Predictive Analytics for Demand Forecasting and Planning Managers – A Big Data Challenge

Hans Levenbach, Delphus, Inc.

ISF2013, KAIST College of Business, Seoul, Korea

Agenda

- Role of Big Data in 'Small" Predictive Analytics
  - Big Data Applications
    - Hospital Patient Care Activity: Geography, Payer Class & Product Lines
    - Multiple Competitor & Product Mix Forecasting
    - Multi-State & Migration Perspectives
    - Physician Activity & Multiple Physician Ranking
    - Hospital Closing and Merger Simulations
  - A Big Data and Reporting Framework For Hospital Management
- A Modeling Pathway For Demand Forecasters/Planners – The PEER Process
  - **P**reparing Healthy Data
  - **E**xecuting Analytic Modeling Methodologies
  - **E**valuating Performance Diagnostics
  - **R**econciling Multiple Modeling Pathways
- .

Big Data Usage Examples

(in petabytes)

- The world's effective capacity to exchange information through two-way telecom networks was 281 petabytes of (optimally compressed) information in 1986, 471 petabytes in 1993, 2,200 petabytes in 2000, and 65,000 (optimally compressed) petabytes in 2007
  - ***this is the informational equivalent to every person exchanging 6 newspapers per day***!

- **Internet**: Google processes about 24 petabytes of data per day
- **Telecoms**: AT&T transfers about 30 petabytes of data through its networks each day
- **Physics**: The experiment in the Large Hadron Collider produce about 15 petabytes of data per year, which will be distributed over the LHC Computing grid
- **Neurology**: It is estimated that the human's brain's ability to store memories is equivalent to about 2.5 petabytes of binary data
- **Climate science**: The German Climate Computing Centre (DKRZ) has a storage capacity of 60 petabytes of climate data[1]
- **Archives**: The internet archive contains about 5.8 petabytes of data as of December 2010. It was growing at the rate of about 100 terabytes per month in March 2009
- **Film**: The 2009 movie Avatar is reported to have taken over 1 petabyte of local storage at Weta Dogital for the rendering of the 3D CGI effects
  - In August 2011, IBM was reported to have built the largest storage array ever, with a capacity of 120 petabytes
  - In January 2012, Cray began construction of the Blue Water Supercomputer, which will have a capacity of 500 petabytes making it the largest storage array ever if realized!

How Big Is a Petabyte?

A petabyte (derived from the SI prefix *peta*) is a unit of information equal to one quadrillion bytes, or 1024 terabytes.

The unit symbol for the petabyte is PB. The prefix *peta* (P) indicates the fifth power to 1000.

According to IBM: Everyday, we create 2.5 quintillion bytes of data– so much that 90% of the data in the world today has been created in the last two years alone

Multiples of bytes V·T·E				
SI decimal prefixes		Binary usage	IEC binary prefixes	
Name (Symbol)	Value		Name (Symbol)	Value
kilobyte (kB)	10^3	2^{10}	kibibyte (KiB)	2^{10}
megabyte (MB)	10^6	2^{20}	mebibyte (MiB)	2^{20}
gigabyte (GB)	10^9	2^{30}	gibibyte (GiB)	2^{30}
terabyte (TB)	10^{12}	2^{40}	tebibyte (TiB)	2^{40}
petabyte (PB)	10^{15}	2^{50}	pebibyte (PiB)	2^{50}
exabyte (EB)	10^{18}	2^{60}	exbibyte (EiB)	2^{60}
zettabyte (ZB)	10^{21}	2^{70}	zebibyte (ZiB)	2^{70}
yottabyte (YB)	10^{24}	2^{80}	yobibyte (YiB)	2^{80}
See also: Multiples of bits · Orders of magnitude of data				

Predictive Analytics For Healthcare Applications

Predictive analytics encompasses a variety of techniques from data mining, statistics, and game theory that analyze current and historical facts to make predictions about future events.

"Although healthcare industry has lagged behind sectors like retail and banking in the use of big data – partly because of concerns about patient confidentiality – it could soon catch up"

McKinsey Company

We have to bring the science of management back into healthcare"

Donald Berwick, MD
Institute of Healthcare Improvement

Step 1: Preparing Healthy Data: Centralized Data Source for NY Hospital Applications - SPARCS

The annual OUTPATIENT.ZIP file from SPARCS is 700 MB s and expands to 18 gigabytes in a text file format.

The annual INPATIENT.ZIP file from SPARCS is 500 MB and expand to 7 GB in a text flat file format.

- How many hospitals?: New York 268, Connecticut over 30, New Jersey over 140
- How many patient records per year?: NY inpatient Records over 2.5 million
- How many lowest level records in the demand table if we had all hospitals SQL with SPARCS data?

Typical Database Sizes in Supply Chain Organizations

Wal-Mart has 5,000 Stores, 100,000 items, and if a typical store carries a full line of items, then you can then potentially expect

500,000,000 lowest level records

If an online Store has 6 Distribution Centers and 35,000 Items, and each distribution center carries a full line of items, one can then expect

210,000 lowest level records

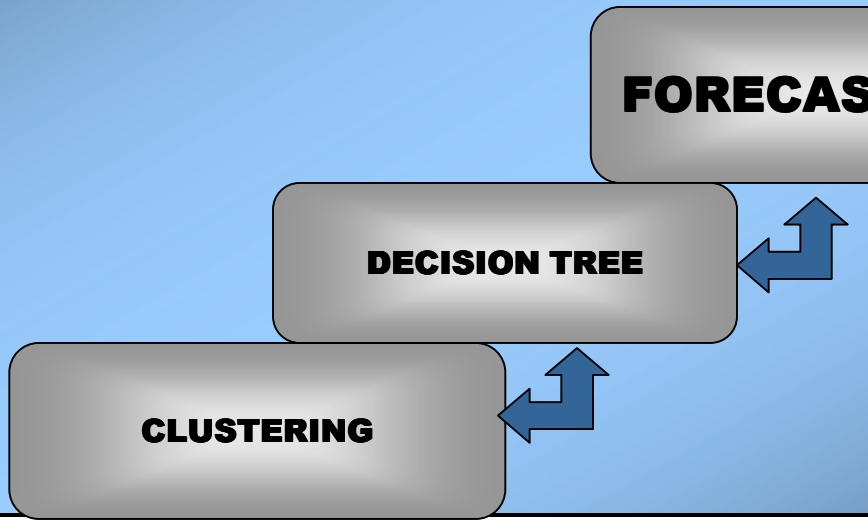
For NY Hospitals, with 5 years of history and approximately 30 competitors, 15 geographic areas, 4 financial classes, 6 services and 23 products we end up with 220,000 lowest level records in the demand table

Step 2: Executing Analytic Modeling:Methodologies

Predictive Analytic Methods Tend To Be Data-Driven

The use of current and past data, in conjunction with statistical, structural or other analytical models and methods, to determine the likelihood of certain future events

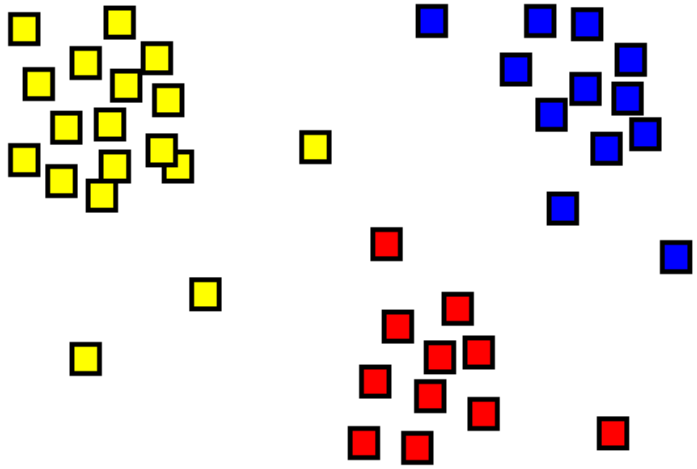
As you move up the steps, the complexity of the approaches increases while the volume of data decreases



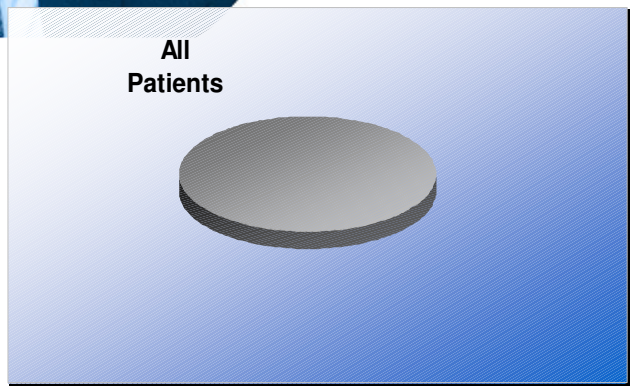
Data framework tends to be structured and connected as DATA QUALITY drives credibility in modeling results

Big Data Techniques
Cluster Analysis

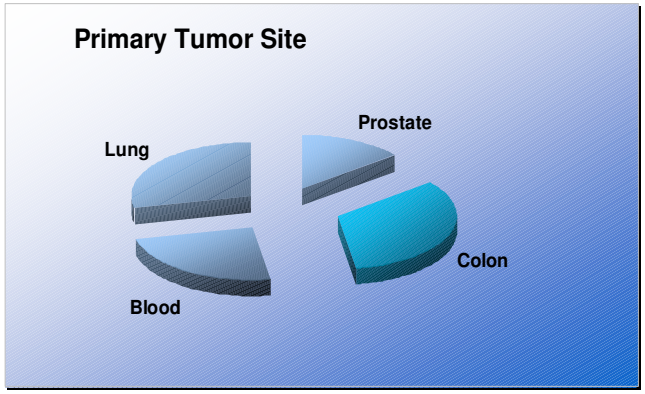
Cluster analysis or clustering is the assignment of a set of observations into subsets (called clusters) so that observations in the same cluster are similar in some sense. Clustering is a method of unsupervised learning, and a common technique for statistical data analysis used in many fields, including machine learning, data mining, pattern recognition, image analysis and bioinformation.

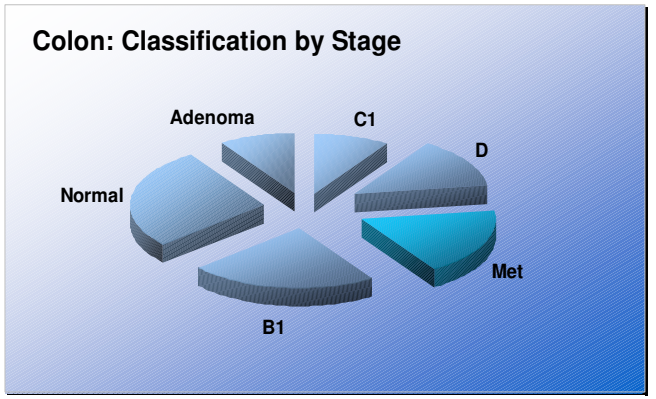


Decision Trees:
Progressive Class Distinction



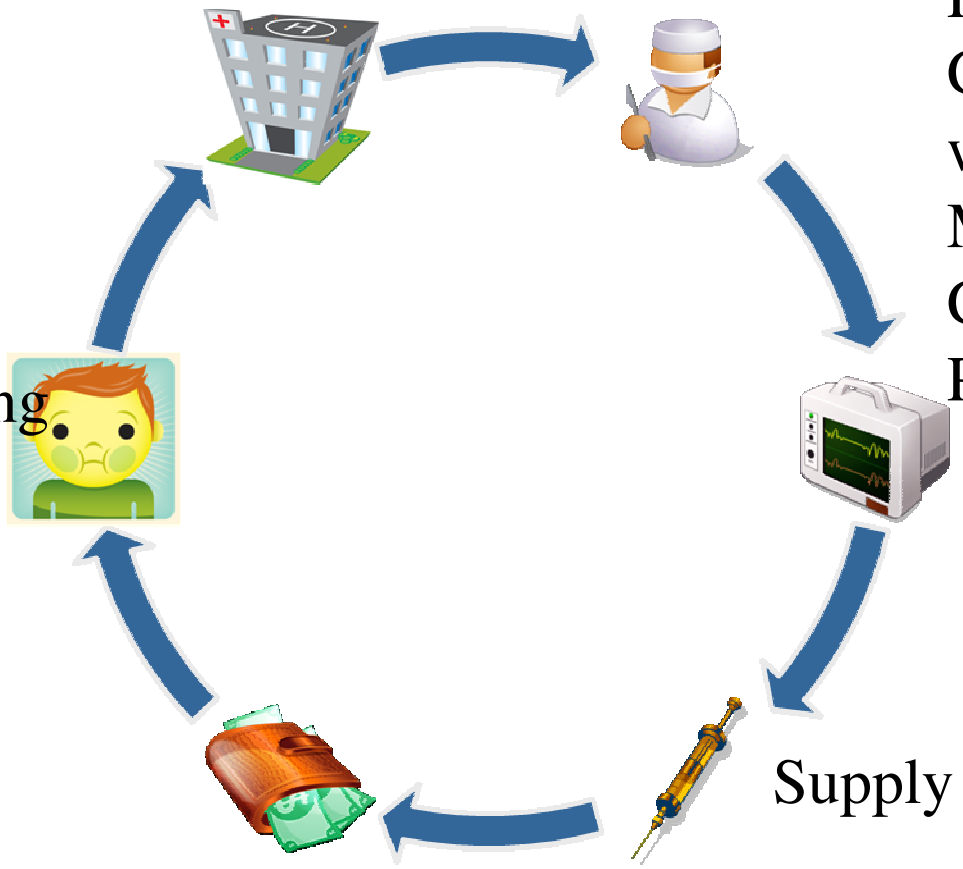
→ Descend tree using any available differentiating attributes; natural, derived or inferred; separately or in combination





→ > 100 product line/service combinations

Patient-Centric Demand Forecasting For Hospital Management



Patient Product/Service Demand Planning

Patient Classification with Co-Morbidities & Clinical Pathways

Supply Components

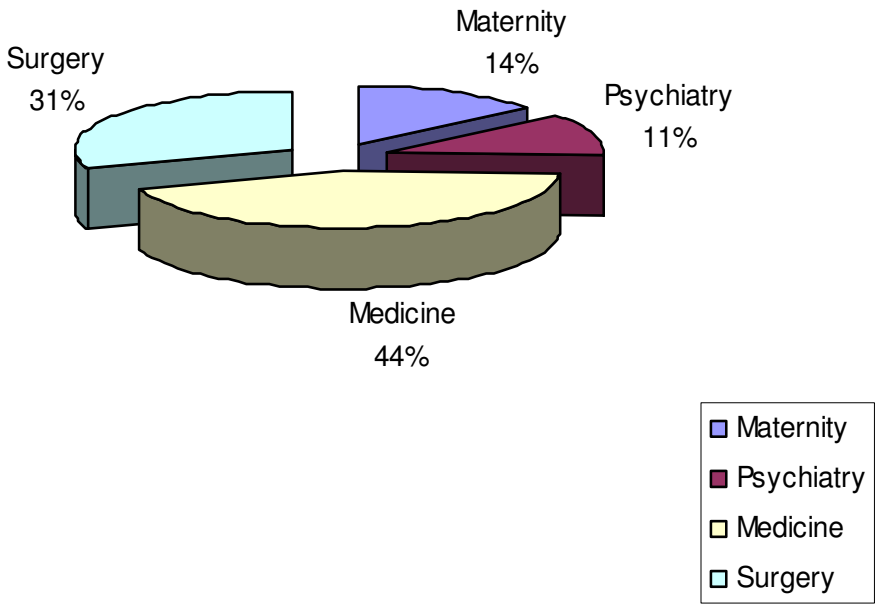
Applications in Healthcare

Understanding medical prevalence, we can predict Admissions --> Co-Morbidities --> Physician Work Flow --> Product/Service Mix --> Supply Chain requirements

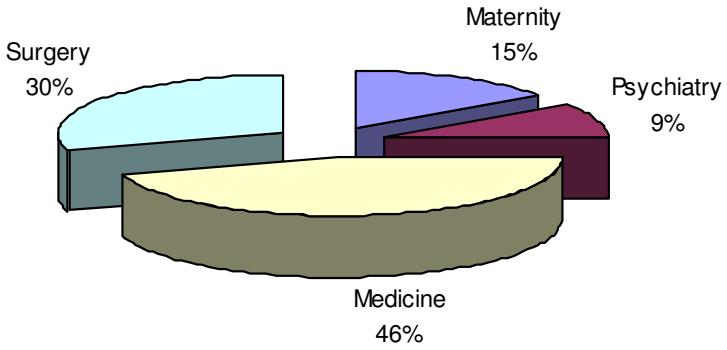
- Patient Care Activity: Geography, Payer Class & Product Lines
- Multiple Competitor & **Product Mix Forecasting**
- Physician Activity & Multiple Physician Ranking
- Multi-State & Migration Perspectives
- Berger Commission Type Simulations – e.g. Hospital Closings and Mergers

Product Mix and Service Need

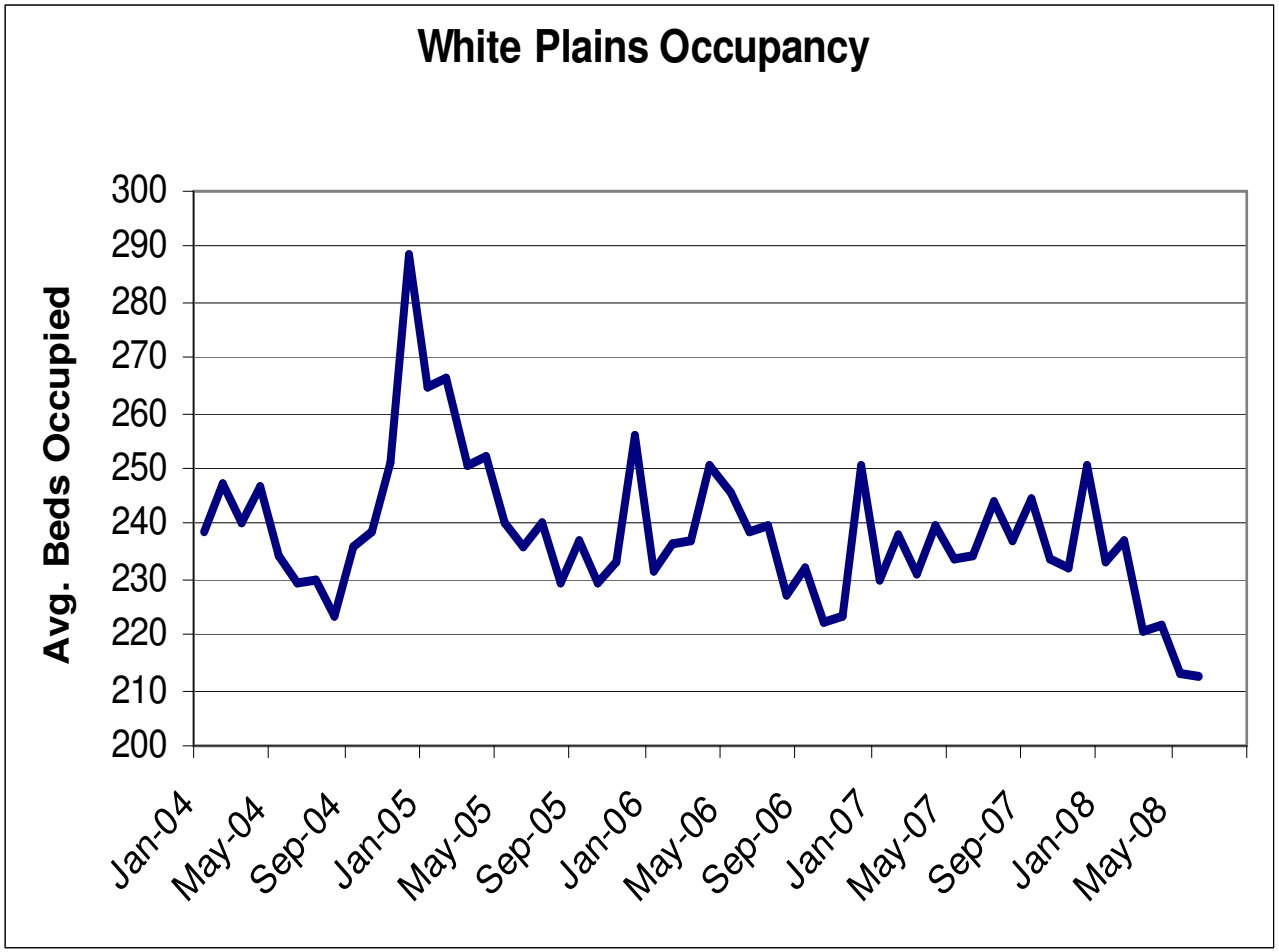
Service Mix for First Year



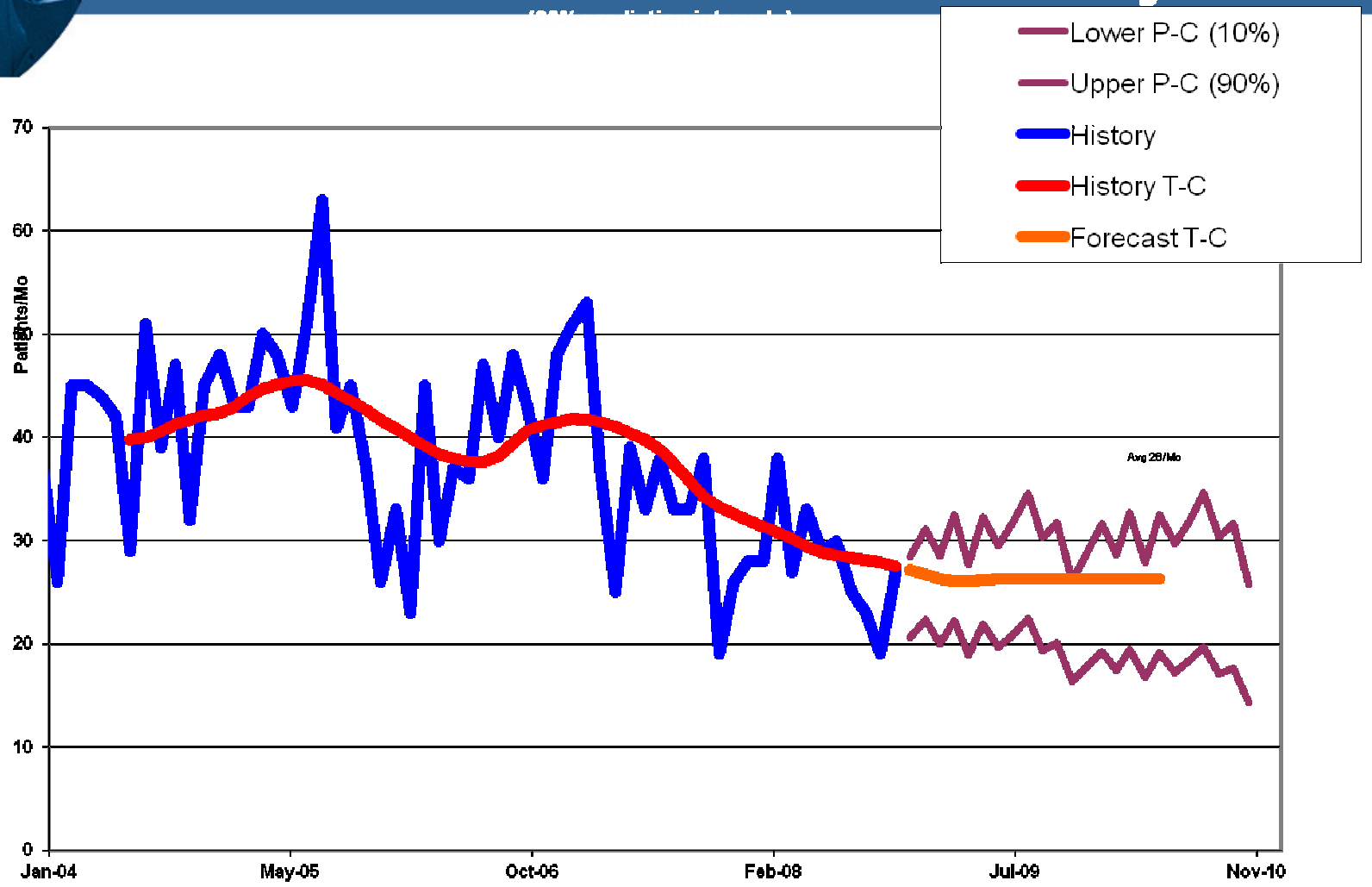
Service Mix for Forecast Year



"Replenishment Planning"

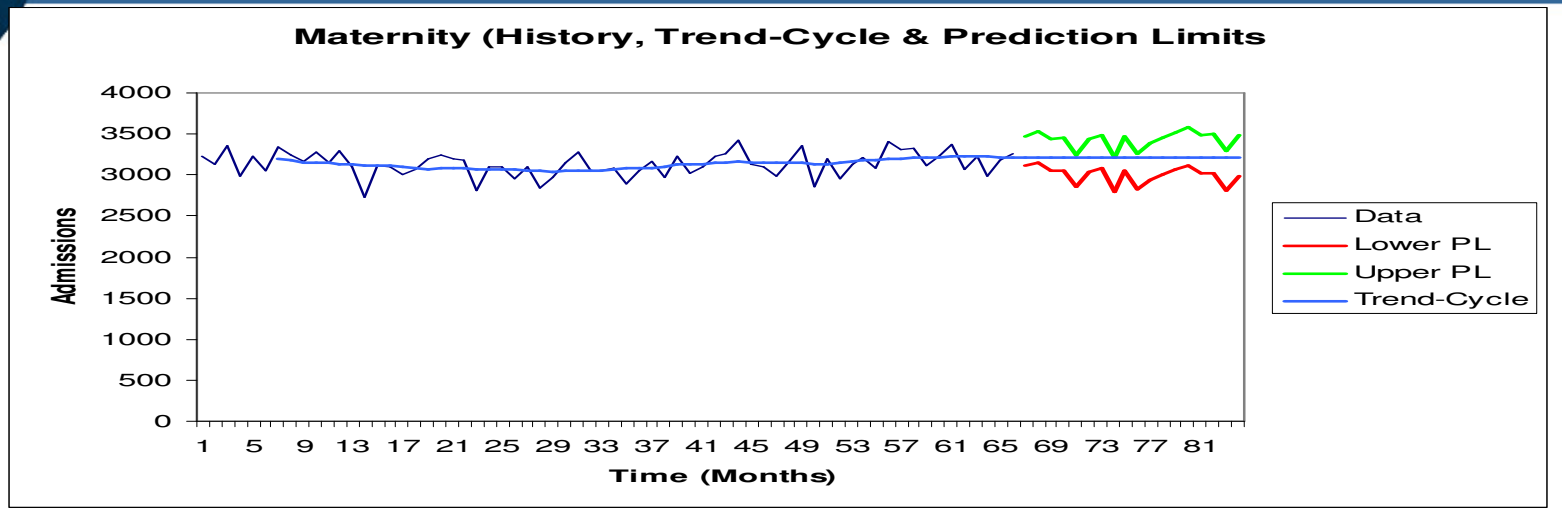


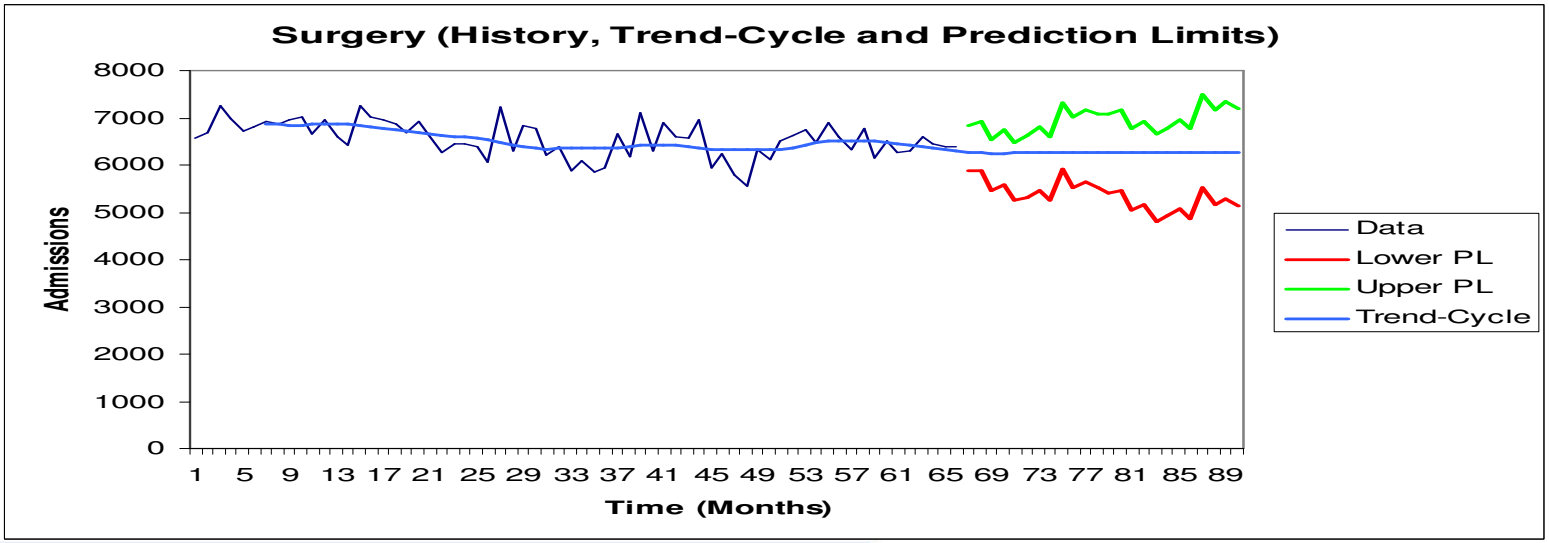
Senior Partner In-Patient Activity



Step 3: Evaluating Performance Diagnostics

Predictive Analytics: Forecasting Service Needs





Predictive Analytics
Development of Methods

The use of current and past data, in conjunction with statistical, structural or other analytical models and methods, to determine the likelihood of certain future events

Forecasting and Strategic Planning methods cover a range from relatively simple time series models to more advanced techniques such as simulation and advising

REPORTING & ADVISING

SIMULATION & SCENARIO PLANNING

FORECASTING

HL/PS, ISF2010
San Diego, CA

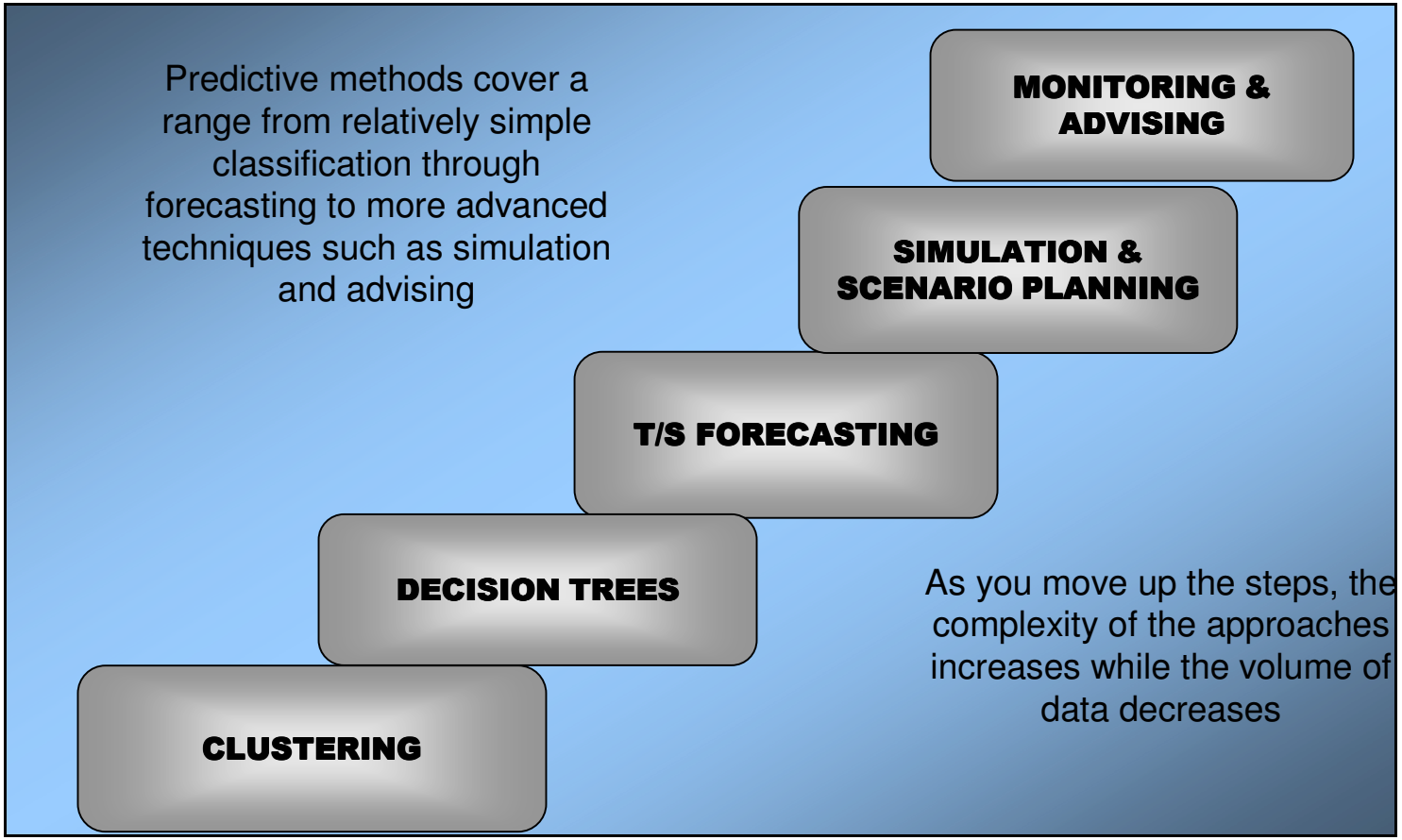
Step 4: Reconciling Multiple Modeling Pathways

Predictive Analytics - Complexity

Data Scale versus Complexity of Methods

The use of current and past data, in conjunction with statistical, structural or other analytical models and methods, to determine the likelihood of certain future events

Predictive methods cover a range from relatively simple classification through forecasting to more advanced techniques such as simulation and advising



As you move up the steps, the complexity of the approaches increases while the volume of data decreases

A Modeling Framework for Hospital Management
(In the Cloud Software-as-a-Service Model)

Data Resource

Multi-year Multi-hospital Information

Mining & Analytics

Product Line Trends

Spatial Analysis

Market Shares

Budget

PEER Planner Dashboard

Forecasting &

Collaborative Planning

Competitive Simulation Model

Interactive

Implications for Researchers and Practitioners

Demand Forecasting Research Opportunities

- Curing Unhealthy Structured Data
  - Intermittent
  - Missing
  - Transitioning new product intros
  - Consistent, . . .
- Demand Modeling with Structured Relational Datasets
  - Product/Service Hierarchies – 23 Product & 6 Service Summary Levels
  - Place Segmentations – Geo Regions, Payor Class,
  - Period Granularities – Months, Weeks, Days, Hours
- New Challenges for Forecasting Decision Support Systems – scalable methods for 'very large' datasets

Questions?

- Contact: Hans Levenbach
- Email: hlevenbach@delphus.com
- URL: www/delphus.com
- and www.cpdftraining.org