Data Mining to Predict Mobility Outcomes for Older Adults Receiving Home Health Care

Bonnie L. Westra, PhD, RN, FAAN, FACMI
Associate Professor
University of Minnesota
School of Nursing

UNIVERSITY OF MINNESOTA
School of Nursing
Driven to Discover℠

Co-Authors

- Gowtham Atluri, PhD-C
- Jacob Cooner, MN, RN
- Connie W. Delaney, PhD, RN, FAAN, FACMI
- Sanjoy Dey, PhD-C
- Joanna Fakhoury, BS
- Gyorgy Simon, PhD
- Michael Steinbach, PhD
- Vipin Kumar, PhD
- Jeremy Weed, BS

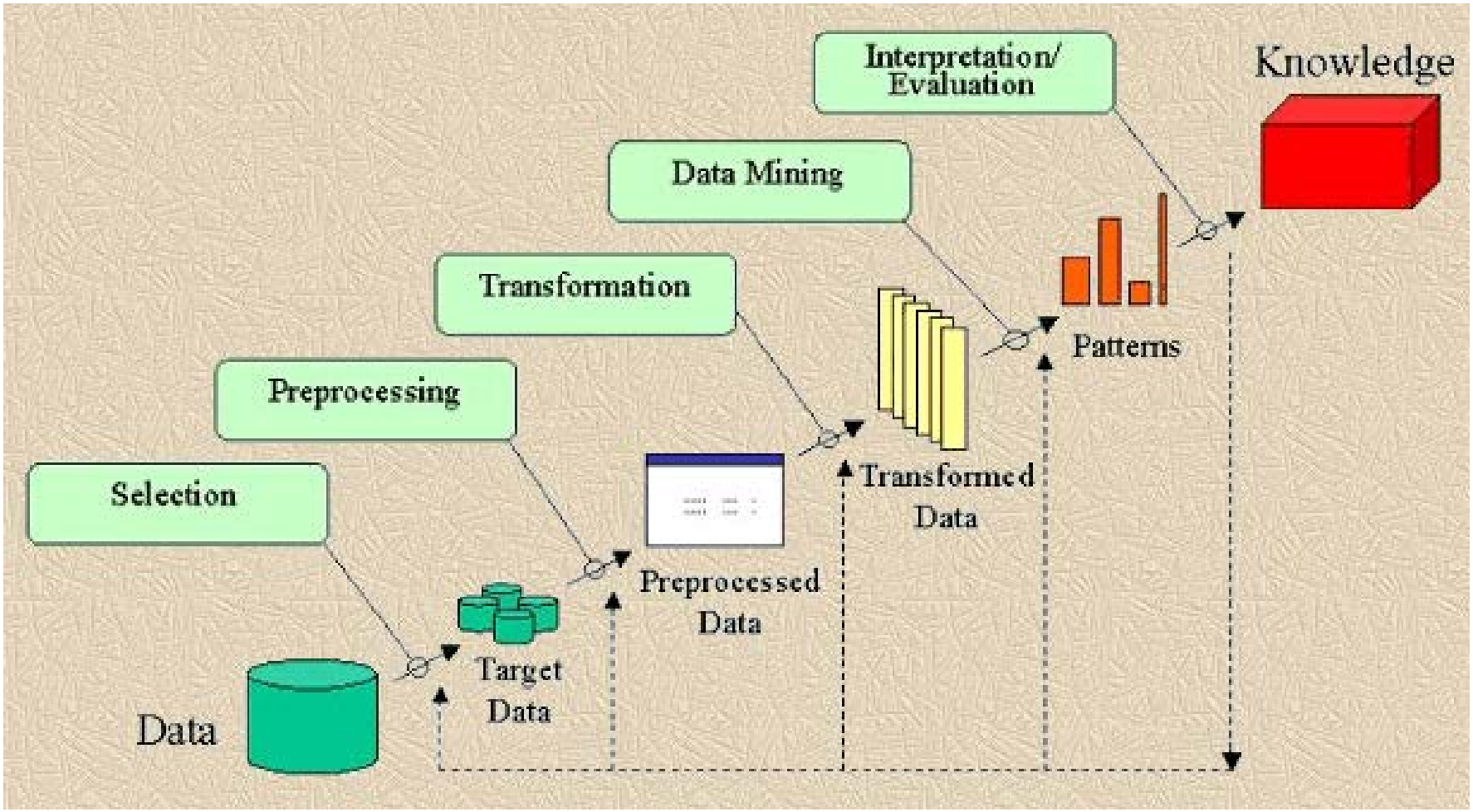
Problem

- In United States, 2010:
  - 4.9 million people required help to complete ADLs
  - 9.1 million people unable to complete IADLs [1]
- Home Healthcare (HHC)
  - Spending in 1980 increased from $2.4 billion to $17.7 billion today
  - Report improved mobility in 46.9% adults before discharge from HHC [2]
- Mobility is one component of functional status
  - Mobility affects functional status and functional disability
  - Less than one-third of older adults recover pre-hospital function [3]
  - Increased risk of falls in home, rehospitalization, disability, social isolation, loss of independence
  - Besides physical issues, also psychosocial issues, comorbidity and death

Purpose

- To discover patients and support system characteristics associated with the improved outcomes of mobility
- 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)
  - We started with group 2 and then compare the observations with other groups
- To compare the predictors across each patient subgroup to find the consistent biomarkers in all subgroups and specific factors in each subgroup

KDD Process



Fayyad UM, Piatetsky-Shapiro G, Smyth P, Uthurusamy R. Advances in knowledge discovery and data mining. Menlo Park, CA: AAI Press/ The MIT Press Press; 1996.

Data Selection

OASIS

- Standard assessment required for all Medicare and Medicaid patients
- Includes
  - 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

Mobility (M0700 Ambulation/ Locomotion) Outcome

Score	Description
0	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	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	Able to walk only with the supervision or assistance of another person at all times.
3	Chairfast, unable to ambulate but is able to wheel self independently.
4	Chairfast, unable to ambulate and is unable to wheel self.
5	Bedfast, unable to ambulate or be up in a chair.

Selection Criteria

- Inclusion Criteria
  - Medicare certified agency – OASIS documentation
  - Minimum of two OASIS records representing an episode
  - Adult, non-maternity clients receiving skilled homecare services
  - No missing data to calculate a change from start to end of an episode for the outcome variables
  - Episode started and completed between 10/1/08 and 12/31/09
- Exclusion Criteria
  - Patients with no mobility problem on admission for outcome variables

Example of Creating a Data Set

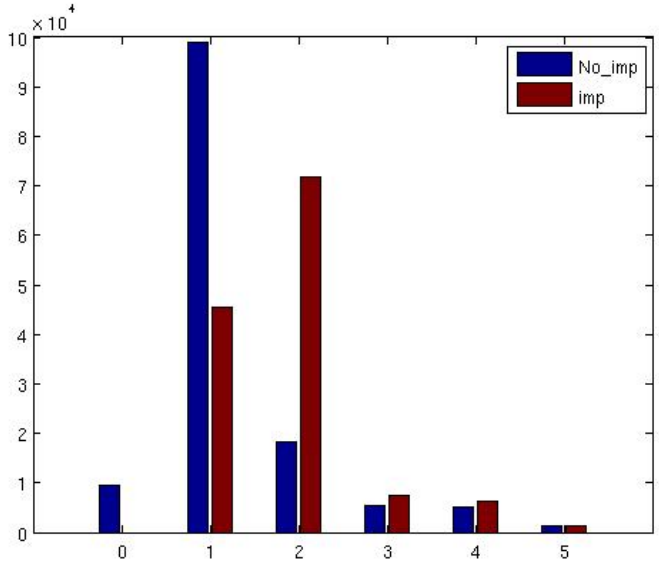
Initial Data Set
808 agencies, 1,560,508 OASIS records, 888,243 patients

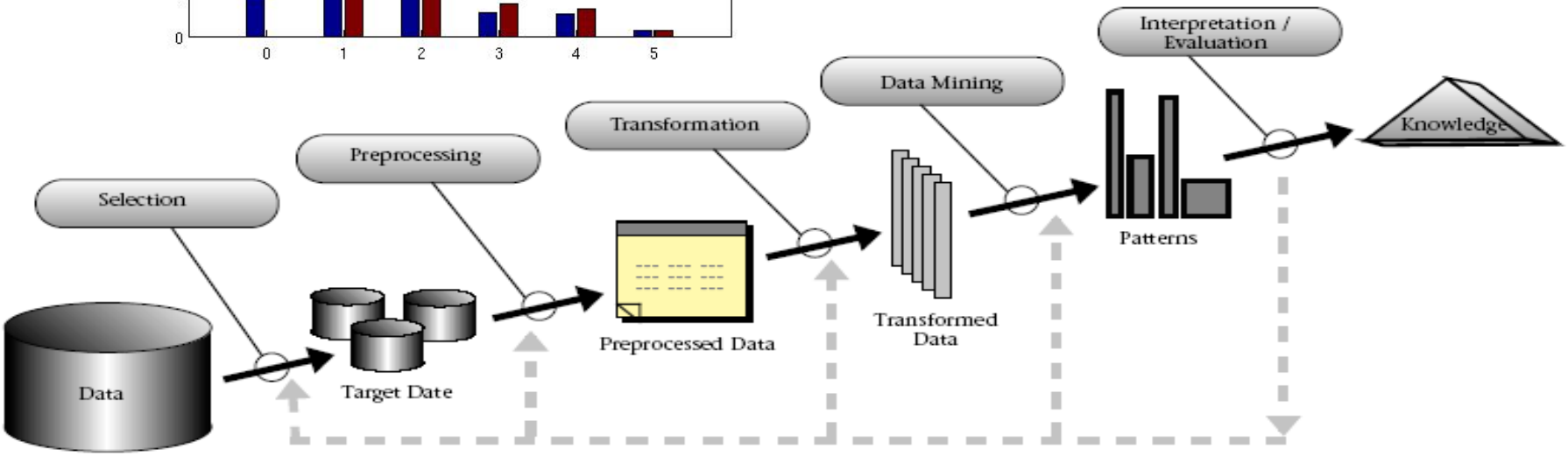
Reason for Removing Records	n
Incomplete episode records	464,485
Assessment outside study dates	125,886
Incorrect type of assessment	51,779
Masked or missing data	16,302
Duplicate records	2,748
Age < 18 or primary dx related to pregnancy/ complications	822

Final Data Set
785 agencies, 447,309 patients,
449,243 episodes of care, 0.6% re-admissions

Overall Steps

OAS
from
patie
Medi
hea





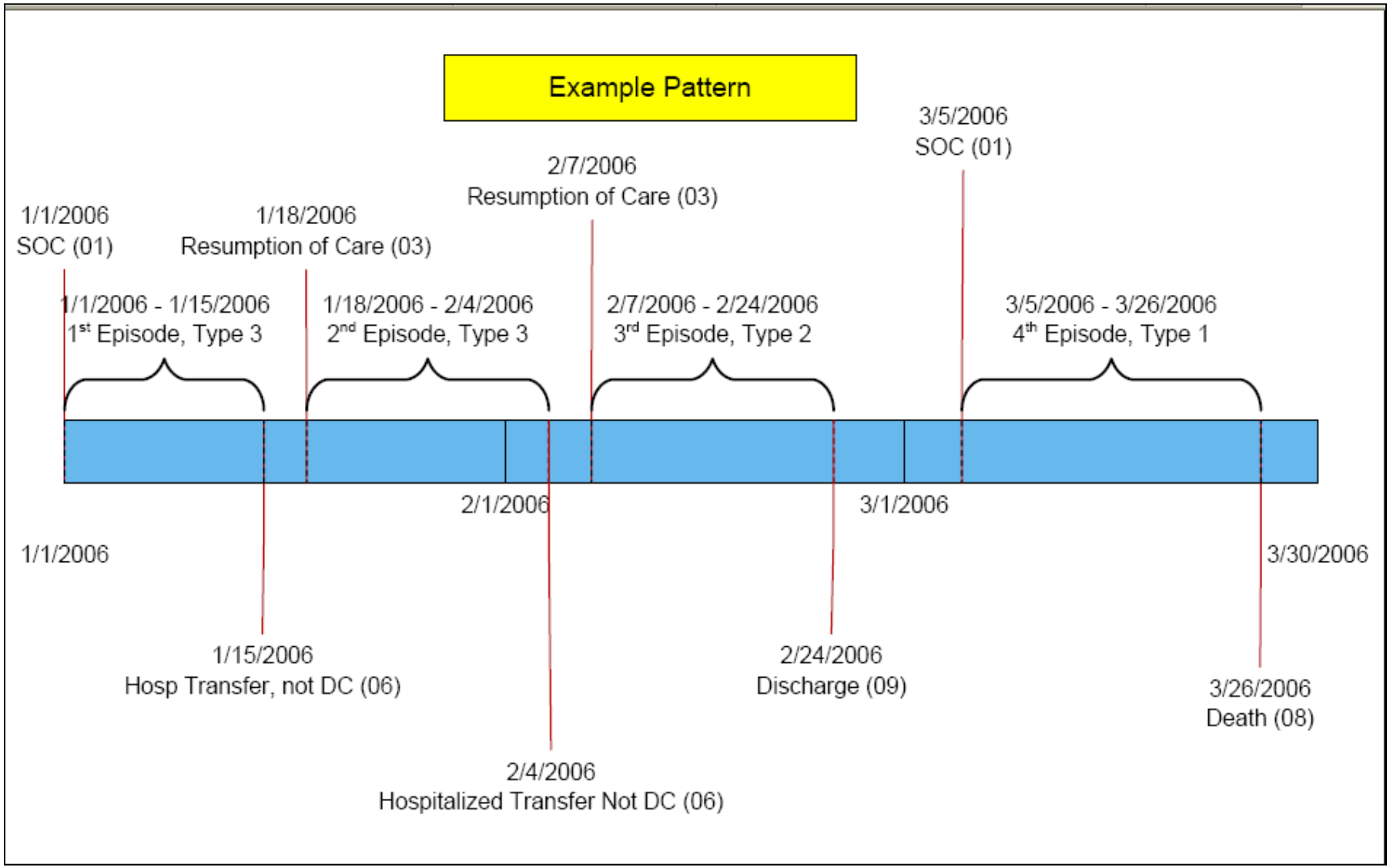
Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, pp. 37 – 54. http://www.kdnuggets.com/gpspubs/aimag-kdd-overview-1996-Fayyad.pdf. P. 41

Data Preparation

Data Preparation

- De-identification of data
- Selecting correct assessment type
- Creating episodes of care
- Removing duplicate records
- Data quality
  - Valid Values
    - Ambulation measured from 0 – 5, but data includes 6 and 7
  - Missing data
    - Just plain incomplete
    - Skip patterns
- Data type needed for analysis

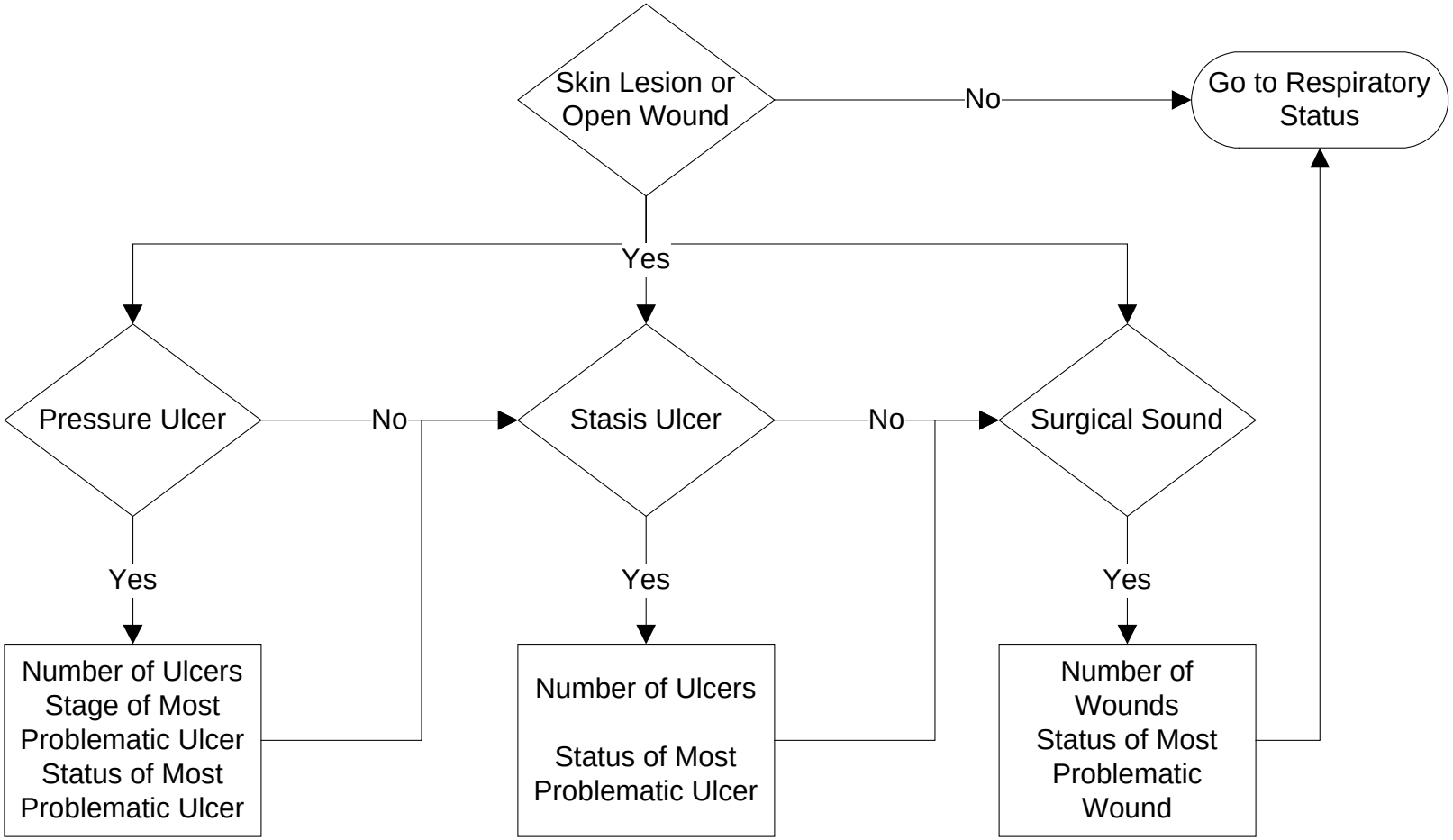
Unit of Analysis



Data 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 with too little variance

Figure 1. OASIS Integumentary Skip Pattern



Recoding Data

Stage of most problematic pressure ulcer

UK = Missing

NA = 0

1 - Stage 1 2 - Stage 2 3 - Stage 3 4 - Stage 4 NA - No observable pressure ulcer	0 No observable pressure ulcer 1 - Stage 1 2 - Stage 2 3 - Stage 3 4 - Stage 4

Data Transformation

- Create new variables
- Data reduction
- Format for to meeting assumptions for analyses
- Increase interpretability of results
- Decrease chaos

Creating Variables

- **Length of stay** = end date of episode – start date of episode + 1 day
  - M0090 Date of Assessment
  - We will need to look at distribution of this variable to determine categories and if there are any patients that are outliers that we might want to drop i.e. < 7 days or > 120 days
    1. < 30 days
    2. 30 – 59 days
    3. 60 – 89 days
    4. 90 – 120 days
    5. > 120 days

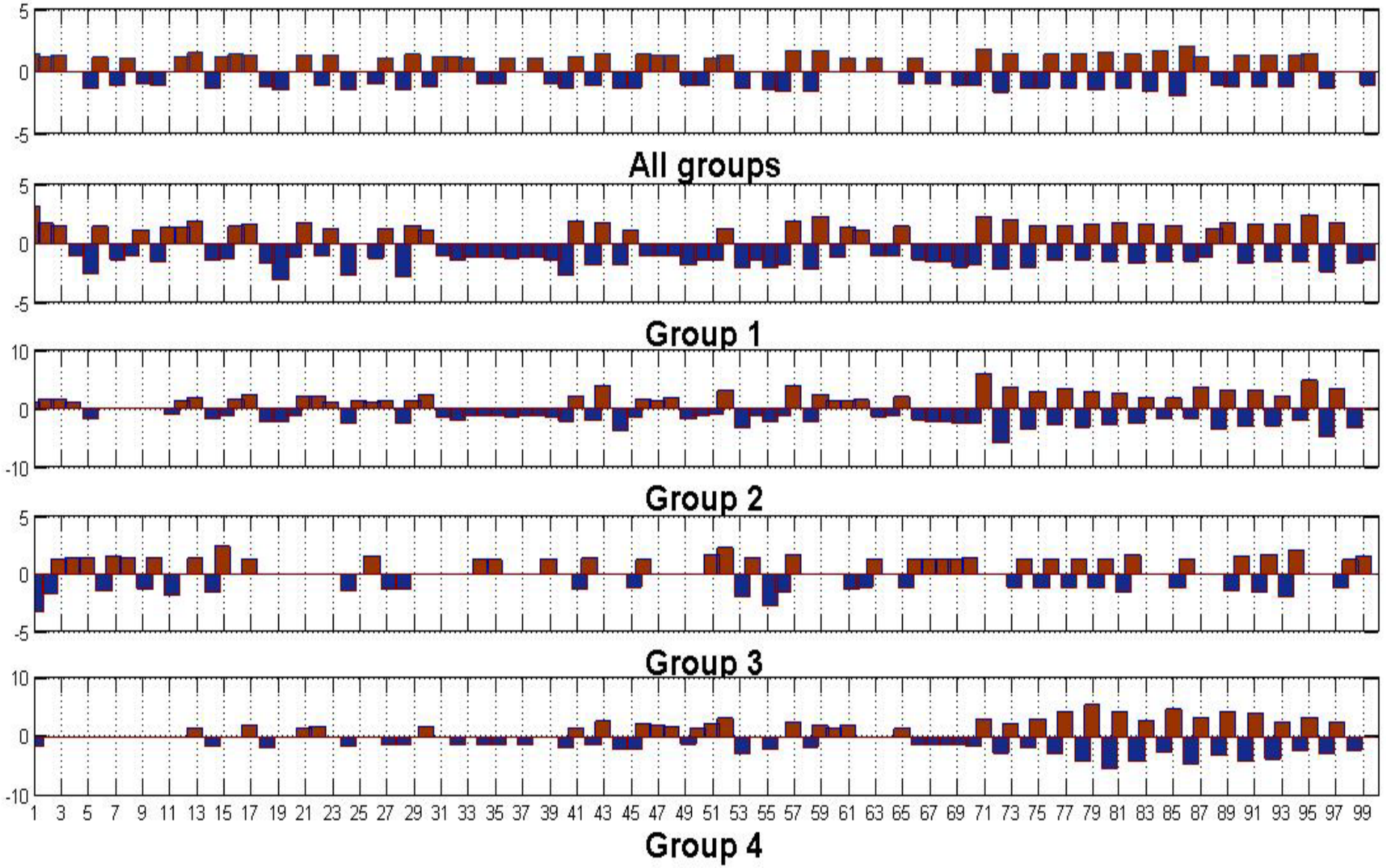
Transformation

- Clinical Classification Software
  - Primary diagnoses and then reduced into 51 smaller groups within 11 major categories
- Charlson Index of Comorbidity
  - Additional medical diagnoses
- Scales
  - Prognosis, Pain, Pressure Ulcer, Stasis Ulcer, Surgical Wound, Respiratory Status

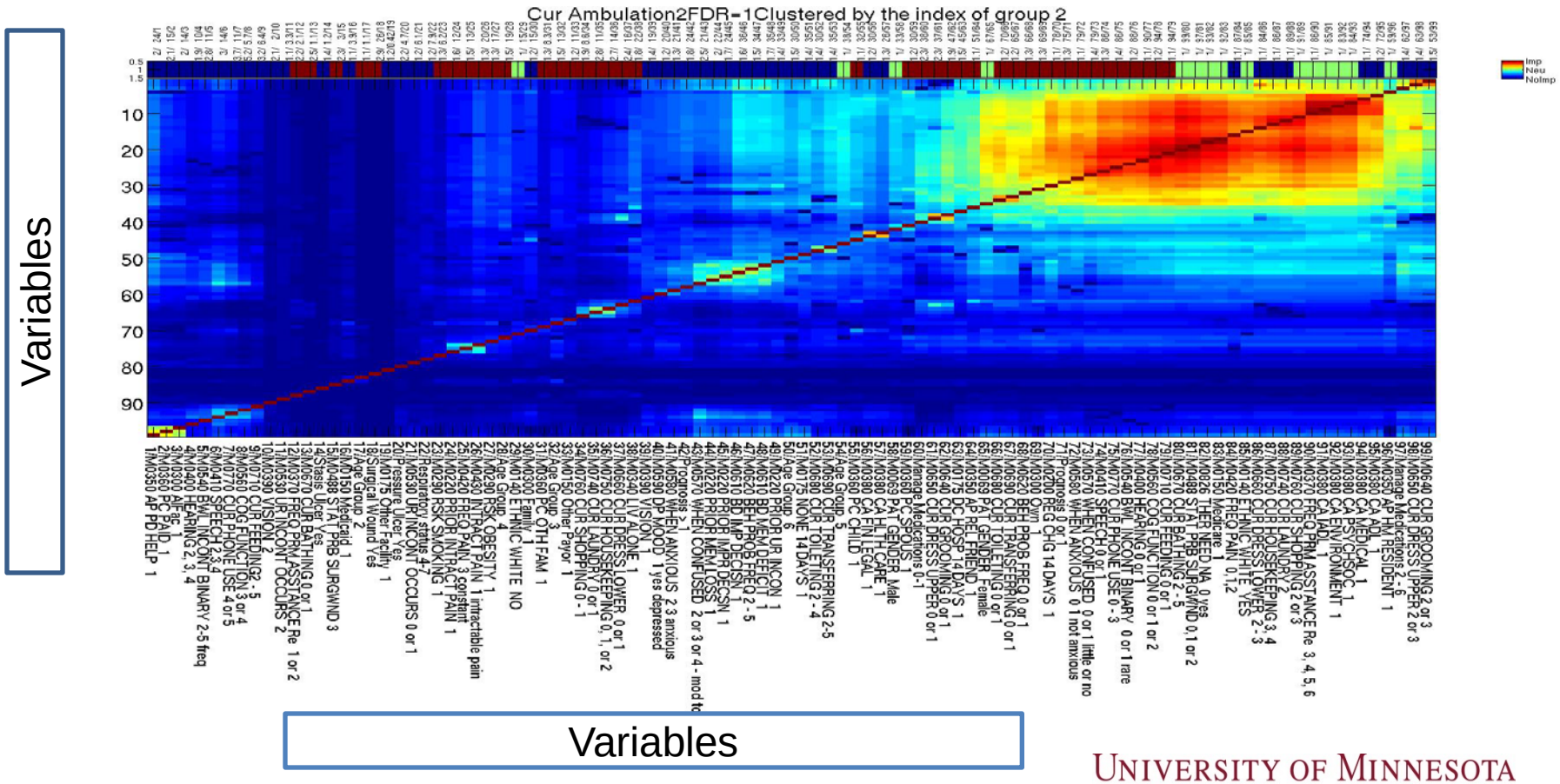
Data Mining Techniques

- We found the risk variables that are significantly associated with mobility outcome vary among the groups
- Group the single predictors based on whether they cover same or different patient group
  - Clustering
    - **Based on similarity of sample space**
    - **Not discriminative**
    - **High frequency variables got merged**
  - Pattern mining based approach
    - **Discriminative**
    - **Coherence (similarity of sample space)**

Subgroup Variability

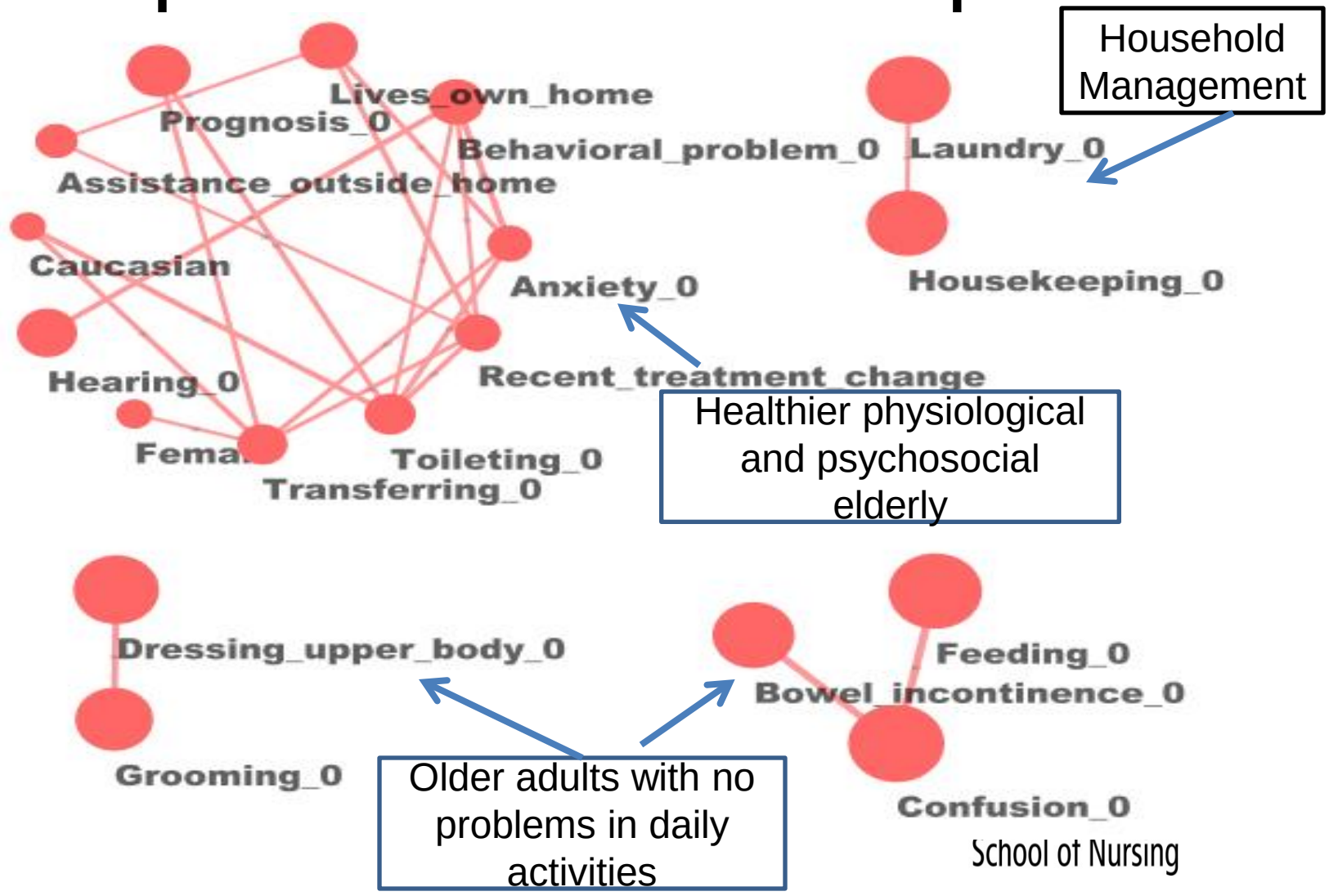


Clustering Variables – Subgroup 2

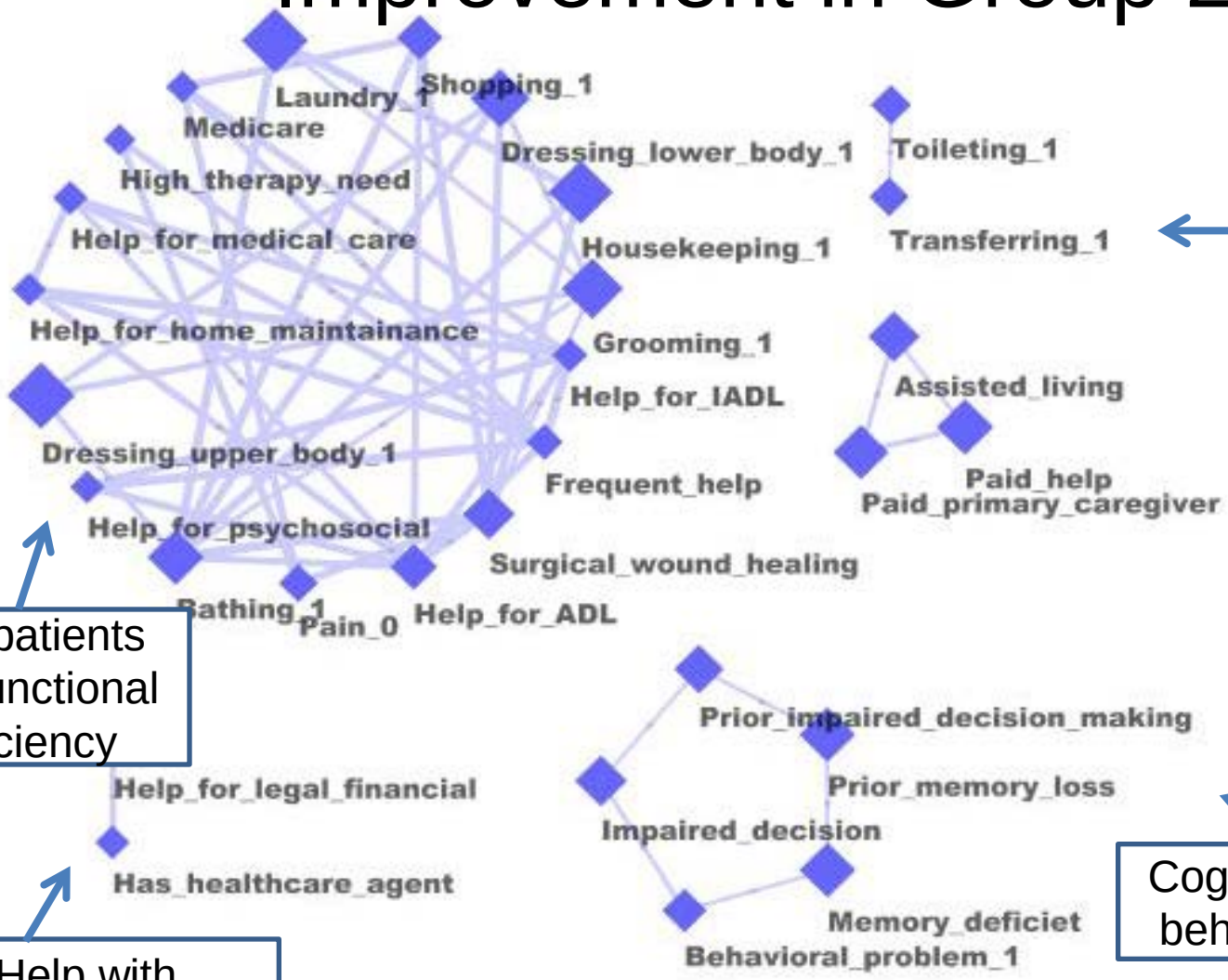


UNIVERSITY OF MINNESOTA
School of Nursing

Patterns Associated with Improvement in Group 2



Patterns Associated with No Improvement in Group 2



Incapable to toilet and transfer

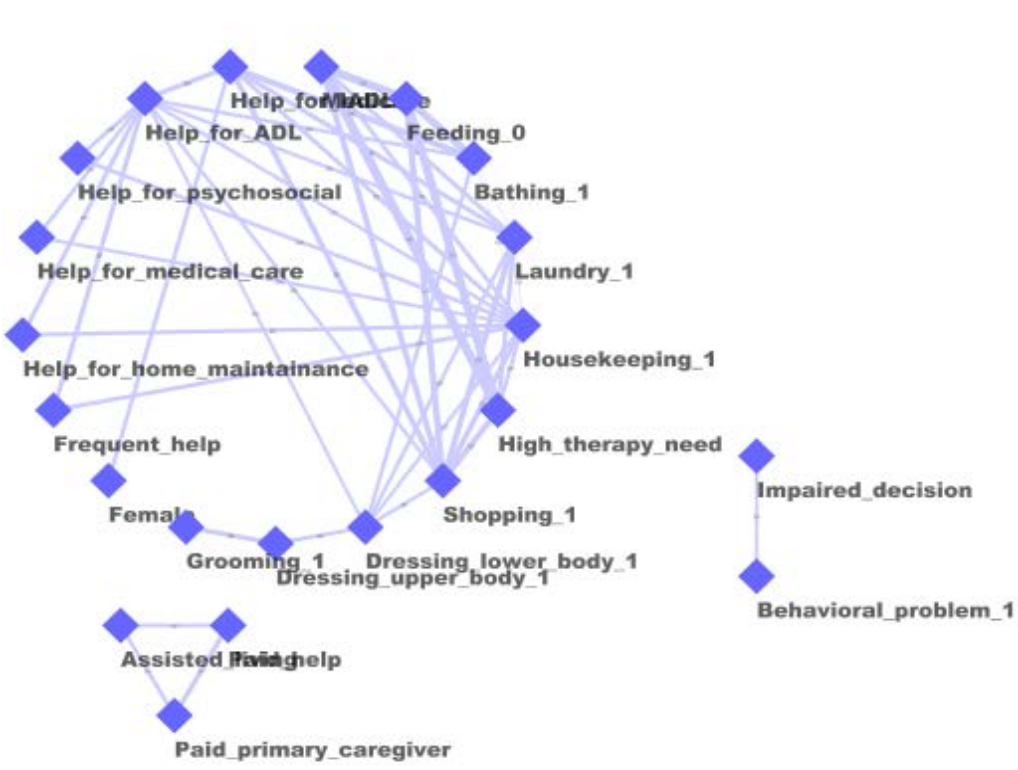
Paid Help

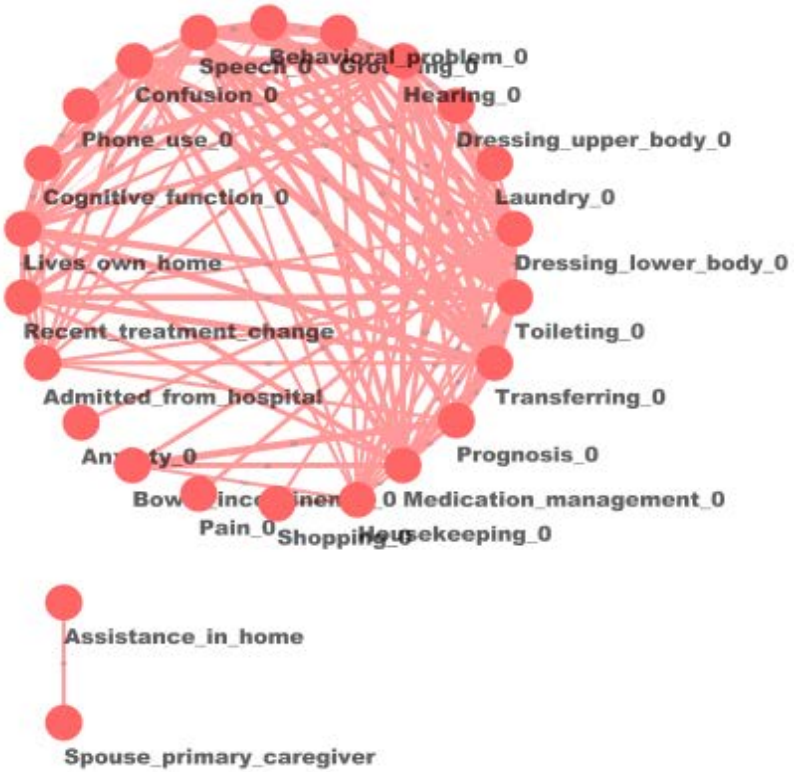
Frail patients with functional deficiency

Cognitive deficits and behavioral problems

Help with financial and legal matters

Patterns Associated with Mobility in Group 1

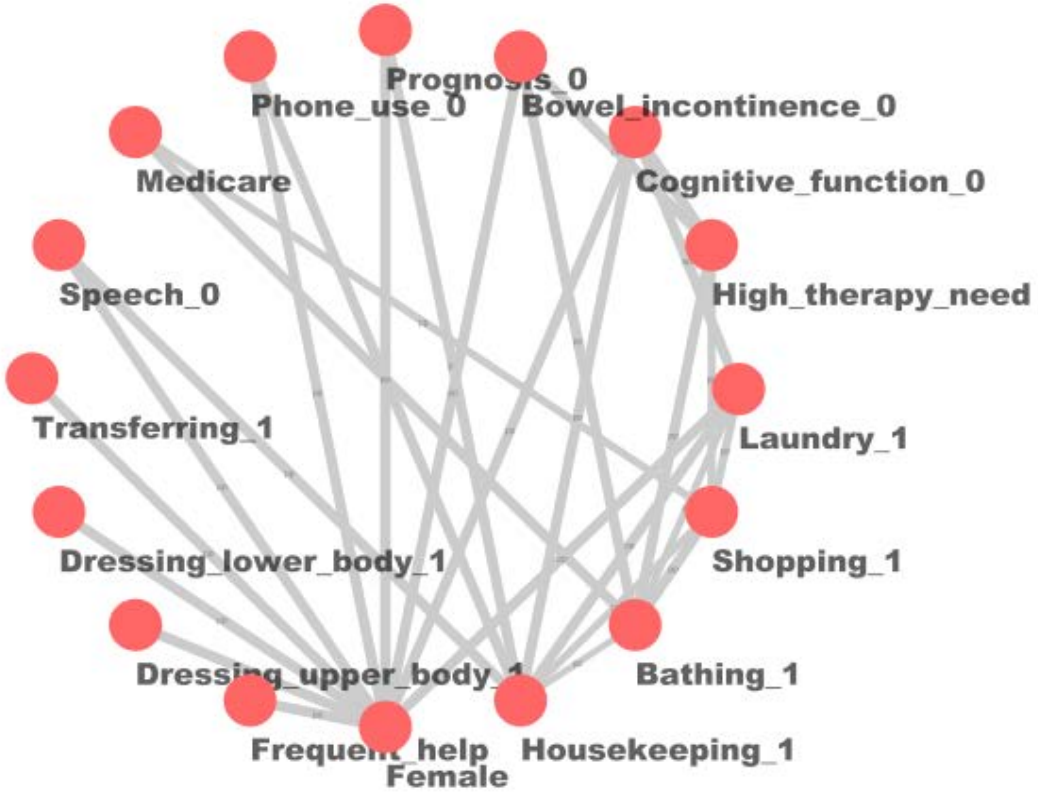




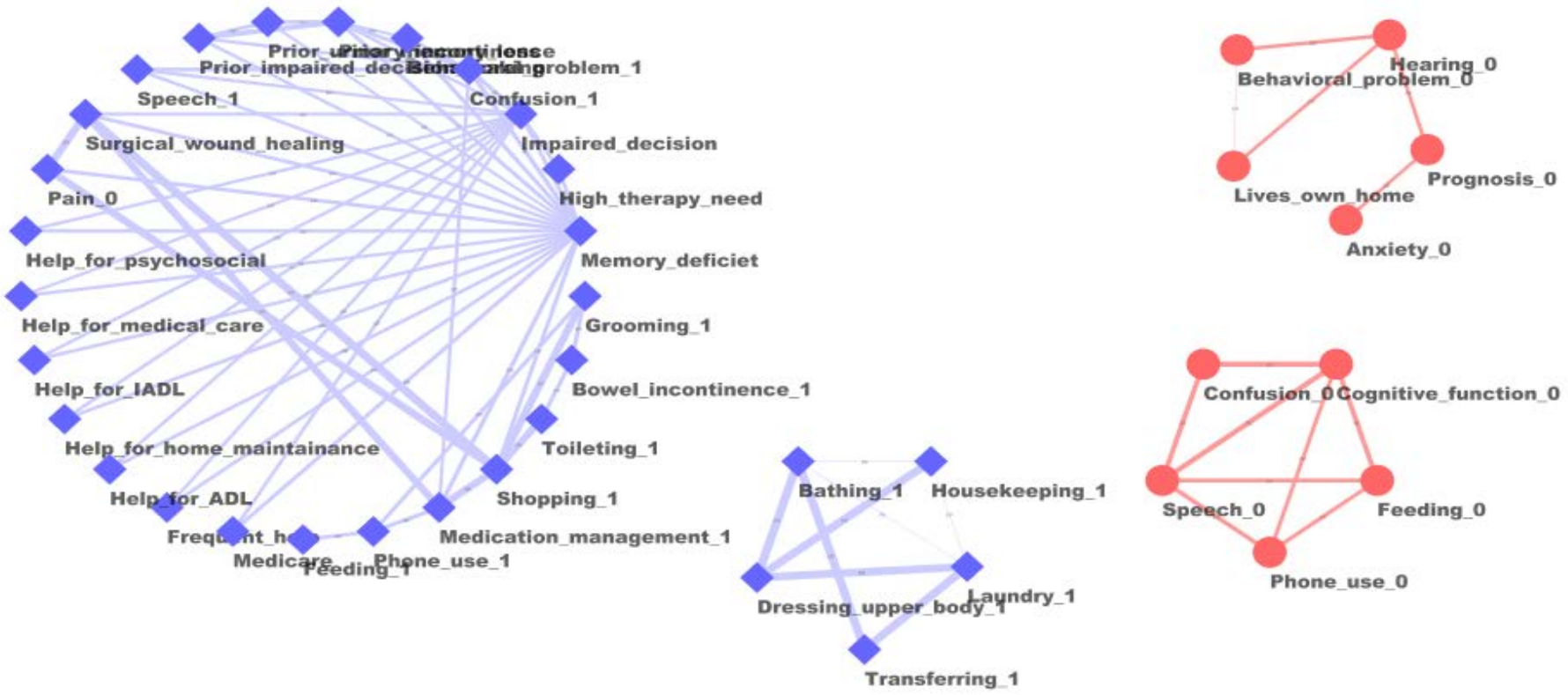
UNIVERSITY OF MINNESOTA

School of Nursing

Patterns Associated with Mobility in Group 3



Patterns Associated with Mobility in Group 4



Discussion

- **Single variables may be less helpful** than patterns of variables – higher categories
- Limitation
  - Large national sample – but not random, may be bias in results
  - Missing interventions due to lack of standardization
  - Length of stay may vary and contribute to findings
- Results are knowledge discovery, not hypotheses testing
- Integrate diagnosis codes (icd-9) and nursing interventions in future to combine factors related to mobility

Discussion

- High prevalence of mobility limitations for HHC patients (97%)
- Mobility status at admission highest predictor of improvement
  - CMS outcome reporting controlling for this, but doesn't look at differences by mobility status
- Variations of predictors within subgroups
- Different clusters point to the need to tailor interventions for subgroups

Next Steps

- Make recommendation to CMS about findings
- Replicate with OASIS C – contains some interventions
- Combine hospital and home care data to determine predictors upstream

UNIVERSITY OF MINNESOTA

School of Nursing