Big Data Analytics for Healthcare

Bonnie L. Westra, PhD,, RN, FAAN, FACMI
Associate Professor and Director,
Center for Nursing Informatics
January 21, 2016

UNIVERSITY OF MINNESOTA
School of Nursing
Driven to Discover℠

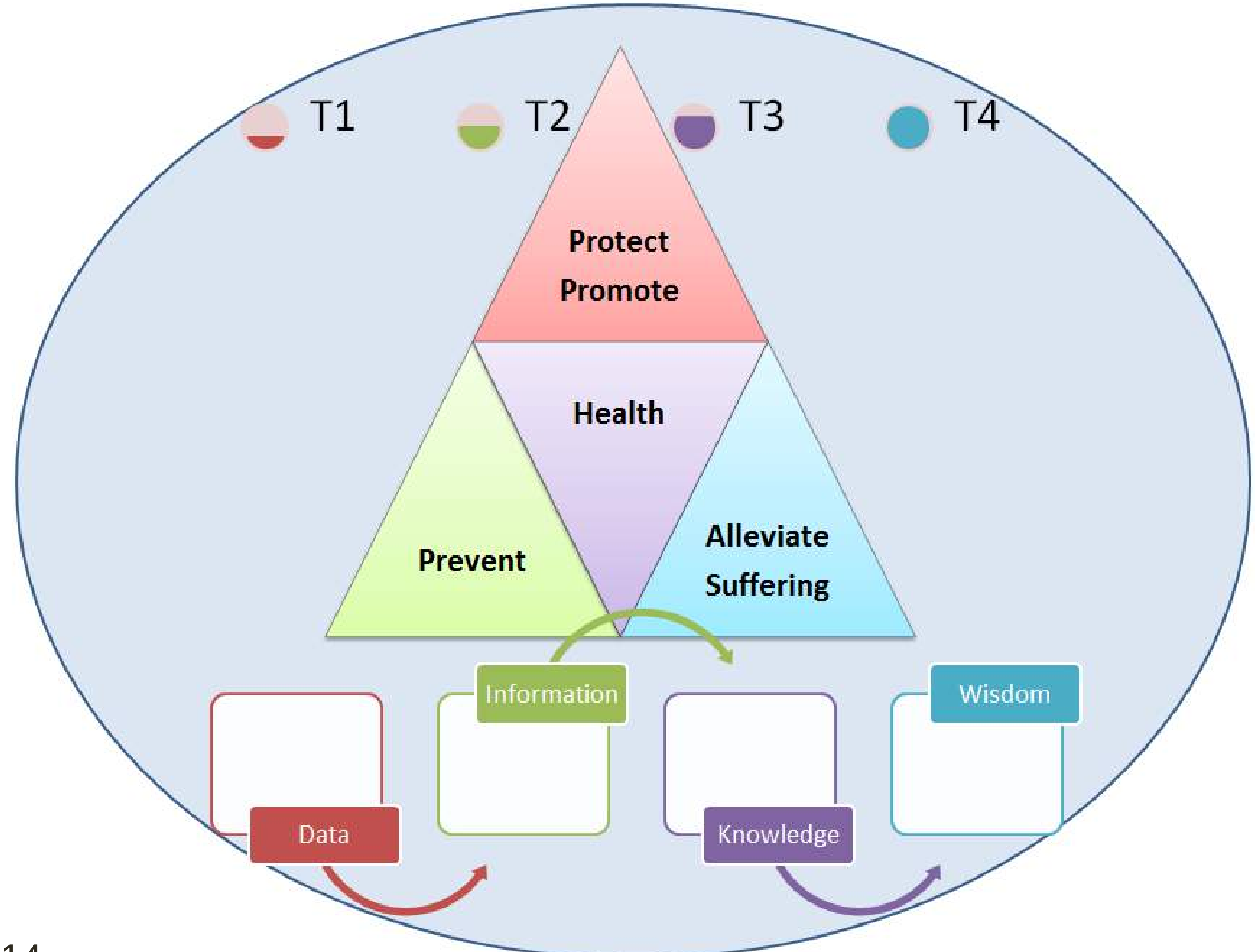
Objectives

- Relate big data and data science to research and quality improvement questions important to practice.
- Identify critical steps to make data useful for big data analytics
- Explore examples big data science research methods and lessons learned.

Big Data Science

- Application of math to large data sets to infer probabilities for associations/ prediction
- Purpose is to accelerate discovery, improve critical decision-making processes, enable a data-driven economy[1]
- Three-legged stool
  - **Data**
  - Technology
  - **Algorithms**

Big Data Analytics



Nursing Informatics & Translational Science

Big Data

- Large volume
- Complex data
- Integration of multiple data sets
- Data over time

Geocodes

https://infocus.emc.com/william_schmarzo/thoughts-on-the-strata-rx-healthcare-conference/

NIH | NIH Big Data to Knowledge (BD2K)

CTSA Clinical & Translational® Science Awards

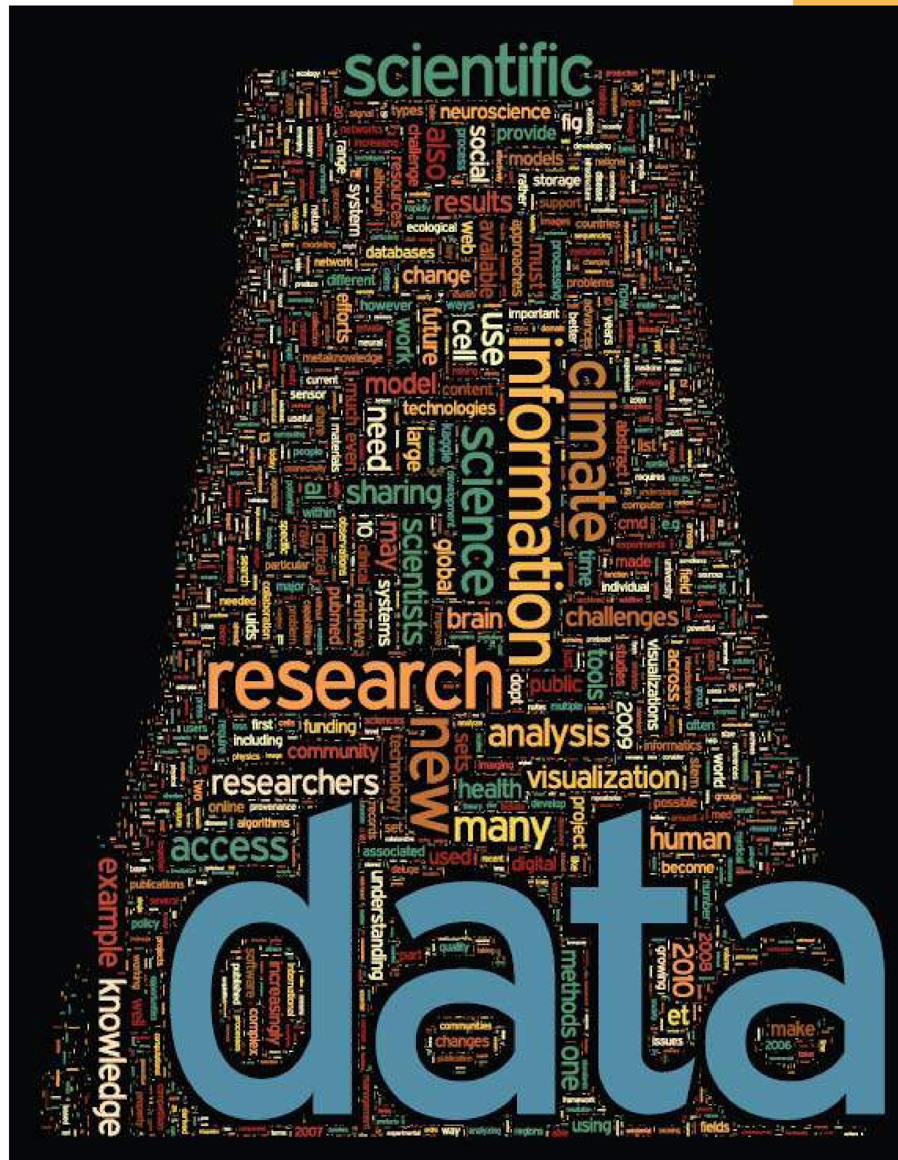
pcori®

NSF Announces Interagency Progress on Administration's Big Data Initiative

Harnessing the EHR for Research

- in areas of eScience such as
    - [data capture],
    - Databases,
    - Workflow management,
    - Visualization
    - Computing technologies.

Nursing Research Journal



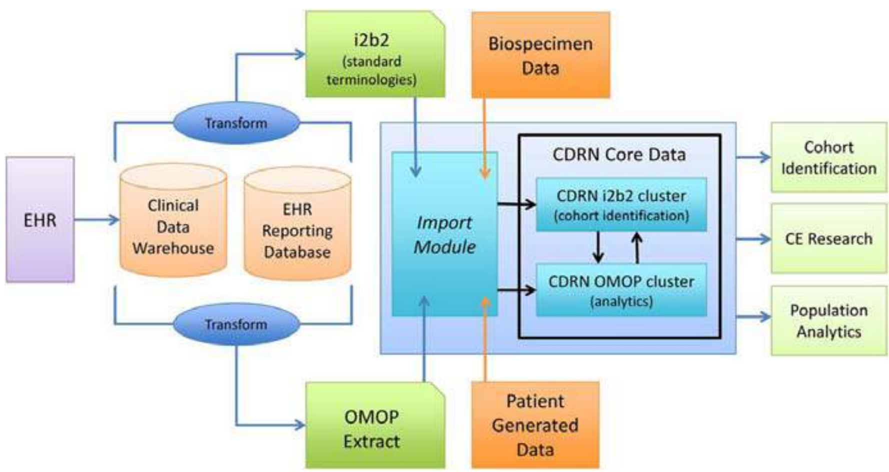
http://www.sciencemag.org/site/special/data/ScienceData-hi.pdf

Data Sources

- CTSA – https://ctsacentral.org/
  - NCATS - https://ncats.nih.gov/
- PCORnet - http://www.pcornet.org/
  - 13 clinical data research networks (CDRNs)
  - 22 patient powered research networks (PPRNs)
- Optum Labs – 140 million lives from claims data + 40 million from EHRs (delaney@umn.edu)
- http://www.data.gov/ - Search over 192,872 datasets

Requirements for Useful Data

- Common data models
- Standardized coding of data
- Standardize queries



PCORnet CDM Domains, v3.0

CONDITION v2.0
A condition represents a patient's diagnosed and self-reported health conditions and diseases. The patient's medical history and current state may both be represented.

DEATH v3.0
Reported mortality information for patients.

DEATH_CAUSE v3.0
The individual causes associated with a reported death.

DEMOGRAPHIC v1.0
Demographics record the direct attributes of individual patients.

DIAGNOSIS v1.0
Diagnosis codes indicate the results of diagnostic processes and medical coding within healthcare delivery.

DISPENSING v2.0
Outpatient pharmacy dispensing, such as prescriptions filled through a neighborhood pharmacy with a claim paid by an insurer. Outpatient dispensing is not commonly captured within healthcare systems.

ENROLLMENT v1.0
Enrollment is a concept that defines a period of time during which all medically-attended events are expected to be observed. This concept is often insurance-based, but other methods of defining enrollment are possible.

ENCOUNTER v1.0
Encounters are interactions between patients and providers within the context of healthcare delivery.

HARVEST v3.0
Attributes associated with the specific PCORnet datamart implementation

LAB_RESULT_CM v2.0
Laboratory result Common Measures (CM) use specific types of quantitative and qualitative measurements from blood and other body specimens. These standardized measures are defined in the same way across all PCORnet networks.

PCORNET_TRIAL v3.0
Patients who are enrolled in PCORnet clinical trials.

PRESCRIBING v3.0
Provider orders for medication dispensing and/or administration.

PRO_CM v2.0
Patient-Reported Outcome (PRO) Common Measures (CM) are standardized measures that are defined in the same way across all PCORnet networks. Each measure is recorded at the individual item level: an individual question/statement, paired with its standardized response options.

PROCEDURES v1.0
Procedure codes indicate the discreet medical interventions and diagnostic testing, such as surgical procedures, administered within healthcare delivery.

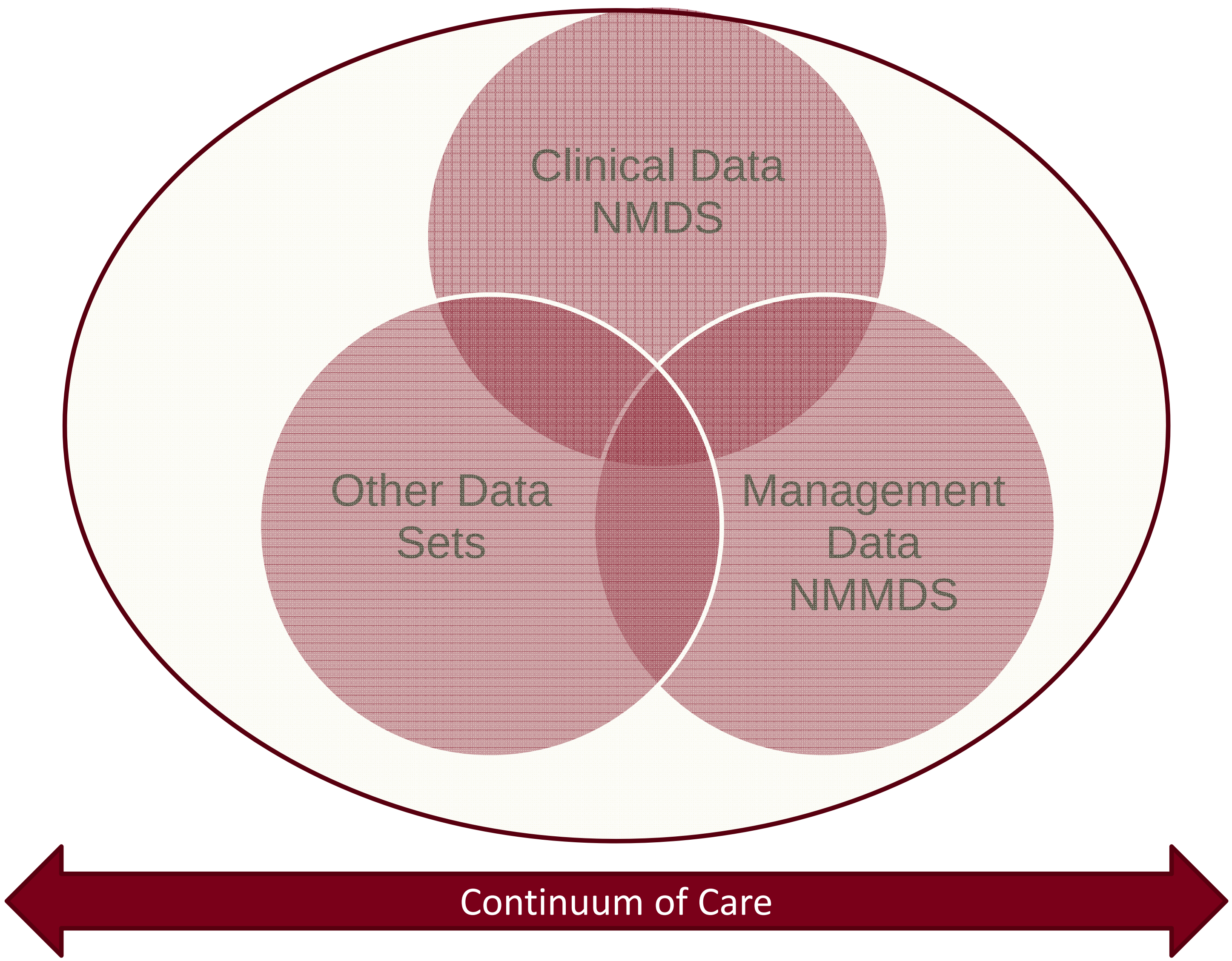
VITAL v1.0
Vital signs (such as height, weight, and blood pressure) directly measure an individual's current state of attributes.

http://www.pcornet.org/resource-center/pcornet-common-data-model/

Data Standardization

- Demographics – OMB
- Medications - RxNorm
- Laboratory data - LOINC
- Procedures – CPT, HCPCS, ICD, SNOMED CT
- Diagnoses - ICD-9/10-CM, SNOMED CT
- Vital status – CDC
- Vital signs - LOINC

Vision – Inclusion of Nursing Data

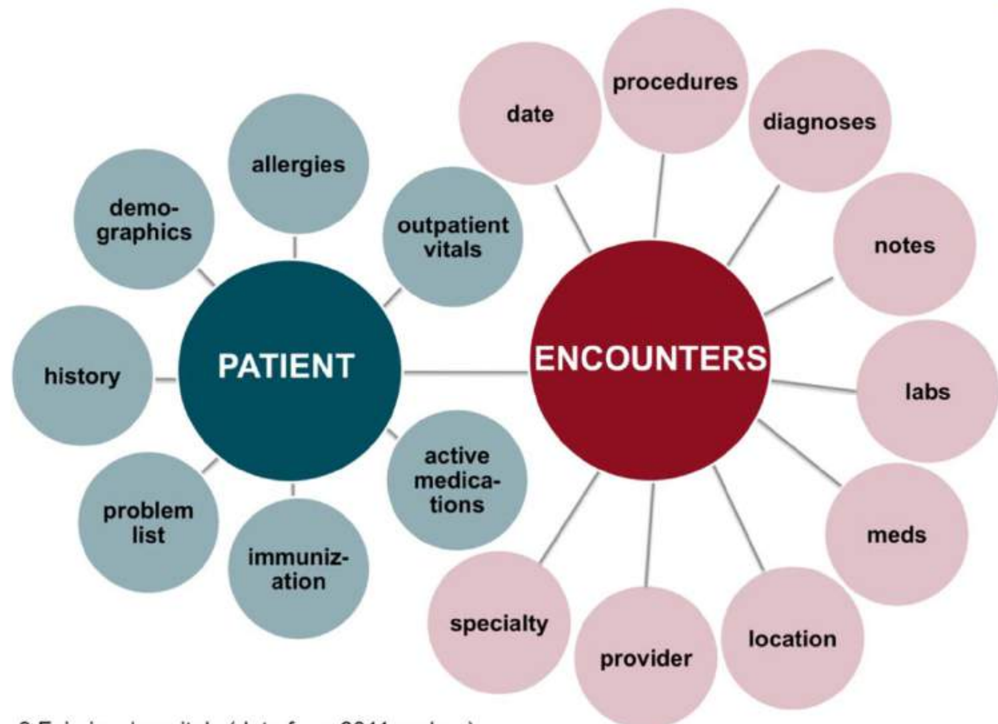


UMN Clinical Data Repository

Cohort discovery /recruitment

Observational studies

Predictive Analytics

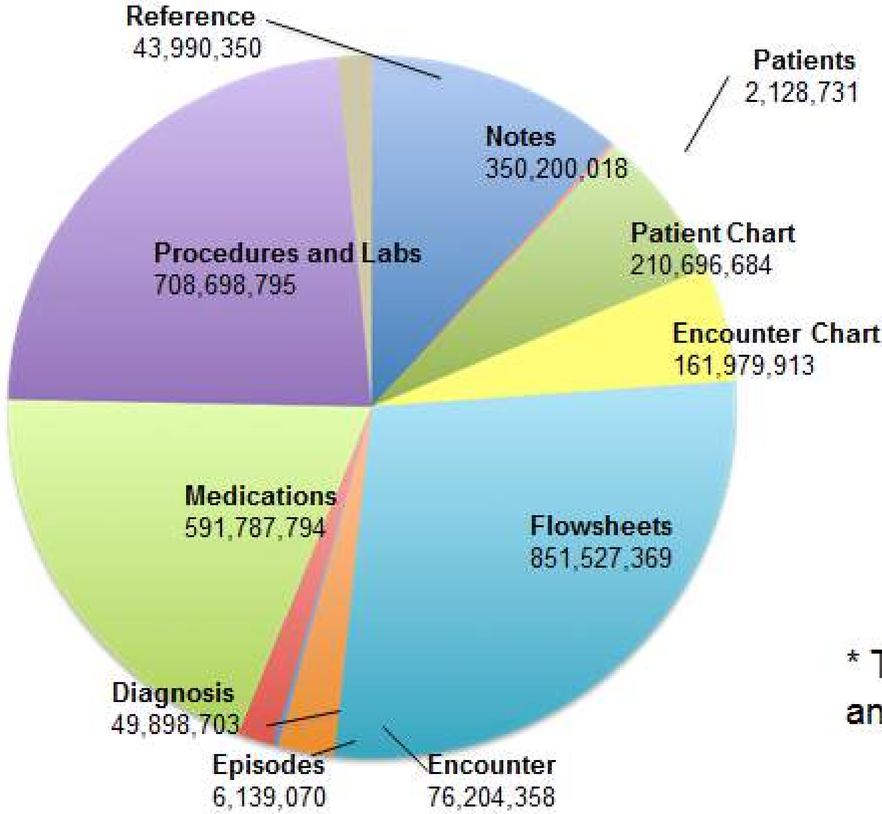


8 Fairview hospitals (data from 2011 and on)
40+ Fairview (from 2005) and UMP clinics (from 2011)

Data available to UMN researchers via the Academic Health Center Information Exchange (AHC-IE)
2+ million patients

MHealth / Fairview Health Services

AHC-IE - acute & ambulatory clinical data
2+ million patients
4+ billion total rows of unique data

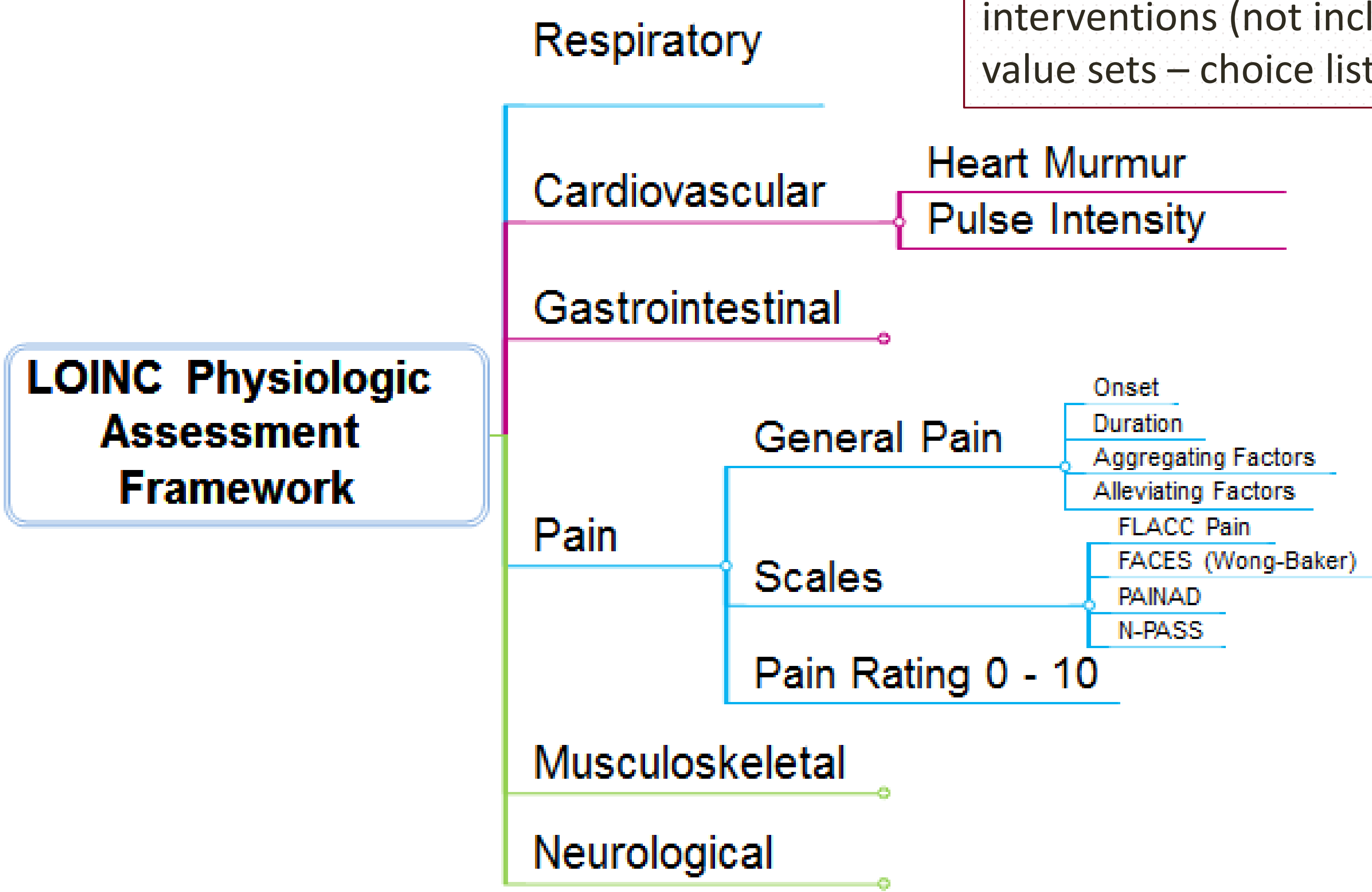


* The number or patients and records changes daily

Making Data Useful

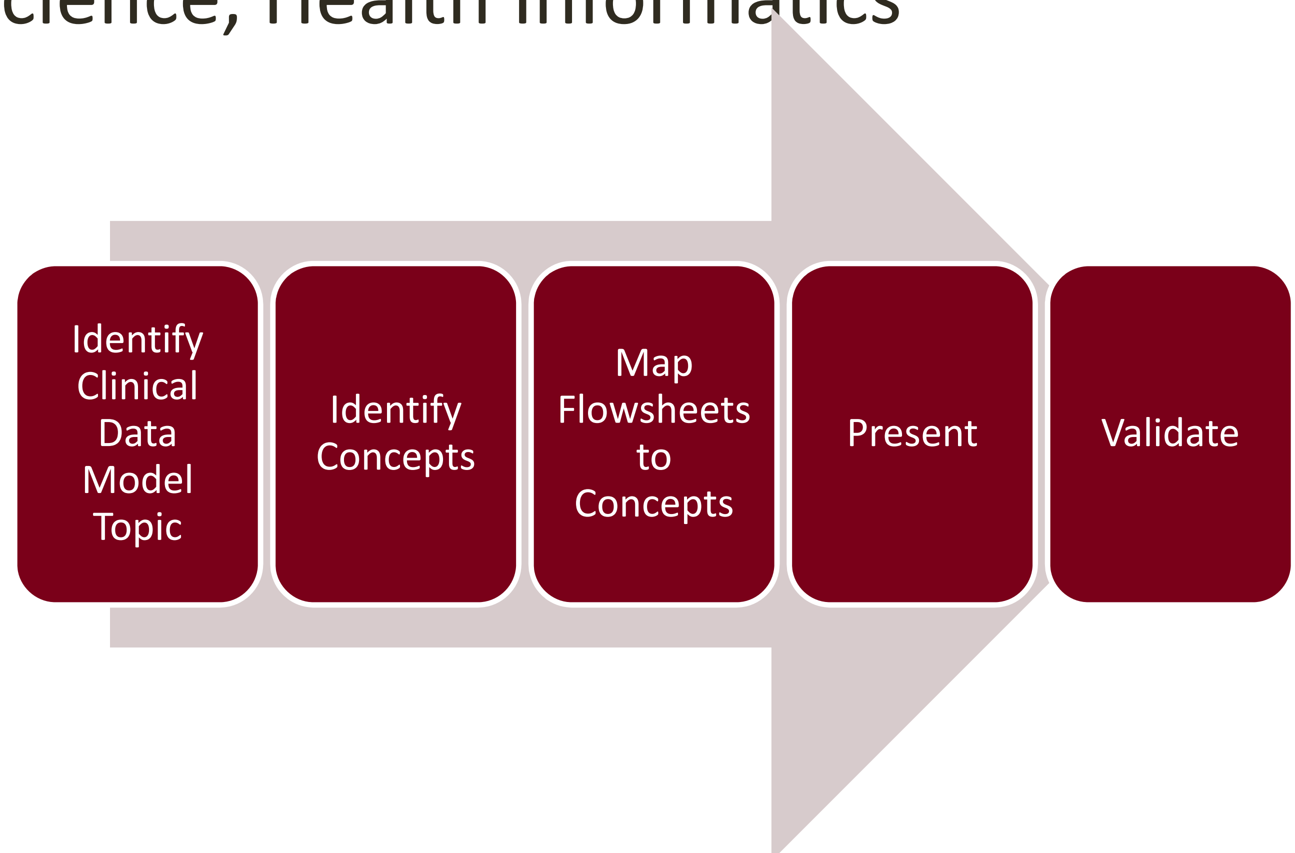
200K Patient Encounter

Pain Information Model –
2137 observations
91 Unique concepts – assessments, goals, interventions (not including value sets – choice lists)



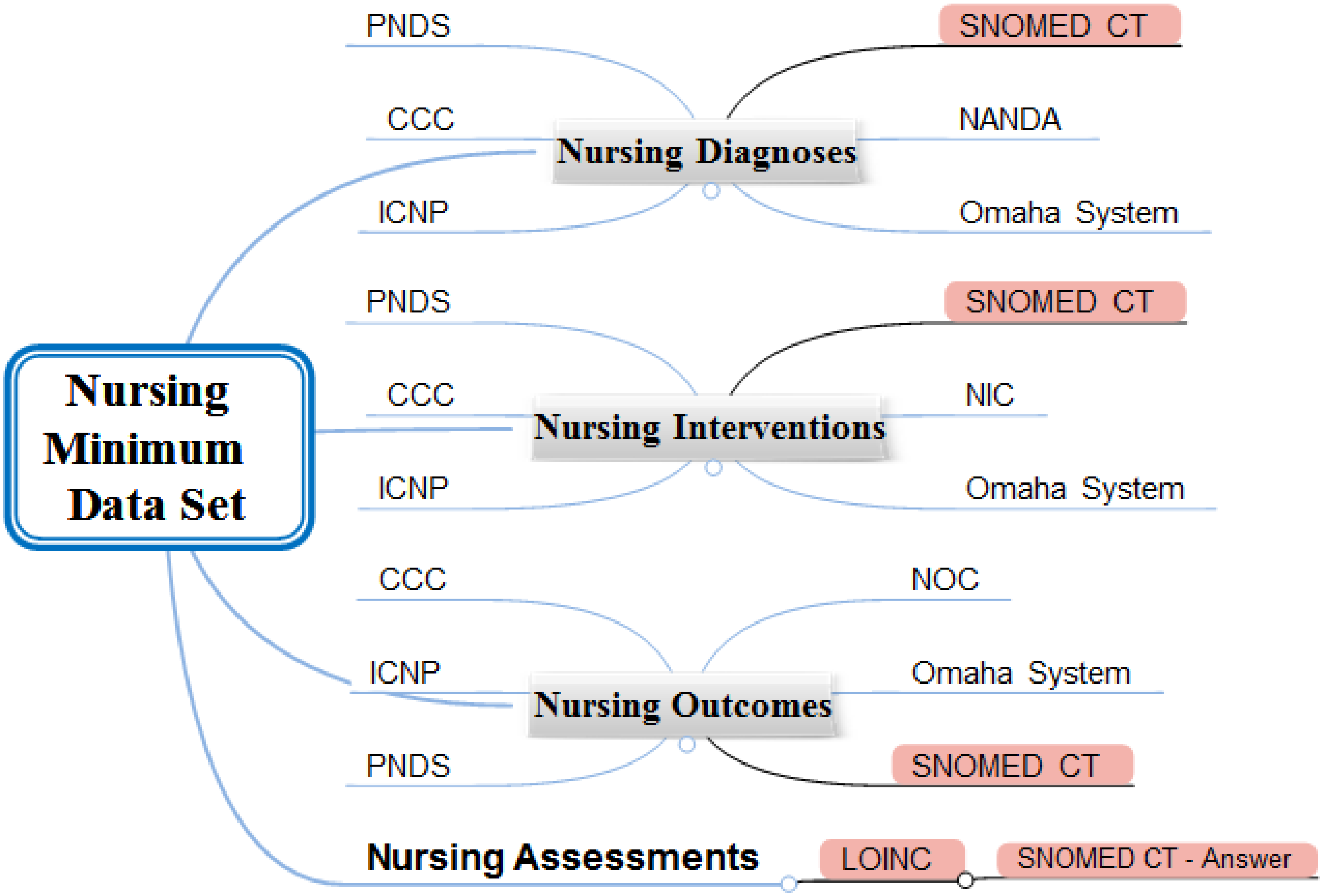
UMN CTSI - Extend CDM

Team: Nursing (DNP/ PhD), Computer Science, Health Informatics



Flowsheet Information Models

BH - Aggression and Interpersonal Violence	Genitourinary System/ CAUTI
BH - Psychiatric Mental Status Exam	Neuromusculoskeletal System
BH - Suicide and Self Harm	Pain
BH - Substance Abuse	Peripheral Neurovascular (VTE)
Cardiovascular System	Pressure Ulcers
Falls/ Safety	Respiratory system
Gastrointestinal System	Vital Signs, Height & Weight



ANA Position Statement – Inclusion of Recognized Terminologies Supporting Nursing Practice within Electronic Health Records and Other Health Information Technology Solutions

http://z.umn.edu/bigdata

Nursing Management Minimum Data Set

NMMDS Data Elements

© Delaney 2015

Environment

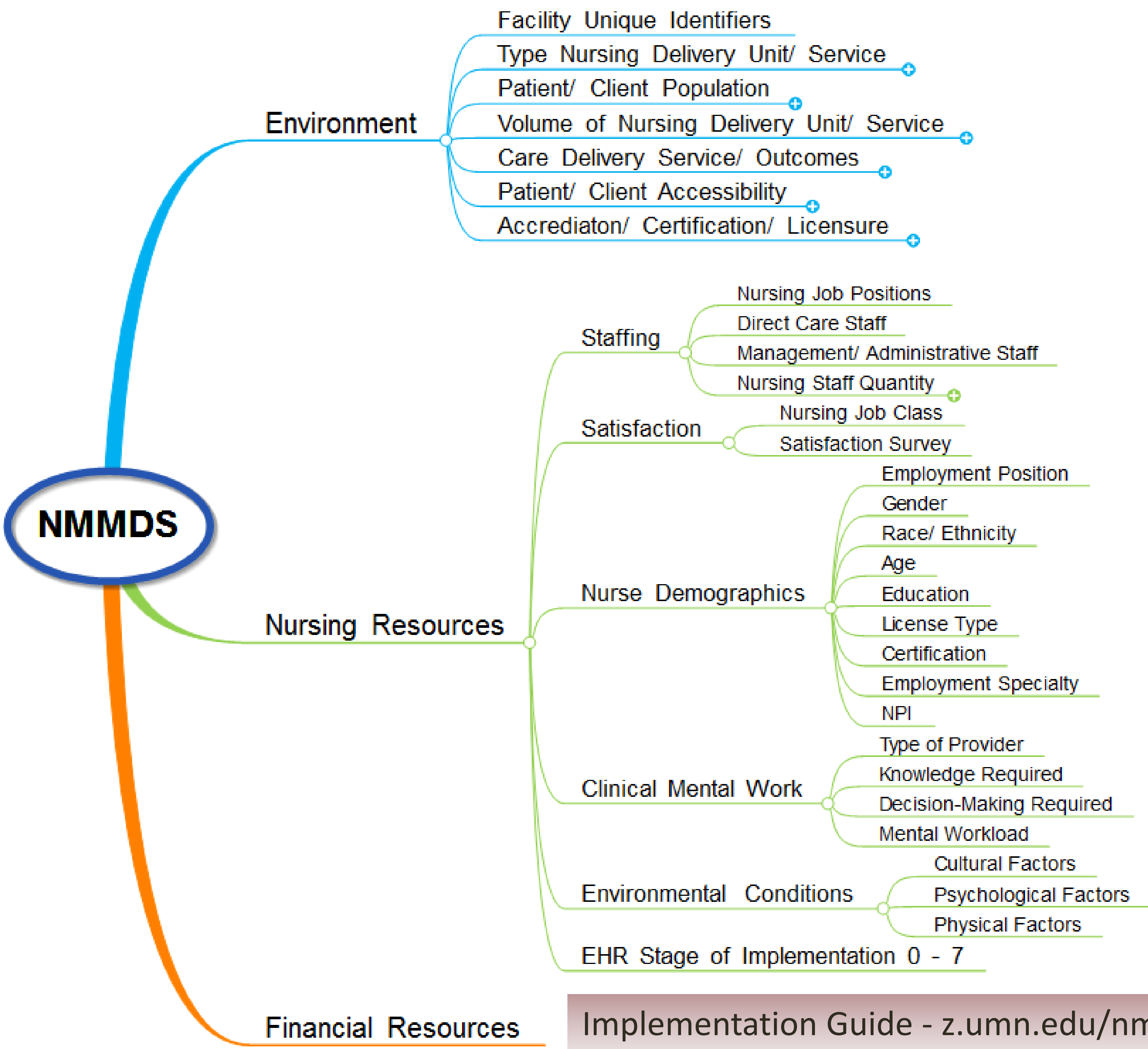
1. Facility Unique Identifier
2. Nursing Delivery Unit/Service
3. Patient/Client Population
4. Volume of Nursing [Care] Delivery Unit/Service
5. Method of Care Delivery
6. Client Accessibility
9. Autonomy
10. Accreditation/ Certification/ Licensure

Nurse Resources

13. Staffing
14. Satisfaction
19. Nurse Demographics per Unit or Service
20. Clinical Mental Workload
21. Environmental Conditions
22. EHR Implementation Stages

~~Financial Resources~~

- Implementation Guide – http://z.umn.edu/nmmds
- LOINC Coding (loinc.org)



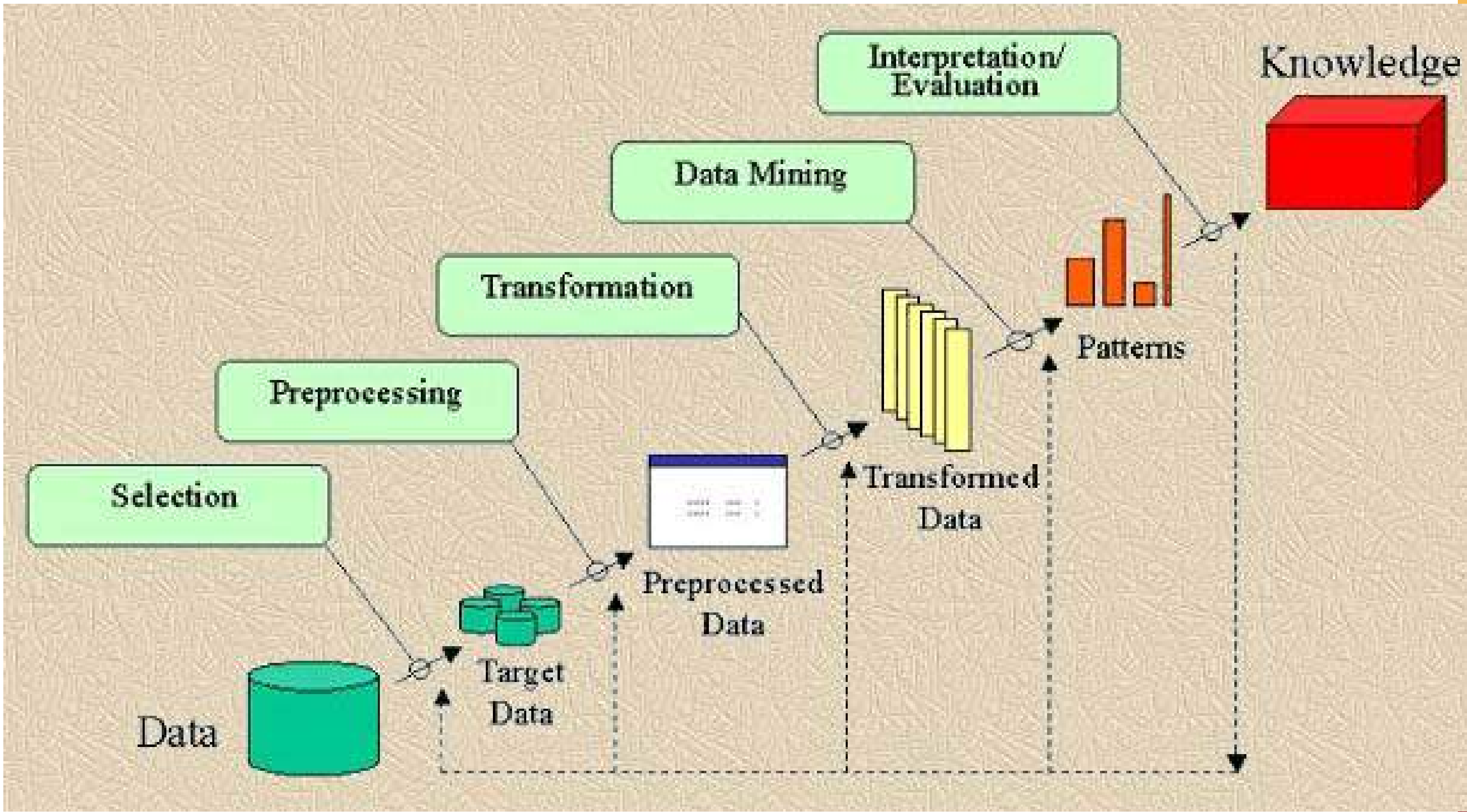
Implementation Guide - z.umn.edu/nmmds

Research Exemplars
Big Data Analytic Methods
Lessons Learned

What is the Question?

- Influence of nursing interventions on patient outcomes?
  - Hospital readmission frail elderly
  - Managing oral medications
  - Urinary and bowel Incontinence
- Influence of Certified WOC Nurses on incontinence & wound outcomes?
- Mobility outcomes by clustering characteristics of patient and support system

Data Analysis Process[2]



Data Preparation - Quality Issues

- Know the Strengths and Limitations of Your Data
- Documentation issues
  - Consistency of processes for documenting
  - Copy forward or copy/paste
  - Incomplete/ inappropriate data in the database
- Rules for data collection
  - Charting by exception
  - Rules i.e. the Joint Commission, CMS, billing
- Database / data model
  - Field type
  - Relationship of fields – how do you link data
- Patient outliers

Data Transformation

- Creating Scales
  - Prognosis, Pain, Pressure Ulcer, Stasis Ulcer, Surgical Wound, Respiratory Status
- Transforming ordinal scale to binary variables
- Combining variables into categories
  - Omaha System interventions – explained later

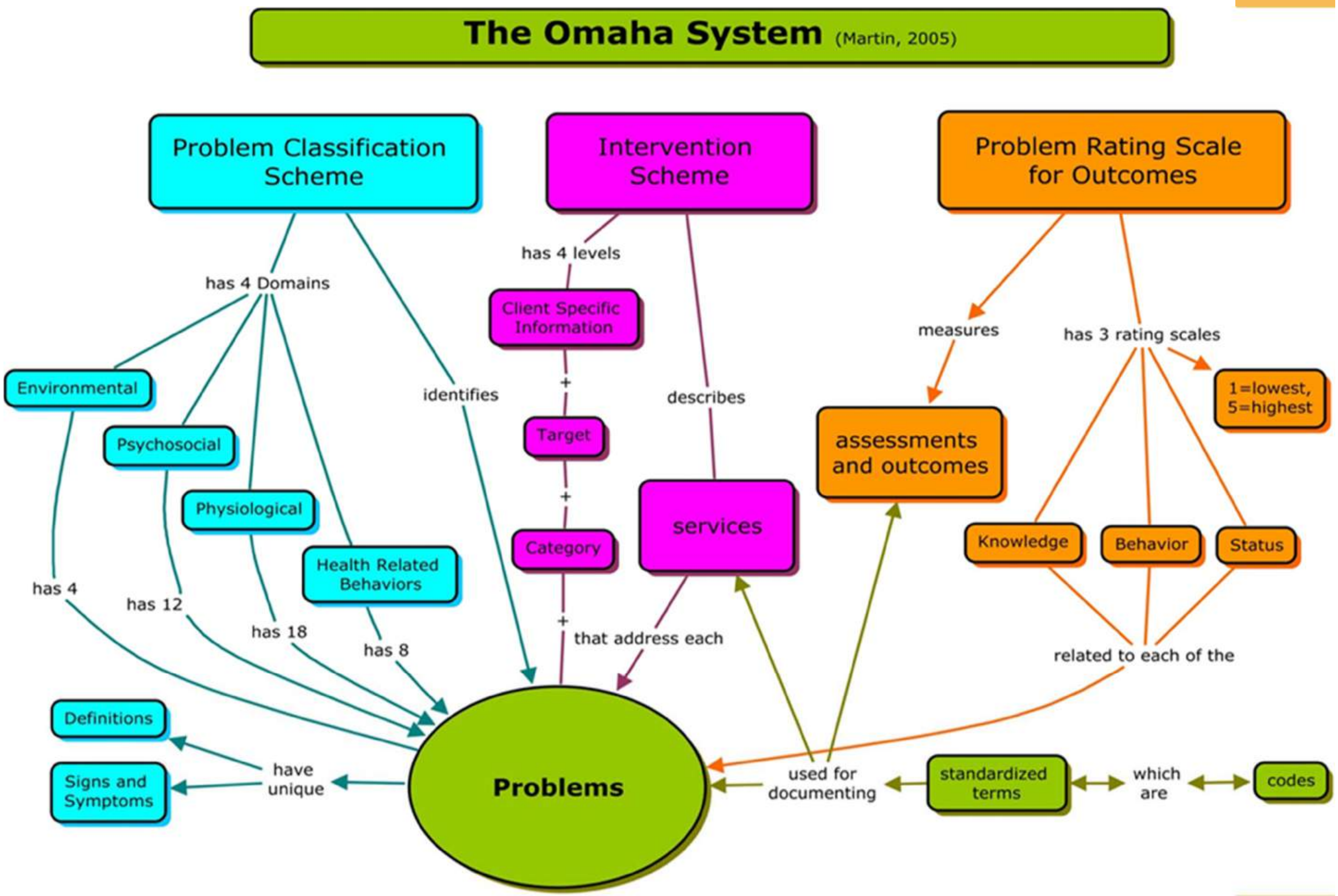
Data Set 1: EHR Homecare

Data Set Description

- Convenience sample of 15 Medicare certified agencies in US
- Obtained de-identified - 2 EHR companies
- 4,244 episodes of care for 2,900 patients
  - Admission and discharge OASIS assessment
- 13,053 Omaha System problems &
- 360,094 interventions, and outcome measures (KBS scores)
- 91,196 Medications

OASIS Data

- Standard assessment required for all Medicare and Medicaid patients
  - Demographic and patient history information
  - Health status
  - Activities of daily living (ADLs) and instrumental activities of daily living (IADLs)
  - Medication and equipment management
  - Service utilization



Monsen Figure 19-2[3]

Medications

- Name
- Dose
- Route
- Frequency
- Instructions
- Start/ end dates

Multiple Studies

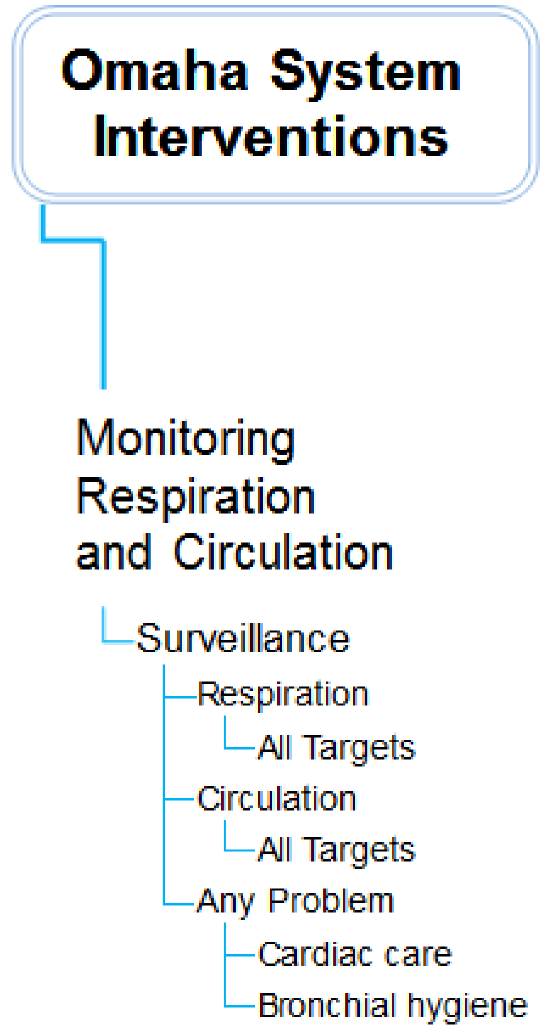
- Modeling interventions
  - Predicting hospitalization
- Medication studies
  - Predicting hospitalization
  - Improvement in managing oral medications
- Urinary and bowel incontinence

Intervention Methods

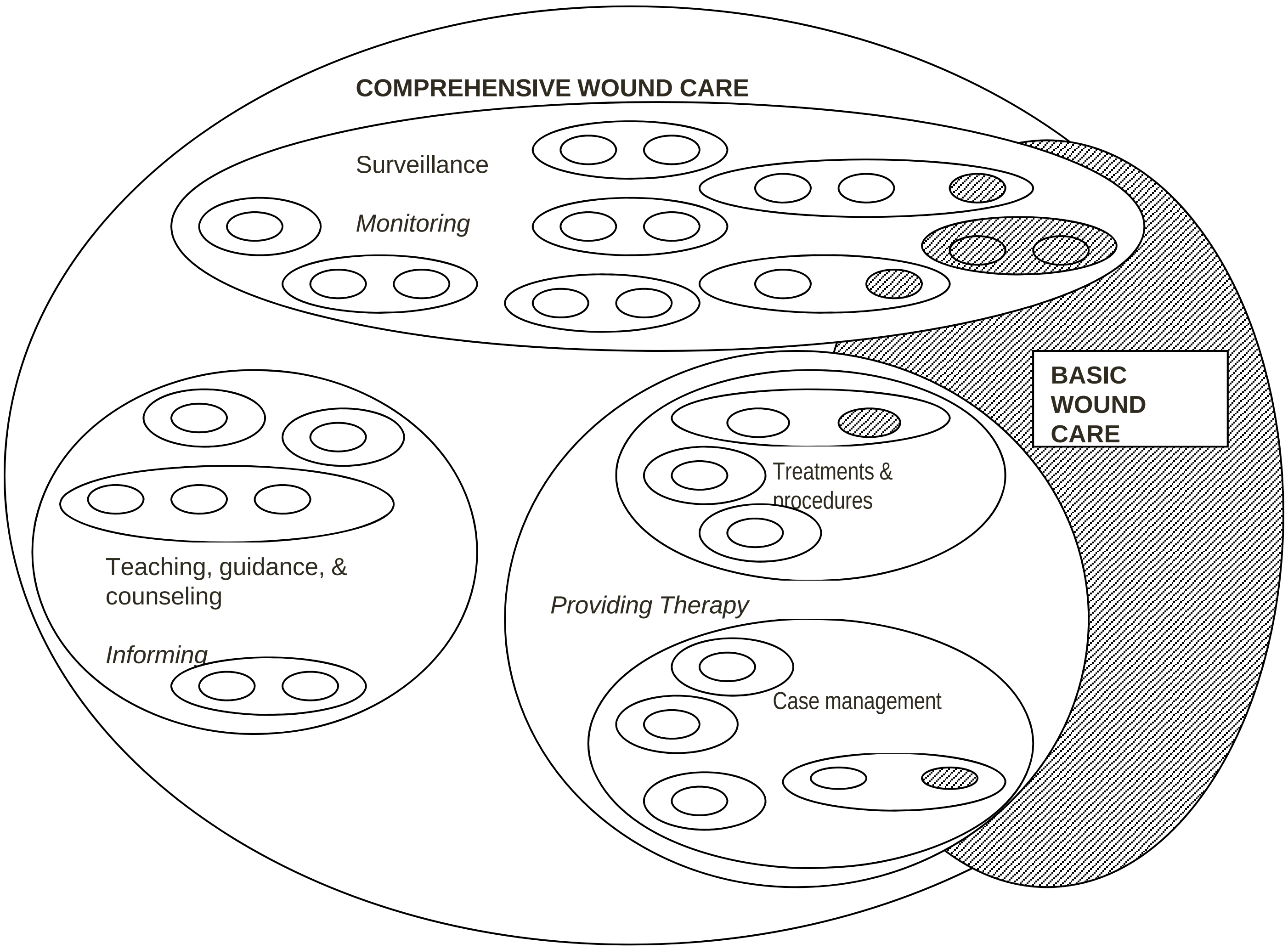
- Feasibility of integrating data across EHR software vendors and home care agencies
- Develop methods of aggregating interventions (1-78 interventions/ patient)
- Three Deductive
    - Classification-based
    - Theory-based
    - Clinical expert consensus
- One Inductive
    - Data-driven

Expert Categorization Omaha System Interventions

	Intervention Type Name
1	Monitoring Respiration and Circulation
3	Monitoring Pain
4	Monitoring Medications
5	Monitoring Injury Prevention
6	Monitoring Skin
8	Coordinating Supplies & Equipment
9	Coordinating Community Resources
10	Coordinating Other
11	Providing Respiration & Circulation Therapy
12	Providing Pain Treatment
13	Providing Medication Treatment
14	Providing Injury Prevention Treatment
15	Providing Wound Care Treatment
16	Providing Bowel and Bladder Treatment
17	Providing Other Treatment
18	Teaching Respiration & Circulation
19	Teaching Medications
20	Teaching Disease Process
21	Teaching Disease Treatment
22	Teaching Emotional & Cognitive Issues
23	Teaching Other



Data Driven Approach[4]



Intervention Methods & Outcomes

- Which intervention management method is associated with hospitalization for frail/ non-frail homecare patients
- Purpose
  - Compare the ability of four intervention data management approaches to explain hospitalization outcomes for frail and non-frail elders separately.
  - Identify intervention groups associated with hospitalization for frail elders and non-frail elders.

Table 2. Interventions Associated With Hospitalization for Frail and Non-Frail Elderly Patients

	OR	Lower 95% CI	Upper 95% CI	*p*-value	*n*	AUC value
Frail elderly patients						
Action category						.593
Treatments and procedures—low	3.67	1.57	8.57	.003	97	
Surveillance—low	2.68	1.06	6.74	.036	115	
Theoretical						.553
Monitoring—low	3.12	1.33	7.34	.009	115	
Clinical expert consensus						.544
Monitoring injury prevention—low	1.99	1.15	3.44	.014	95	
Data-driven						.627
Assist meds and homemaking—high	11.92	2.64	53.85	.001	29	
Medication management—high	.16	.04	.62	.008	79	
Providing injury prevention treatment—low	2.96	1.39	6.33	.005	104	
Non-frail elderly patients						
Action category						.584
Teaching, guidance, and counseling—high	.36	.14	.91	.031	359	
Case management—low	2.76	1.49	5.13	.001	191	
Theoretical						.526
Informing—high	.36	.14	.92	.032	365	
Clinical expert consensus						.603
Coordinating other—medium	2.17	1.01	4.66	.048	141	
Providing medication treatment—medium	3.72	1.47	9.38	.005	49	
Data-driven						.545
None						

OR, odds ratio; CI, confidence interval; AUC, area under curve.

Medication Studies

High Risk Medication Regimen (HRMR)[6]

- High Risk Medication Regimen & Re-Hospitalization for Elderly
  - Aim 1: Describe polypharmacy, potentially inappropriate medication use, and medication regimen complexity.
  - Aim 2: Determine what combination of factors (polypharmacy, potentially inappropriate medications, medication regimen complexity) compose the concept of high risk medication regimens.
  - Aim 3: Evaluate the extent to which high risk medication regimens, as a mediating variable between comorbidity and hospital readmission, account for variance in hospital readmission.
- Used OASIS and medication data
- Mapped instruments to EHR data

High Risk Medication Regimen

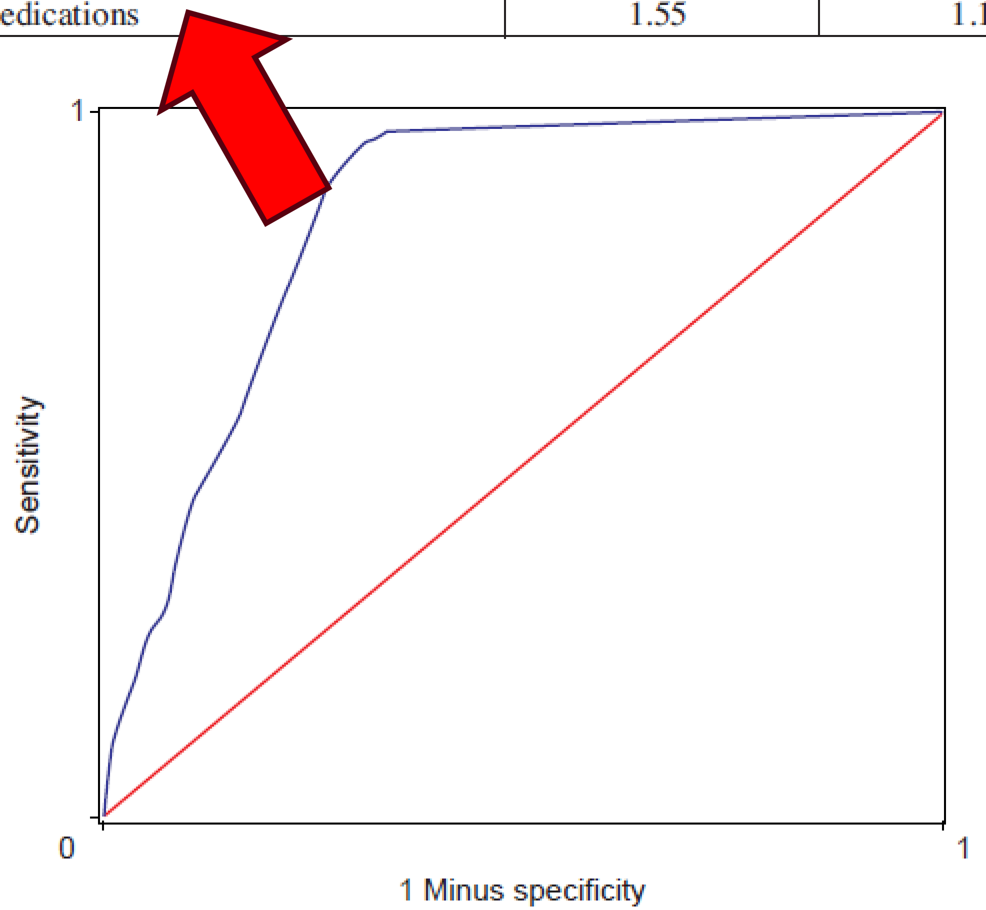
- HRMR Measures
  - Polypharmacy
  - Potentially Inappropriate Medications (PIM) (Beers' criteria)
    - Specific medications, medication class/ disease
  - Medication Regimen Complexity Index (MRCI) –
    - # route, dosing frequency, additional directions or preparation
- Analysis – descriptive/ correlational analysis, factor analysis, structured equation modeling
  - Three unique components to HRMR
- Results – HRMR uniquely predicted 10% of rehospitalization, performed as well as the Charlson Index of Comorbidity

Managing Oral Medications[7]

- Improvement in oral medication management for home care patients
- Compared 3 methods to develop predictive rules that are parsimonious and clinically interpretable
- OASIS & Omaha System data

Table 1. Results of logistic regression

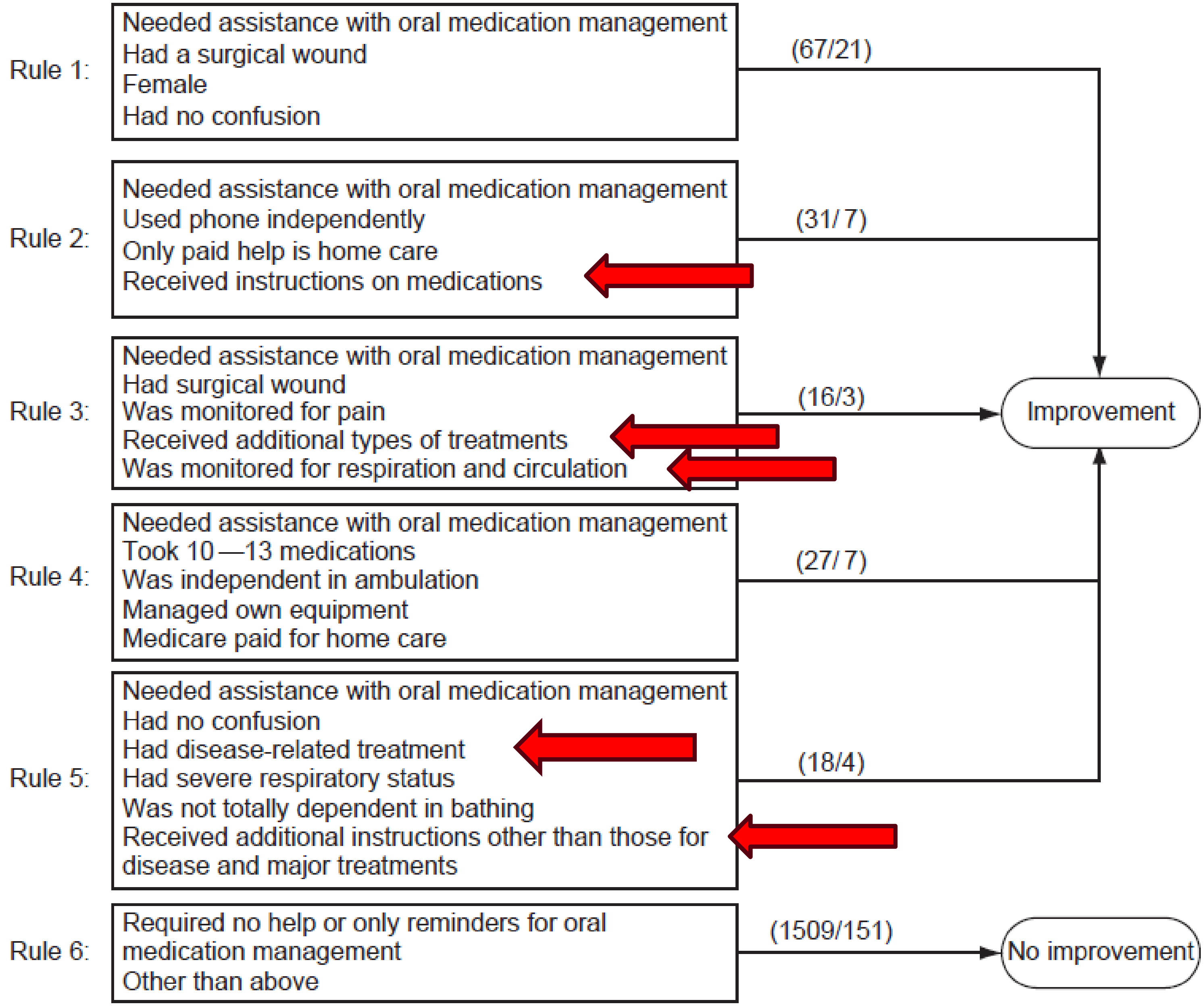
Predictor Variable	Odds Ratio	95% Confidence Intervals
No prior inpatient stay previous 14 days	0.32	0.20–0.51
Prepare light meals	0.61	0.48–0.78
Oral-medication management at admission	8.50	6.27–11.52
Teaching medications	1.55	1.12–2.14



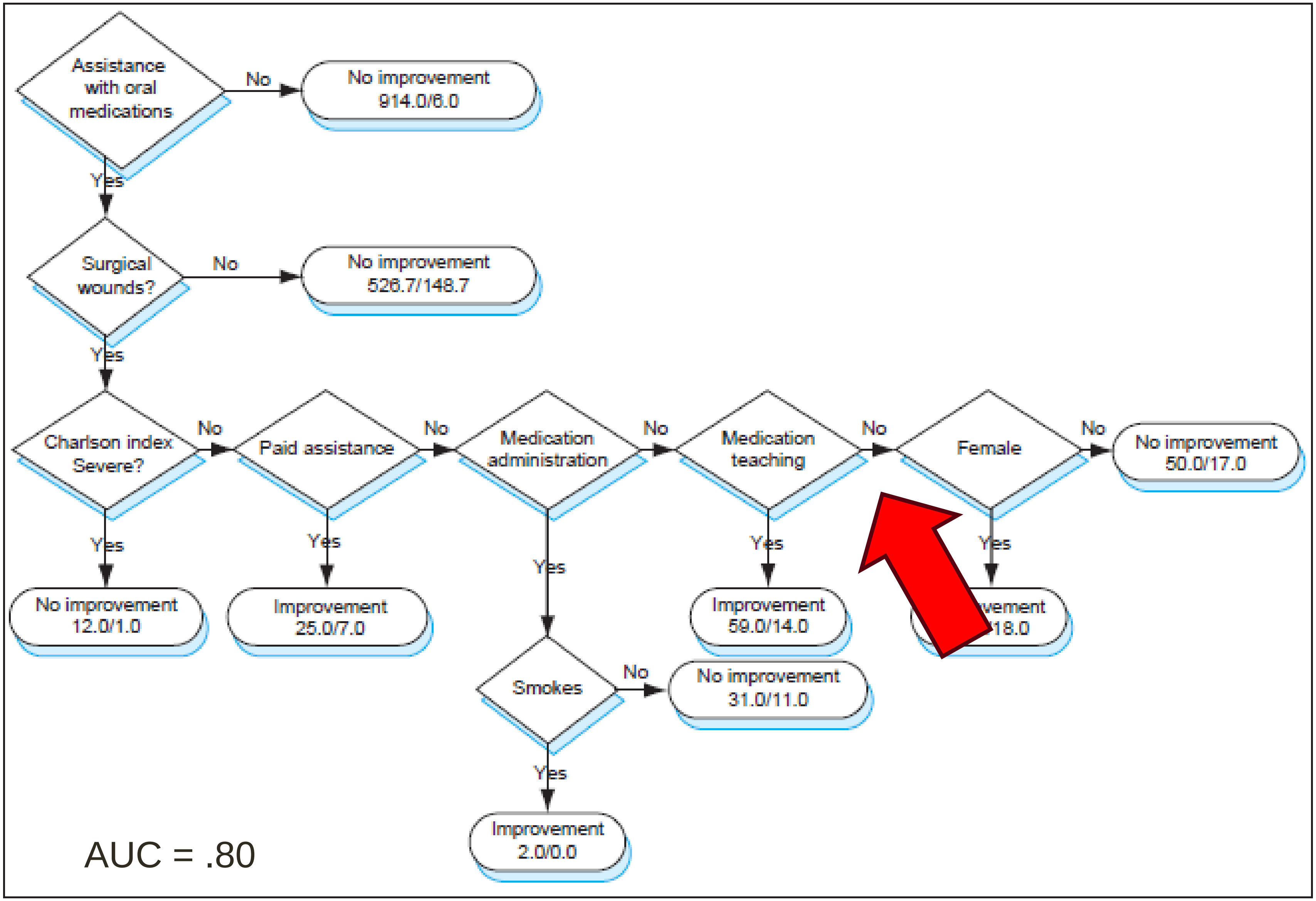
AUC = .85

Ripper Rules Classification

AUC = .81



Decision Tree



Conclusions/ Lessons Learned

- Interventions contributed to all models
- Step-wise logistic regression
  - Produced a more parsimonious clinically interpretable model, while classification rules better reflected the complex decision making
  - Manual entry of variables into model – stepwise effects
- Data mining
  - Problem with an imbalanced class (outcome)
  - Fully automated for DM methods

Data Set 2: EHR Homecare

Home Care EHR De-Identified Data[8, 9]

Initial Data Set

808 agencies, 1,560,508 OASIS records, 888,243 patients

List of patients with and without WOC Nurse

Reason for Removing Records	n
Incomplete episode records	464,485
Assessment outside study dates	125,886
Incorrect type of assessment	51,779
Masked or missing data	16,302
Duplicate records	2,748
Age < 18 or primary dx related to pregnancy/ complications	822

Final Data Set

785 agencies, 447,309 patients,

449,243 episodes of care, 0.6% re-admissions

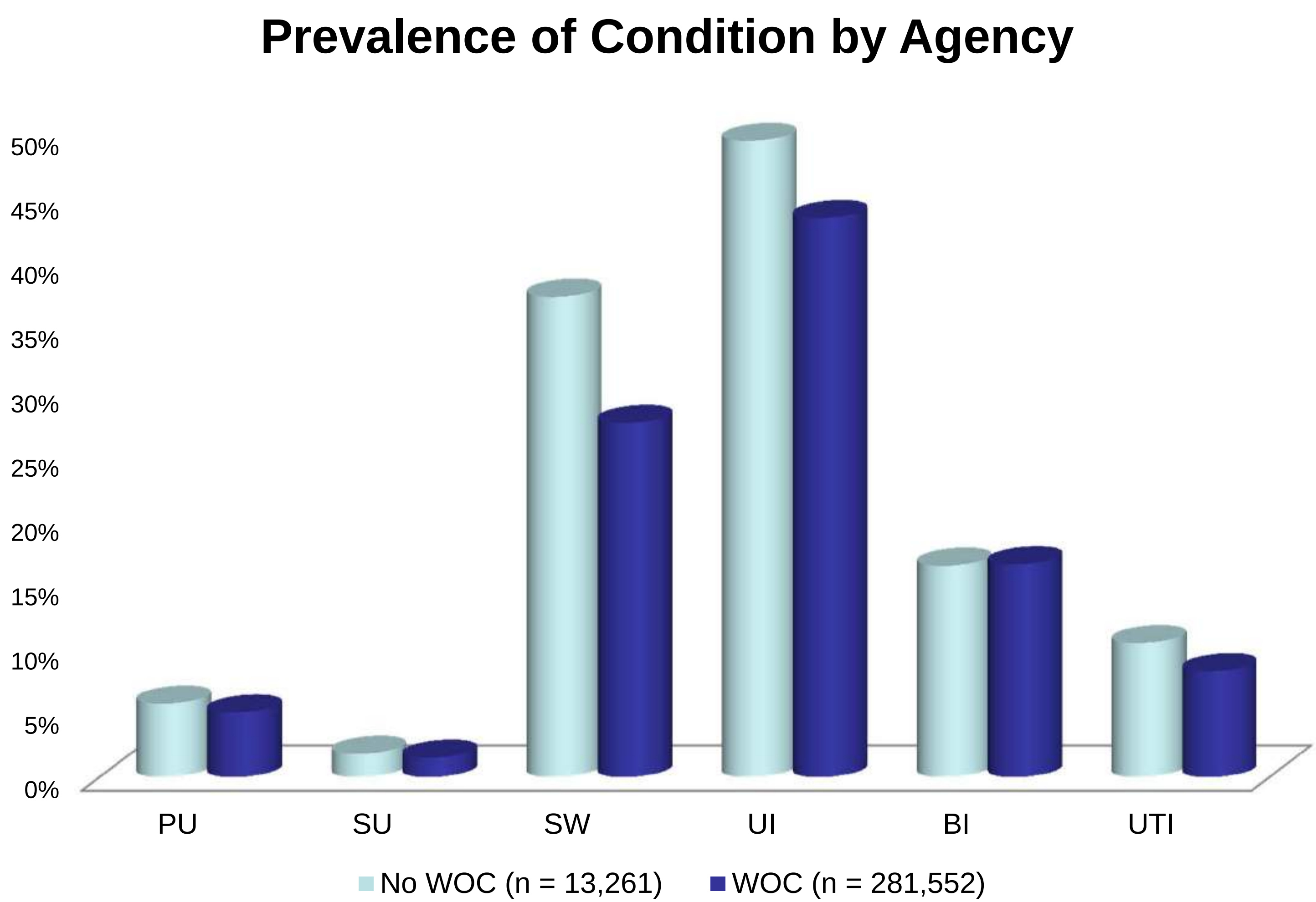
Certified WOC Nurses - Incontinence & Wounds

Outcome Variables	Description
Pressure Ulcers	Total number of pressure ulcers (M0450 a-e)
Stasis Ulcers	Total number stasis ulcers (M0470/ M0474)
Surgical Wounds	Total number of surgical wound (M0484/ M0486)
Urinary Incontinence	Presence/management of urinary incontinence or need for a catheter (M0520)
Urinary Tract Infection	Treated for UTI in past 14 days (M0510)
Bowel Incontinence	Frequency of bowel incontinence (M0540)

Improved/ Not Worse (Stabilize) Outcomes

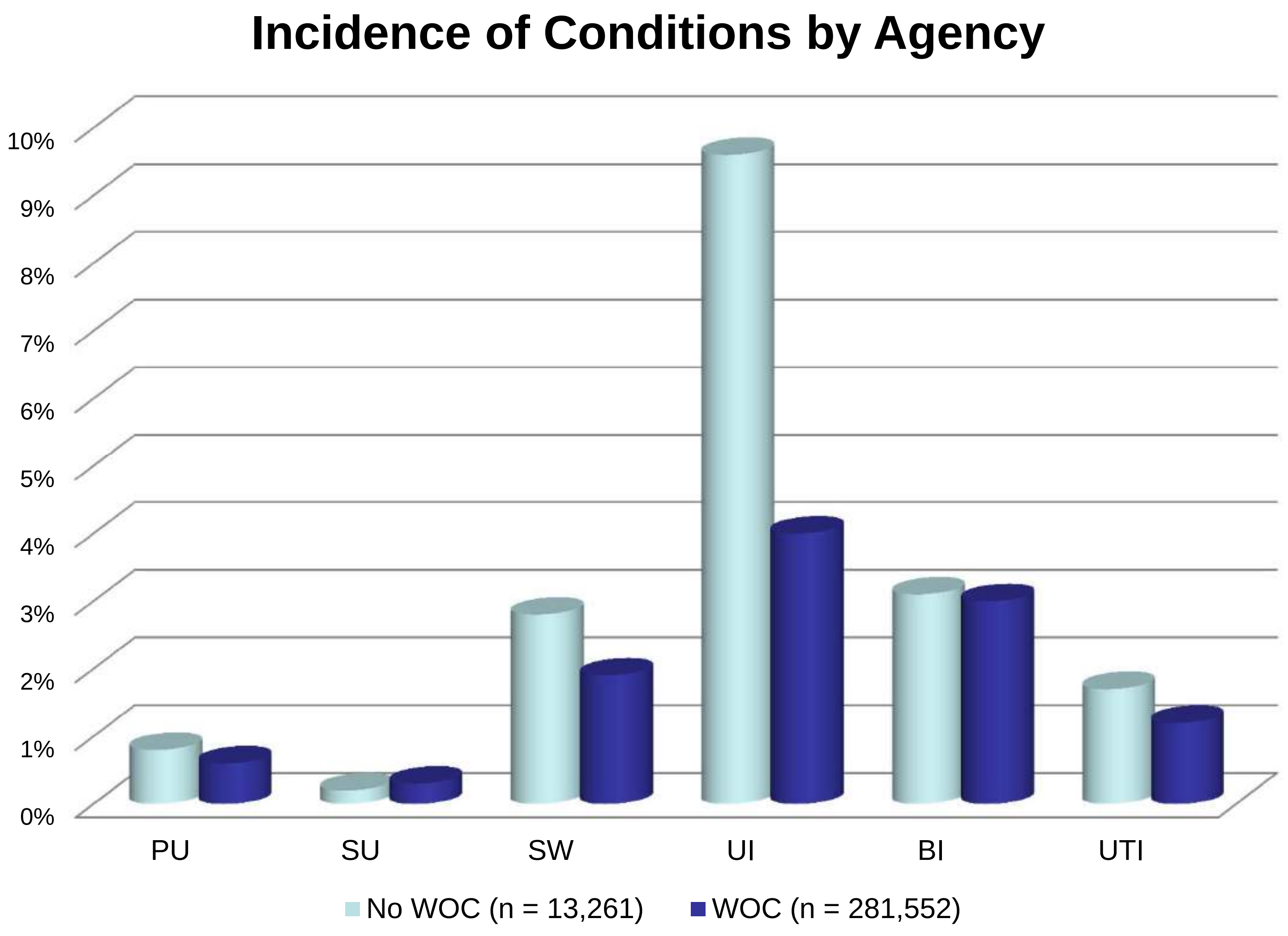
Score	Bowel Incontinence Frequency	Improved	Not Worse (Stabilize)
0	Very rarely /never has BI or has ostomy for bowel elimination		
1	Less than once weekly		
2	One to three times weekly		
3	Four to six times weekly		
4	On a daily basis		
5	More often than once daily		

Aim 1: Prevalence



Pressure Ulcer (PU), Stasis Ulcer (SU), Surgical Wound (SW),
Urinary Incontinence (UI), Bowel Incontinence (BI), Urinary Tract Infection (UTI)

Aim 2: Incidence



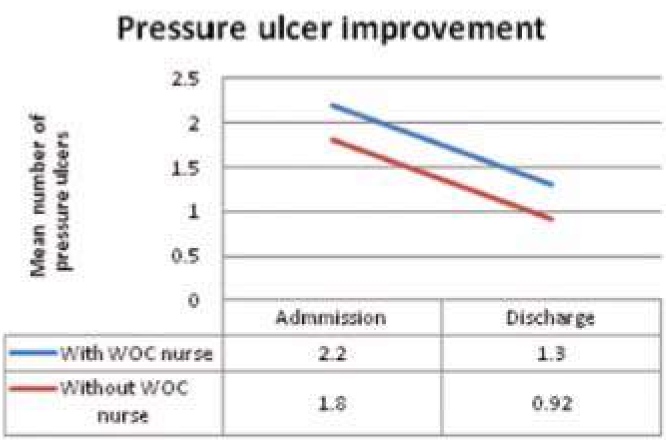
Effect of WOC Nurses on Agency Outcomes

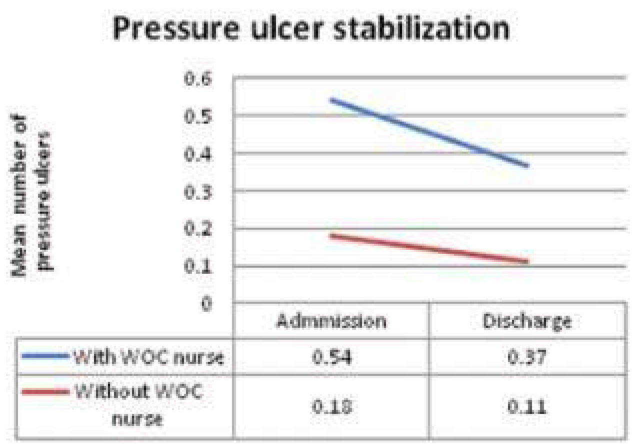
Outcomes Comparing Agencies With and Without a WOC Nurse[a]

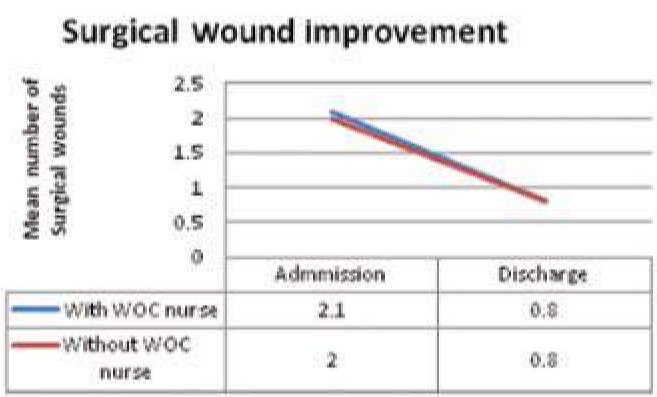
	Improvement		Stabilization	
Outcome Concept	**OR**	**95% CI**	**OR**	**95% CI**
Pressure ulcers	1.9	1.8-2.0	1.29	1.21-1.37
Urinary incontinence	1.4	1.38-1.43	2.3	2.26-2.4
Urinary tract infections	1.4	1.38-1.43	1.2	1.16-1.27
Surgical wounds	1.39	1.36-1.42	1.5	1.46-1.57
Stasis ulcers	1.2	1.1-1.3	Unable to model[b]	
Bowel incontinence	1.14	1.11-1.2	1.16	1.23-1.9

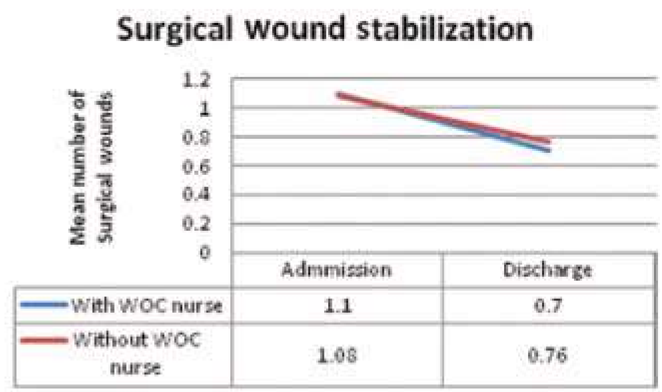
Abbreviations: CI, confidence interval; OR, odds ratio.
[a]ORs weighted by the propensity score for having a WOC nurse.
[b]Unable to model due to more than 99% stabilization across all subjects.

Individual Patient Outcomes



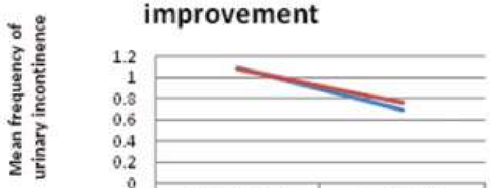






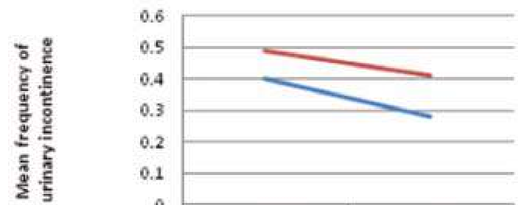
Individual Patient Outcomes

Urinary incontinence improvement



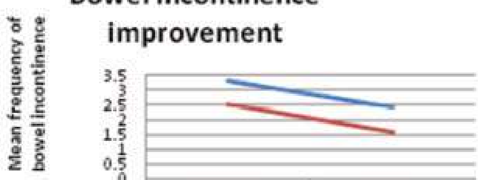
	Admission	Discharge
With WOC nurse	1.1	0.7
Without WOC nurse	1.08	0.76

Urinary incontinence stabilization



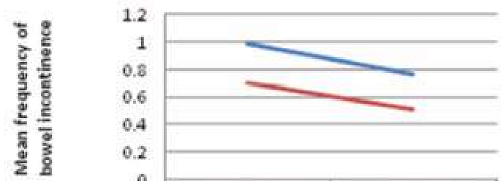
	Admission	Discharge
With WOC nurse	0.4	0.28
Without WOC nurse	0.49	0.41

Bowel incontinence improvement



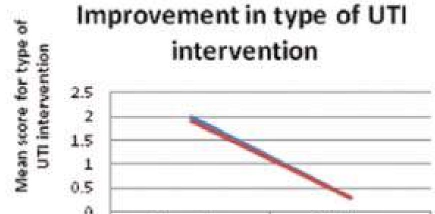
	Admission	Discharge
With WOC nurse	3.3	2.4
Without WOC nurse	2.5	1.6

Bowel incontinence stabilization



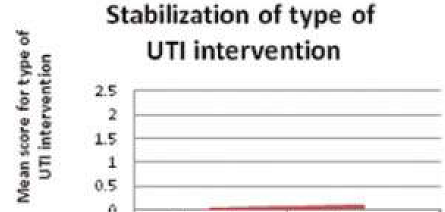
	Admission	Discharge
With WOC nurse	0.99	0.77
Without WOC nurse	0.70	0.51

Improvement in type of UTI intervention



	Admission	Discharge
With WOC nurse	2	0.3
Without WOC nurse	1.9	0.28

Stabilization of type of UTI intervention



	Admission	Discharge
With WOC nurse	0.01	0.05
Without WOC nurse	0.01	0.06

Lessons Learned

- Obtaining data
- Tracking WOC nurse patient visits
- Data quality
  - Matching patients start and discharge
  - Duplicate patient records
  - Encrypted data
  - Missing data
- Selecting variables - theory and domain expertise
- Type of analysis - Research question, structure of the data

Mobility Outcomes [10]

- Discover patients and support system characteristics associated with the mobility outcomes
- Find new factors associated with mobility besides current ambulation status during admission (OR = 5.96)
- In each subgroup of patients defined by current ambulation status during admission (1-5)
- To compare the predictors across each patient subgroup to find the consistent biomarkers in all subgroups and specific factors in each subgroup

Mobility Outcome

TABLE 1. Mobility Scores

Score	Label	Description
0	INDP	Able to independently walk on even and uneven surfaces and climb stairs with or without railings (i.e., needs no human assistance or assistive device)
1	DEVICE	Requires use of a device (e.g., cane, walker) to walk alone or requires human supervision or assistance to negotiate stairs or steps or uneven surfaces
2	SUPERV	Able to walk only with the supervision or assistance of another person at all times
3	CHAIR_I	Chairfast, unable to ambulate but is able to wheel self independently
4	CHAIR_NI	Chairfast, unable to ambulate and [not independent] to wheel self
5	BED	Bedfast, unable to ambulate or be up in a chair

Note. Scores are based on Outcome and Assessment Information Set question M0700 Ambulation/Locomotion.

Comparison of Outcomes by Group

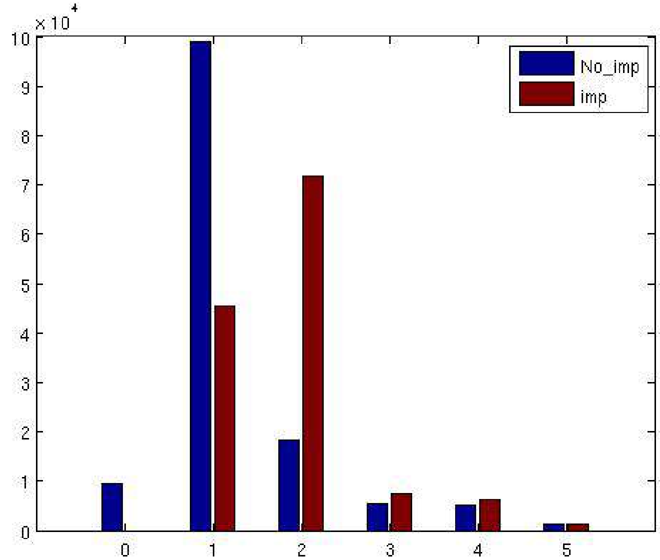
TABLE 2. Mobility Scores at Admission and by Change in Mobility at Discharge From Home Healthcare

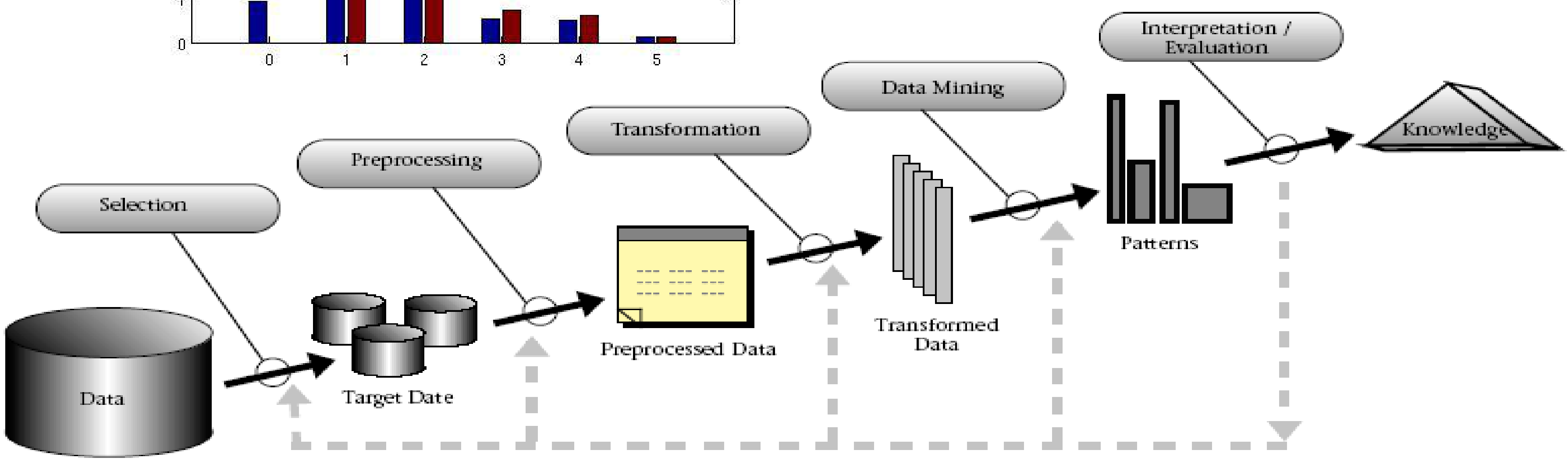
		Total (N = 261,035)		No improvement[b] (n = 128,920)		Improvement[c] (n = 132,115)	
Score[a]	Label	n	(%)	n	(%)	n	(%)
1	INDP	144,615	(55.4)	99,119	(68.5)	45,496	(31.5)
2	DEVICE	89,860	(34.4)	18,129	(20.2)	71,731	(79.8)
3	SUPERV	12,669	(4.9)	5,322	(42.0)	7,347	(58.0)
4	CHAIR_I	11,339	(4.3)	5,163	(45.5)	6,176	(54.5)
5	CHAIR_NI	2,552	(1.0)	1,187	(46.5)	1,365	(53.5)
All		261,035	(100.0)	128,920	(49.4)	132,115	(50.6)

[a]Scores are based on Outcome and Assessment Information Set question M0700 Ambulation/Locomotion. [b]Mobility outcome = 0.
[c]Mobility outcome = 1.

Overall Steps

OASIS
EHRs f
serve
certif



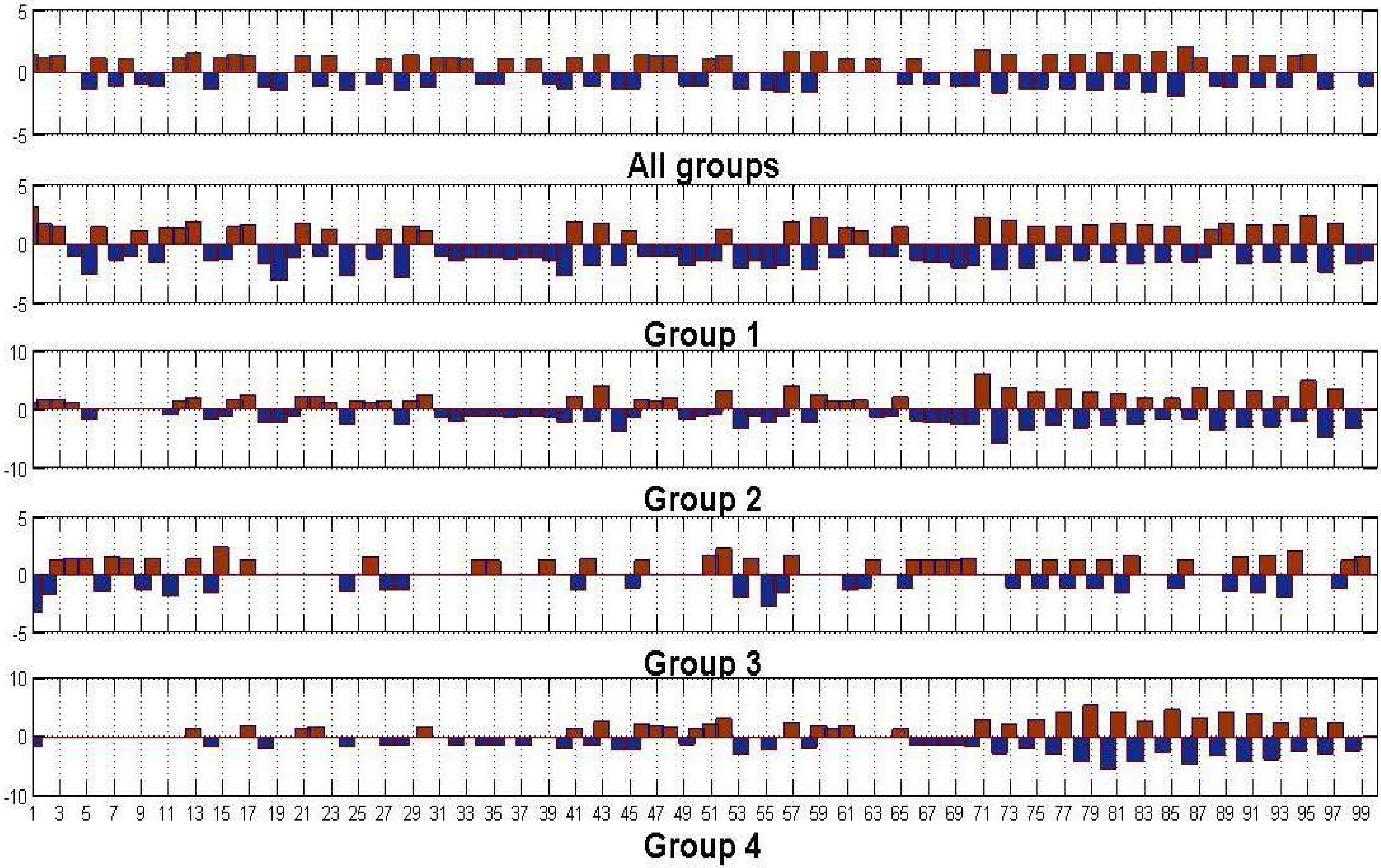


Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, pp. 37 – 54. http://www.kdnuggets.com/gpspubs/aimag-kdd-overview-1996-Fayyad.pdf. P. 41

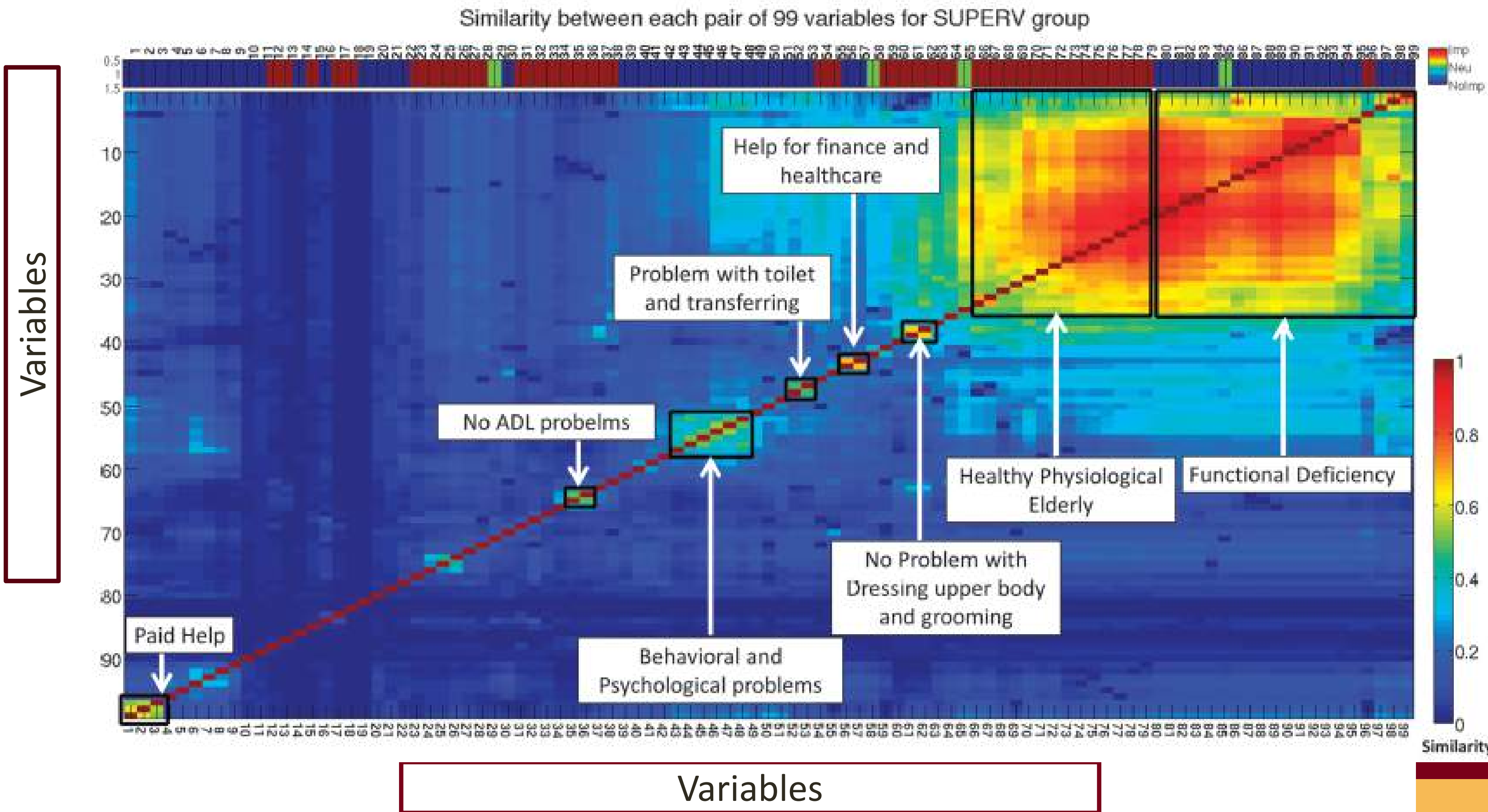
Data Mining Techniques

- Identify risk variables significantly associated with mobility outcomes - varied among the groups
- Group the single predictors based on whether they cover same or different patient group
  - Clustering
    - Based on similarity of patients
    - Not discriminative
    - High frequency variables got merged
  - Pattern mining based approach
    - Discriminative
    - Coherence (similarity of patients)

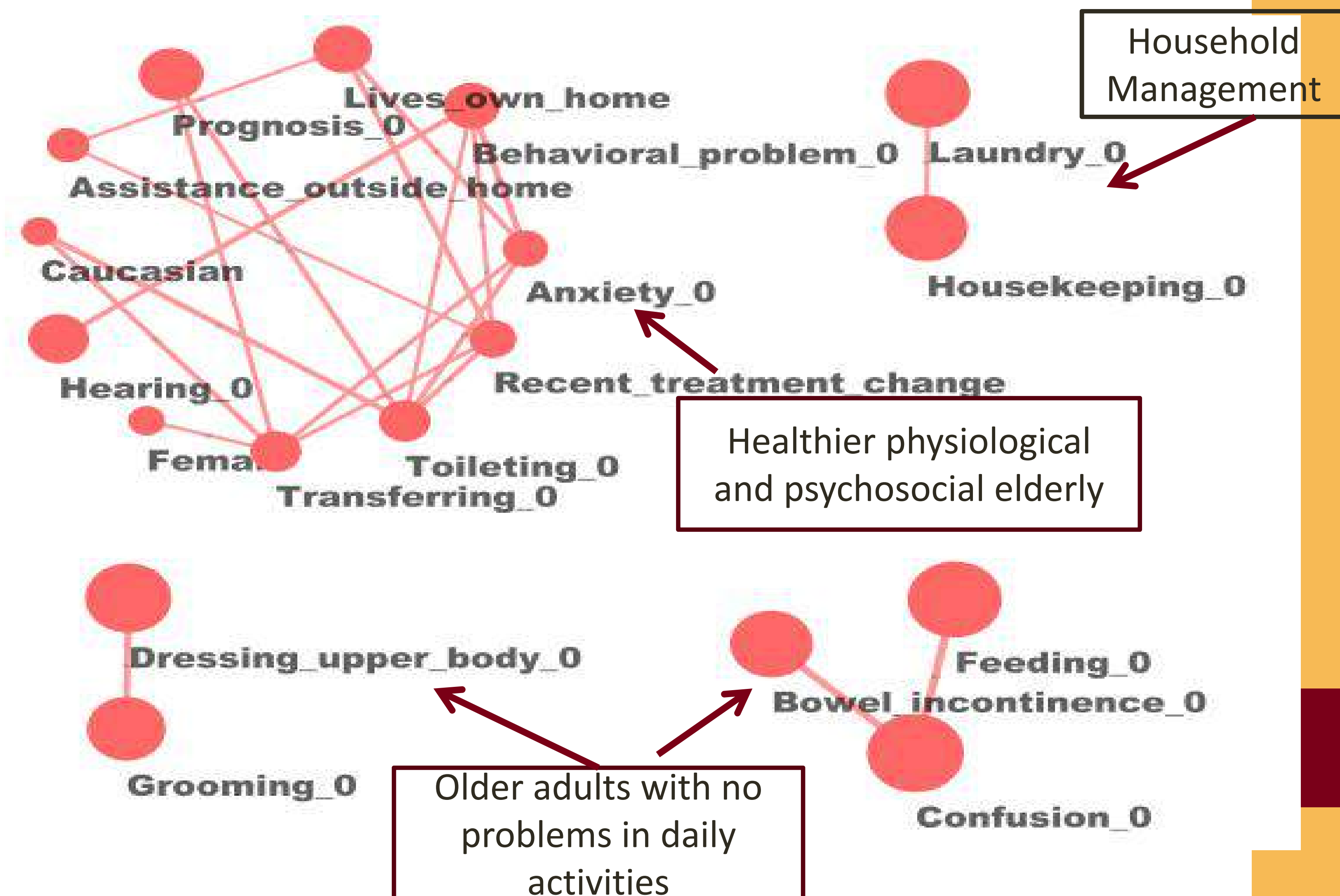
Subgroup Variability



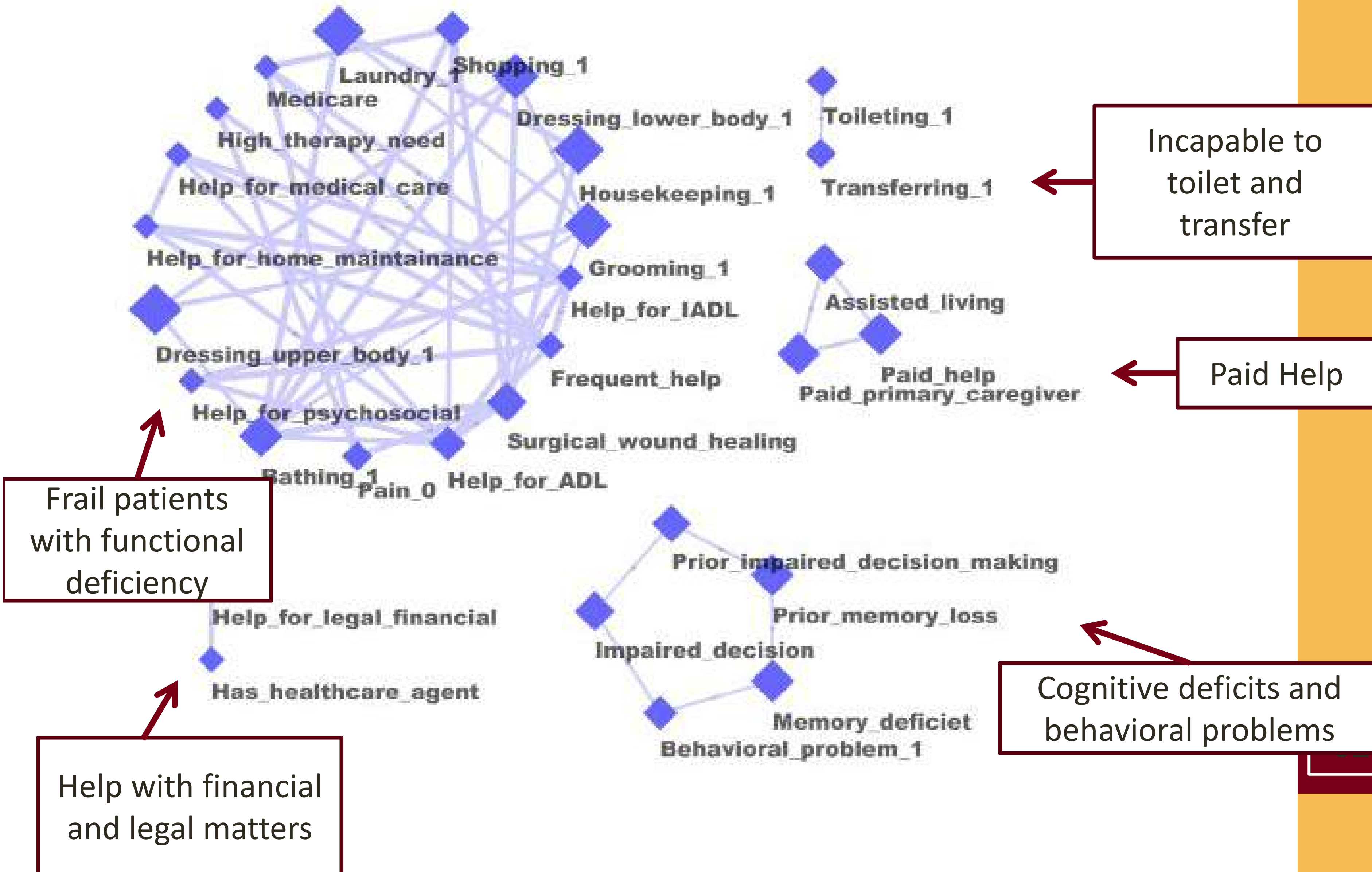
Clustering Groups



Improvement Group 2



No Improvement Group 2



Lessons Learned

- Transform data into binary variables
- Selection of variables – remove if
  - Too little variation or high inter-correlations of predictors
- Medical diagnoses used to describe patients, not predict
- Analysis by subgroup
- Interpretation of results is critical – requires domain experts
- Different clusters point to the need to tailor interventions for subgroups
- **Lack of standardized interventions** precluded understand how care provided effects outcomes

Summary

- Big data is increasing
- Existing and newer methods for data analysis
- Big data science useful to address practice questions
- Lessons learned
  - Data quality – originates in practice
  - Standardized data and common data / information models needed for usable data
- "Takes a village" – combined expertise important

Acknowledgments

- Multiple funding for studies :
- Grant Number 1UL1RR033183 from the National Center for Research Resources (NCRR) of the National Institutes of Health (NIH) to the University of Minnesota Clinical and Translational Science Institute (CTSI).
- Two NSF grants: NSF IIS-1344135 and NSF IIS-0916439 as well as a University of Minnesota Doctoral Dissertation Fellowship.
- Wound, Ostomy, Continence Nursing Association

Questions?

Bonnie L. Westra, PhD, RN, FAAN, FACMI
westr006@umn.edu

See Handout for References

References – Big Data Analytics for Healthcare

Bonnie L. Westra, PhD, RN, FAAN, FACMI

1. Brennan, P. & Bakken S. (2015) Nursing Needs Big Data and Big Data Needs Nursing, **J**ournal of Nursing Scholarship, 47(5): 477-484.

2. Fayyad UM, Piatetsky-Shapiro G, Smyth P, Uthurusamy R. (1996). Advances in Knowledge Discovery and Data Mining. AAI Press/ The MIT Press, Menlo Park, CA

3. Westra, B.L., Monsen, K.A., & Delaney, C.W. (2015). Big Data in Nursing Research. S. Henley, (Ed.). Routledge international handbook of advanced quantitative methods in nursing research. (pp. 280-293). Routledge/Taylor & Francis, Abingdon, UK.

4. Monsen, K., Westra, B.L., Yu, F., Ramadoss, V.K. & Kerr, M.J. (2009). Data Management for Intervention Effectiveness Research: Comparing Deductive and Inductive Approaches, Research in Nursing and Health, 32(6) 647-656.

5. Monsen, K., Westra, B.L., Oancea, S.C., Yu, Fang, Kerr, M. J. (2011). Linking Home Care Interventions and Hospitalization Outcomes for Frail and Non-Elderly. *Research in Nursing & Health,*34,160-168.

6. Dierich, M.T., Mueller, C., & Westra, B.L. (2011). Medication Regimens in Elderly Home Care Patients . Journal of Gerontological Nursing. 37 (12): 45-55.

7. Westra, B.L., Dey, S., Fang, G., Steinbach, M., Kumar, V., Savik, K., Oancea, C., & Dierich, M. (2011). Interpretable Predictive Models for Knowledge Discovery from Homecare Electronic Health Records. Journal of Healthcare Engineering, 2 (1), 55-74

8. Westra, B.L., Bliss, D.Z., Savik, K., Hou, Y., Borchert, A. (2013). Effectiveness of Wound, Ostomy, and Continence Nurses on Agency Level Wound and Incontinence Outcomes in Home Care. Journal of Wound, Ostomy, and Continence Nursing, 40(1), 25 – 53

9. Bliss, D.Z., Westra, B.L., Savik, K., Hou, Y. (2013). Effectiveness of Wound, Ostomy, and Continence Certified Nurses on Individual Patient Outcomes in Home Health Care. Journal of Wound, Ostomy, and Continence Nursing. 40, 135–142

10. Dey, S., Cooner, J., Delaney, C.W., Fakhoury, J., Kumar, V., Simon, G., Steinbach, M., & Westra, B.L. (2015). Data Mining to Predict Mobility Outcomes in Home Health Care, Nursing Research. 64(4), 235-245.