Knowledge Discovery
Data Analytics Methods:
Evaluating and Optimizing Healthcare Value Using Big Data Methods

Karen A. Monsen, PhD, RN, FAAN
December 7, 2015
mons0122@umn.edu
Three Major Shifts of Big Data Research

• The first is the ability to analyze vast amounts of data about a topic rather than be forced to settle for smaller sets.
• The second is a willingness to embrace data's real-world messiness rather than privilege exactitude.
• The third is a growing respect for correlations rather than a continuing quest for elusive causality.


“More, Messy, Good Enough” (p. 12, Mayer-Schonberger & Cukier, 2013)

• ‘Big Data’ changes the scientific paradigm
  • More data means less sampling error
  • Messy means we can try to understand and account for the biases inherent within observational data
  • Good enough means we can let go of our fixation on causation
    • Description
    • Pattern Discovery
  • Hypothesis generation
    • Letting the data give voice to nurses and patients


Nursing Data

**Nursing Minimum Data Set**

- a minimum set of elements of information with uniform definitions and categories concerning the specific dimensions of nursing, which meets the information needs of multiple data users in the health care system.

- Client characteristics & outcomes
- Nursing assessments & interventions


**Nursing Context Data**

**Nursing Management Minimum Data Set**

- core essential data needed to support the administrative and management information needs for the provision of nursing care. The standardized format allows for comparable nursing data collection within and across organizations.

- Nurse and health system characteristics
- Nurse and health system credentials


---

**Recognized Nursing Terminologies**

**American Nurses Association**

<table>
<thead>
<tr>
<th>Terminology</th>
<th>Nursing Problem</th>
<th>Nursing Interventions</th>
<th>Nursing Specific Items from the NMDS</th>
<th>Nursing Outcome</th>
<th>Nursing Intensity</th>
</tr>
</thead>
<tbody>
<tr>
<td>NANDA (1992)</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NIC (1997)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NOC (1997)</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The above three terminologies must be used together to obtain information about the nursing problem (diagnosis), intervention and outcome. The below terminologies all have terms for the nursing problem, intervention, and outcome.

<table>
<thead>
<tr>
<th>Terminology</th>
<th>Nursing Problem</th>
<th>Nursing Interventions</th>
<th>Nursing Specific Items from the NMDS</th>
<th>Nursing Outcome</th>
<th>Nursing Intensity</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCC (HHCC) (1992)</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>INDS (1997)</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ICNP (2000)</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Interdisciplinary Terminologies

<table>
<thead>
<tr>
<th>Terminology</th>
<th>Nursing Problem</th>
<th>Nursing Interventions</th>
<th>Nursing Specific Items from the NMDS</th>
<th>Nursing Outcome</th>
<th>Nursing Intensity</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNOMED-CT (1999)</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SNOMED-CT Nursing Subset</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOINC (2002)</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Omaha System (1992)</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


http://dlthede.net/informatics/Chap16Documentation/chap16.html
Big Data Laboratory

• 2010: Dean Delaney invited the Omaha System Partnership for Knowledge Discovery and Healthcare Quality within the University of Minnesota Center for Nursing Informatics
  – Scientific teams
  – Affiliate members
  – Data warehouse
Using a Logistical Mixed-effects Model with Nursing Data

- How do nurses and interventions contribute to variability in patient and population health?

- Nurse (17%)
- Client (50%)
- Problem (17%)
- Intervention (17%)

Client age was significantly positively associated with knowledge benchmark attainment in all models.

This research is partially supported by the National Science Foundation under grant # SES-0851705, and by the Omaha System Partnership.


Using Data Visualization to Detect Client Risk Patterns

Each image (sunburst) was created in d3 from public health nursing assessment data for a single patient. Data were generated by use of the Omaha System signs and symptoms and Problem Rating Scale for Outcomes.

Key:
- Colors = problems
- Shading = risk
- Rings = Knowledge, Behavior, and Status
- Tabs = signs/symptoms

Documentation patterns suggest a comprehensive, holistic nursing assessment.

Kim et al. found that the presence of mental health signs and symptom tends to be associated with more diagnostic problems and worse patient condition.

Population Perspectives on Assessment

• Our assessments may be guided by evidence and policy
  • Required documentation protocols
  • Population characteristics
    • Maternal-child health
    • Diabetes
    • Tuberculosis
    • Community-dwelling elders

Diabetes
Tuberculosis

Community Dwelling Elders
Using Generalized Estimating Equations for Cohort Comparison

- Mothers with intellectual disabilities have twice as many problems as mothers without intellectual disabilities
- Receive more public health nursing service
  - Twice as many encounters and interventions
- Show improvement in all areas
  - Do not reach the desired health literacy benchmark in Caretaking/parenting


Using Graphing Methods with Multilevel Kway Partitioning to Form Non-Overlapping Intervention Clusters

This research was supported by a Midwest Nursing Research Society New Investigator Seed Grant. Monsen, K. A., Banerjee, A., & Das, P. (2010). Discovering client and intervention patterns in home visiting data. Western Journal of Nursing Research, 32(8), 1031-1054. doi:10.1177/0193945910370970
Using Kaplan-Meier Curves to Depict Problem Stabilization

Using Inductive and Deductive Approaches to Create Overlapping Intervention Groups

Relationships between four intervention grouping/clustering methods for wound care.

Using Receiver Operating Curves to Understand Model Fit

- Comparison of Intervention Modeling Approaches and Hospitalization Outcomes for Frail and Non-frail Elderly Home Care Patients

Using Logistic Regression to Associate Home Care Interventions and Hospitalization Outcomes

- Too little care may result in hospitalization when patients have more intensive needs
  - Frail elders are more likely to be hospitalized if they have low frequencies of four skilled nursing intervention clusters
Using Pattern Comparison Pre- and Post-Intervention to Demonstrate Intervention Effectiveness

Knowledge scores across problems over time
- Pre-intervention, patterns by race/ethnicity
- Post-intervention, patterns by problem


Toward Personalized Algorithms to Improve Nursing Care Quality and Efficiency

- Using data from PHN documentation for 1,618 low income/high risk clients including 113,989 interventions, determine the feasibility of data-driven personalized care approaches using a machine learning approach to answer resource-related research questions:
  1. which clients will need more services,
  2. which clients will have better outcomes if they receive more services
  3. which intervention patterns are used to address the Oral health problem
  4. how to personalize interventions to maximize client outcomes based on client characteristics.

Chih-Lin Chi, PhD, MBA; Brady Alsaker, MN, RN; Karen A. Monsen, PhD, RN, FAAN
Manuscript under review
Methods

Question 1 (intervention need) method employed support vector machines in Matlab software.
Input: Client characteristics (demographics and signs/symptoms from first encounter).
Output: Intervention resource across all clients (compared to 50th and 75th percentiles).

Question 2 (responsiveness) method employed support vector machines in Matlab software with sensitivity analysis.
Input: Client characteristics (demographics and signs/symptoms from first encounter).
Output: Responsive score (sensitivity) based on personal characteristics.

Question 3 (care planning) method employed simple cluster analysis in Excel using round-up or round-down technique using proportion of interventions by category observed in the data.

Question 4 (outcome optimization) method employed support vector machines in Matlab software with optimization.
Input: Client characteristics (demographics and signs/symptoms from first encounter) and intervention patterns for each client.
Output: Any KBS improvement from admission to discharge.

Algorithm 1: Predicting Intervention Need

• Approximately 20% of clients received over 70% of interventions. The intervention needs of clients were identified in two scenarios (50% of clients and 25% of clients).

Model 1: 50% of clients received 92% of interventions.
Accuracy: 60%
Area Under Curve: 75%

Model 2: 25% clients received 76% of interventions.
Accuracy: 74%
Area Under Curve: 77%
Algorithm 2: Predicting Responsiveness

- Improvement was twice as great for clients who were predicted to exhibit high responsiveness than for those who are predicted low responsiveness.

Algorithm 3: Predicting Personalized Care

- Round up-round down techniques identified eight combinations of interventions that were used to personalize care.
Algorithm 4: Optimizing Outcomes

• For clients with improvement space, the probability of improvement using predicted personalized care was 49% (p < .001) based on a client's characteristics, needs, and responsiveness.

• Improvement space (average of the difference between the highest and lowest predicted outcome improvement probabilities across all clients): 1.18%

• The relative outcome improvement was 49% (p<0.001), which was calculated by comparing probability increase (0.58%) to the largest possible improvement space of 1.18%.

Next Steps

• Standardize the process and develop data-mining pipeline for other Problems
• Validate with world-wide structured nursing data

<table>
<thead>
<tr>
<th>Problems Data Source</th>
<th>Oral health Problem</th>
<th>Caretaking (parenting)</th>
<th>Growth and development</th>
<th>Pregnancy</th>
<th>Other Problems</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dakota County, MN</td>
<td>Personalized intervention</td>
<td>To be done</td>
<td>To be done</td>
<td>To be done</td>
<td>To be done</td>
</tr>
<tr>
<td>Other County, MN</td>
<td>To be done</td>
<td>To be done</td>
<td>To be done</td>
<td>To be done</td>
<td>To be done</td>
</tr>
<tr>
<td>Other County, Other State</td>
<td>To be done</td>
<td>To be done</td>
<td>To be done</td>
<td>To be done</td>
<td>To be done</td>
</tr>
<tr>
<td>Other Country</td>
<td>To be done</td>
<td>To be done</td>
<td>To be done</td>
<td>To be done</td>
<td>To be done</td>
</tr>
</tbody>
</table>
Further Research

Monsen, K. A. et al., 2014

• Examine the impact of evidence-based care
  • What is the impact of evidence-based care effect relative to patient assessment, intervention tailoring, and patient outcomes?
  • What is the value of evidence-based care?
• Examine associations between intervention patterns and client outcomes
  • Is there differential effectiveness of intervention patterns for similar client profiles?
  • What is optimal intervention tailoring?
  • What is the nurse effect?
• Evaluate patterns across agencies and programs
  • Do patterns persist across agencies?
  • What aspects of nursing interventions are similar across programs and populations?
  • How does individualized care relate to evidence-based practice?


Recommendations

• Expand the development of data analytics methods to incorporate nursing and interprofessional datasets and encompass all standardized terminologies and structured data.
• Continue to develop and test new methods on existing datasets.