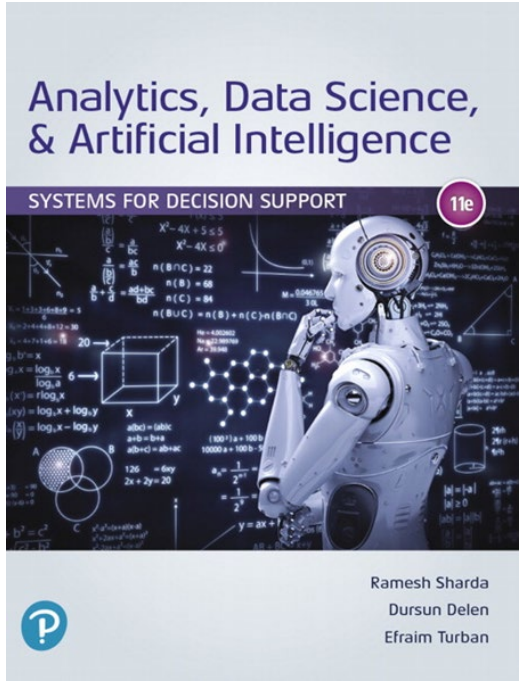
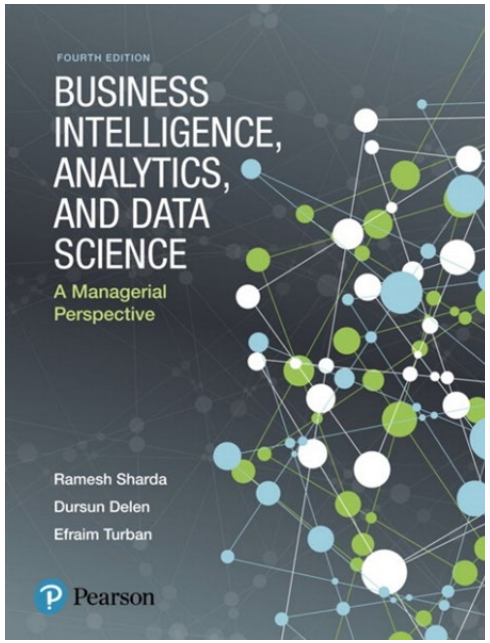
Network Science and Health Analytics Research



Ramesh Sharda, PhD
Vice Dean/Regents Prof.
Oklahoma State University
https://business.okstate.edu/sharda
Ramesh.sharda@okstate.edu

Aalto University Fulbright
Distinguished Chair (Spr 2023)
Ramesh.sharda@aalto.fi



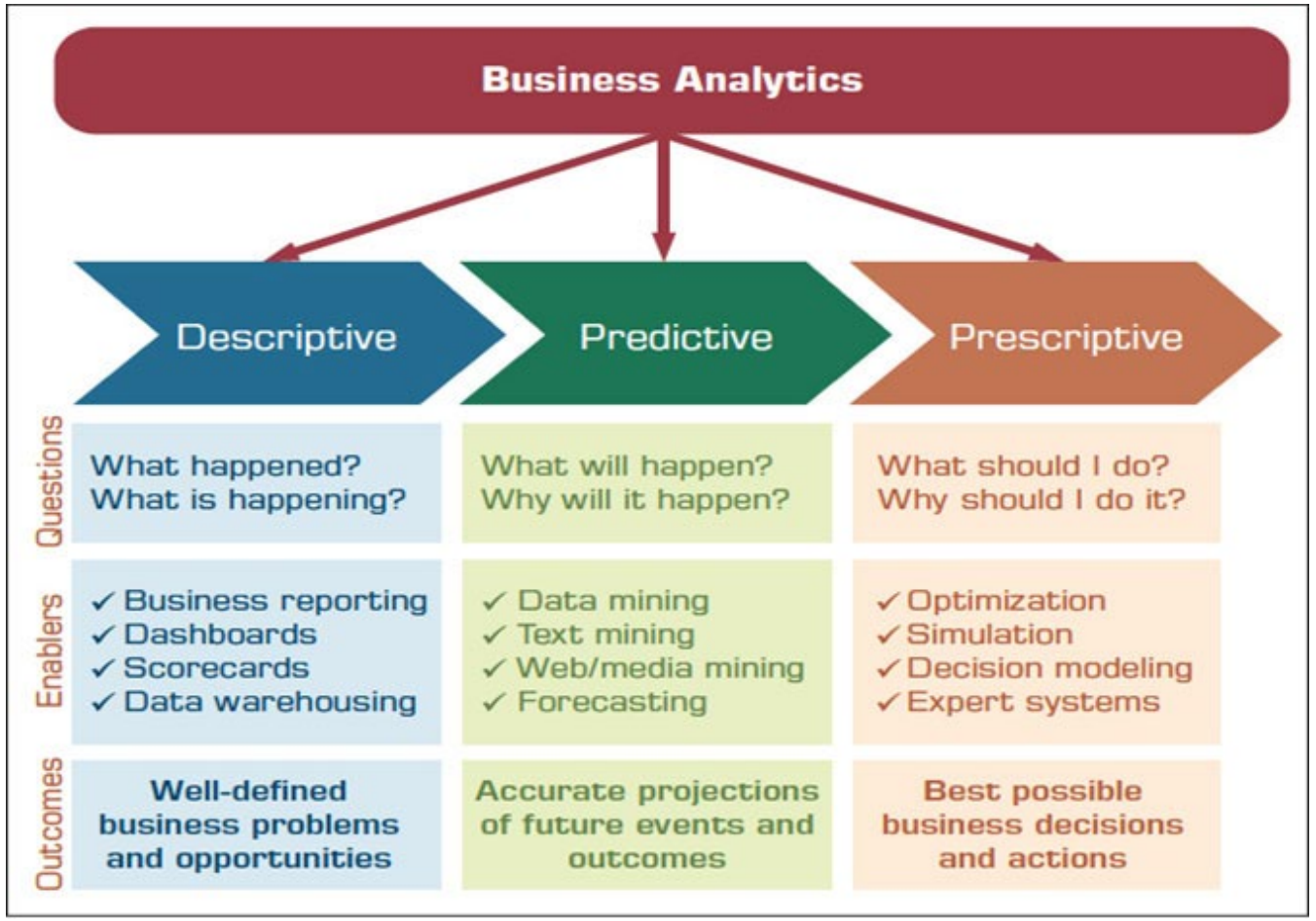
FULBRIGHT FINLAND FOUNDATION

Aalto University
School of Business

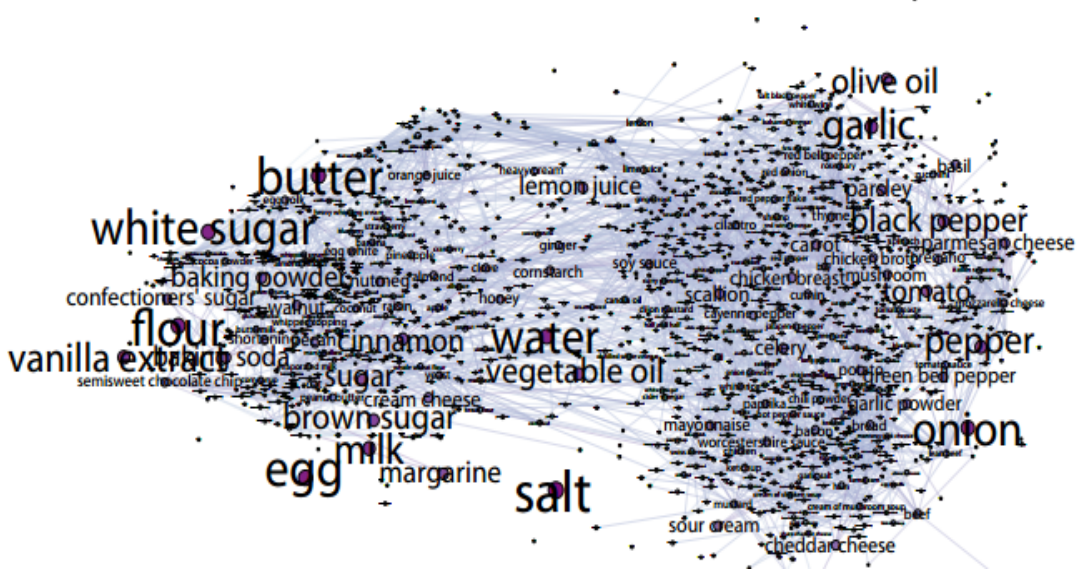
Research Theme – Impactful Applications

- Operations Research Modeling
- Design Science
- Decision Support Systems
- Analytics/Data Science/Machine Learning

Types of Analytics

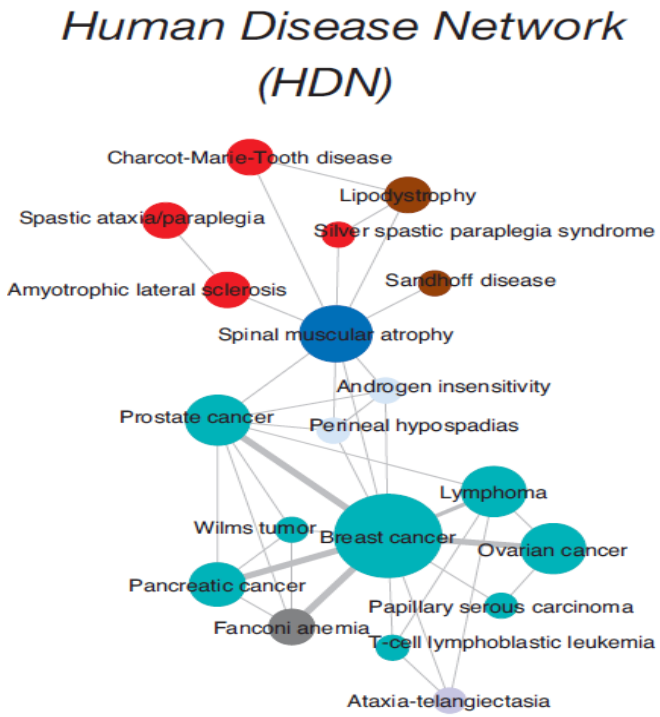


Network Science



- Network method – nodes and relationships
  - Explicit Networks (Facebook Network)
  - Implicit/Inferred Networks (Comorbidity Network)

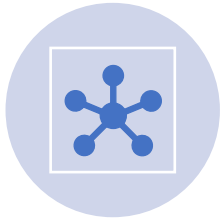
- Implicit Network
- Relationships are inferred – similarity index



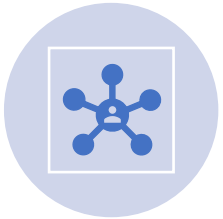
Network analytics research

Link predictions

Information diffusion



Sample the network



Impact of network on its nodes, which are internal to the network

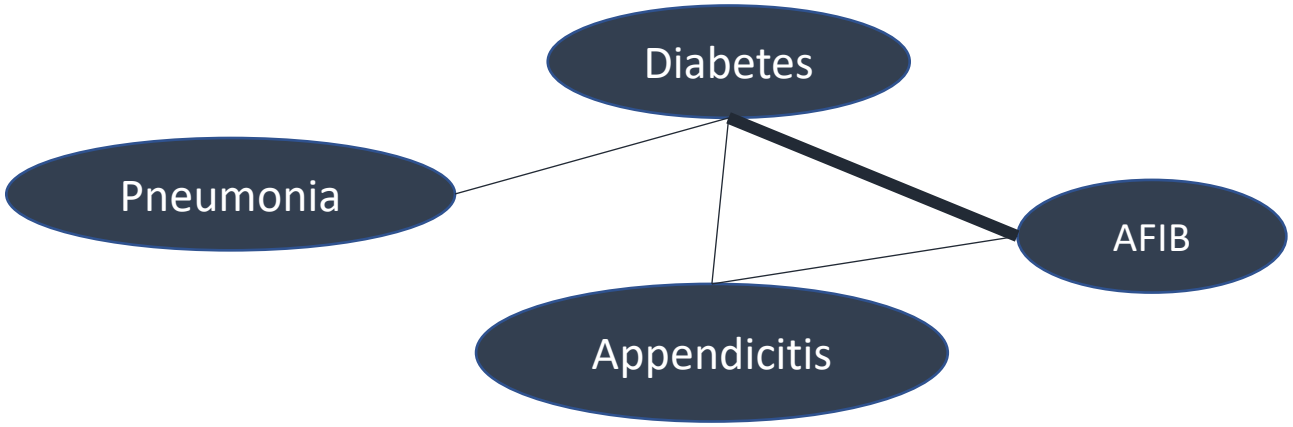
Recommendation engines

Limited work on understanding the impact of network on exogenous outcomes

Network of diseases

- Network of diseases – comorbidity network
  - diseases linked to each other based on co-occurrences in patients (comorbidity)

  - Node: a disease (diagnosis)
  - Edge: comorbidity/co-occurrence



Comorbidity Network

Kalgotra, P., Sharda, R., & Luse, A. (2020). Which similarity measure to use in network analysis: Impact of sample size on phi correlation coefficient and Ochiai index. *International Journal of Information Management*, 55, 102229.

- Network – nodes and edges
- Edges – similarity index
  - $Pearson's\ Correlation\ Coefficent$

$$PCC_{ij} = \frac{(c_{ij}{*}N) - (c_i{*}c_j)}{\sqrt{(c_i{*}c_j)(N - c_i)(N - c_j)}}$$

  - *Salton Cosine Index*

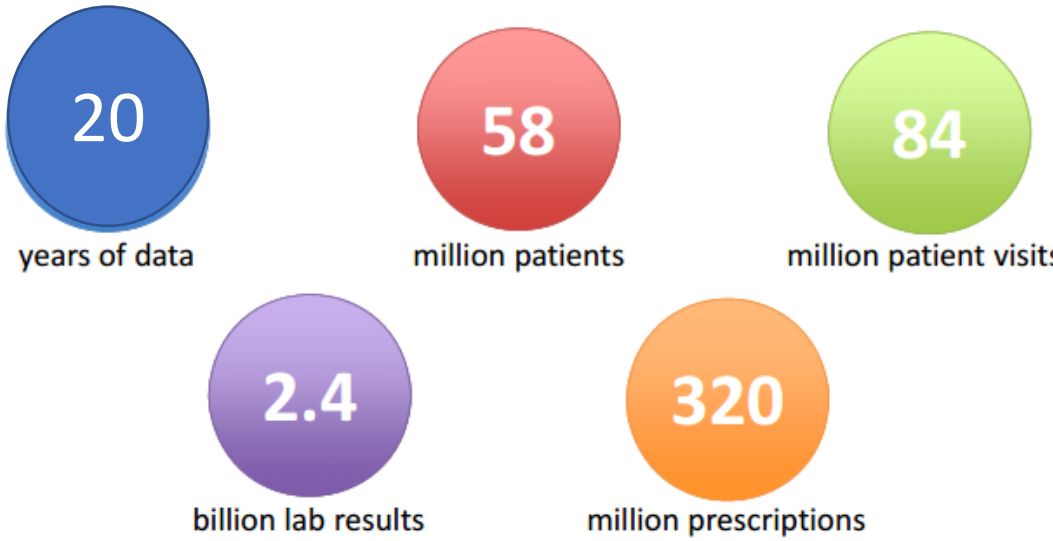
$$SCI_{ij} = \frac{(c_{ij})}{\sqrt{(c_i{*}c_j)}}$$

- *N* – number of transactions
- (c_i) - prevalence of node *i*
- (c_j) - prevalence of node *j*
- (c_{ij}) - prevalence of node *i and j* together

Our Data Asset

- *https://business.okstate.edu/chsi/*
- Electronic medical records (EMR) – Patient history
- Cerner database
  - more than 50 million patients' visits
- Disease is measured in ICD-9-CM codes

428 Heart failure
- 428.0 Congestive heart failure, unspecified
- 428.1 Left heart failure
- 428.2 Systolic heart failure
  - 428.20 Systolic heart failure, unspecified
  - 428.21 Acute systolic heart failure
  - 428.22 Chronic systolic heart failure
  - 428.23 Acute on chronic systolic heart failure



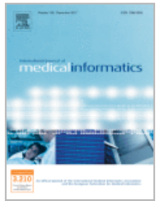
Descriptive Analytics Applications

How do healths differ across genders/races? Looking at comorbidities

ELSEVIER

International Journal of Medical Informatics

Volume 108, December 2017, Pages 22-28



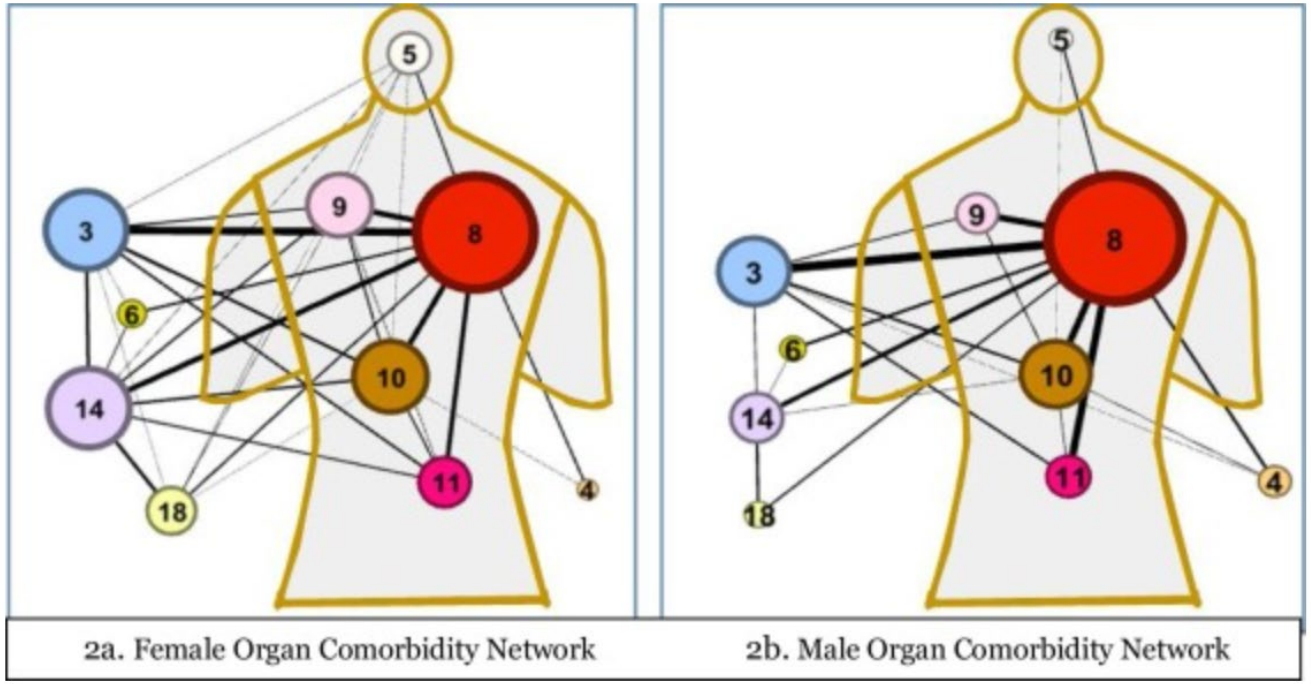
Research Paper

Examining health disparities by gender: A multimorbidity network analysis of electronic medical record

Pankush Kalgotra [a], Ramesh Sharda [b], Julie M. Croff [c]

Show more

Descriptive Analytics



2a. Female Organ Comorbidity Network 2b. Male Organ Comorbidity Network

3	Endocrine, nutritional and metabolic diseases, and immunity disorders	240–279
4	Diseases of the blood and blood-forming organs	
5	Mental disorders	290–319
6	Diseases of the nervous system	320–359
8	Diseases of the circulatory system	390–45
9	Diseases of the respiratory system	460–519
10	Diseases of the digestive system	520–579
11	Diseases of the genitourinary system	580–629
14	Diseases of the musculoskeletal system and connec tissue	710–739
18	Injury and poisoning	800–999

Discussion

- **Differences**
  - **more comorbidities of mental disorder in females**
  - **same with disorders of genitourinary system and musculoskeletal system**
  - **however, disorders of metabolic and immunity systems are related to blood-forming in males**
  - **chronic kidney and heart conditions also more connected in males**
- **Obesity and HIV infections more in males**
- **Care seeking behavior may increase the likelihood of diagnoses in females**
- **Public health implications for research and policy**

Descriptive Analytics: *Extending the analysis to study differences in groups based on race*

scientific reports

Explore content ˅ About the journal ˅ Publish with us ˅

nature › scientific reports › articles › article

Article | Open Access | Published: 11 August 2020

Examining multimorbidity differences across racial groups: a network analysis of electronic medical records

Pankush Kalgotra ✉, Ramesh Sharda & Julie M. Croff

Scientific Reports **10**, Article number: 13538 (2020) | Cite this article

3673 Accesses | **27** Citations | Metrics

Abstract

Descriptive Analytics

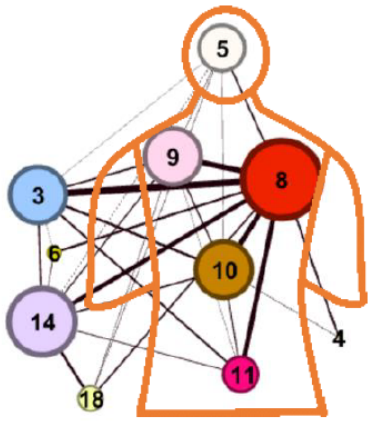


Fig 2a. White

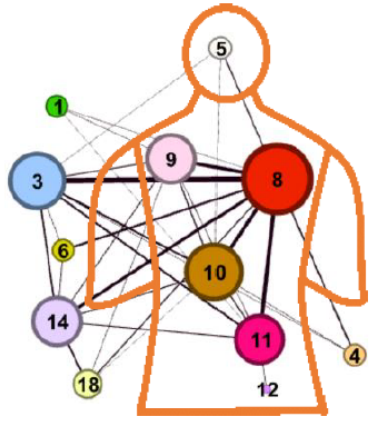


Fig 2b. African-American

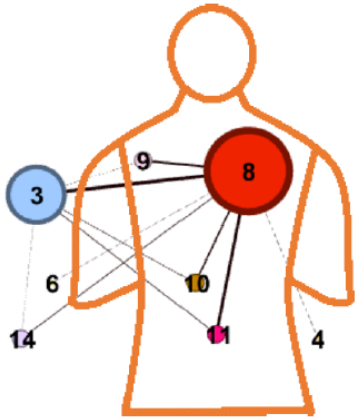


Fig 2c. Asian Network

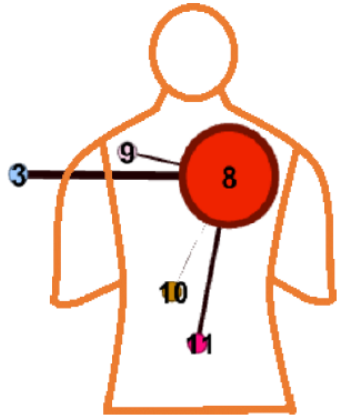


Fig 2d. Hispanic

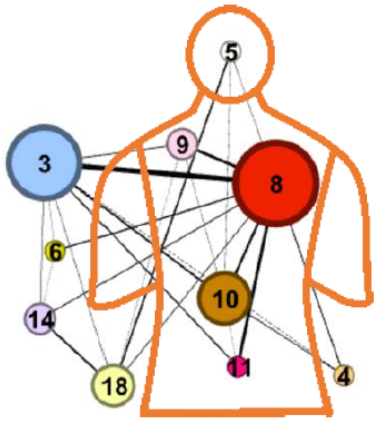


Fig 2e. Native American

Fig 2. Multimorbidity Network by Race at Organ System Level

Predictive Analytics Applications

Journal of Management Information Systems ›
Volume 38, 2021 - Issue 4: Special Issue: Fake News on the Internet

Journal homepage

Enter keywords, authors, DOI, ORC

Research Article

When will I get out of the Hospital? Modeling Length of Stay using Comorbidity Networks

Pankush Kalgotra & Ramesh Sharda

Pages 1150-1184 | Published online: 02 Jan 2022

Download citation https://doi.org/10.1080/07421222.2021.1990618

Predictive Analytics

- Predicting hospital length of stay at the time of admission
- Used a network approach to create a construct of “probable” comorbidity
- Multidimensional comorbidity comprising historical and probable diseases

$$\begin{bmatrix} LOS & Pchar_1 & \dots & Pchar_k & d_{o1} & d_{o2} & \dots & d_{on} \end{bmatrix} \begin{bmatrix} d_{h1} & d_{h2} & d_{h3} & \dots & d_{hn-1} & d_{hn} \\ d_{p1} & d_{p2} & d_{p3} & \dots & d_{pn-1} & d_{pn} \end{bmatrix}$$

Predictive Analytics

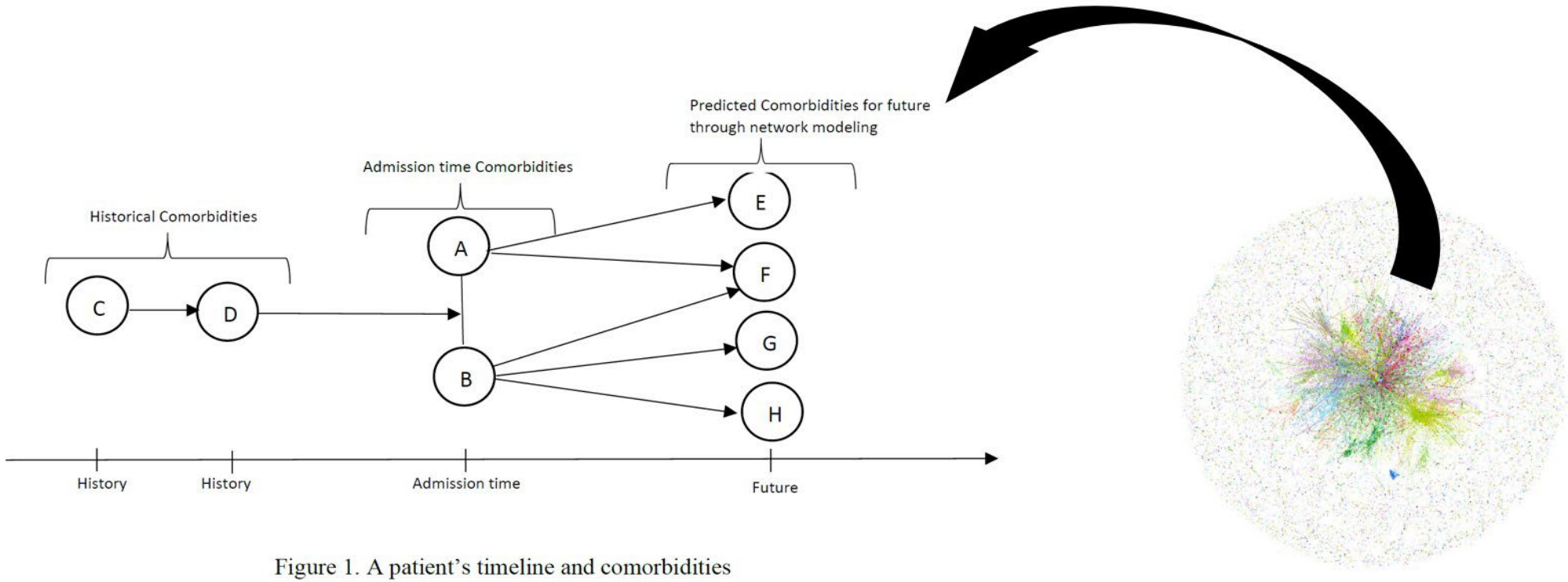


Figure 1. A patient's timeline and comorbidities

$$\begin{bmatrix} LOS & Pchar_1 & \dots & Pchar_k & d_{o1} & d_{o2} & \dots & d_{on} & \begin{matrix} d_{h1} & d_{h2} & d_{h3} & \dots & d_{hn-1} & d_{hn} \\ d_{p1} & d_{p2} & d_{p3} & \dots & d_{pn-1} & d_{pn} \end{matrix} \end{bmatrix}$$

Predictive Analytics

- **Different LSTM architecture for different categories of diseases**

Table 6. Deep Learning Modeling Results

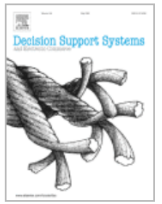
Category of admission	No. of visits	Avg. LOS, Std. Dev	*MPres* MAE in days (Baseline)	*MuCOM* MAE in days	*MuCOM* MAPE	Improvement		Exact Prediction with Tolerance[5]	
						Percentage	No. of days	0-Day	1-Day
1-Infectious and parasitic diseases	432,531	2.89, 3.45	1.16[1]	**1.129**[2]	26.9%	2.67	13,408	52.3%	69.7%
2-Neoplasms	255,256	4.34, 4.09	2.06	**2.025**	56%	1.70	8,934	27.9%	59.4%
3-Endocrine, nutritional and immunity disorders	674,413	3.59, 3.5	1.62[1]	**1.579**	46%	2.53	27,651	35.6%	67.9%
4-Diseases of the blood and blood-forming organs	256,962	4.26, 3.95	2.0[1]	**1.963**[3]	53%	2.0	9,508	28.1%	59.3%
5-Mental disorders	676,160	3.72, 4.08	1.65[1]	**1.599**	38.3%	3.09	34,484	45.4%	68.8%
6-Diseases of the nervous system	316,279	3.22, 3.52	1.44[1]	**1.40**	42%	2.78	12,651	45.6%	71.1%
7-Diseases of the sense organs	305,932	1.61, 1.97	0.45[1]	**0.41**[1]	13.5%	8.89	12,237	81.8%	92.1%
8-Diseases of the circulatory system	1,007,110	3.59, 3.34	1.67	**1.655**	51%	0.90	15,107	33.7%	66%
9-Diseases of the respiratory system	1,124,062	2.46, 2.86	0.90	**0.899**	27%	0.11	1,124	62.7%	81.4%
10-Diseases of the digestive system	955,490	3.01, 3.18	1.274[1]	**1.25**	36%	1.88	22,932	49.8%	73.5%
11-Diseases of the genitourinary system	719,537	2.56, 2.91	0.998[1]	**0.965**	27%	3.31	23,745	59.7%	81.2%
12-Complications of pregnancy, childbirth, and the puerperium	699,886	2.44, 1.75	0.658[1]	**0.655**[1]	24%	0.46	2,100	58%	91.4%
13-Diseases of skin and subcutaneous tissue	363,362	2.12, 2.55	0.736[1]	**0.695**[2]	19%	5.57	14,898	70.7%	85.6%
14-Diseases of the musculoskeletal system and connective tissue	970,747	1.96, 2.32	0.625[1]	**0.618**[1]	19%	1.12	6,795	73%	87.8%
15-Congenital anomalies	55,909	4.29, 4.56	2.03	**2.017**[4]	50%	0.64	727	34.7%	63.5%
16-Certain conditions originating in the perinatal period	84,375	4.26, 4.93	1.69	**1.68**	35%	0.59	844	46.5%	74.4%
17-Symptoms, signs, and ill-defined conditions	1,200,000	1.91, 2.28	0.653[1]	**0.636**	21%	2.61	20,400	71.2%	87.4%
18-Injury and poisoning	1,200,000	1.68, 2.16	0.47[1]	**0.467**[1]	14%	0.64	3,600	81.8%	91%
19-External causes of injury and supplemental classification	1,028,695	2.82, 3.03	0.92[1]	**0.907**	26.8%	1.41	13,373	60.9%	85.8%
Total		2.25[7]	1.04[8]	1.02[8]	29.8%[8]	1.9%[8]	244,518[6]	58%[9]	79.7%[9]

[1]Adam optimizer was used. [2]The outputs from LSTM layer are 20, which feed to fully connected layer of 100 neurons followed by another 100 neurons. [3]The output from LSTM layer is one, which feeds to fully connected layer of 100 neurons followed by another 100 neurons. [4]The output from LSTM layer is 10, which feed to fully connected layer of 10 neurons followed by another 10 neurons. [5]Number of visits with an error of zero and one day. [6]Sum. [7]Average across all visits. [8]Average adjusted by number of records. [9]Percentage of visits across all types

Predictive Analytics

- Small but significant increase in predictive accuracy of LOS
- Method innovation - use of network properties in a predictive model
- Deep Learning

Decision Support Systems
Volume 168, May 2023, 113929



Quantifying disease-interactions through co-occurrence matrices to predict early onset colorectal cancer

Pankush Kalgotra [a], Ramesh Sharda [b], Sravanthi Parasa [c]

https://doi.org/10.1016/j.dss.2023.113929

Introduction

- Colorectal cancer (CRC) is the third most common cause of cancer incidence and death
- The median age of CRC diagnosis was 72 years in 2001-2002, which decreased to 66 years in 2015-2016 (Siegel et al. 2020)
- Rising number of CRC cases in populations less than 50 years of age in the past few years
- According to American Cancer Society (ACS) - 51% increase in CRC among those under age 50 since 1994

In our Data

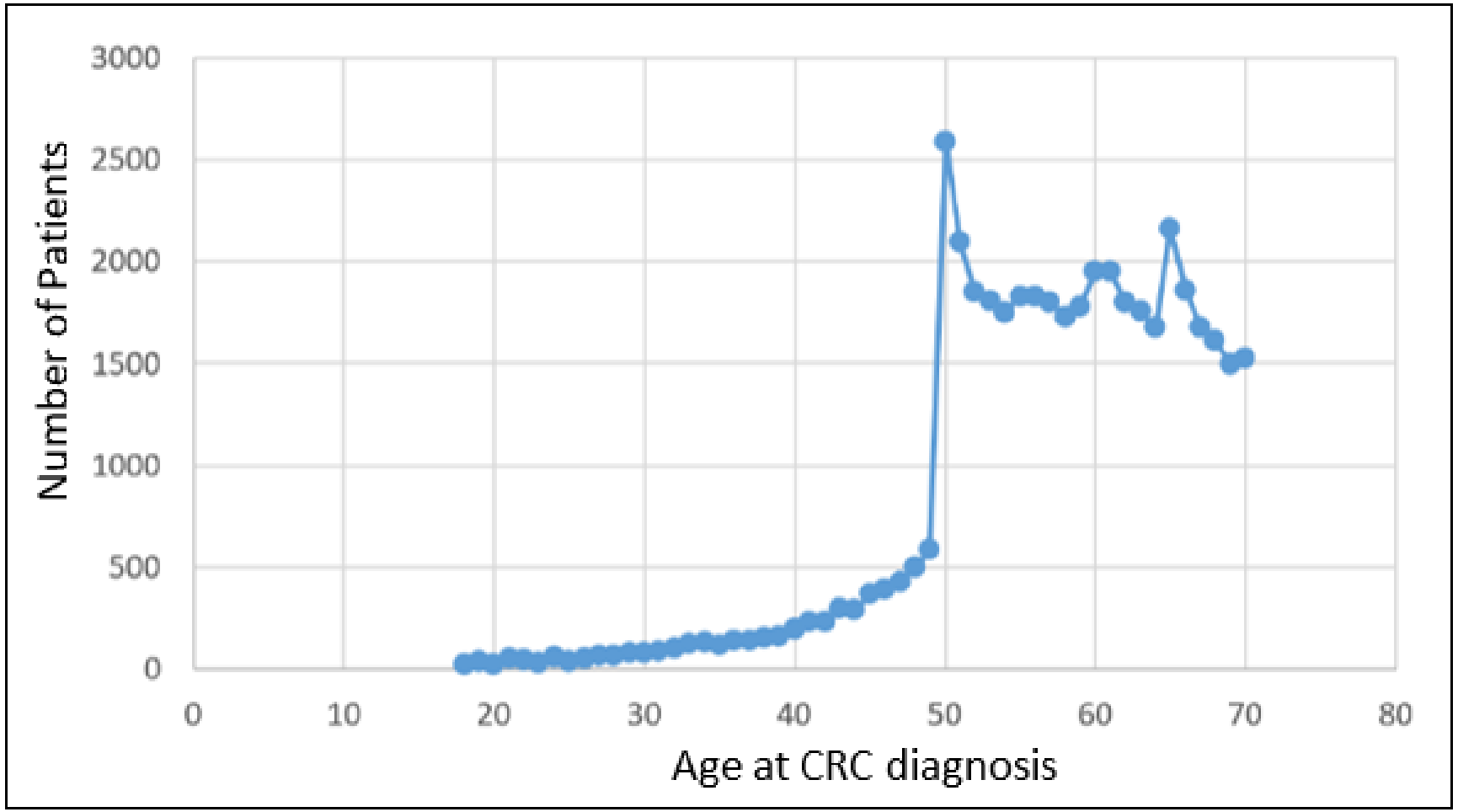


Figure 1. Age distribution at diagnosis visit

Objectives

- **Predictive Models for a population < 50 years**
- **To develop a method to classify patients based on their historical disease patterns**

Problem Formulation

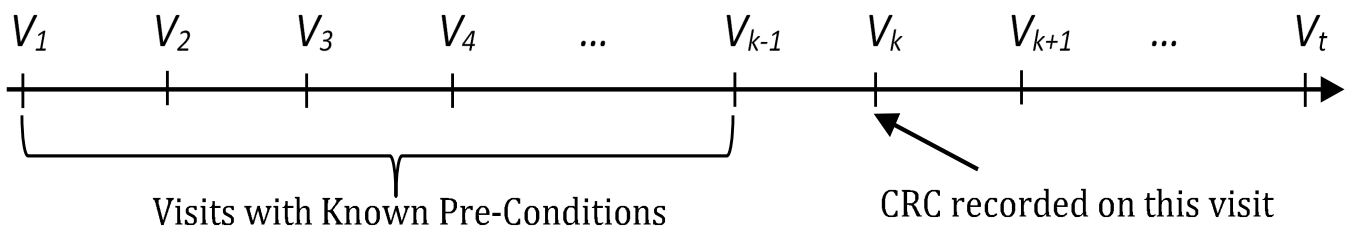


Figure 1. A Patient Timeline

- Given the universal disease set *D* containing *n* diseases, each visit has a set of diseases, S_i, recorded on a visit V_i where $i := 1, 2 .. t$ and $S \subseteq D$.

- All pre-conditions (R_j) known for a patient *j* before the k^{th} visit in Figure 1 were used for creating c-occurrence matrix and predicting CRC

$$R_j = S_1 \cup S_2 \cup S_3 \cup S_4 \cup \ldots S_{k-1}$$

Co-occurrence Matrix for CRC patients

- To compute the strength of a disease-pair comprising two diseases *x* and *y*, where {*x*, *y*} ∈ *D*, c_{xy} is computed by Jaccard's Index as

$$c_{xy} = \frac{|X \cap Y|}{|X \cup Y|}$$

$$C = \begin{bmatrix} 0 & c_{12} & \dots & c_{1n} \\ c_{21} & 0 & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & \dots & & 0 \end{bmatrix}$$

Co-occurrence Matrix for non-CRC patients

- A transaction comprises of all diseases of a patient recorded prior to the last known visit

$$c'_{xy} = \frac{|X \cap Y|}{|X \cup Y|}$$

$$C' = \begin{bmatrix} 0 & c'_{12} & \dots & c'_{1n} \\ c'_{21} & 0 & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ c'_{n1} & \dots & & 0 \end{bmatrix}$$

Disease-Interaction Variables for a New Patient

- A new patient coming with a set of diseases *R* with a cardinality *r* and $R \subset D$
- With *r* diseases, total number of combinations possible is $(r(r\text{-}1)/2)$
- For each disease-pair $\{g, h\} \in R$ with $g<h$, the co-occurrence matrices provide c_{gh} and c'_{gh} where $g=x$ and $h=y$

$$SumC = \sum_{a \in g} \sum_{b \in h} c_{ab}$$

$$AverageC = \frac{\sum_{a \in g} \sum_{b \in h} c_{ab}}{\frac{r(r-1)}{2}}$$

$$SumC' = \sum_{a \in g} \sum_{b \in h} c'_{ab}$$

$$AverageC' = \frac{\sum_{a \in g} \sum_{b \in h} c'_{ab}}{\frac{r(r-1)}{2}}$$

Variable Label	Description	CRC (18-49)	No CRC (18-49)
Cardiovascular risk factors	1 – Presence, 0 – Absence	40.9%	20.0%
Diabetes Mellitus	1 – Presence, 0 – Absence	14.9%	5.5%
Personal history of digestive problems	1 – Presence, 0 – Absence	3.3%	0.46%
Flatulence, peristalsis, abnormal bowel sounds, fecal incontinence	1 – Presence, 0 – Absence	2.9%	0.67%
Abnormal feces	1 – Presence, 0 – Absence	0.08%	0.03%
Diarrhea	1 – Presence, 0 – Absence	11.1%	3.9%
Other digestive symptoms	1 – Presence, 0 – Absence	2.2%	0.25%
Abdominal pain	1 – Presence, 0 – Absence	37.2%	20.0%
Iron deficiency anemia	1 – Presence, 0 – Absence	3.8%	1.48%
Other Anemias	1 – Presence, 0 – Absence	11.4%	6.1%
Rectal bleeding	1 – Presence, 0 – Absence	7.1%	0.75%
GI hemorrhage	1 – Presence, 0 – Absence	6.9%	1.18%
Diverticula of colon	1 – Presence, 0 – Absence	4.7%	0.75%
Intestinal obstruction	1 – Presence, 0 – Absence	0.84%	0.19%
Anal/Rectal polyp	1 – Presence, 0 – Absence	0.5%	0.03%
Gender	1 – Men, 0 – Women	Men-41.03%	Men-36.9%
Race*	Five binary variables representing Whites, African Americans, Native Americans, Hispanics and Asians	W-68%; AA-13.4%; NA-2.8%%; H-1.05%; A-1.2%	W-61.35%; AA-16.5%; NA-1.3%; H-2.1%; A-1.7%
Age	Latest known age; Continuous variable	40.2 years	32.8 years
Number of diseases (3-digit ICD9)	Continuous variable	11.8	8.1
Number of distinct organ systems	Continuous variable	5.6	4.1
SumC	Continuous variable	Average – 8.9	Average – 4.3
SumC'	Continuous variable	Average – 6.1	Average – 3.4
AverageC	Continuous variable	Average – 0.07	Average – 0.069
AverageC'	Continuous variable	Average – 0.05	Average – 0.057

Results

Table 4. Modeling Results

Model	Accuracy (%), SD	AUC, SD	Sensitivity, (%), SD	Specificity, (%), SD
Model 1: Age	67.5, 0.5	0.73, 0.01	71.1, 1.8	64.0, 1.8
Model 2: Age, *Dem*	68.5, 0.6	0.74, 0.01	72.1, 1.7	64.9, 1.7
Model 3: Age, *Dem*, *Sym*	72.5, 0.6	0.80, 0.01	73.6, 1.1	71.5, 1.5
Model 4: Age, *Dem*, *DI*	70.5, 0.6	0.77, 0.01	73.4, 0.8	67.6, 1.1
Model 5: *Sym*	67.5, 0.6	0.73, 0.01	60.6, 2.2	**74.4, 2.6**
Model 6: *DI*	67.0, 0.5	0.72, 0.01	65.7, 1.0	68.3, 1.4
Model 7: Age, *Dem*, *Sym*, *DI*	**73.2, 0.4**	**0.81, 0.01**	**75.3, 0.6**	71.1, 0.8

SD-Standard Deviation

A Prescriptive Analytics Application
(Work in progress)

Traps: Identifying Mortality-related Cliques in a Comorbidity Network – in progress

- **To model combinations**
- **To develop a method to find hidden combinations related to an outcome in a situation**

Solution

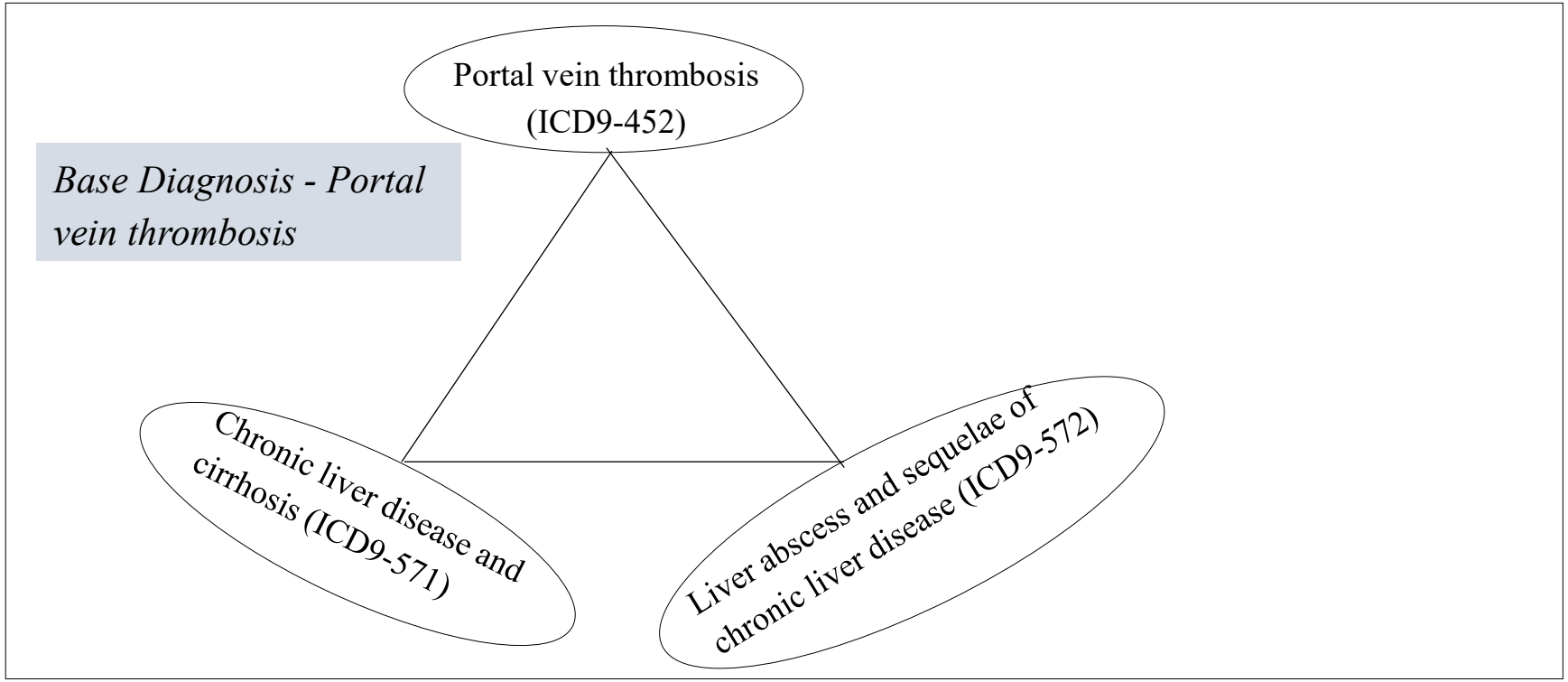
- **Network Theory to model combinations**
  - **Latent Network**
  - **Identifying Cliques in the network, which are highly related to an outcome**

Demonstration: Mortality related Cliques

- **Causes of mortality – diseases, actually multiple diseases**
- **Direct and indirect interactions of diseases may be related to mortality**
- **Which combinations of diseases are critical?**

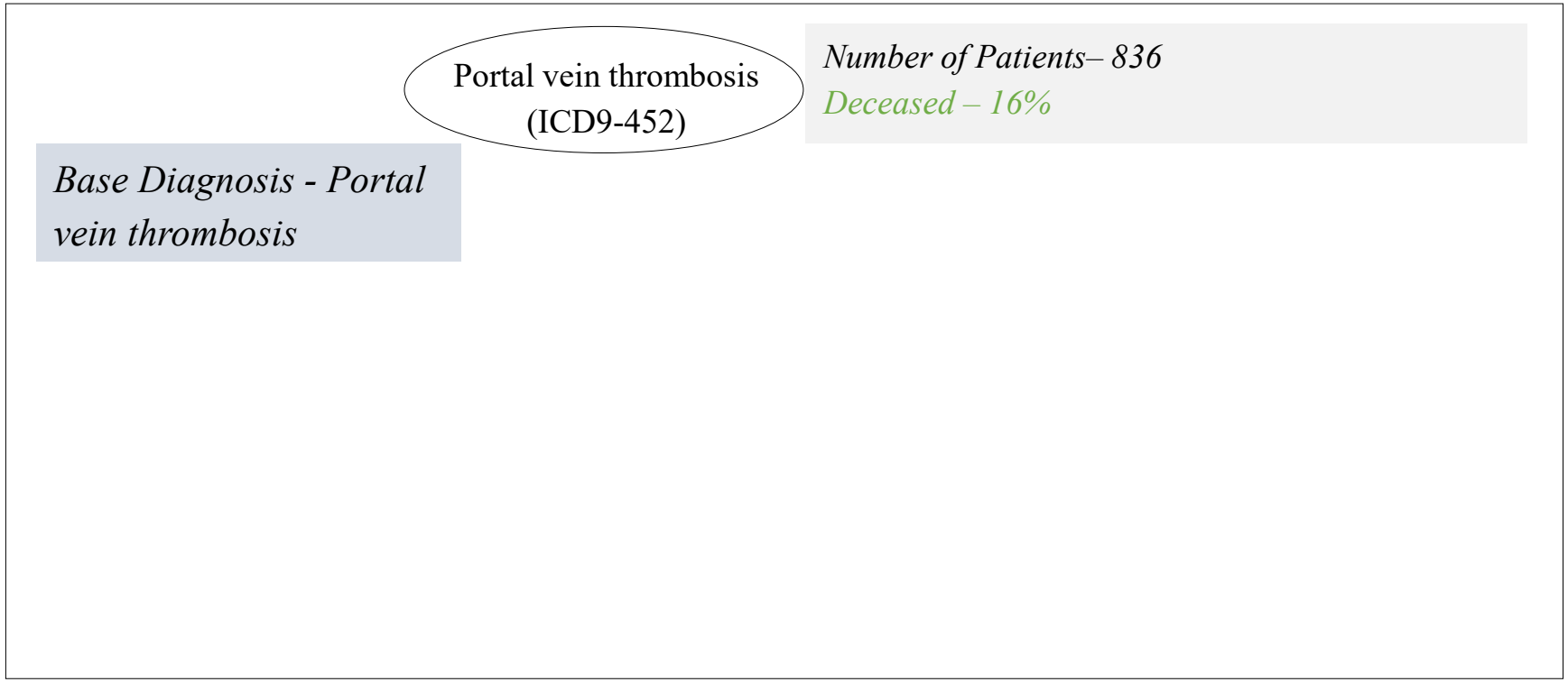
Demonstration: Mortality related Cliques

- Identify cliques of diseases with a higher mortality rate
- Objective function – non-linear

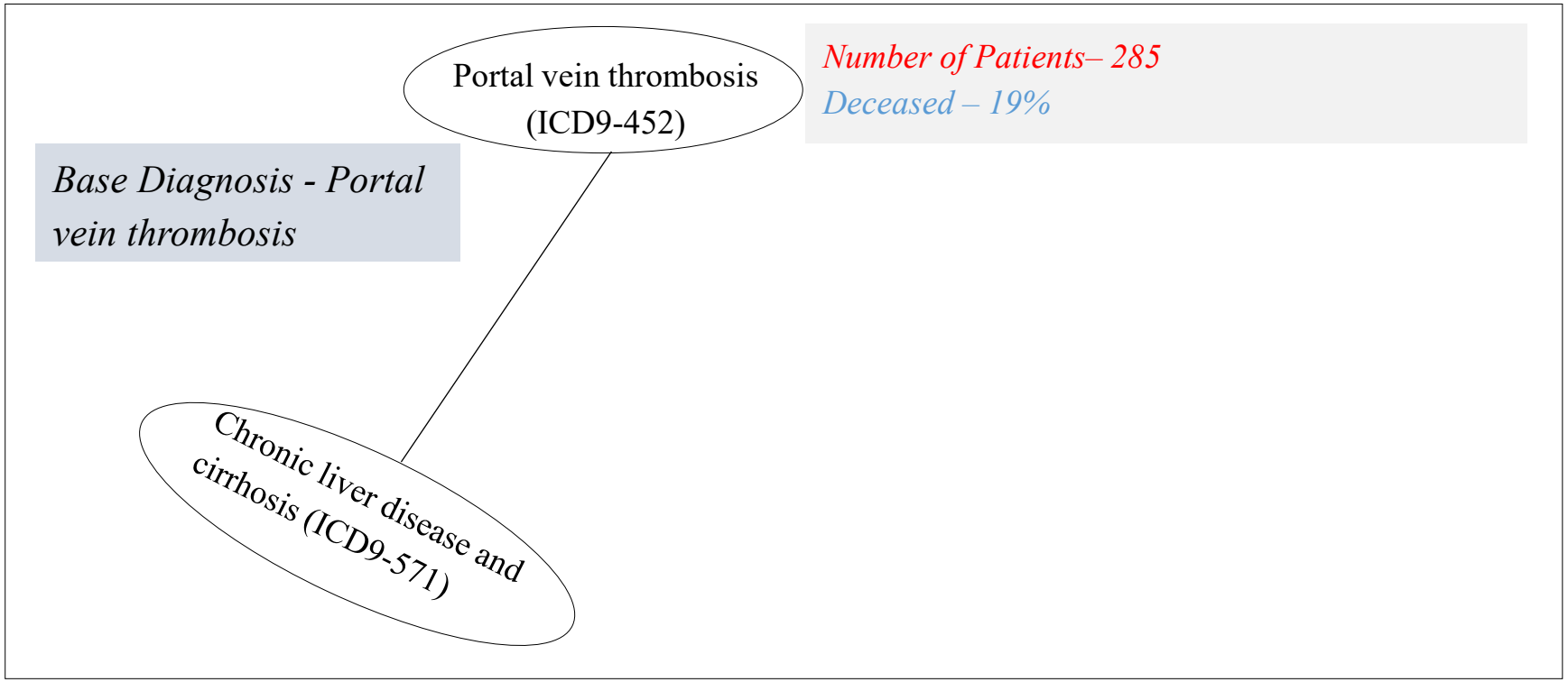


Example - A clique/triangle of three diseases with their joint impact on mortality

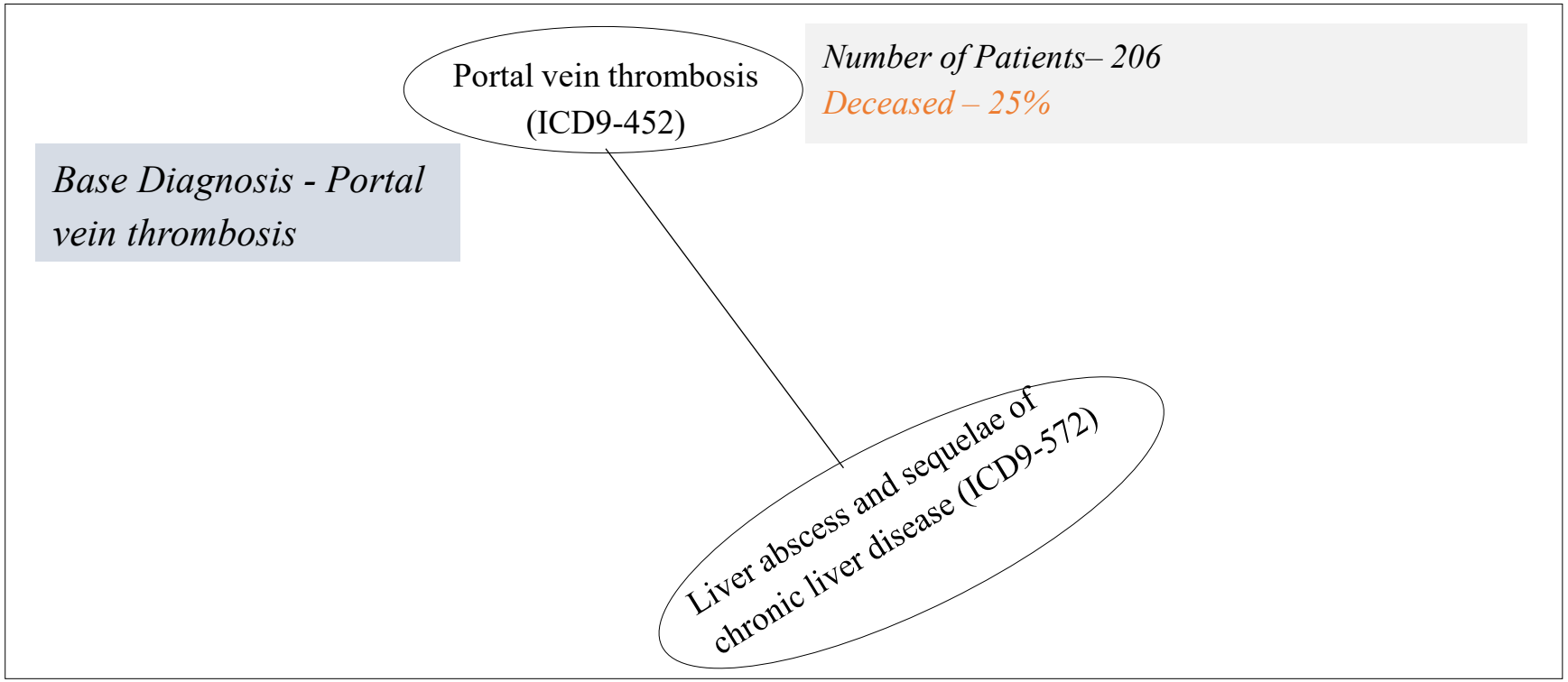
Results


Portal vein thrombosis
(ICD9-452)
Number of Patients– 836
Deceased – 16%
Base Diagnosis - Portal vein thrombosis

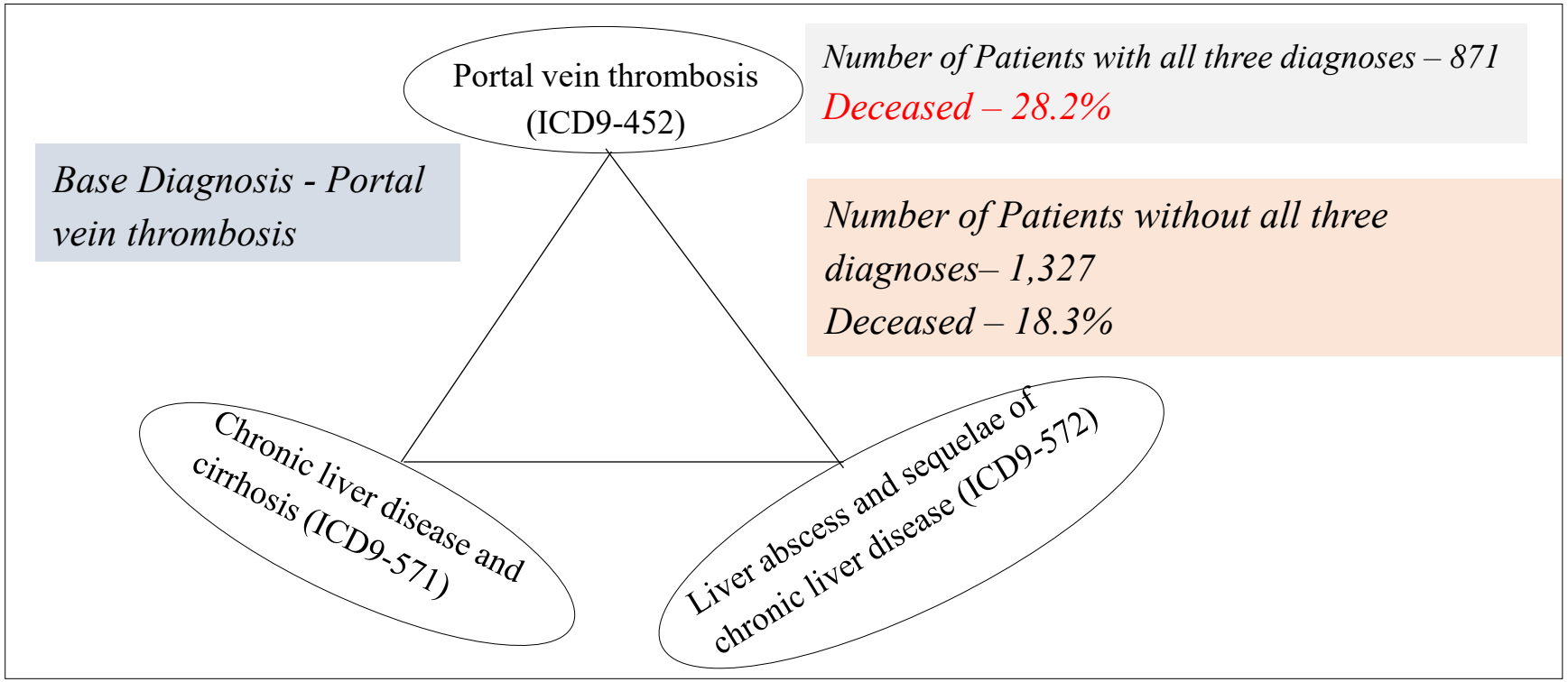
Results


Portal vein thrombosis (ICD9-452)
Number of Patients– 285
Deceased – 19%
Base Diagnosis - Portal vein thrombosis
Chronic liver disease and cirrhosis (ICD9-571)

Results


Portal vein thrombosis (ICD9-452)
Number of Patients– 206
Deceased – 25%
Base Diagnosis - Portal vein thrombosis
Liver abscess and sequelae of chronic liver disease (ICD9-572)

Results



A clique/triangle of three diseases with their joint impact on mortality

Potential Contributions

- Clique/combination of event as a trap
- Outcome related cliques - algorithmic contribution
- Theorize phenomenon using network properties
- Method applicable to problems where latent interactions affect an outcome

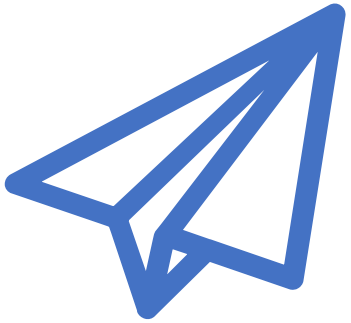
Acknowledgements

This work was conducted with data from the Cerner Corporation’s Health Facts database of electronic medical records provided by the Oklahoma State University Center for Health Systems Innovation (CHSI). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Cerner Corporation.

Other Recent Interests

- **Other Network Science Applications**
- **Interruptions management**
- **Citizen science – employing analytics to make a societal impact**
- **Your interests !**

Looking forward to Collaborations!



Email: ramesh.sharda@okstate.edu

Thank You