Mining Healthcare Systems to Personalize Patient Care & Improve Clinical Decisions

R. Bharat Rao, PhD
Knowledge Solutions
Siemens Medical Solutions Inc.
Aug 17, 2010
Outline

- Healthcare Challenges
- Knowledge-Based Medicine
- KBM in practice
  - Quality
  - Targeted decision support
  - Knowledge discovery
- Conclusion
Global Healthcare Market Trends

- Increasing Healthcare Costs (73% over the next 9 years)\(^1\)
- Increased consumerism & consumer awareness
- Knowledge explosion
  - Increased creation and dissemination of evidence based guidelines
- Data overload
  - Increased digitization of data
- New Technology
- Personalized Medicine

\(^1\)“National Health Expenditures...Calendar Years 2009—2018” - Centers for Medicare & Medicaid Services, Office of the Actuary
Clinical decisions are more complex

Clinicians need support to manage these challenges. Data mining and personalization of care can help.
Outline

- Healthcare Challenges
- Knowledge-Based Medicine
- KBM in practice
  - Quality
  - Targeted decision support
  - Knowledge discovery
- Conclusion
Knowledge-based Medicine

Integrate patient data with medical knowledge to improve health outcomes.

- Access existing patient data from disparate sources
- Manage and integrate knowledge for decision support
- Impact workflow & outcomes
Impacting Clinical Workflow

Knowledge Solutions

Data → Knowledge Solutions → Clinical Answers

Knowledge
Two aspects of Knowledge Based Medicine

- **Knowledge Utilization**
  - Knowledge-driven decision support
  - Data aggregation from disparate sources – **using existing data**
  - Inject knowledge (e.g., guidelines) via Probabilistic inference
  - Present conclusive findings

- **Knowledge Discovery**
  - Combine imaging, clinical, IVD and genetic information
  - Discover combined diagnostic signatures
  - ... from multi-institution data
  - Enable increasing personalization of care

*Not considered a Medical device.*
REMIND Knowledge Platform*: Architecture
Reliable Extraction & Meaningful Inference from Nonstructured Data

*Not considered a Medical device.
Outline

- Healthcare Challenges
- Knowledge-Based Medicine
- KBM in practice
  - Quality
  - Targeted decision support
  - Knowledge discovery
- Conclusion
“Soarian Quality Measures is an essential, time-saving solution that automates the abstraction of critical patient data from both structured and unstructured free text from clinical narratives. It enables our teams to focus on providing the best quality of care possible.”

Janene Yeater, Vice President of Quality and Planning MedCentral Health System - July 2009

Quality Measurement & Improvement

Data → REMIND Platform → Clinical Answers

Knowledge

Existing guidelines
- CMS & Joint Commission guidelines

Automated Quality Measurement & Improvement

Labs

Demographics

Free text Pt. Factors

Medication
Outline

- Healthcare Challenges
- Knowledge-Based Medicine
- KBM in practice
  - Quality
  - Targeted decision support
  - Knowledge discovery
- Conclusion
Personalized dose recommendation
Non Small-Cell Lung Cancer

Current plans based on “average dose response curves” appropriate for a population

REMIND Platform: Prediction models for lung cancer patients: www.predictcancer.org

Create personalized dose response rate for individual patients for models created by mining large database of patient data and external medical knowledge

*This information about this product is preliminary. The product is under development and not commercially available for sale in the U.S. and its future availability cannot be ensured.
Prediction of survival outcomes and side effects

**Patient clinical factors**
- Age: 76
- Gender: female
- Height (cm): 179.00
- Weight (kg): 72.00
- BMI: 22.47
- Weight loss (during the last 3 months): stable
- Smoking: current smoker
- Pack Yrs:
- WHO Perf Scale: 1.00
- Charlson Comorbidity Index: 0.00
- Lung function FEV1, 1 second (lit): 
- Lung function FEV1, 1 second (lit):
- Hemoglobin level [mmol/l]: 7.70

**Tumor Information**
- Stages: 11lb
- T-Stage: T4
- N-Stage: N3
- M-Stage: M0
- Pathology: 
- Histology: squamous cell ca

**Image Information**
- Gross Tumor Volume (cc): 83.78
- Maximum Standard Uptake Value, primary tumor from FDG PET: 133.00
- Num. of pos. lymph node stations: 1.00

**Treatment Plan**
- Mean Lung Dose [Gray]: 15.82
- V20 [V]: 24.00
- Treatment regime: sequential before RT
- Overall treatment time [days]: 41.00
- Treatment dose [Gray]: 60.00
- Fraction size [Gray]: 2.00
- Fractions/day: 1.00
- Number of cycles [chemo]: 3.00
- Chemotherapy type:

**Treatment Monitoring**
- Mean % deviation of given dose [from plan]:
- Mean shift of radiation fields [mm]:

**Prediction**
- Probability of Survival at 2 yrs: 34%
- Probability of pneumonitis: 11%

*This information about this product is preliminary. The product is under development and not commercially available for sale in the US and its future availability cannot be ensured.*
**REMINDD Platform CDS**

**Patient Information**
- Age: 66
- Gender: female
- Height (cm): 179.00
- Weight (kg): 72.00
- BMI: 22.47
- Weight loss (during the last 3 months): stable
- Smoking: current smoker
- Pack Yrs: 
- WHO Perf Score: 1.00
- Charlson Comorbidity Index: 0.00
- Lung function FEV1 second (lit): 
- Lung Function FEV1 second (%): 
- Hemoglobin Level [mmol/lit]: 7.70

**Tumor Information**
- Stage: IIib
- T-Stage: T4
- N-Stage: N3
- M-Stage: M0
- Pathology: 
- Histology: squamous cell ca

**Image Information**
- Gross Tumor Volume (cc): 83.78
- Maximum Standard Uptake Value, primary tumor (from FDG PET): 133.00
- Num. of pos. lymph node stations: 1.00

**Treatment Plan**
- Mean Lung Dose [Gray]: 15.82
- V20 [%]: 24.00
- Treatment regime: sequential before RT
- Overall treatment time (days): 41.00
- Treatment dose (Gray): 60.00
- Fraction size (Gray): 2.00
- Fractions/day: 1.00
- Number of cycles (chemo): 3.00
- Chemotherapy type: 

**Treatment Monitoring**
- Mean % deviation of given dose (from plan): 
- Mean shift of radiation fields (mm): 

**Prediction**
- Probability of Survival at 2 yrs: 34%
- Probability of pneumonitis: 11%

*This information about this product is preliminary. The product is under development and not commercially available for sale in the US and its future availability cannot be ensured.
Increasing treatment dose increases the probability of survival, but also increases likelihood of side effects.
Further increasing treatment dose increases likelihood of side effects, but does not improve probability of survival.
Targeted Decision Support at Point of Care
Personalized therapy selection for Lung cancer

REMIND Platform

Data → REMIND Platform → Clinical Answers

- **Labs**
- **Images**
- **Treatment Plans**
- **Genomics**

Knowledge

**Existing guidelines**

**Learned knowledge**
- Predictive models discovered from patient data

**Treatment impact**
- Adjust dose based on outcomes & side effects

REMIND Platform: Prediction models for lung cancer patients: [www.predictcancer.org](http://www.predictcancer.org)
Outline

- Healthcare Challenges
- Knowledge-Based Medicine
- KBM in practice
  - Quality
  - Targeted decision support
  - Knowledge discovery
- Conclusion
Discover Combined Diagnostics
Models that span Imaging, IVD, Genomic, Molecular & Clinical Biomarkers

Knowledge Discovery

- Images
- Proteomics
- Genomics
- Labs
- Treatment Plans
- Patient Factors
Combining clinical data from disparate sources improves prediction accuracy

LOO ROC Plot for S2y (82pts, P/N: 24/58)

- AUC: 0.65 (Clinic)
- AUC: 0.76 (Clinic + Image)
- AUC: 0.85 (Clinic + Image + Biomarker)

What about data from multiple institutions?

Predicting 2 year Lung Cancer Survival

- 455 inoperable NSCLC patients, stage I-IIIB, referred to the MAASTRO clinic (Netherlands) to be treated with (chemo)radiotherapy
- 112 patients from Gent hospital (Belgium)
- 40 patients from Leuven hospital (Belgium)

How do we use data from different institutions while preserving patient privacy?
Example: Cox Regression for Survival Analysis

- **Cox regression**, or the **Cox propositional-hazards model**, is one of the most popular algorithms for survival analysis
  - Defines a semi-parametric model to directly relate the predictive variables with the real outcome (i.e. the survival time)
  - Assumes a linear model for the log-hazard:

\[
\lambda(t|x_i) = \lambda_0(t) \exp[w^T x_i]
\]

- We propose privacy-preserving Cox regression (PPCox) which is based on **random projection**
  - Provides accurate classification
  - Does not reveal private information

\[
\lambda_{\text{PPCox}}(t|x_i) = \lambda_0(t) \exp[w^T B^T x_i]
\]
We are using real data from several institutions across Europe to build models for survival prediction for non-small-cell lung cancer patients while addressing the potential privacy preserving issues.

Example showing the improvement in performance of a model trained using all the data available from multiple sites against models learned only using local available data.
Which asymptomatic patients will develop hypertension within 5 years?
Collaboration with Univ of Greifswald

- Based on data from large study, >4000 patients over 10 years
  - Genetic (SNPs), clinical, lab, imaging, …
- Analyzed 100s of potential predictive variables to identify the most informative subset
- Learned a Bayesian Network
  - predicts patients who are at high risk to become hypertensive (AUC >0.85).

Creating Medical Knowledge
Mining large patient databases from multiple institutions

Data → REMIND Platform → Predictive Models

Knowledge:
- Existing guidelines for cancer therapy
- Learned knowledge: Predictive models discovered from patient data

REMIND Platform

Images
Labs
Medication
Genomics
Patient Factors
Automated Identification of Patients for Clinical Trials**

Knowledge Discovery can accelerate and improve the selection process of appropriate patients for clinical studies**

5 million patients in hospital database 85 found by REMIND SmartTrials 44 eligible (52% eligible) 3 patients enrolled

Greatly Reducing Research Nurse Time

Manual (current methodology)
- 1 eligible patient found per 3 hours of nurse time
- 1 patient enrolled per 75 hours of nurse time

Using REMIND SmartTrials
- 1 eligible patient found per 12 minutes of nurse time
- 1 patient enrolled per 3 hours of nurse time

Greatly Reducing Elapsed Time

Manual (current methodology)
- 7 patients enrolled in 9 months

Using REMIND SmartTrials
- 3 patients enrolled in 1 week

* This solution is not considered a medical device. **This information about this product is preliminary. The product is under development and not commercially available for sale in the U.S. and its future availability cannot be ensured.
Outline

- Healthcare Challenges
- Knowledge-Based Medicine
- KBM in practice
  - Quality
  - Targeted decision support
  - Knowledge discovery
- Conclusion
The Road to Personalized Medicine

Future

Knowledge-Driven Workflow
Clinical decisions using all available electronic data & medical knowledge for Personalized Medicine, Quality Improvements, and Workflow Management.

Key Enablers
Application of Knowledge
Integration of Data

Current

Focused Tasks
“Does this patient have a history of smoking?”
“Find a lung nodule in a chest CT”
Integrate existing patient data with medical knowledge to improve health outcomes.

- **Key components**
  - Access *existing patient data* from disparate sources
  - Manage and integrate *knowledge* for decision support
  - Impact workflow & *outcomes*

- **Healthcare Impact**
  - Improve quality of care
  - Support personalization of care via targeted decision support
  - Continuous improvement by learning new medical knowledge
Knowledge-based Medicine

Integrate **existing** patient data with medical knowledge to improve health outcomes.

**Key components**
- Access **existing patient data** from disparate sources
- Manage and integrate **knowledge** for decision support
- Impact workflow & **outcomes**

**Healthcare Impact**
- Improve quality of care
- Support personalization of care via targeted decision support
- Continuous improvement by learning new medical knowledge
Thank You

R. Bharat Rao
bharat.rao@siemens.com
Knowledge Solutions
Siemens Healthcare
Patient Identification for Clinical Trials enabled by the REMIND Platform

Computerized Patient Data

Images
Patient Factors
Proteomics
Genomics
Treatment Plans

Knowledge Discovery
Clinical Decision Support
High-Quality Structured Clinical Patient Data

Copyright © 2010 Siemens Medical Solutions USA, Inc. All rights reserved.
Patient Identification for Clinical Trials enabled by the REMIND Platform

- Two use cases
  1. Clinician: Assess what clinical trials for which current patient is eligible
  2. Researcher: Assess which patients are eligible for trial
     - Contact clinicians or patients to consider trial
     - Feasibility study for trial (will there be enough patients?)

Automated Identification of Patients for Clinical Trials**

REMIND SmartTrials* accelerates and improves the selection process of appropriate patients for clinical studies**

5 million patients in hospital database → 85 found by REMIND SmartTrials → 44 eligible (52% eligible) → 3 patients enrolled

Greatly Reducing Research Nurse Time

Manual (current methodology)
- 1 eligible patient found per 3 hours of nurse time
- 1 patient enrolled per 75 hours of nurse time

Using REMIND SmartTrials
- 1 eligible patient found per 12 minutes of nurse time
- 1 patient enrolled per 3 hours of nurse time

Greatly Reducing Elapsed Time

Manual (current methodology)
- 7 patients enrolled in 9 months

Using REMIND SmartTrials
- 3 patients enrolled in 1 week

* This solution is not considered a medical device. **This information about this product is preliminary. The product is under development and not commercially available for sale in the U.S. and its future availability cannot be ensured.