



# Exploration and Visualization of Post-Market Data

Jianying Hu, PhD

Joint work with

David Gotz, Shahram Ebadollahi, Jimeng Sun, Fei Wang, Marianthi Markatou

Healthcare Analytics Research

IBM T.J. Watson Research Center, Hawthorne, NY

Contact: [jyhu@us.ibm.com]

## Disclosure Information

FDA Public Workshop on Study Methodology for Diagnostics in the Post-Market Setting  
May 12, 2011

Jianying Hu, Ph.D., IBM T. J. Watson Research Center

I have no financial relationships to disclose

I will not discuss off label use and/or investigational use in my presentation

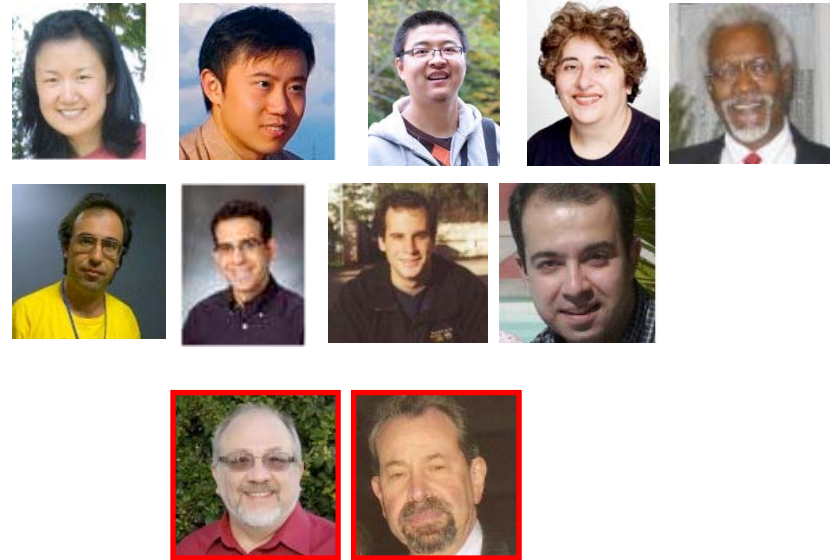
# Healthcare Analytics Research Group

## **Mission:**

- Enable and provide the I/T and scientific underpinnings for Healthcare Transformation by conducting world class research in the area of health informatics.

## **Areas of Expertise:**

- Data Mining, Machine Learning, Pattern Recognition
- Optimization, Simulation
- Bio-statistics
- Health Economics
- Multimedia Content Analysis
- Multimedia Information Visualization
- Visual Analytics
- Medicine



## Evidence & Insight Generation

- Patients' Similarity
- Comparative Effectiveness & Outcomes Research
- Disease Modeling
- Predictive Analytics

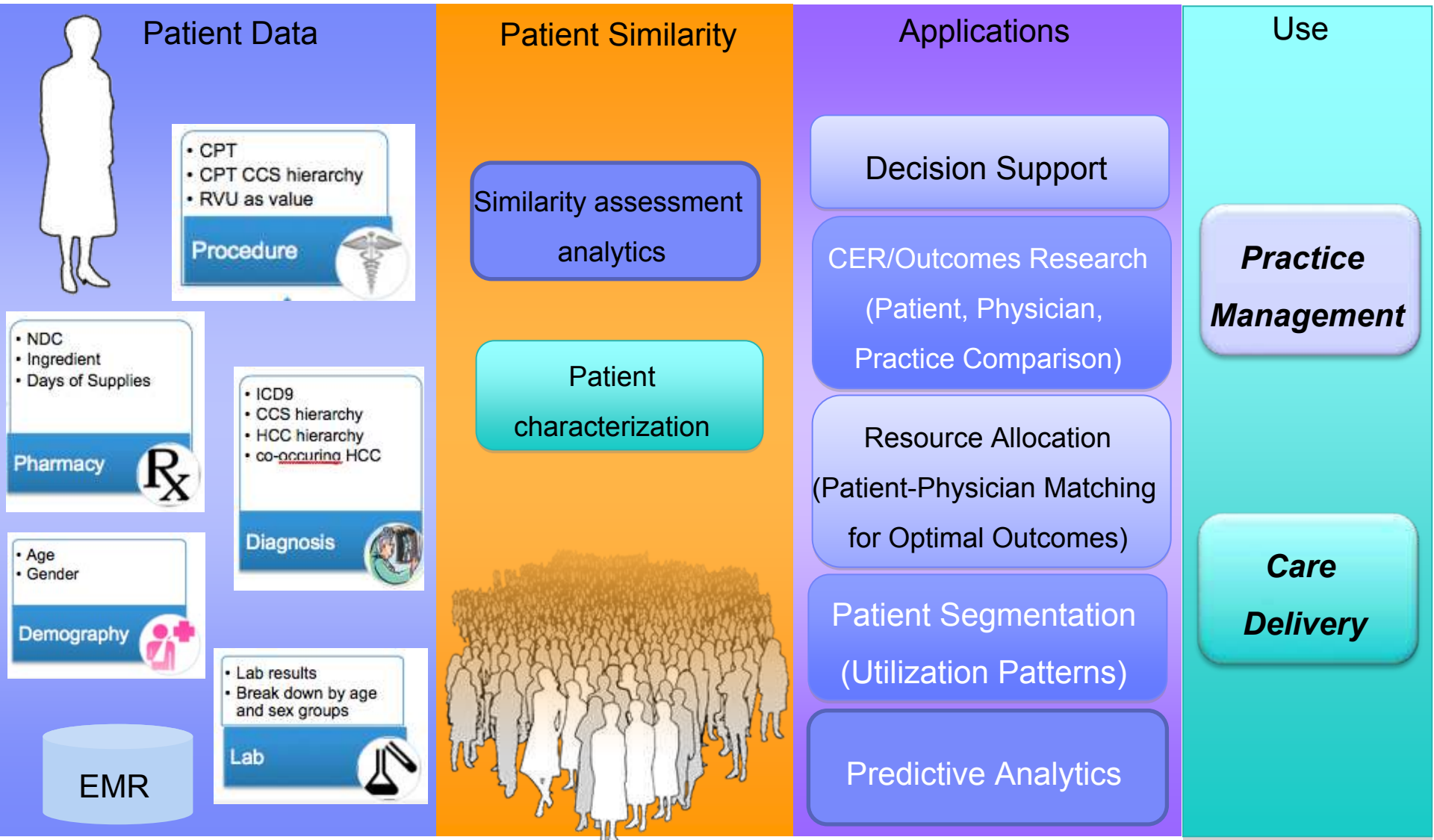
## Evidence & Insight Use

- Longitudinal Patient Data Presentation
- Context-aware Patient Information Selection & Visualization
- Visual Analytics

## Health Policy & Payment

- Evidence-Informed Outcome-Based Payment
- Simulation for Payment

# Analytics for Personalized Decision Making - Overview



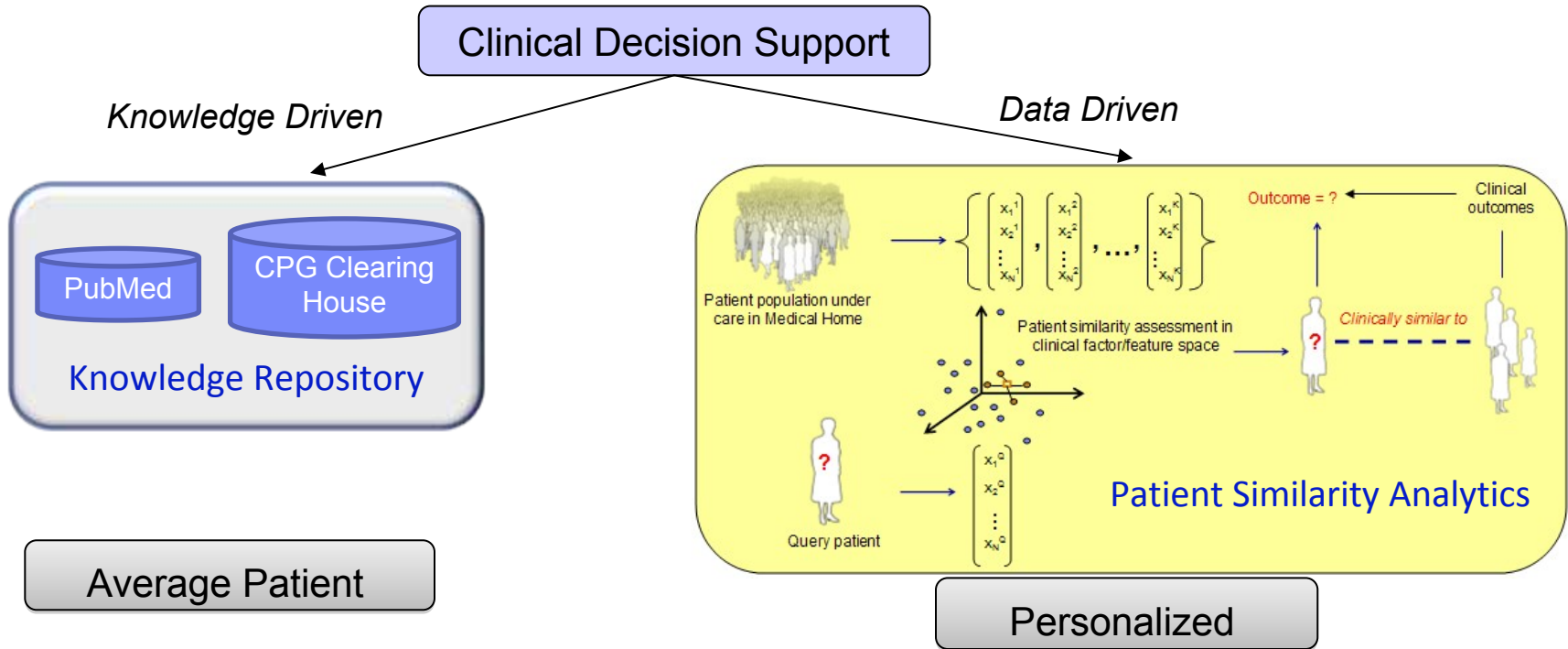
# Patient Similarity Assessment

## Objective

Given an index patient, find clinically similar patients for decision support and Comparative Effectiveness

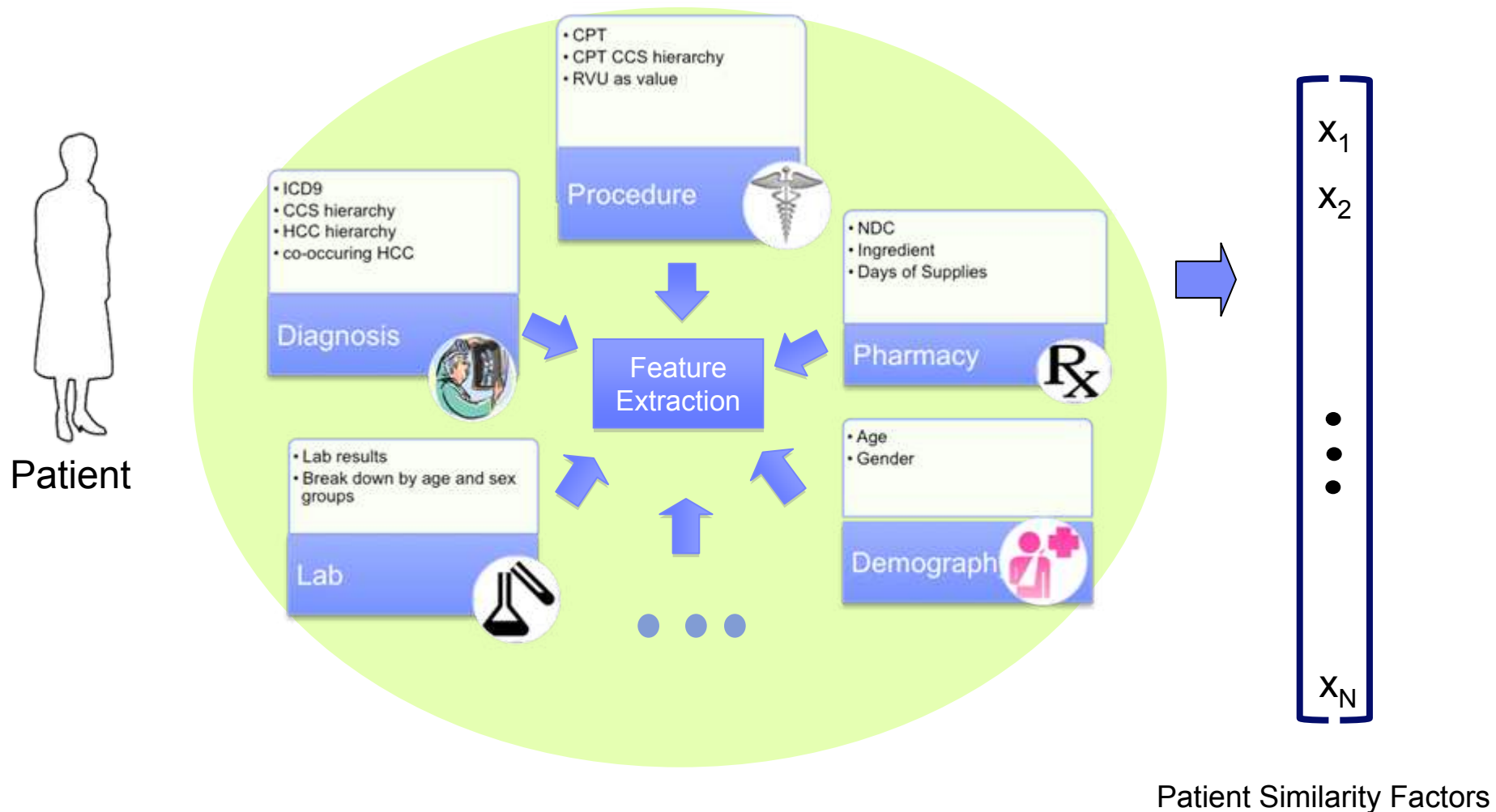
## Value

- Provider: personalized clinical and care management decision support based on insights most relevant to index patient
- Payer: intelligent patient cohort segmentation for comparison and evaluation of treatments, practices and physicians





# Representing Patients using Information Obtained from Multiple Sources of Data



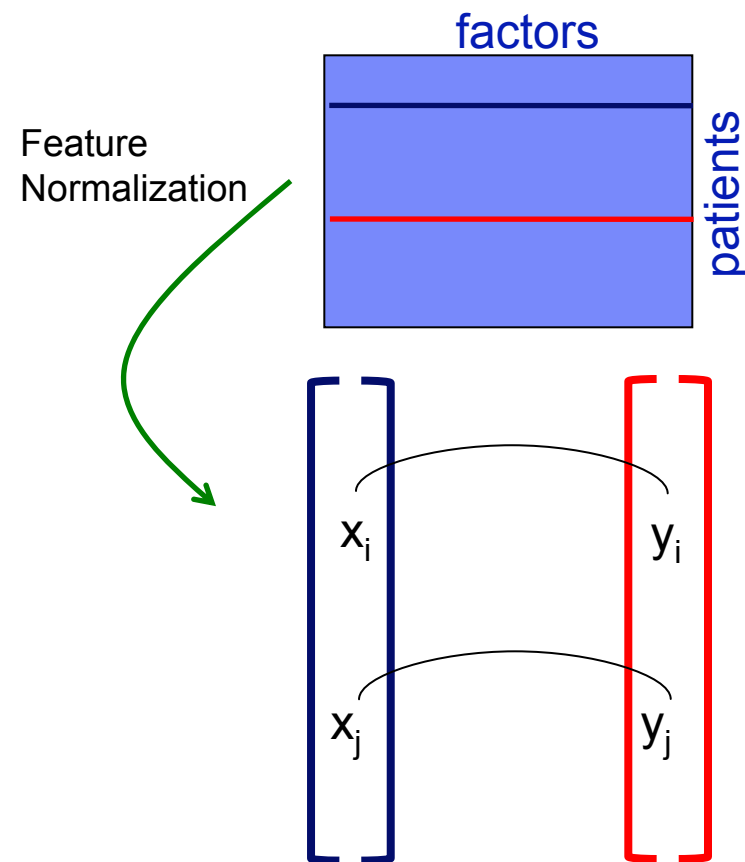
# Similarity Metric Incorporating Multiple Sources

## How to combine different categorical features?

- Feature normalization
  - ▶ Start with the number of occurrences of the factor in the current patient
  - ▶ Adjust using number of occurrences of that factor in whole population
  - ▶ Give higher weight for matching on rare factors and lower weight for matching on common factors

## How to enable flexible matching?

- Categorical attributes such as ICD9, CPT, NDC are converted into hierarchical representation
- Similarity scores are measured at multiple level of the hierarchy in order to enable flexible matching on the categorical values



$$S = \sum_{i=1, \dots, K} x_i \cdot y_i$$

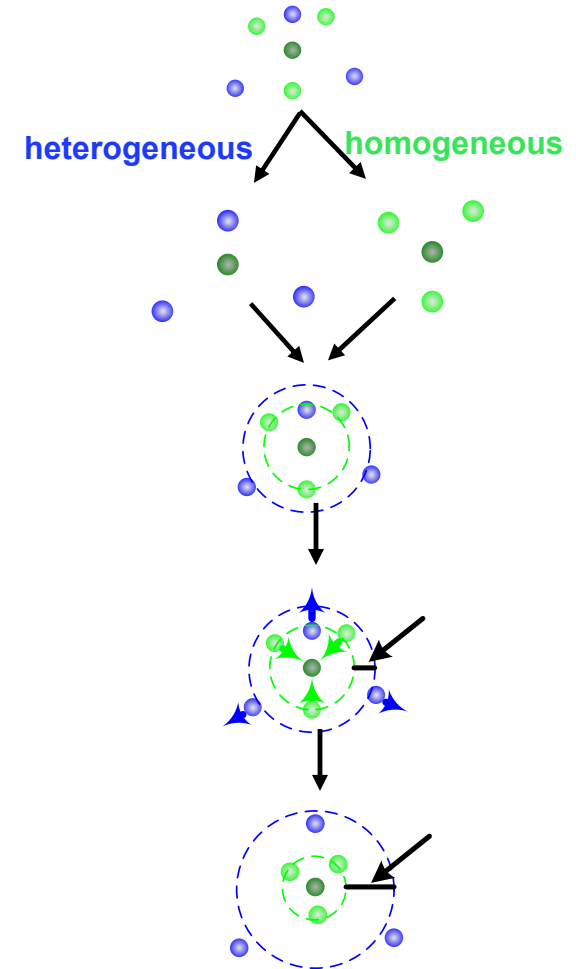
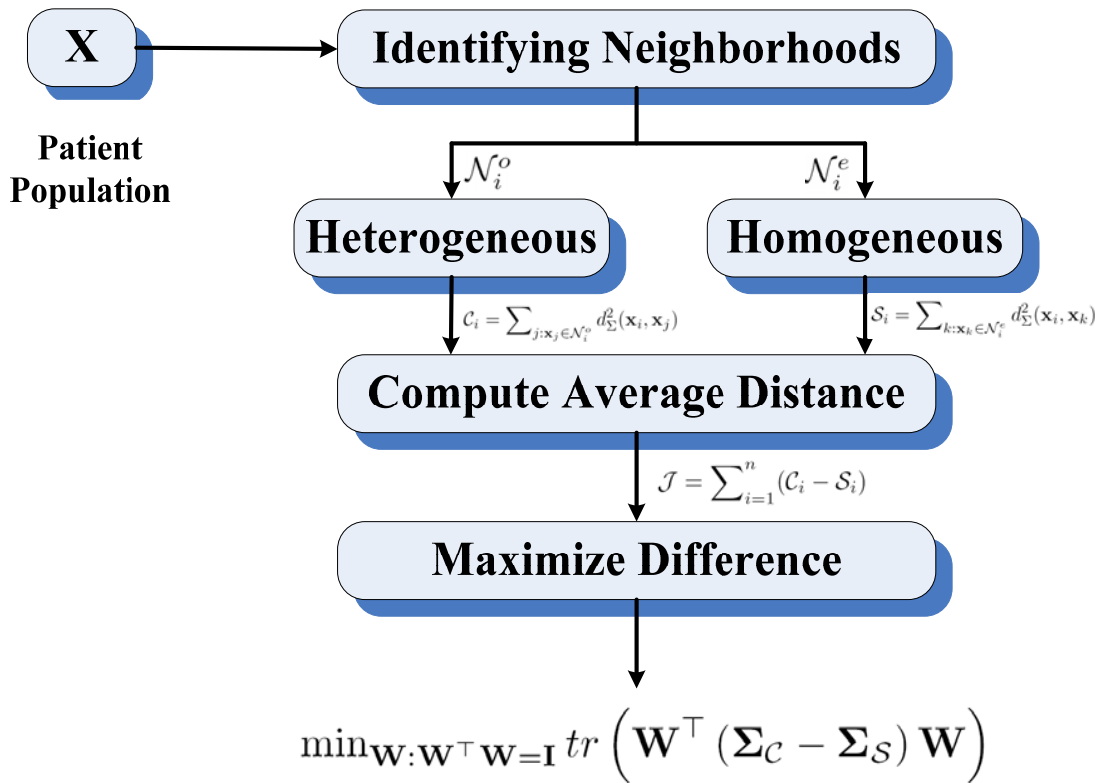
Baseline Metric: factors combined using expert defined weights  
Customized Metric: context and end point specific distance metric

# Customized Distance Metric: Locally Supervised Metric Learning (LSML)

- Goal: Learn a generalized Mahalanobis distance :

$$d_{\Sigma}(x_i, x_j) = \sqrt{(x_i - x_j)^{\top} \Sigma (x_i - x_j)}$$

$$\Sigma = WW^{\top}$$



Applied to: near term prognostication



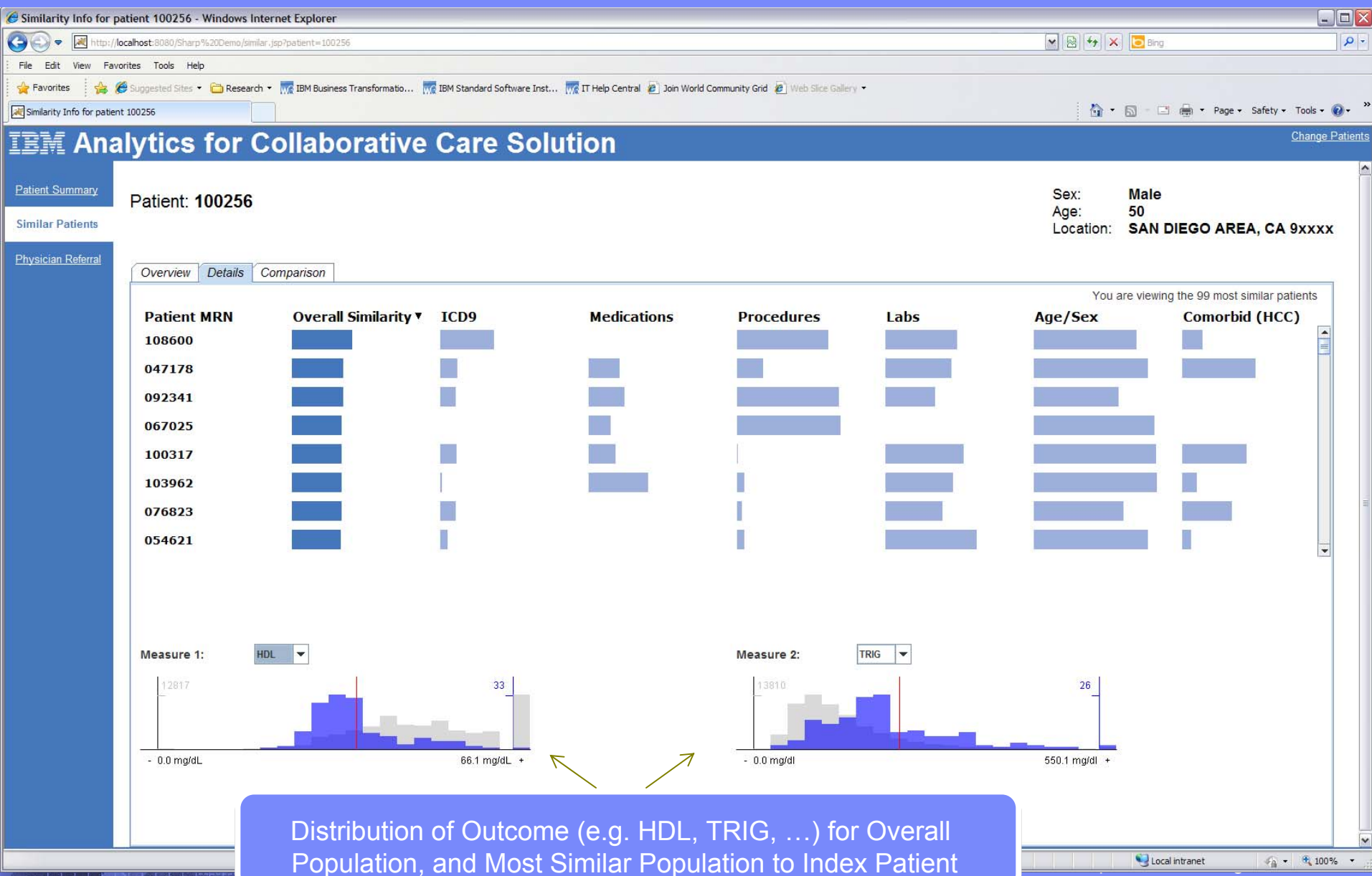
## Visualization of Health Data

- Analytics designed to produce **actionable** insights
- Actions often taken by people
  - **Doctors**
  - **Patients**
  - **Administrators**
- Visualization can help people **consume** and **refine** results of analytics
  - Provides intuitive presentation for quick comprehension by domain experts
    - Little or no statistical training
    - Users are decision or task oriented
  - Can address many data challenges
    - Large volume of data
    - High-dimensional data
    - Poor semantics
      - e.g., “Why are these patients clustered together?”





# Visualization: More Detailed View of Similarity & Distributions of Outcomes



# Visualization: Treatment Outcome Comparison over Similar Patients

Patent Summary

Similar Patients

Physician Referral

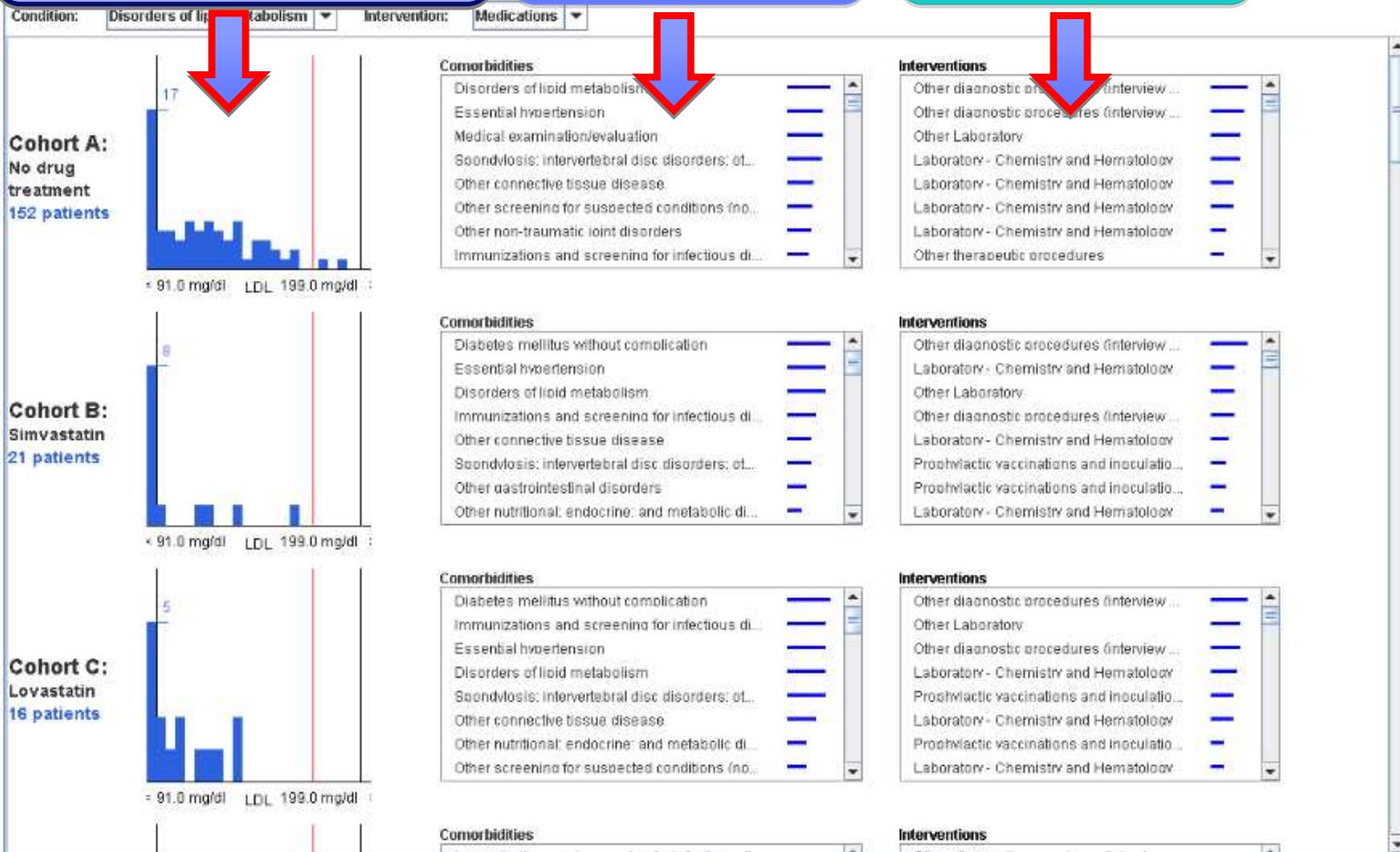
Patient: 407-725-614-522

Sex: Female  
Age: 56  
Location: 7

Distribution of outcomes under Different treatment regimens

Distribution of patient's co-morbidities

Distribution of patient's other interventions





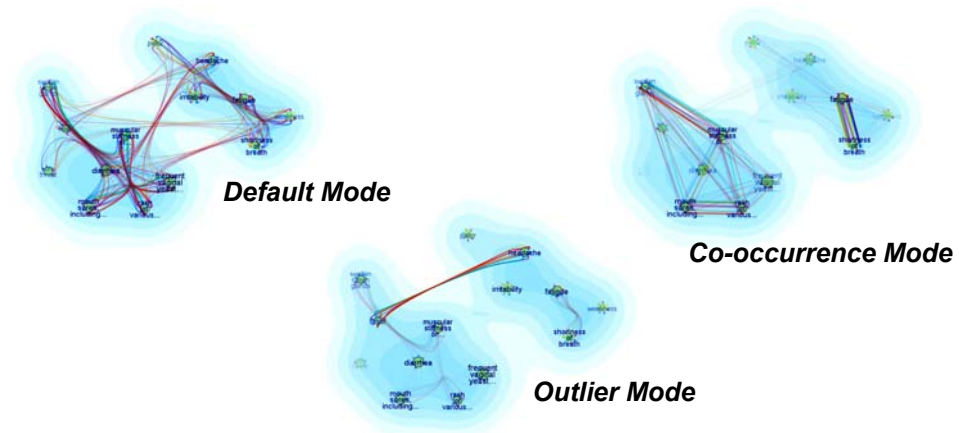
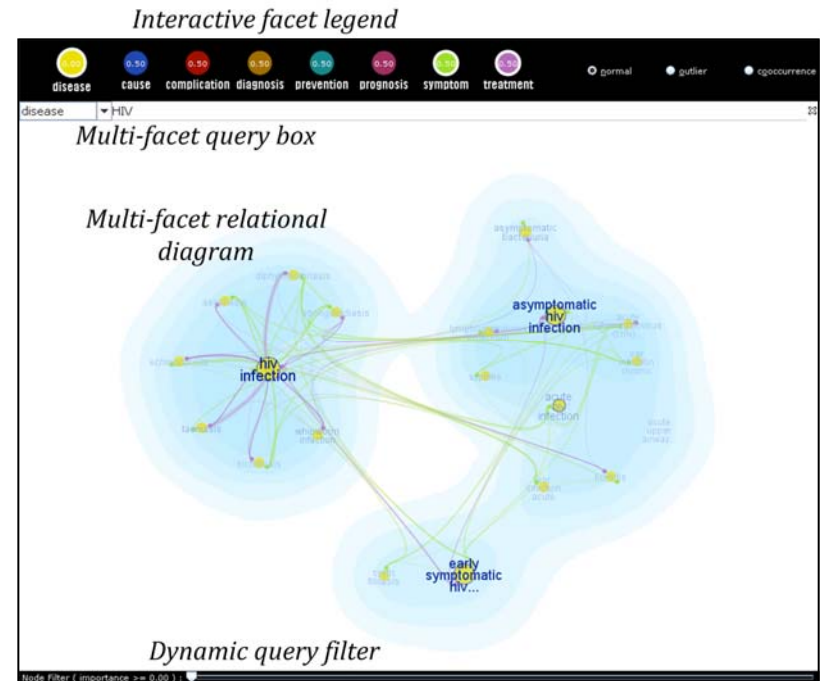
# FacetAtlas

## ■ Key Features

- Simultaneous view of:
  - global context (density map)
  - local relationships (overlaid graphs)
- Dynamic facet-based context switching
- Integrated search

## ■ Applications

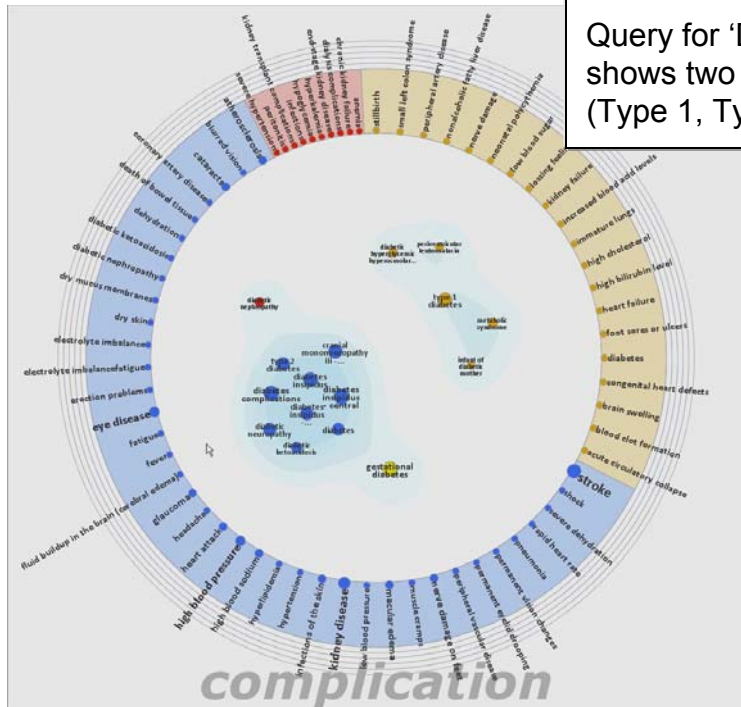
- Prototype applied to [Google health](#) documents
- Extended for similar patient cohorts



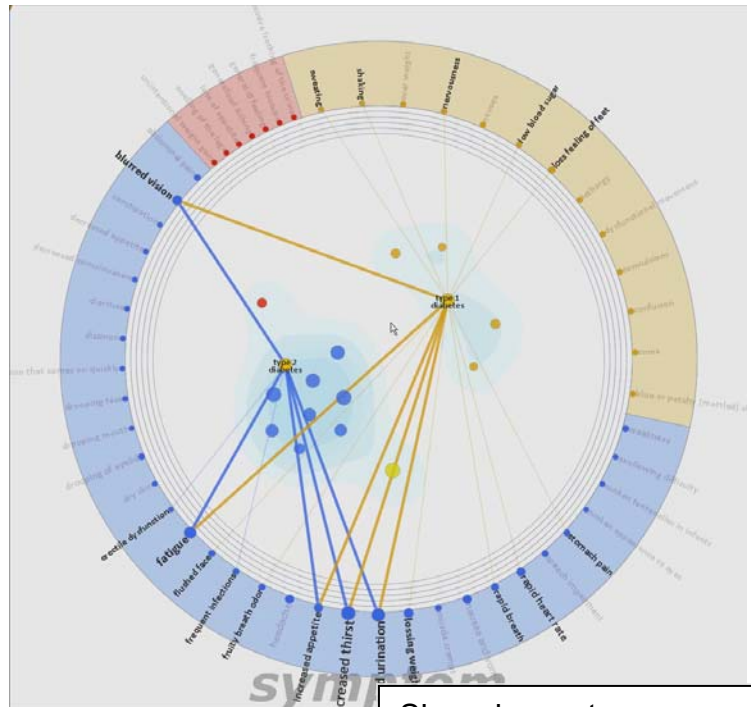




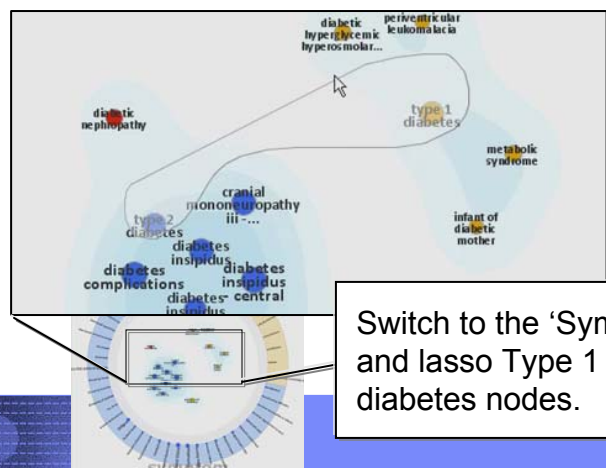
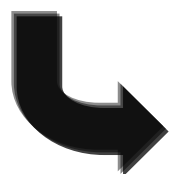
# Example: Shared Symptom Analysis



Query for 'Diabetes' shows two main clusters (Type 1, Type 2)



Shared symptoms are highlighted with heavy edges while lighter edges represent distinguishing symptoms.



Switch to the 'Symptom facet' and lasso Type 1 and Type 2 diabetes nodes.

