Application of Integrated Health Analytics in a Regulatory Environment

National Predictive Modeling Summit
December 5, 2012

Morgan Crafts, MBA, PMP
Director of Technology
Health IT, Northrop Grumman Information Systems
The Need for Data-driven Predictive Analytics

• Transformation’s Three Part Aim:
  – Improve care
  – Improve population health
  – Reduce costs

• The Challenge:
  – Different members of the health community have access to different sets of data
  – Few have been able to look across the data sets to get a real and timely sense of the health ecosystem
  – Transformation could be accelerated through integrated health analytics information to inform strategy, guidance, operations, evaluation

• The Need:
  – A flexible and scalable predictive analytics platform that can rapidly provide integrated insights to a broad range of health decision makers
• Security and privacy are the most important priorities among health IT managers. Seventy-three percent rank security and privacy among their agency’s top three priorities, and the same amount say it will be the top priority in the next 12 months.

• Digital Health Records and Increased access to health data are next most important priorities.

• Mobile applications and social media are seen as having increased focus in the coming year.

Security and Privacy top current HIT Priorities - Social Media and Mobility Expected to Increase in Importance
Respondents were asked to identify challenges within their respective agencies; No item dropped below 79 percent, indicating relative parity among items listed.

Ninety-five percent of respondents indicated that leveraging existing IT systems is a challenge; The second largest challenge is increasing mobile access at 91 percent.

Seventy-nine percent of federal managers indicated that identity fraud was a challenge.

Percentage of respondents, n=168
Respondents who indicated “Don’t Know” were not included in the analysis.
Federal managers were asked to select the tools that improve their agency’s health outcomes.

Health information exchange (HIE) systems (68 percent) and electronic health records (65 percent) are the most helpful tools when it comes to improving health outcomes.

Data analytic tools, mobile technology and continuous monitoring systems are high on the list.

**Tools That Can Improve Health Outcomes**

- Health information exchange systems: 68%
- Electronic health records: 65%
- Data analysis tools: 46%
- Mobile technology: 38%
- Continuous monitoring systems: 38%
- Imaging technologies/GIS: 34%
- Cloud computing: 26%
- Other: 7%

Percentage of respondents, n=205
Respondents could select more than one answer.

NG technology focus/tool development in line with survey observations.
## Applying Prediction: Sample Current Health Challenges

<table>
<thead>
<tr>
<th>Focus Areas</th>
<th>Unsolved Problems</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Health Systems Management</strong></td>
<td>Predictive Models for Fraud Prevention, Personalized Medicine: Analytics from global benchmarks to personal life indicators, Disability analytics: tracing life events for the population (plus migration of legacy code)</td>
</tr>
<tr>
<td><strong>State Health &amp; Human Services</strong></td>
<td>Complex business rules management for eligibility across programs, Multi-state reusable solutions to control costs, improve speed to deploy: SaaS</td>
</tr>
<tr>
<td><strong>Public Health</strong></td>
<td>Heterogeneous data mining at national, state and personal levels, Data security; secure protection and sharing, Multi-platform affordability: from cloud to mobile, international to personal</td>
</tr>
<tr>
<td><strong>Military Health</strong></td>
<td>Modernization of complex legacy (Gordian knot), Data security; secure protection and sharing, Multi-platform affordability: clinician and citizen data analysis &amp; access</td>
</tr>
</tbody>
</table>
Analytics-Supported Decision Making

Management
- Planning
- Administration
- Regulation
- Legislation

Resource Production
- Workforce
- Facilities
- Commodities
- Knowledge

Organization of Programs
- Public agencies
- Private market
- Voluntary agencies
- Enterprises

Service Delivery
- Prevention
- Primary, specialty
- Secondary, Tertiary, Long-Term

Economic Support
- Private Insurance
- Social Security
- Governmental

Decisions & Actions

Health System*

Disparate Data

Surveillance
- Registries
- Community Needs

Resources
- Interventions
- Risk Factors

ED
- Outpatient
- Inpatient

Emerg Srv.
- Home Health
- Mental Health

Long-Term
- Self-Ins
- Private Ins.

CHIP
- Medicaid
- Medicare

Integrated Insights

Probable fraudulent claims
Projected Medicaid costs
Cost of care outliers

County health service utilization
Resource allocation change impacts
Market-wide expenditures

ACO performance
Practice performance
Regional health outcomes


NORTHROP GRUMMAN PRIVATE / PROPRIETARY LEVEL I
Types of Health Analytics and Markets

**Health Research**
- NIH, FDA, DOE, Pharma

**Biomedical Research - Data Standards**
- Clinical Data Interchange Standards Consortium (FDA, NIH)
- “Common Data Elements”
  Implication: Integrate conventional clinical data with imaging data and various genome-based data

**Generative Modeling (vs. Statistical Modeling)**
- Simulated data vs. real world data
- Local risk events modeled at transactional and individual level
  Implication: New models for health surveillance; shifts focus away from traditional analytic methods

**Market Size: Federal market of ~$4-6B in during the next 5 years**
(CMS, CDC, FDA, NIH)

**Medical Informatics - Patient Code Data**
- EHRs based on integrated data
- Replace ICD-9 with ICD-10
- Health Information Exchange
  Implication: Track lifetime spending by individuals, new risk management models for payers and Providers

**Clinical Data Collection**
- Integration of data across healthcare and federal systems to monitor individual and population health
- Implantable sensors automatically report health data on real-time data streams
  Implication: Big data and data management challenges; real-time decision support

**Genome-based Data**
- Pre-symptomatic diagnostics
- Personalized medicine
  Implication: Potential of predictive health

**Focus areas to drive down health costs**

**Healthcare**
- CMS, VA, DOD/MHS

**Public Health Analytics**
- Biosurveillance, geographic information systems, disease modeling, prevention, health impacts, policy formation
  Implication: Linking public health data to other types of health data provides a understanding of population health

**Public Health**
- CDC, SAMHSA, DoD, Other HHS

**Health Innovation**
- NIH, FDA, Pharma, Industry, DARPA, Other DoD, VA

**Types of Health Analytics and Markets**

NORTHROP GRUMMAN PRIVATE / PROPRIETARY LEVEL I
Integrated Health Analytics: “Analytics Fan” Layered Framework

- Data Sources
- Data Security
- Data Standards
- Need Analysis
- Data Governance

- Web Service
- Data Warehouse
- Data Cleansing
- Ontology
- Encryption/Decryption

- Predictive Modeling
- Statistical Data Model
- Semantic Integration
- Schema Matching

- Proactive fraud detection
- Care seeking behavior of populations
- Estimations of future costs
- Program Integrity

- Service cost comparisons, outliers

- Health Analytics
- Public Health Surveillance
- Systems Analyst
- Clinical Informatics

- Geospatial
- Statistical

- Policy and Process

- Subject Matter Expertise
## Conceptual Framework

| Case Management               | • Mobile Devices, Portals, Web Applications  
|                              | • Data Feeds, Workflows, Champion-Challenger  |
| Data Visualization           | • Dashboards, Displays, Pattern Recognition  
|                              | • Exploratory Data Analysis, Interactive Iteration  |
| Analytics & Integration      | • Predictive Modeling, Simulation, Neural Nets  
|                              | • NLP, CRISP-DM, Business Rules  |
| Virtual Data Layer           | • Enterprise Data Sharing & Integration  
|                              | • Data Federation, Linking, Matching  |
| Data Preparation             | • Data Quality, Cleaning, Transformation  
|                              | • Synthetic Variable Generation, Data Cubes  |
Analytics Maturity Model

**Foundational / Tactical**
- Limited data governance
- Limited quality assurance
- Analyses are typically ad hoc and reactive
- Inconsistent use of BI tools
- More detailed reports require laborious data gathering and aggregation

**Strategic Enablers**
- Formal data management exists for critical projects
- Enterprise reporting with standard BI tools is established for relevant centrally controlled data sources
- Decision makers still depend on data mining specialists for more detailed information
- Advanced analytics and predictive models periodically available to provide decision support

**Highly Strategic**
- Holistic systems approach to data governance
- Automatically available analyses of key performance indicators
- Power users can run additional ad hoc queries and reports
- Data mining system allows users to apply analytical tools without deep expertise
- Continuous real-time monitoring and alerts with drill-down capabilities
- Rich visualization tools
- BI integrated with business process management in a closed-loop to improve results

Provide an improved path to evidence-based decision-making
Development Process via CRISP-DM

Integrated Health Analytics Lifecycle*

Business Understanding
- Determine business objectives
- Identify desired insights
- Assess environments
- Form project plan

Data Understanding
- Review data sources
- Verify data quality and completeness
- Form analytics plan
- Identify needed reference arch elements

Platform & Data Prep
- Construct tailored platform
- Access data sources
- Preprocess data
- Format and integrate data

Exploration
- Apply analytics techniques
- Generate initial insights
- Describe findings

Evaluation
- Evaluate results
- Assess alignment with business objectives
- Plan for ongoing access
- Determine next steps

Production
- Add new analytic views
- Sustain platform
- Monitor and maintain data source access

*Adapted from: Cross Industry Standard Process for Data Mining (CRISP-DM), Visual Guide by Nichole Leaper
Predictive Analytics Capabilities

• Provide flexibility to work with pre-existing architecture as well as new architectures

• Reduces costs and time for integration among different data sources

• Offers robust analytics, visualizations and reporting customized to customer needs managing “big data”

• Cuts operational costs

• Generates resources and support for evidence-based decision-making within “big data”

Partnering opportunities provide a win-win situation for Northrop Grumman and its partners.
Data-driven Modeling for Health Condition, Service, and Population Costs

Average payment per beneficiary in age groups for state, regional, and national levels ➔ Identify high-cost sub-populations

Variations among nearby states in average healthcare costs by disease conditions (diabetes, heart disease, stroke) ➔ Locations with high risk groups as well as positive results

Contrast average healthcare costs in nearby states accounting for number of risk factors (disease burden) ➔ Impact of multiple risk factors on healthy population

Trends in number of child Medicaid beneficiaries over last 10 years ➔ Increased eligibility & reimbursement requirements for ARA

State differences in average healthcare costs by number of risk factors (disease burden) ➔ National trends in cost impact of risk factors
Fraud Predictive Analytics Visualizes Suspicious Information for Investigation

**Anomaly Detection**: Payment per Medicare beneficiary by hospital type of service code ➔ Identify services and individual cases with extreme values

**Cluster Analysis**: Clusters of high average costs vs. low average costs in Medicare patients ➔ Investigation of patient groups & procedures

**Predictive Modeling**: Predicting number of child Medicaid beneficiaries from last 10 years ➔ Increased eligibility & reimbursement requirements for ARA
Hospital Acquired Infections (HAI) Data

- Two million patients/yr acquire HAI’s – approx $28-$33B dollars in healthcare spending
- HHS goal to reduce HAIs by 25-50% in 5 yrs.
- RADC conducted exploratory analysis of HAI data from the National Hospital Discharge Survey
- Found that HAI reporting significantly increased over the study period (regulations and acceptability increasing)
- Identified data that supported known conditions:
  - Longer hospital stays lead to increased likelihood of infection
  - Hospital size scales concordantly with hospital stay
- Showed CMS customer that over 50% of HAI costs were paid by Medicare and Medicaid
Fraud Analytics Workstation: Anomaly Detection

- User interface for FAW – used to demonstrate different fraud scenarios
- Outliers Detection tab connects to SAS product for identifying anomalies
- Using CMS PUF of over 9.7 million rows of claims data sample from 2008.
- Subset of claims by ICD-9 coding for diabetics.
- Identifies the high cost outliers for different type of service codes
- Several kinds of charts can be output for user.

Key Point
High cost outliers for specific types of service codes are identified among diabetic claims,
Medicaid Eligibility Projections*  

\[
\text{Summary of Fit} \\
\text{RSquare} \quad 0.781461 \\
\text{RSquare Adj} \quad 0.754144 \\
\text{Root Mean Square Error} \quad 1482.604 \text{ (1)} \\
\text{Mean of Response} \quad 82504.9 \text{ (1)} \\
\text{Observations (or Sum Wgts)} \quad 10 \text{ (1)} \\
\]

\[
\text{Analysis of Variance} \\
\text{DF} \quad 6 \quad 8 \quad 9 \quad \text{Sum of Squares} \quad 62880881 \quad 2198114.5 \quad 80465797 \quad \text{Mean Square} \quad 28.6067 \quad 0.0007^* \quad 0.0007^* \\
\text{F Ratio} \quad 0.0007^* \\
\text{Prob} > F \quad 0.0007^* \\
\]

\[
\text{Parameter Estimates} \\
\text{Term} \quad \text{Estimate} \quad \text{Std Error} \quad \text{t Ratio} \quad \text{Prob}>|t| \\
\text{Intercept} \quad 1831633.3 \quad 5.60 \quad 0.0005^* \\
\text{YEAR} \quad -873.03636 \quad 5.35 \quad 0.0007^* \\
\]

*Excludes expansion population
This user interface tab shows a flash file of a bubble chart that displays the percent of Medicaid eligibles and percent of population on SNAP (food stamps) over time.

Bubbles float to show changes: population, percentage of SNAP recipients as well as percentage of Medicaid eligibles over time for the counties shown.
# Applied Predictive Analytics: In Process

<table>
<thead>
<tr>
<th>Macro-Level Research</th>
<th>Member &amp; Patient Analytics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identification of Avoidable Expenses</td>
<td>Gaps in care</td>
</tr>
<tr>
<td>Population-based Analysis</td>
<td>High ED Utilization</td>
</tr>
<tr>
<td>Geographic-based Analysis</td>
<td>Unfilled Prescriptions</td>
</tr>
<tr>
<td>Cost and Performance Trends</td>
<td>High Risk Members</td>
</tr>
<tr>
<td>Procedural Effectiveness</td>
<td>High Prescription Utilization</td>
</tr>
<tr>
<td>Preventative Campaign Effectiveness</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Measures and Benchmark Reports</th>
<th>Patient Centered Medical Home Analytics</th>
<th>Quality Analytics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identification of Avoidable Expenses</td>
<td>Single Patient Visit Report</td>
<td>HEDIS Measures</td>
</tr>
<tr>
<td>Cost Measures</td>
<td>Prioritized Patient Panel Report</td>
<td>Affordable Care Act Measures</td>
</tr>
<tr>
<td>Quality Measures</td>
<td>Complete Patient Panel Report</td>
<td>AHRQ Measures</td>
</tr>
<tr>
<td>Meaningful Use Reporting</td>
<td>Non-engaging Patient Report</td>
<td>Bayou Health Measures</td>
</tr>
<tr>
<td>Key Performance Indicators</td>
<td>Population Performance Report</td>
<td></td>
</tr>
<tr>
<td>Operations Reporting</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hospital Average Length of Stay</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hospital Readmission Rate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hospital Infection Rate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Procedure Effectiveness</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cost per Incidence of Care</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Evidence-Based G/L Compliance</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Section 4241 of the Small Business Jobs Act of 2010 (SBJA) mandates that CMS implement a predictive analytics system to analyze Medicare claims to detect patterns that present a high risk of fraudulent activity, and enables CMS to employ real-time, pre-payment claims analysis to identify emerging trends of potentially fraudulent activity.
FPS Operation Flow
National Fraud Prevention Program
Two-pronged Approach

Take quick administrative action to prevent improper payments

Take quick action to remove bad actors from Medicare

Identify bad actors and prevent them from enrolling in Medicare

Take quick action to remove bad actors from Medicare
Advantages of Fraud Prevention Program

- Prevent the payment of claims that have been identified as potentially fraudulent
- Integrate predictive modeling as part of an end-to-end solution that triggers effective, timely administrative actions by CMS
- Assure that analytics are effective (minimize false positives), risk-based, and efficient
- Real-time access to high quality data
- Integration of multiple data sources (e.g., complaint data, compromised numbers)
- Understand patterns of care and utilization
- Understand payment errors and program vulnerabilities
- Identify areas for policy changes and provider education
- Allow interventions to be made before issues grow
Limitations and Challenges for Predictive Modeling

- Need accurate data to validate predictions.
- Access to certain variables limits measurement.
- Linking data is challenging.
- Correlation is not causation. Lots of confounding variables can get in the way.
- Difficult to predict from one group or situation to another environment.
- Experiments and trials to test relationships can be expensive and lengthy.
Personalized Medicine: A “Game Changing” Focus for Health

Uses information about a person's genes, proteins, and environment to prevent, diagnose, and treat disease – this includes such areas as genomics, proteomics, and metabolomics (“omics”)

Paradigm shift that moves the current medical model of being reactive to disease to one that is more proactive for treatment and prevention

What will enable this revolution in healthcare?

• **Rapid reduction in cost** of genomic & proteomic analysis to become clinically affordable

• **Massive amounts of analyzed data** will help reduce cost of developing new cures, and allow better targeting of treatment/lifestyle recommendations to keep people healthier

• **Large, secure, reliable national Health IT systems** to support widespread clinical value and use

“Omics” discovery, development, and translation will be one of the major disruptive health advances of the 21st Century
Predictive Analytics

• Explosive data is a great asset and necessary
  – But, it hampers data integration, quality, standardization

• Data Scientists growth in developing comprehensive rules and algorithms

• The maturation of predictive analytics is creating the visibility to enable
  – Improved Health Outcomes
  – And Reduced Costs