iCeNSA
Interdisciplinary Center for Network Science & Applications

Bringing Big Data to Personalized Healthcare: A Patient-Centered Framework

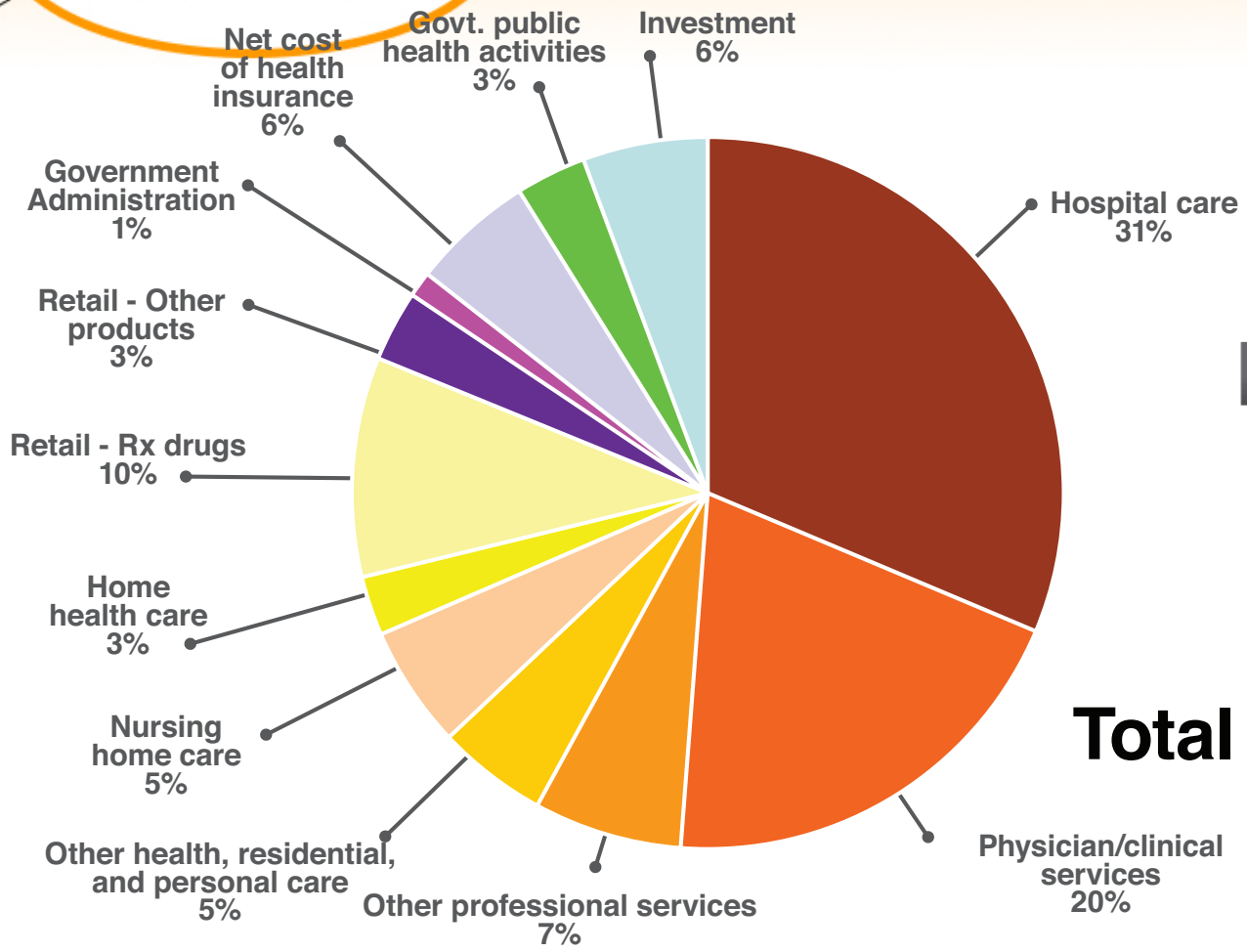
Nitesh Chawla, PhD

Frank Freimann Collegiate Associate Professor

Computer Science and Engineering

Director, iCeNSA

National Health Expenditures, 2010



Total - $2.594 trillion

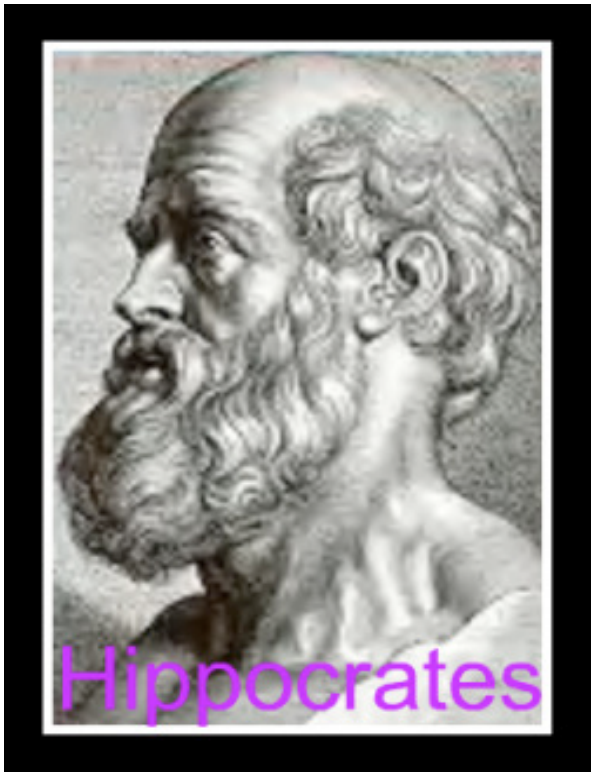
Source: Martin A.B. et. al., "Growth In US Health Spending Remained Slow in 2010: Health Share of Gross Domestic Products Was Unchanged from 2009," *Health Affairs*, 2012

Patient empowerment. Electronic health records will help empower patients to take a more active role in their health and in the health of their families

"Health care has been evolving away from a 'disease-centered model' and toward a 'patient-centered model.' In the older, disease-centered model, physicians make almost all treatment decisions based largely on clinical experience and data from various medical tests. In a patient-centered model, patients become active participants in their own care and receive services designed to focus on their individual needs and preferences, in addition to advice and counsel from health professionals." AHRQ.GOV

Two thousand years ago..

"It is far more important to know what person the disease has than what disease the person has,"
Hippocrates



What are <u>my</u> disease risks? *A Personalized Approach*

"Determine <u>individual</u> risk of developing specific diseases, detect the disease's <u>earliest</u> onset, and <u>prevent</u> or intervene early enough to provide maximum benefit"

Similarities and shared experiences matter

NETFLIX

amazon.com

iTunes

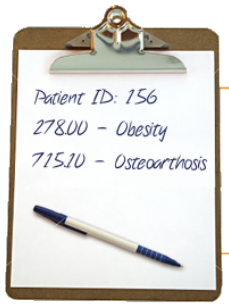
CARE: Collaborative Assessment and Recommendation Engine

Patent No. 8,504,343

Empowering the patient and physician with the inferences drawn from millions of other patients

CARE

Patent No. 8,504,343



Individual Medical History

Other Patients' Medical Histories

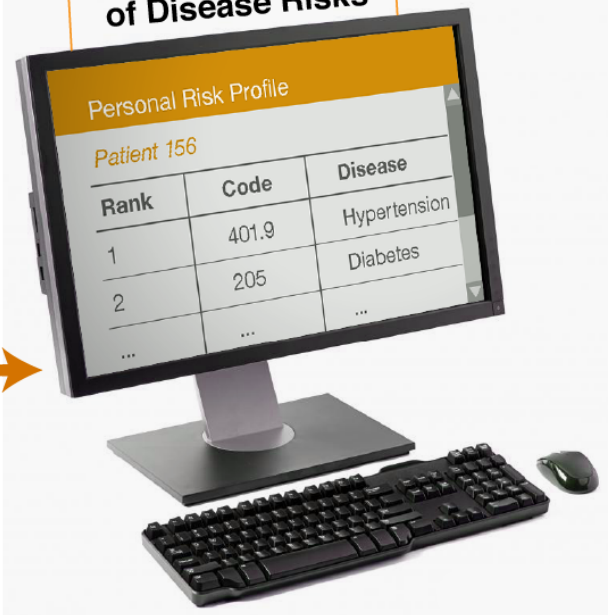
Similarity Constraint (J_a disease constraints)

Collaborative Filtering

- Inverse Frequency
- Vector Similarity
- Best Match Subset

ICARE Ensemble

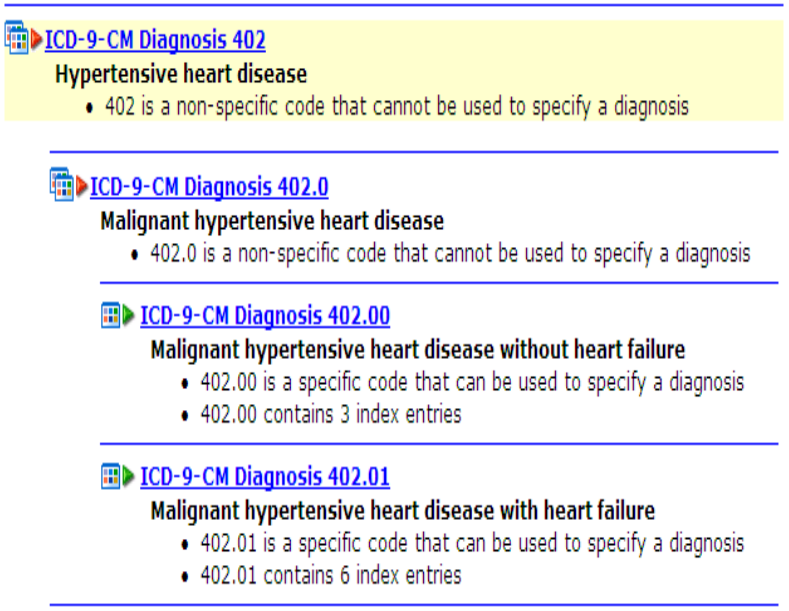
Ranked List of Disease Risks



"Data is a vital raw material"

Partnership is THE Enabler

- 13 Million ICD-9-CM data (Medicare)
  - In-patient
  - 32 Million visits over 4 years
- 500,000 out-patient data from a regional health system
  - Spread over 12 years



Details

$$P_{a,j} = \bar{v}_j + K_a \sum_{i \in I_j} w(a,i)(v_{i,j} - \bar{v}_i)$$

i	- training patient
j	- disease
a	- the active patient
I	- set of all training patients
I_j	- set of all training patients with disease *j*
$w(a,i)$	- the similarity between user a and user *i*

ICARE – an Iterative Version

- Problem: Even with IFVS, common diseases overwhelm rare diseases, but tend to be more obvious and less interesting
- Solution: Disease-focused ensembles

$$\max_{c \in G} \left(\overline{v}_{j,c} + \kappa(1 - \overline{v}_{j,c}) \sum_{i \in I_{j,c}} w(a,i) \right)$$

iCeNSA
Interdisciplinary Center for Network Science & Applications

Time Sensitive iCARE

- Patients should be more similar is disease progression is similar, not just diagnoses
- Many medical events
- Want to further limit data to the specific period of interest in each patient's life

Time Sensitive iCARE

- Goal: Find best match (highest VS) subset of consecutive visits from training patient's record
- Special version of maximum subsequence problem
  - Linear scan solution

Experimental Setup

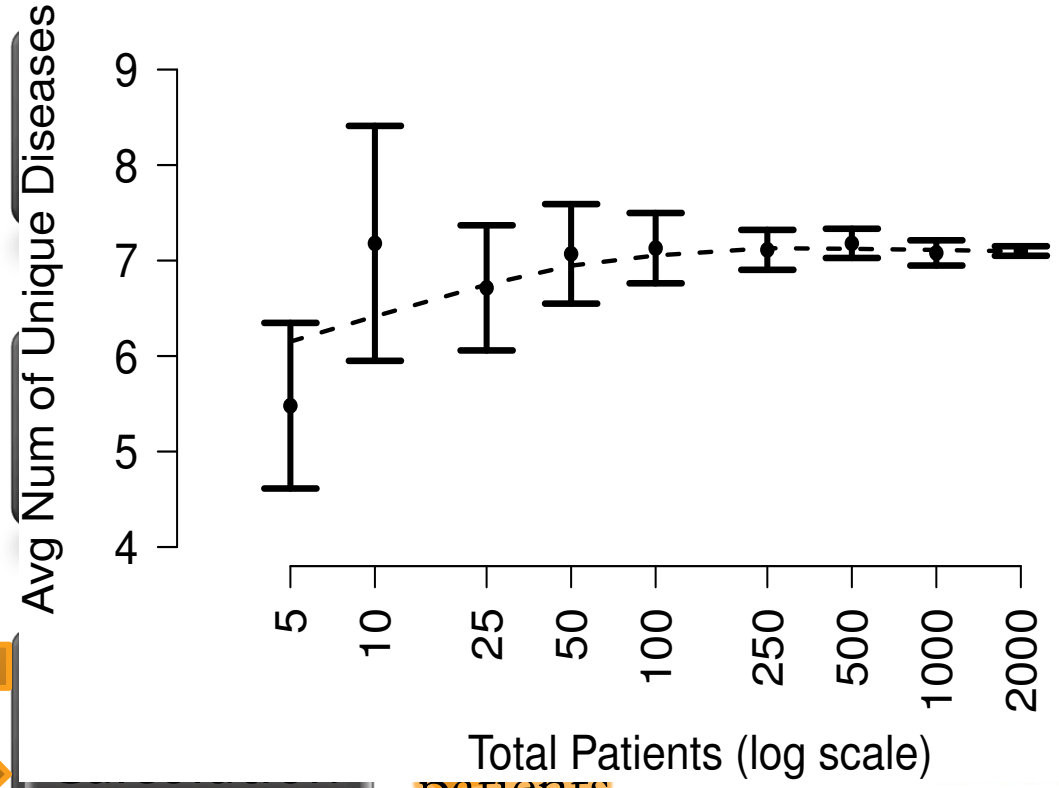
- Predictions are only for future diseases
- Patient must have at least 5 visits

Round	Training	Testing
1	Visit 1	Visits 2-5
2	Visits 1-2	Visits 3-5
3	Visits 1-3	Visits 4-5
4	Visit 1-4	Visit 5

iCARE Results

	Baseline	3-digit ICARE
Top 20		
Coverage	0.321	0.513
Average Rank	7.326	5.668

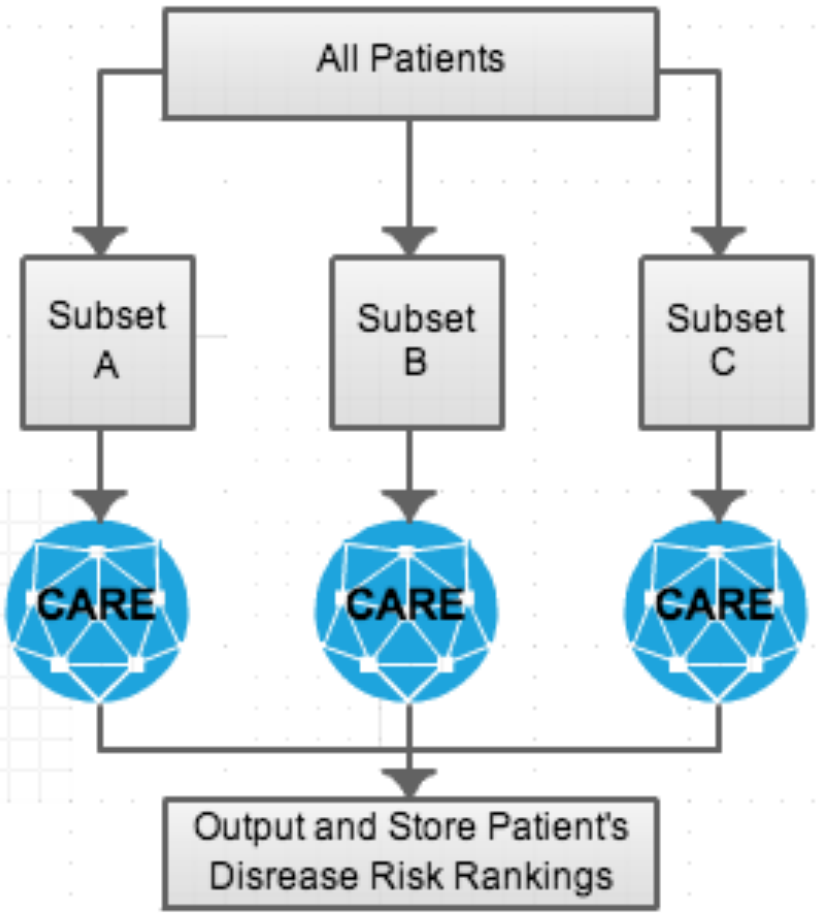
Computational Challenges



Iterate Over patients

iCeNSA
Interdisciplinary Center for Network Science & Applications

Distributed Model



"Data is a vital raw material"

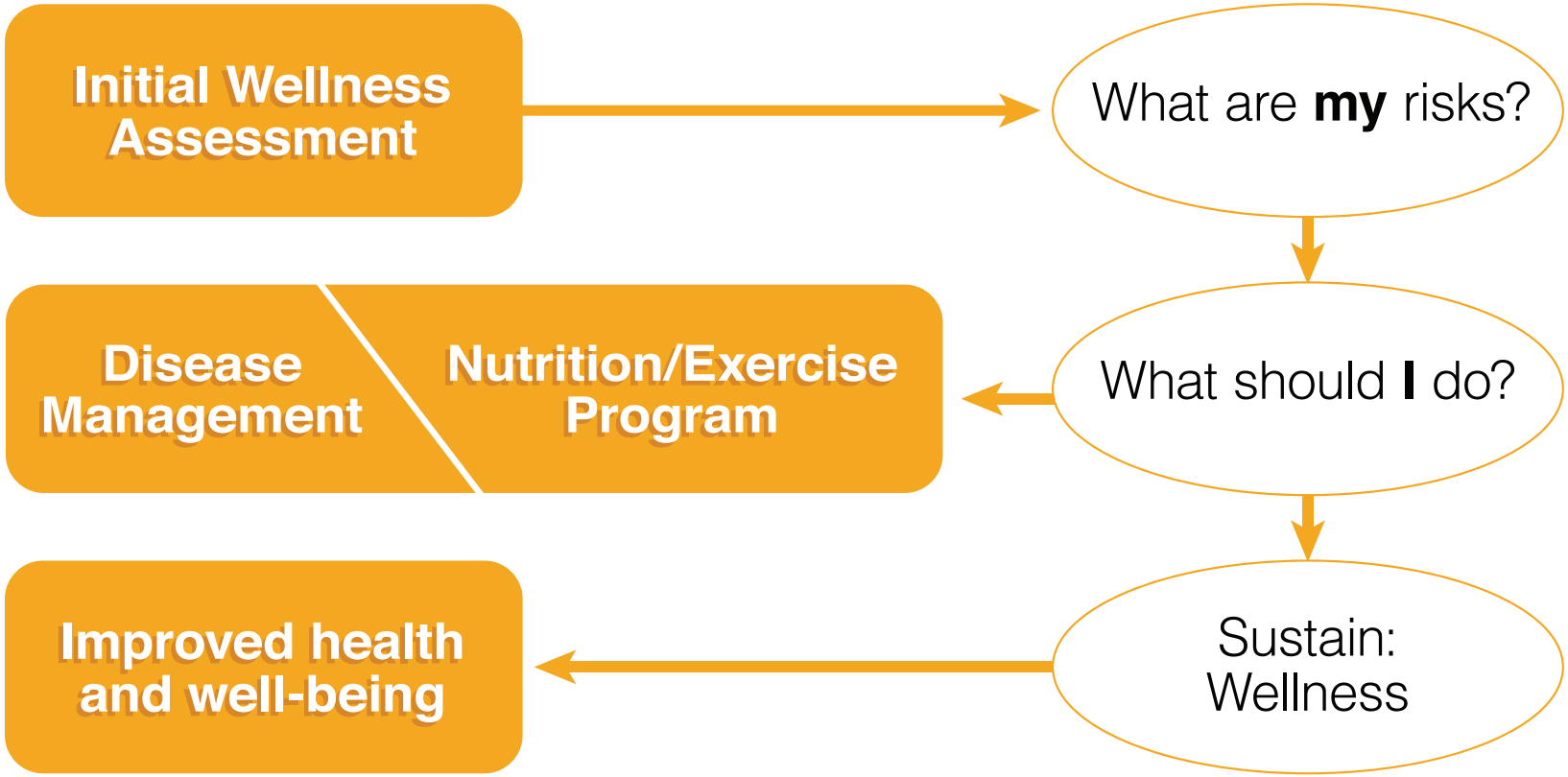
Partnership is the Enabler

Leverage for:

1) personalized and prospective healthcare
2) improved chronic disease management
3) reduction in medial errors and ADEs
3) reduced healthcare costs
4) best practices and efficiencies in delivery of healthcare
5) educational opportunities
6) exciting research opportunities and grand challenge problems for data mining and computer science

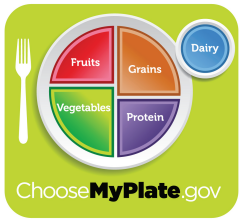
iCeNSA
Interdisciplinary Center for Network Science & Applications

Technology solve: Empower, personalize and sustain



iCeNSA
Interdisciplinary Center for Network Science & Applications

Me



Empower | **Personalize**

Connect | **Sustain**

References

- Feldman, K. and Chawla, N. V.., "Scaling personalized healthcare with Big Data," *International Conference on Big Data Analytics in Healthcare (BDAH)*, 2014.
- Dasgupta et al., "Integrated Digital Care Framework for Successful Aging," *IEEE EMBS-BHI Conference*, 2014.
- Chawla, N. V. and Davis, D., (2013) "Bringing Big Data to Personalized Healthcare: A Patient-Centered Framework," *Journal of General Internal Medicine,* 28 (3), 660 – 665.
- Hoens, T., Blanton, M., Steele A., and Chawla, N. V., (2013) "Reliable Medical Recommendation Systems with Patient Privacy," *ACM Transactions on Intelligent Systems and Technology (ACM TIST),* 4 (4), 31 pages.
- Rider, A. and Chawla, N. V., "An Ensemble Topic Model for Sharing Healthcare Data and Predicting Disease Risk," *ACM International Conference on Bioinformatics, Computational Biology and Biomedical Informatics (ACM BCB),* 2013
- Regola, N. and Chawla, N. V. (2013) "Storing and Using Health Data in Amazon Virtual Private Cloud," *Journal of Medical Internet Research,* 15(3):e63.
- Davis, D. Chawla, N. V., et al. (2010). "Time to CARE: A recommendation system for prospective health care," *Data Mining and Knowledge Discovery Journal,* 20(3), 388 – 415.

Thank you

- nchawla@nd.edu
- http://icensa.nd.edu

Data, Inference, Analysis and Learning Lab @ ND

UNIVERSITY OF NOTRE DAME

iCeNSA
Interdisciplinary Center for
Network Science & Applications