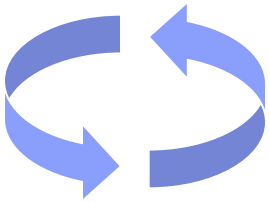
Knowledge Analytics Synergy in Clinical Decision Support



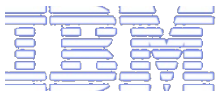
MIE 2012

Noam Slonim, Boaz Carmeli, Abigail Goldsteen, Oliver Keller, Carmel Kent, Ruty Rinott

IBM Haifa Research Labs

Machine Learning and Data Mining (MLDM) group

Cli-G group

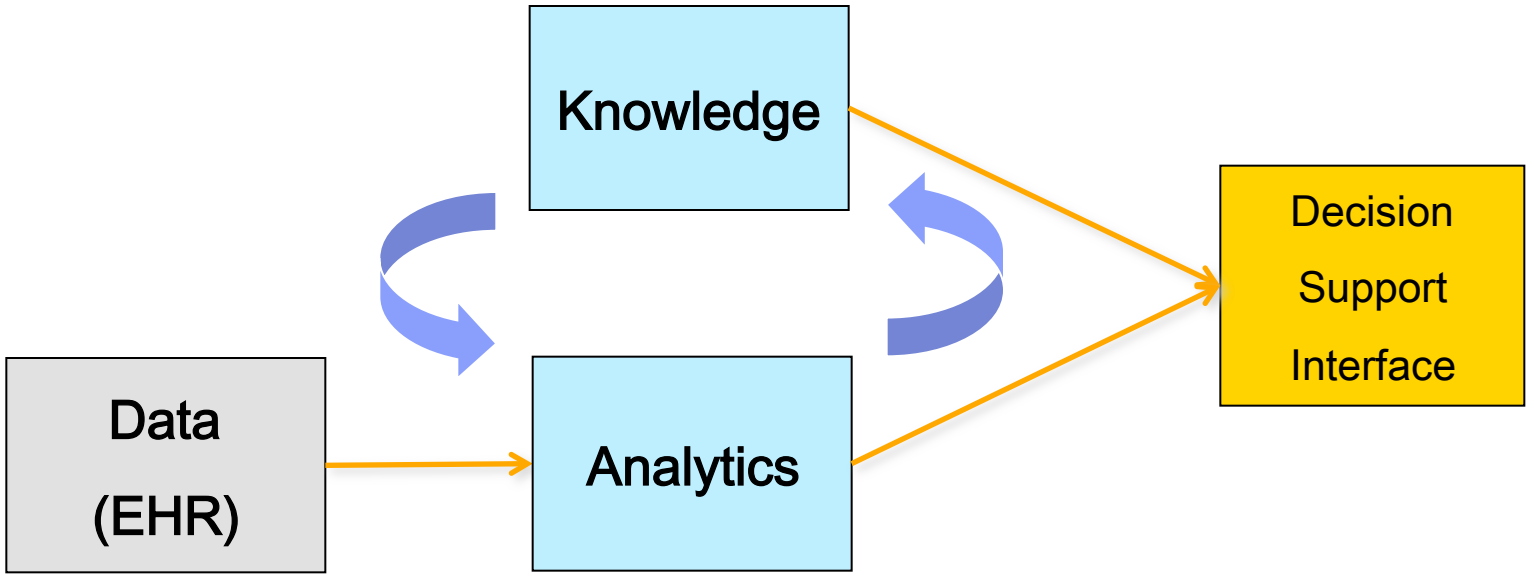


Clinical Genomics IT

Introduction

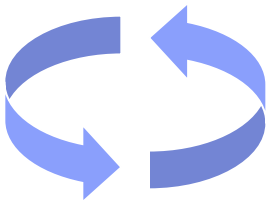
- Clinical Decision Support (CDS) have great potential
  - Improving clinical care
  - Reducing associated costs
- Most existing tools are knowledge-based CDS
  - Clinical trials → Clinical guidelines → simple CDS rules
  - Well established
- Emerging alternative paradigm – analytics-based CDS
  - EHR data → Analytics (Machine Learning) → Statistically-based CDS
  - No need to define/maintain long list of rules
  - Allows to take new "omic" data into account quickly
  - Address groups not participating in clinical trials (co-morbidities, elderly, etc.)

Both paradigms are complementary



- Analytics techniques may benefit from domain knowledge
- Domain knowledge may benefit from analytics

- CDS knowledge-Analytics Synergy (KAS) paradigm



The proposed CDS KAS Paradigm

- Synergistically combine relevant clinical knowledge with analytics applied to EHR data, improving overall CDS quality
- Knowledge and analytics components mutually feed and enhance each other to achieve better results.
- Implemented as part of a generic CDS system – Cli-G
  - Cli-G underlying architecture explicitly supports the KAS paradigm
  - See also Evicase presentation (Wed. 1000, Fermi)

Data examined to demonstrate KAS potential

- 3 cohorts of hypertensive patients:*
  - EPOGH
    - 1,149 patients
    - 15 clinical features
  - IMA_Sardinia
    - 278 patients
    - 45 clinical features
    - Genotype data for 1M SNPs
  - Immidiet
    - 258 patients
    - 45 clinical features
    - Genotype data for 1M SNPs

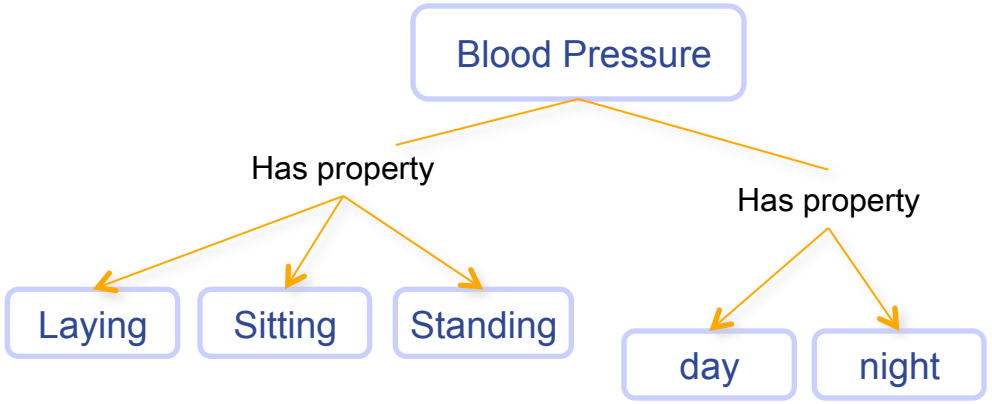
HYPERGENES
European Network for Genetic-Epidemiological Studies

* http://www.hypergenes.eu/

cli-g

The Knowledge-Management (KM) component

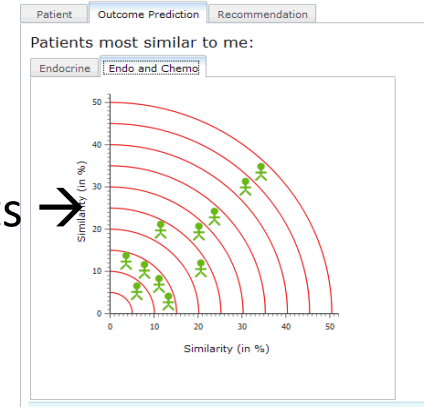
- Represents and streams relevant clinical knowledge to other components.
- Declarative Knowledge (DK)
  - Factual information: hierarchy of concepts with properties and relations

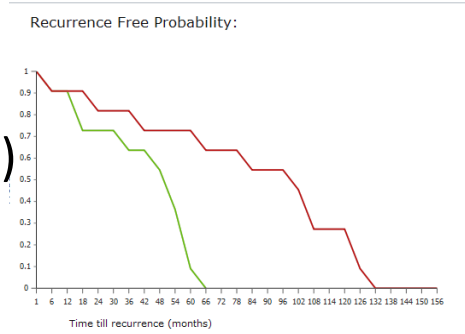


- Procedural Knowledge (PK)
  - The knowledge on how to operate upon the DK concepts
    - Clinical Guidelines
    - Rules defining Deductive Concepts to be added to the DK model; e.g.,
      - Risk Group based on Gender, Weight, Co-morbidities, etc
- May support various CDS applications
  - Guidelines based treatment recommendations
  - Rule-based alerts for adverse-drug events

The Analytics component

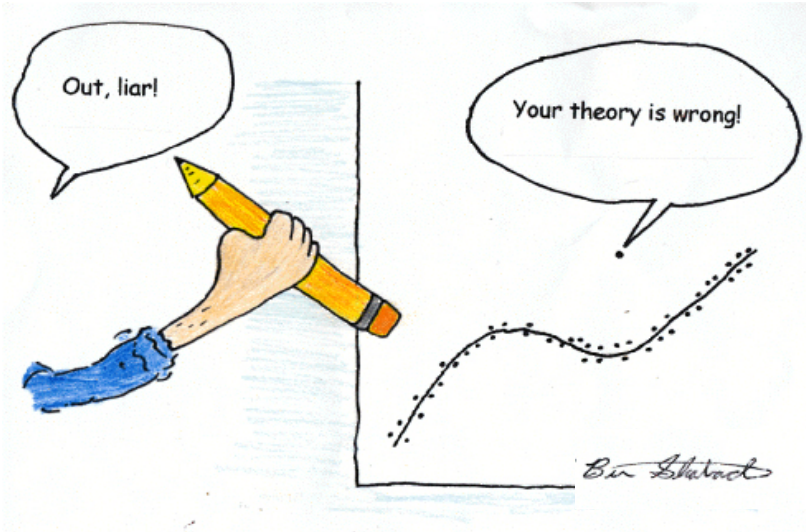
- Encompasses an arsenal of ML algorithms for various CDS tasks
- Feature ranking/selection
  - Automatically highlight clinical features of potential interest to the clinician
- Estimating patient similarity
  - Automatically reveal meaningful groups of similar patients → recommendations at Point Of Care
- Supervised Machine Learning algorithms
  - Predicting most common treatment (IHI 2012)
  - Predicting the outcome of candidate treatment (EuResist)
  - More…

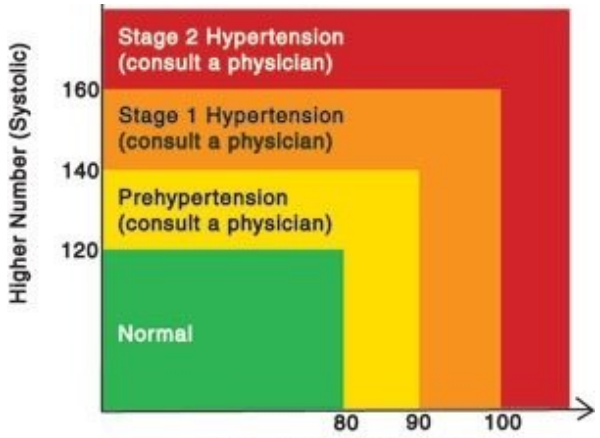




KAS: using KM component to enhance Analytics

- Procedural Knowledge rules for **Data Cleansing**:
  - Outliers / non-valid data removal
    - Remove Height measurements < 140
    - Remove Serum Cretanine > 2
  - Unifying data reported in different scales
    - Serum Insulin: µIU/mL or pmol/L
  - Quantizing continuous data
    - Systolic Blood Pressure:
      - < 120 → Normal
      - 120-139 → Pre Hypertension
      - 140-159 → Stage-1
      - > 160 → Stage-2

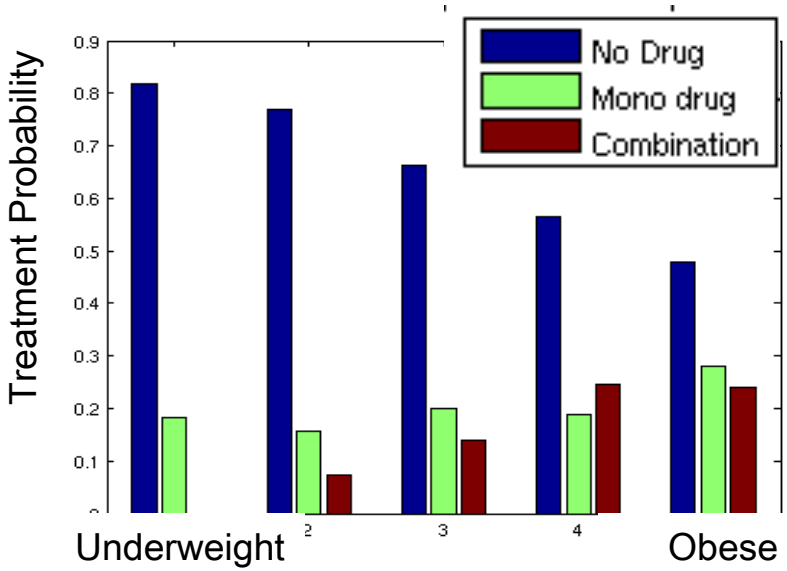




KAS: using KM component to enhance Analytics

- **Data enrichment**
  - Use domain knowledge to pinpoint relevant feature combinations to explore
    - BMI
    - [Serum-Glucose*Serum-Insulin]/25 → HOMAI
    - More...

- Some Deductive Concepts were later identified by feature-ranking as associated with clinicians treatment decision;
  - BMI in the EPOGH cohort → insights regarding treatment-selection process*



* Rinott *et. al* IHI 2011

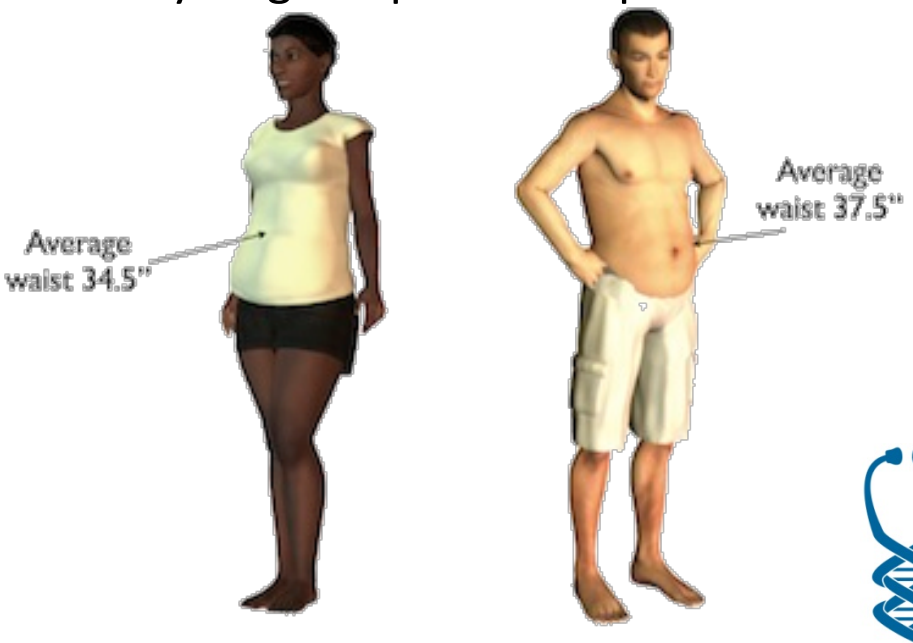
KAS: using KM component to enhance Analytics

- **Data filtering**
  - Use domain knowledge to filter irrelevant features, focusing analytics on important features
    - Find similar patients based on genetic SNP data
    - Using all 1M SNPs available
      - Time costly
      - Low signal to noise ratio
    - Pre-select 5,717 SNPs according to domain knowledge - potentially related to molecular pathways associated with hypertension
      - Reduce run time by more than 2 orders
      - Clustering patient based on similarity results in meaningful clusters

KAS: using Analytics to enhance KM component

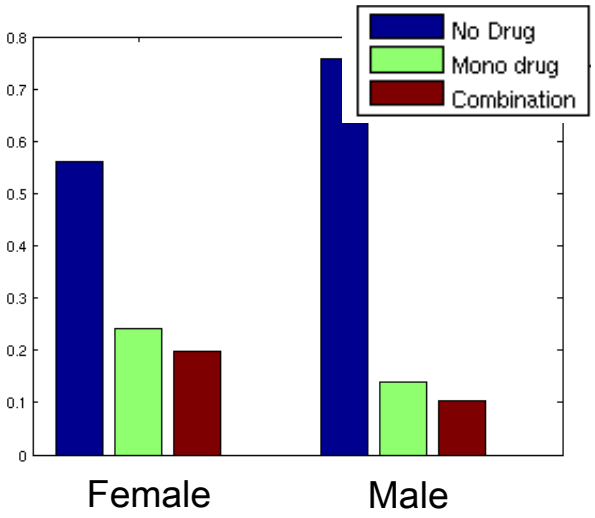
- **Enhancing Declarative Model**
  - EHR often includes free-text data
  - Analytics may reveal the importance of a particular free-text term; e.g.,
    - Drug-name → added to the DK Model

- Deductive Concepts
  - Exploring all feature pairs → Reveal pairs with synergistic predictive power → TBA as Deductive Concepts to the DK model
  - Waist circumference & Gender has synergistic predictive power for selected treatment in EPOGH data



KAS: using Analytics to enhance KM component

- **Enhancing Procedural Knowledge**
  - Reveal clinicians "common practices" → TBA to the PK model → supporting rule-based CDS applications
  - Feature ranking applied to identify features associated with treatment selection
  - Gender found to be significantly associated with treatment selection
  - Gender highlighted in PK model as related to "Common Practices" in "Treatment Selection"

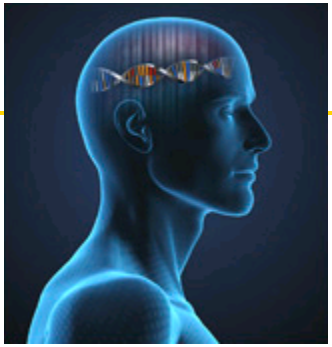


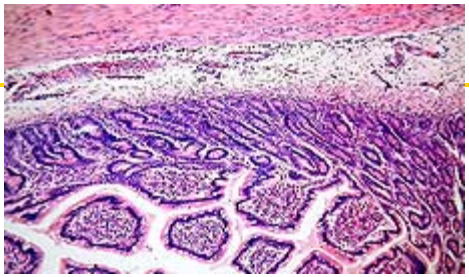
- Guidelines refinement
  - Pinpoint SNPs of potential clinical value to enhance PK and DK → guidelines refinement (Not implemented yet)

Summary

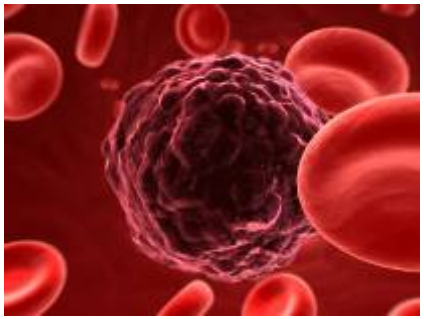
- Presented the Knowledge Analytics Synergy paradigm for CDS
  - Demonstrated via a generic tool (Cli-G) that relies on the KAS Architecture
- Presented concrete examples over clinical and genomic data:
  - Knowledge can enhance Analytics
  - Data-driven analytics can enhance Knowledge
- The synergisytic approach holds great potential for CDS tools

cli-g





Thank You



cli-g

Introduction

- Clinical Decision Support (CDS) have great potential
  - Improving clinical care
  - Reducing associated costs

- Most existing tools are *knowledge-based CDS*
  - Clinical trials → Clinical guidelines → simple CDS rules

- Emerging alternative paradigm – *analytics-based CDS*
  - EHR data → Analytics (Machine Learning) → Statistically-based CDS rules

Results – Analytics enhancement – **Data Filtering**

- Patients were clustered using Patient Similarity defined according to SNP data
  - Emphasizing relative similarity of patients sharing identical rare SNPs
- Similarity estimated from 5,717 SNPs (from 1M) selected by domain knowledge
  - Filtering reduced computation time by more than two orders of magnitude
  - Filtering focused analysis in advance on potentially most relevant SNPs
- Statistical significance of clinical features was estimated in each cluster
  - c10 (30 people)
    - Avg. Systolic BP: 153, vs. 140 in remaining population (P-val 0.004)
  - c5 (23 people)
    - Avg. Triglyceride: 95.3, vs. 132 in remaining population (P-val 0.0002)
  - More…
- → Encouraging results that should be further explored …

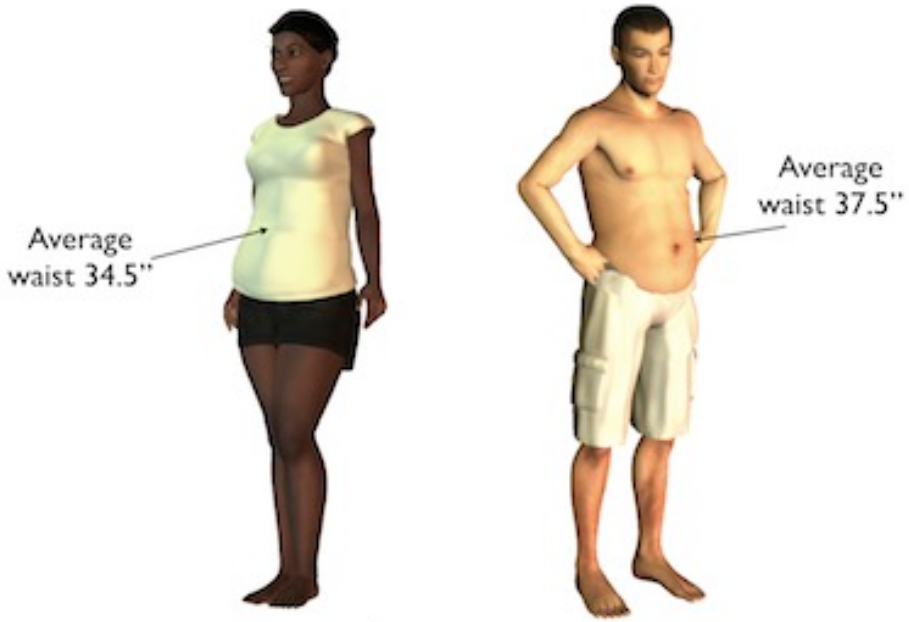
Both paradigms are potentially complementary

- Analytics-based Insights extracted from EHR data may better reflect the relevant population
- Analytics-based CDS require no need to define/maintain long list of rules
- Inherently stochastic statistical signals may better reflect clinical world than deterministic rules
- Being blind to the relevant clinical knowledge is obviously hazardous
  - Contraindications
- Analytics techniques may benefit from domain knowledge

CDS knowledge-Analytics Synergy (KAS) paradigm

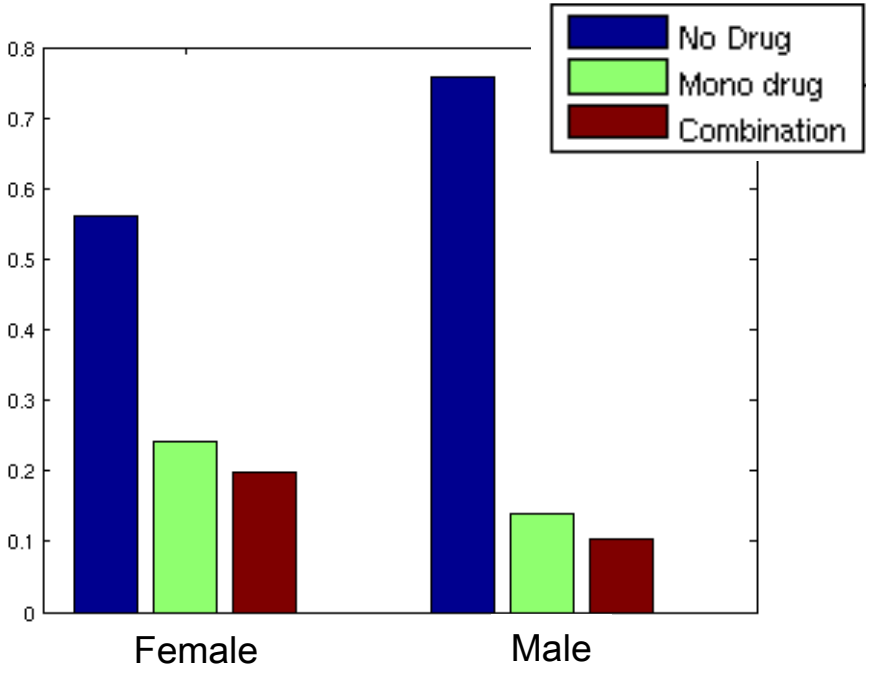
Results – Knowledge enhancement

- Searched for feature-pairs with synergistic predictive power for selected treatment in EPOGH data
  - Several pairs were detected as significantly associated with treatment allocation:
    - Waist circumference & Gender → added to DK model as Deductive Concepts



Results – Knowledge enhancement

- Feature ranking applied to identify features associated with treatment allocation
  - Gender found to be significantly associated with treatment selection
  - → Gender highlighted in PK model as
    - related to "Common Practices" in "Treatment Selection"
  - Relevant reference exist
    - → Accompany analytics-based evidence with pointer to relevant literature…?

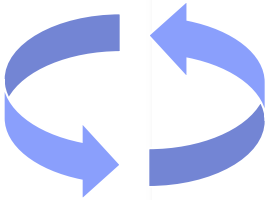


Both paradigms are complementary

- Being blind to the relevant clinical knowledge is obviously hazardous
  - Contraindications
  - Rare conditions
- Analytics techniques may benefit from domain knowledge

- Analytics-based insights may highlight important features not taken into account by current guidelines
- Analytics-based Insights extracted from EHR data may better reflect population
  - Address groups not participating in clinical trials (co-morbidities, elderly, etc.)
  - Consider new genetic information
- Domain knowledge may benefit from analytics

CDS knowledge-Analytics Synergy (KAS) paradigm



KAS CDS High level architecture