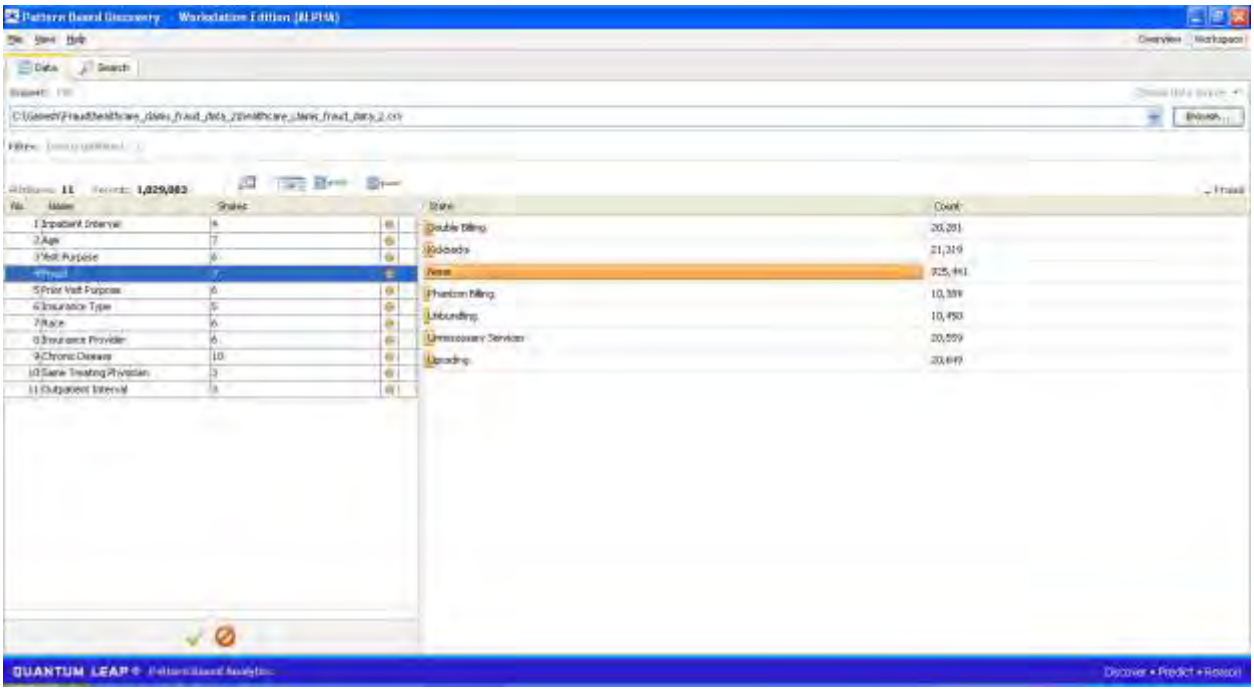
Key questions to be answered include:

a. What are the strongest patterns that associate with each type of Fraud?

b. Are there informative statistics/clusters within the data subsets associated with the strongest patterns that can be used as a basis for proactive monitoring of fraud?

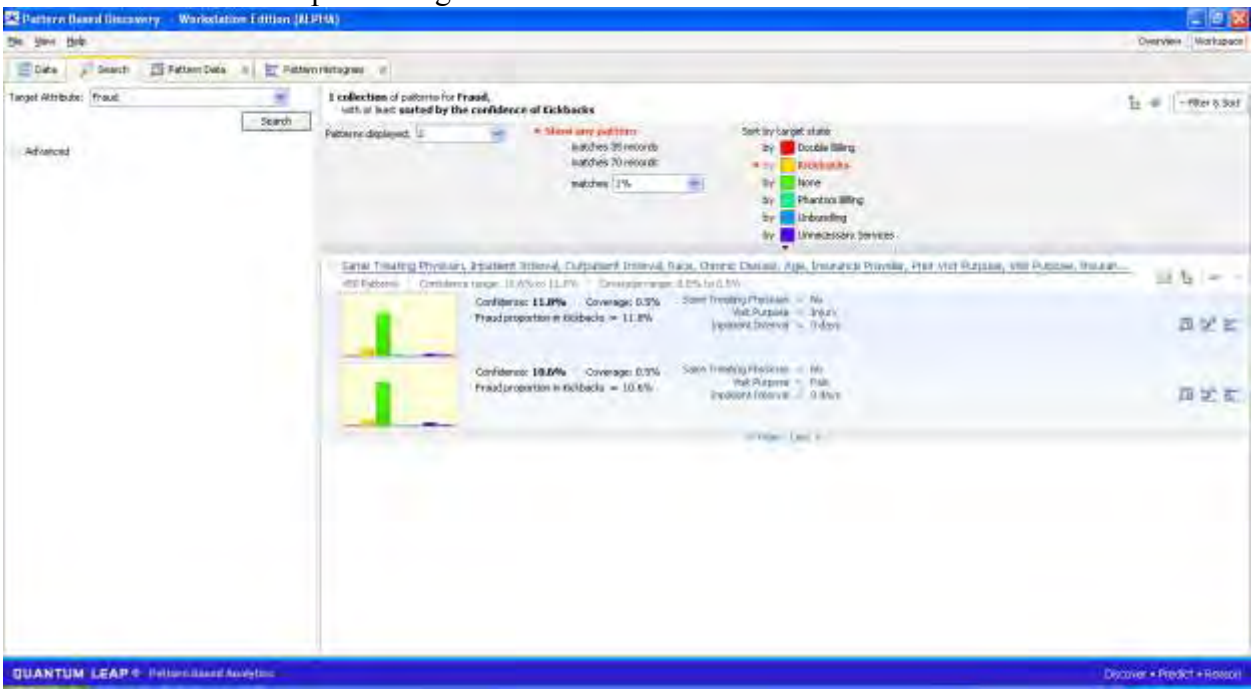
In the following tutorial, we walk the user through the use of the Discovery product to address these types of questions.

A. Load Data ("healthcare_claims_fraud_data_2.csv") into Discovery:



The attributes that make up the data are shown on the left. The "Fraud" attribute is highlighted and a histogram of the corresponding distribution of fraud types is displayed on the right.

B. Search for patterns against FRAUD:

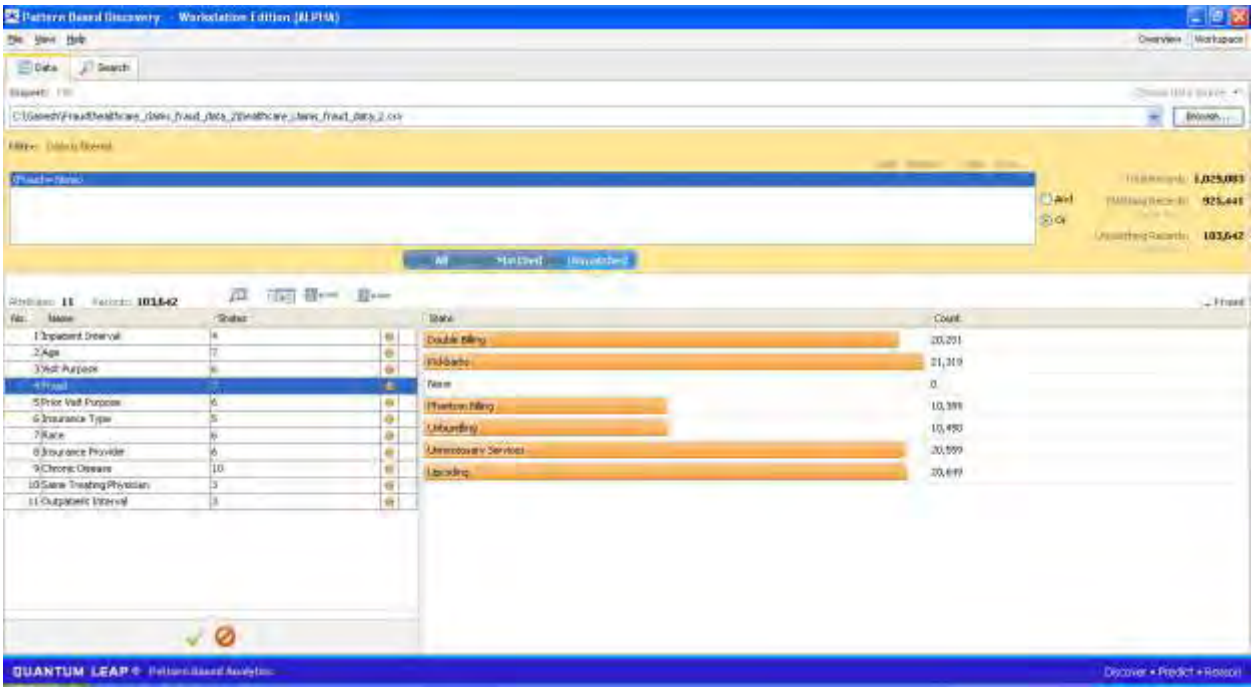


The search term in the top left of the screen is "Fraud". The top two patterns associated with "Kickbacks" are displayed on the right with the confidence level for Kickbacks shown as 11.8% next to the bar graph. Note that the dominant confidence level is

associated with Fraud type "None" shown in green in the bar graph accompanying the pattern. This is due to the dominance of "None" within the data.

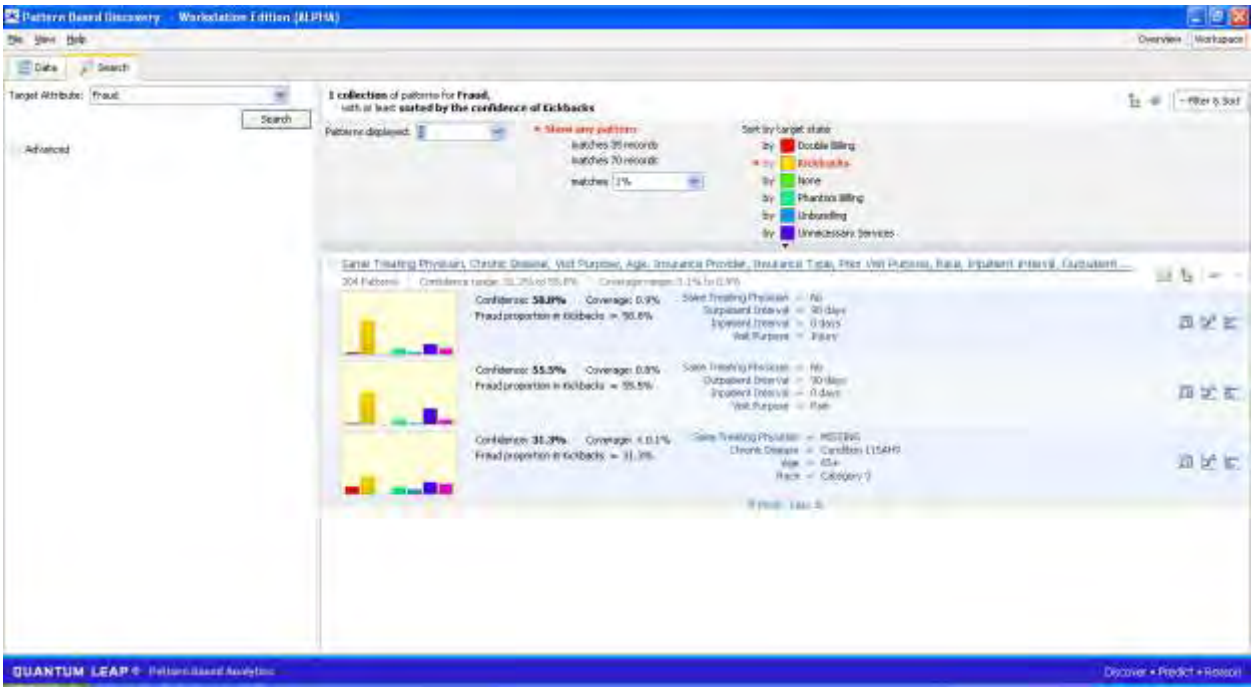
In order to "zoom in" on actual Fraud patterns, it will be useful to filter out the dominant Fraud type of "None". We can do this by returning to the Data screen to apply a filter.

C. Go back to Data screen and filter data to exclude Fraud type "None":



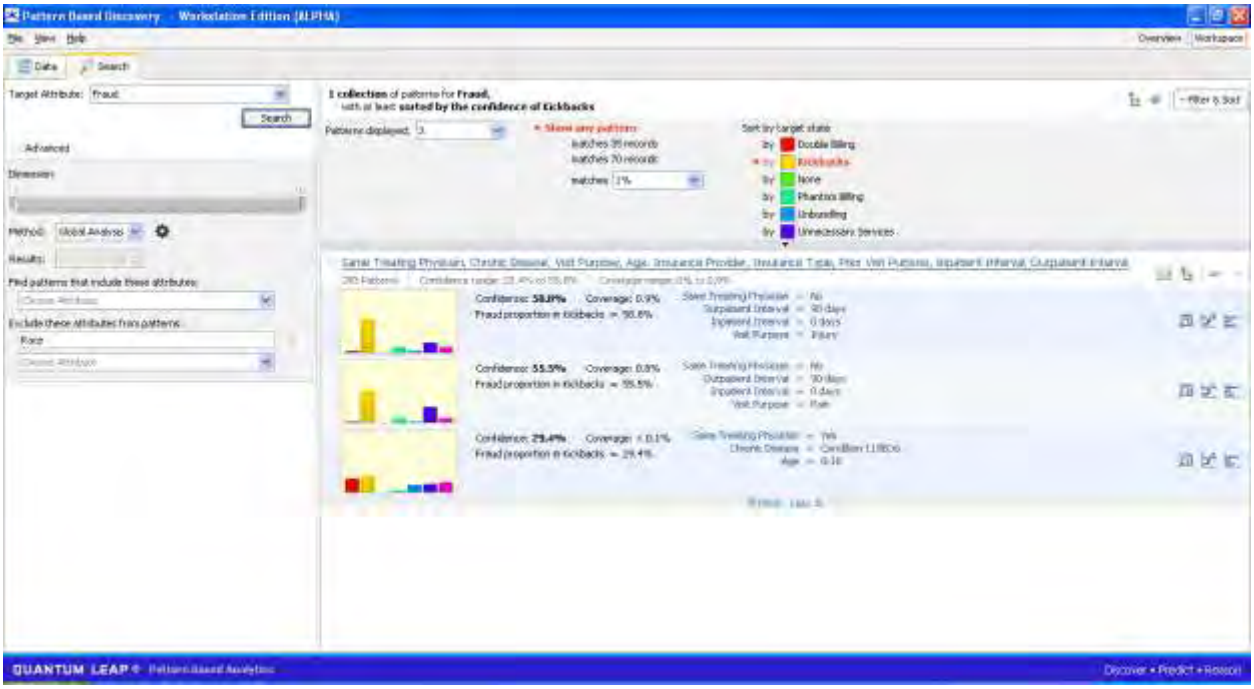

Note the highlighted Filter (Fraud = None) within the Filter Window and the reduced number of data records (103642 records versus the original data size of 1,029,083 records) that remain after applying the filter. This filter was added by clicking on the "Add" button within the Filter window.

D. Search the filtered Fraud data to discover patterns against “Fraud”:




We note that the confidence level for Fraud type “Kickbacks” has now increased dramatically to 58.8%. Another interesting observation is the inclusion of the attribute “Race” in the third pattern. To exclude Race as an attribute for Pattern Discovery, we can perform an Advanced Search where we can customize or refine our search.

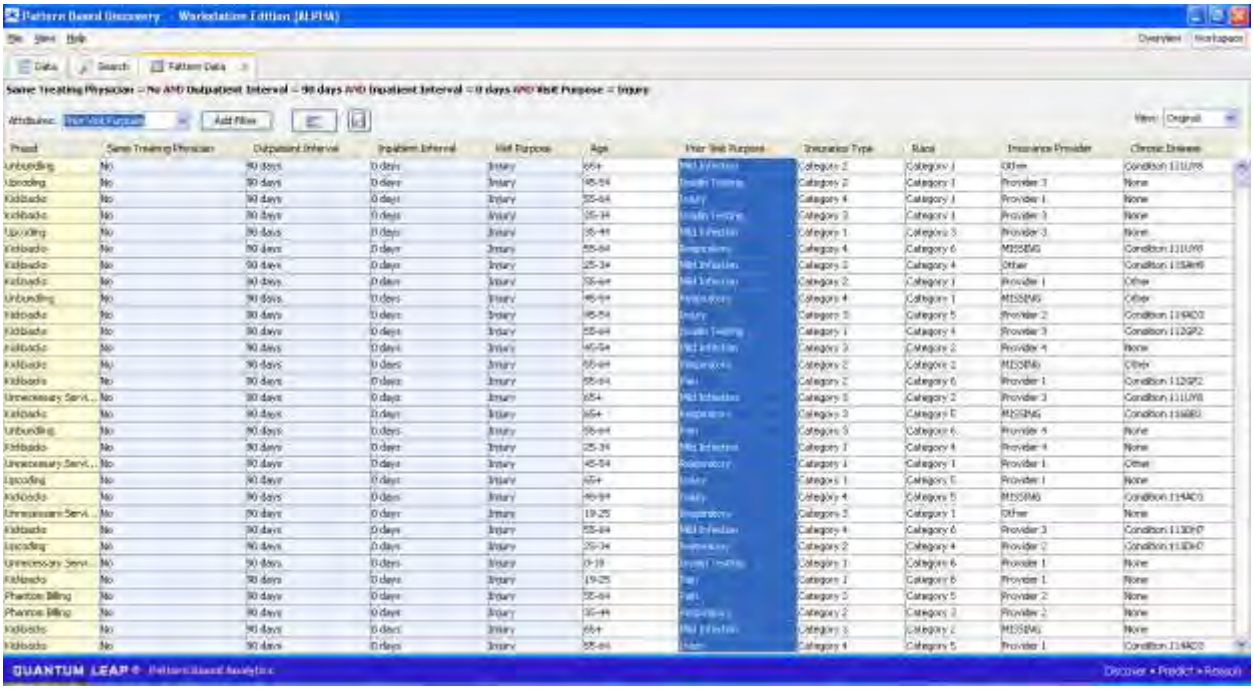
E. Enter “Advanced” Search to exclude Race from pattern discovery:




The excluded attribute “Race” is listed under the Advanced window on the lower left. Note that the resulting patterns on the right no longer include the Race attribute. We can

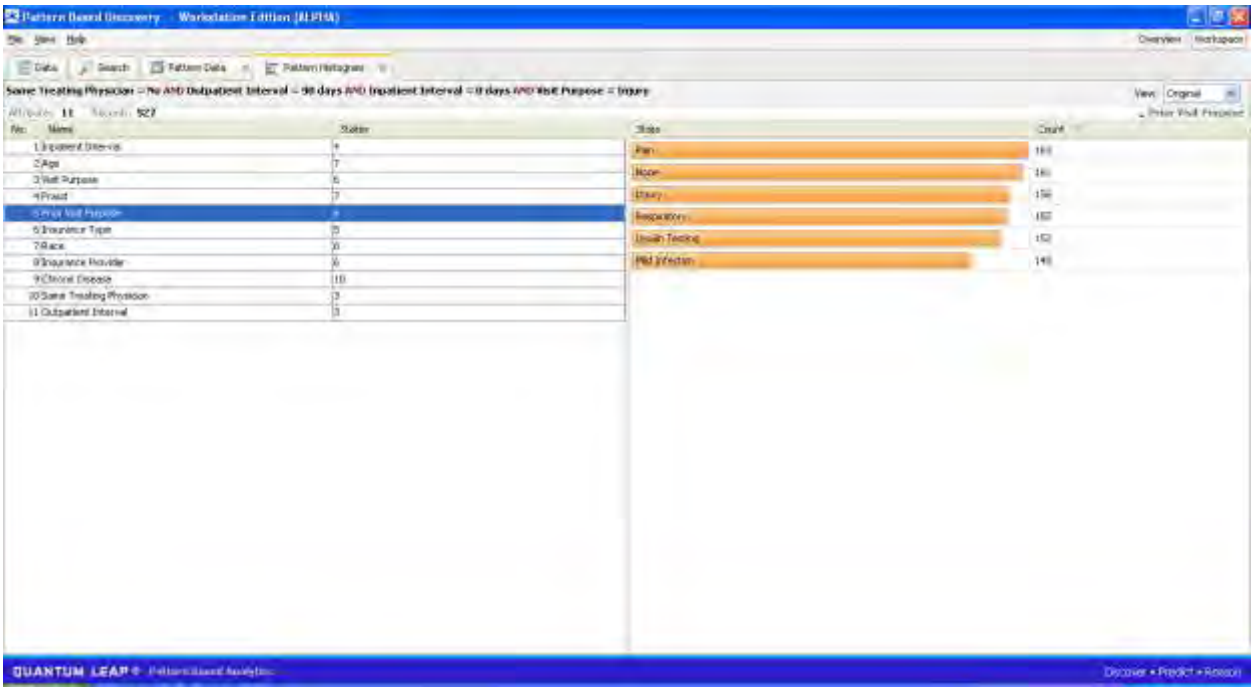
now click on the "Show data table" icon on the right of the visual summary of each pattern to examine the data associated with the top pattern in more detail:

F. Examine data table associated with top pattern for examining the "Prior Visit Purpose" distribution within this pattern:



The data table shows the target attribute ("Fraud") on the far left, highlighted in yellow. The attributes that make up this pattern are shown in light blue, followed by the remaining data associated with the pattern. The selected "Prior Visit Purpose" attribute is highlighted in dark blue. A histogram of the "Prior Visit Purpose" distribution for the data described by this pattern can be displayed by clicking on the Histogram icon next to the "Add Filter" button on the top left. The data table and the pattern can be saved by clicking on the Save Data tab immediately to the right of the Histogram icon.

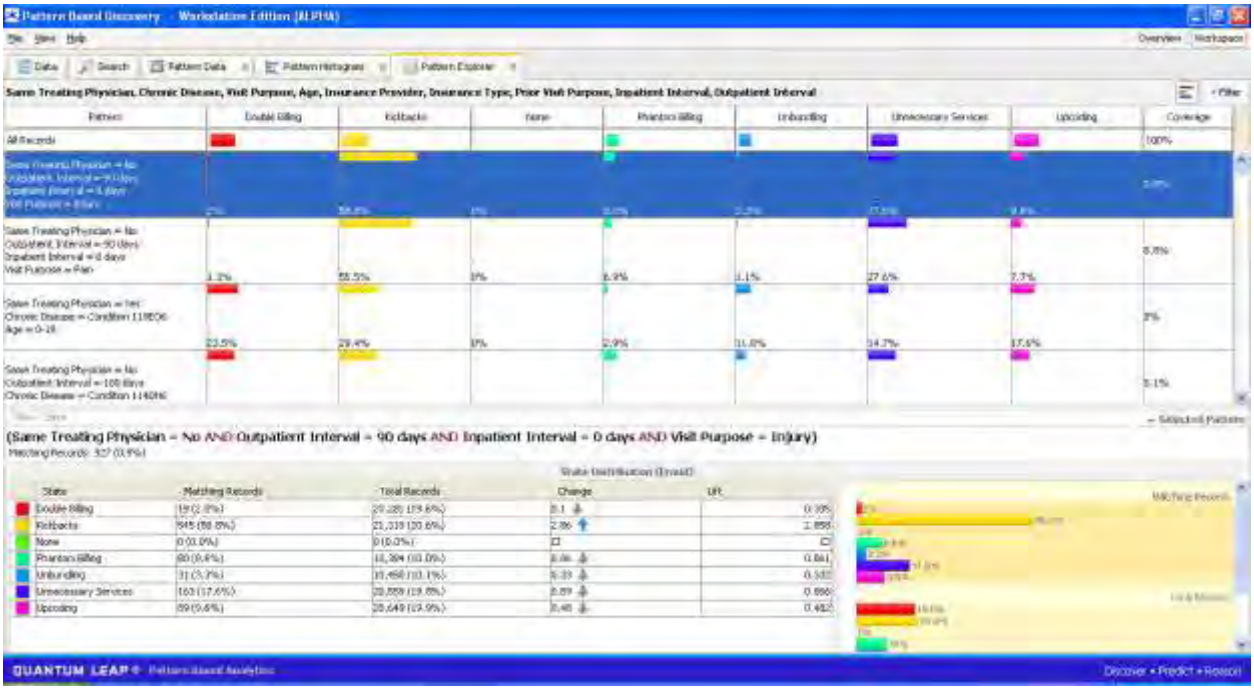
G. Plot Histogram of "Prior Visit Purpose" Distribution:




The histogram shows that the most frequently occurring "Prior Visit Purpose" category within this pattern is "Pain".

We may further be interested in examining all 293 patterns in (E) to get a global understanding of the patterns. We can visually examine all the patterns in the collection using Pattern Explorer.

H. Visually examine all 293 patterns using Pattern Explorer:

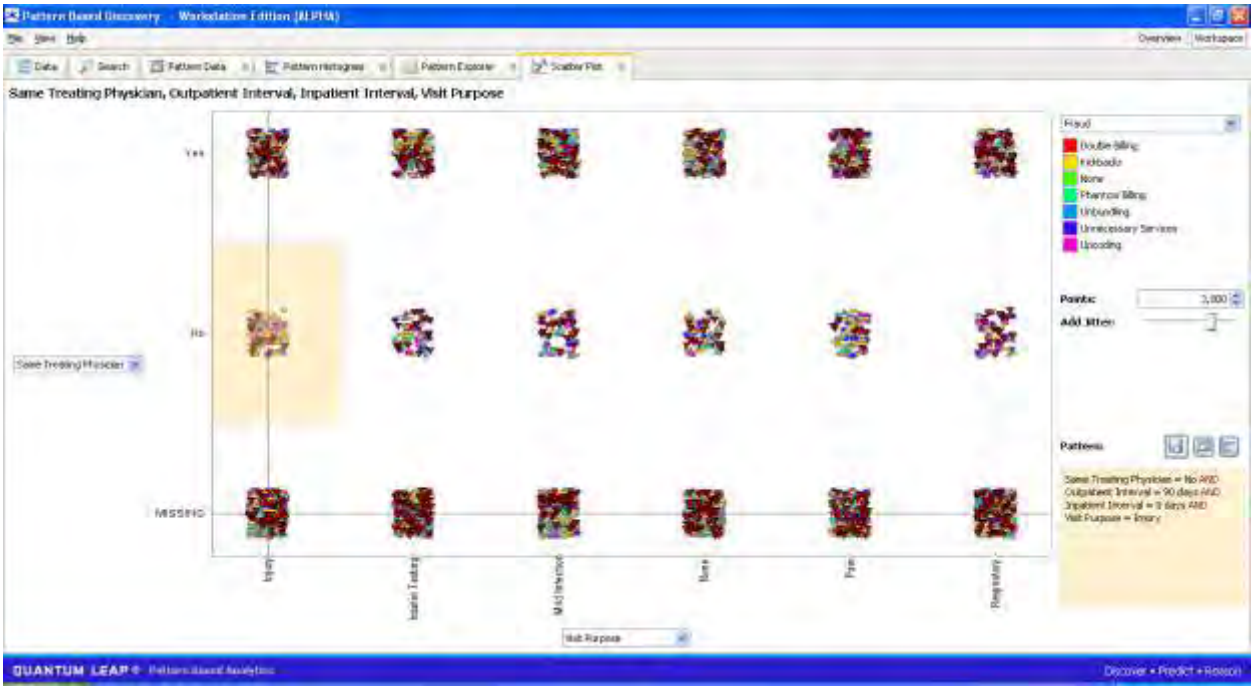


Pattern Explorer aggregates all the patterns in the collection for global pattern analysis. In this example, the patterns have been sorted in descending order by the maximum confidence level for "Kickbacks". Summary statistics on the highlighted pattern are displayed at the bottom.

The user can further examine specific relationships in data described by a pattern using the scatter plot feature where any two attributes can be plotted against each other. We demonstrate this feature using the top pattern displayed in (E).

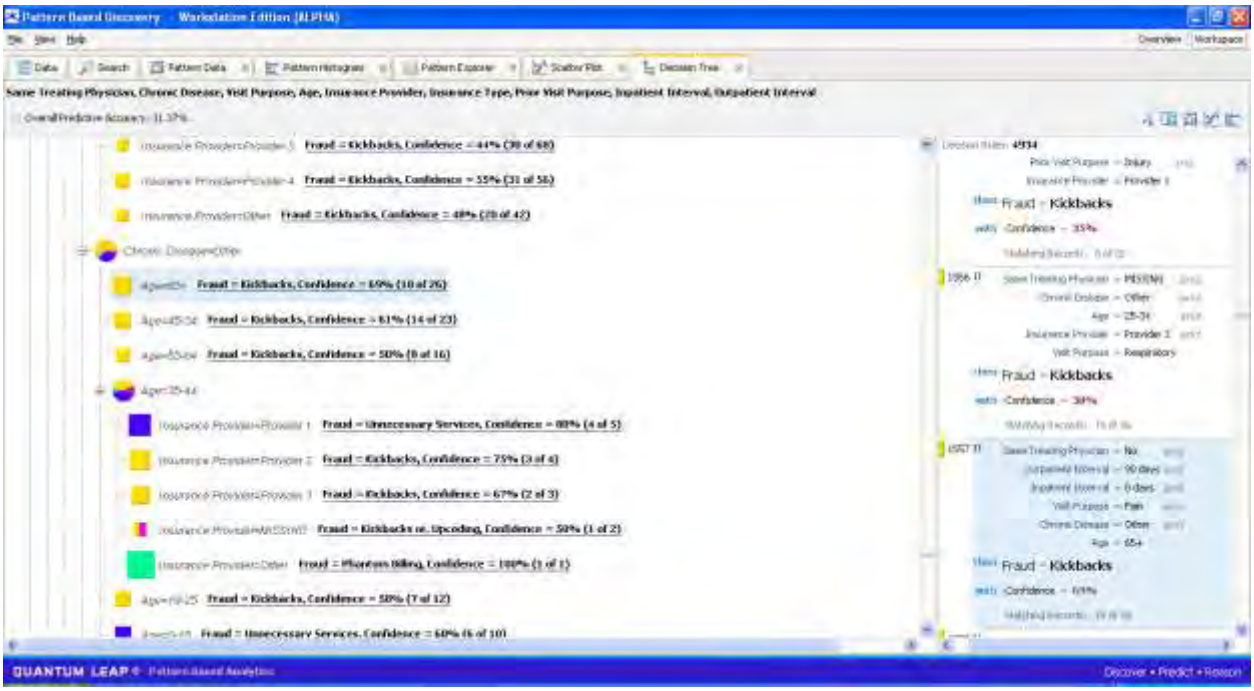
I. Displaying x-y relationships using a scatter plot:

The top pattern in (E) involves both "Visit Purpose" and "Same Treating Physician". When the scatter plot icon is clicked and adjusted, the user sees the scatter plot below where the user can select the attributes to be displayed from those that define the pattern and the query. In this example, we display "Same Treating Physician" versus "Visit Purpose". The shaded yellow rectangle indicates the portion of the entire data described by the selected pattern. Note that the selected data subset shows a greater density of yellow icons representing "Kickbacks". For discrete data, we have added "jitter" as can be seen at the right to separate overlying data. In addition, we randomly sampled 3000 data points to reduce data density.




Finally, the user may be interested in examining more detailed relationships between all the patterns associated with the attributes that form the collection of patterns. We can "zoom in" on patterns involving this collection in more detail using the Decision Tree tab associated with the collection of patterns.

J. Displaying Inter-Pattern relationships using a Decision Tree:




When the user clicks on the Decision Tree tab associated with the collection, a detailed tree is generated. Note that the highlighted pattern is even stronger than the best pattern from our search list in (g)! This is because the resolution of this local decision tree was set to a very high level using the "Pruning" icon on the top right of the screen.

This example highlights how the user can explore a collection of attributes in more detail to reveal further insights using a local decision tree whose resolution can be controlled by the user.

NOTE: The visualizations throughout the example can be saved to clipboard by right clicking on the visual. Items of interest can then be copied for example to Word to generate a report.

## Appendix A: Summary of Data Characteristics

The states associated with each attribute can be examined using the data table in Screenshot 1.

**Attributes**

Inpatient Interval
Age
Visit Purpose
Prior Visit Purpose
Insurance Type
Race
Insurance Provider
Chronic Disease
Same Treating Physician
Outpatient Interval
Fraud (Double Billing, Kickbacks, Phantom Billing, Unbundling, Unnecessary Services, Upcoding)

**Customer Engagements:**

During the current reporting period, significant effort has been spent on exposing the Quantum Leap Pattern Based Analytics Platform to the broader Analytics community. The Pattern Based Analytics Linked In Group has grown to several hundred members. In addition, our team has been focusing their efforts in launching a beta program for users to test and evaluate the platform, details of which will be reported on during the next reporting period. In addition, there have been several ongoing discussions with enterprise customers on testing the platform to provide value for their respective businesses.

# NEXT STEPS:

During the next reporting period, further efforts will continue in expanding awareness of the Pattern Based Analytics platform. In addition, work will begin on integrating Pattern Based Prediction (to enable Predictive Analytics) with Pattern Based Discovery.

# FINANCIAL SUMMARY:

**Contract Activity**

| | |
|---|---|
| QLI Contract N00014-10-C-0363<br>Award date: 07/01/2010 | $2,987,891 |
| ACTUAL: Expenditures Invoiced to the Government through September 30, 2011 | $1,717,303 |

57% of Contract Value has been spent as of September 30, 2011