Big Data at the Interdisciplinary Level to Improve Care

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

Associate Professor and Director, Center for Nursing Informatics
Summer Institute for Nursing Informatics
July 20, 2016

http://umn.edu/bigdata

Purpose

• Identify information documented by nurses and interprofessional providers in flowsheet data for big data and data science
• Discuss the challenges of using flowsheet data for research
• Demonstrate use of information models with standardized data to create usable flowsheet data
• Examine an exemplar of interprofessional research that incorporates flowsheet data.

Big Data & 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
• Three-legged stool
  • Data
  • Technology
  • Algorithms

Example Flowsheet

<table>
<thead>
<tr>
<th>Adult Assessment</th>
<th>Vital Signs</th>
<th>Input</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pain Location</td>
<td>Temperature</td>
<td>Heart Rate</td>
<td>Blood Pressure</td>
</tr>
<tr>
<td>Headache</td>
<td>Blood Glucose</td>
<td>Respiration</td>
<td>Blood Pressure</td>
</tr>
<tr>
<td>Neck Pain</td>
<td>Blood Glucose</td>
<td>Respiration</td>
<td>Blood Pressure</td>
</tr>
<tr>
<td>Chest Pain</td>
<td>Blood Glucose</td>
<td>Respiration</td>
<td>Blood Pressure</td>
</tr>
<tr>
<td>Abdominal Pain</td>
<td>Blood Glucose</td>
<td>Respiration</td>
<td>Blood Pressure</td>
</tr>
<tr>
<td>Jaw Pain</td>
<td>Blood Glucose</td>
<td>Respiration</td>
<td>Blood Pressure</td>
</tr>
<tr>
<td>Shoulder Pain</td>
<td>Blood Glucose</td>
<td>Respiration</td>
<td>Blood Pressure</td>
</tr>
<tr>
<td>Upper Arm Pain</td>
<td>Blood Glucose</td>
<td>Respiration</td>
<td>Blood Pressure</td>
</tr>
<tr>
<td>Lower Arm Pain</td>
<td>Blood Glucose</td>
<td>Respiration</td>
<td>Blood Pressure</td>
</tr>
<tr>
<td>Elbow Pain</td>
<td>Blood Glucose</td>
<td>Respiration</td>
<td>Blood Pressure</td>
</tr>
<tr>
<td>Wrist Pain</td>
<td>Blood Glucose</td>
<td>Respiration</td>
<td>Blood Pressure</td>
</tr>
<tr>
<td>Finger Pain</td>
<td>Blood Glucose</td>
<td>Respiration</td>
<td>Blood Pressure</td>
</tr>
</tbody>
</table>

Example flowsheets and data visualizations are used to illustrate the application of big data and big data science in nursing informatics.
Flowsheets – Interdisciplinary Care

Flowsheet Data Challenges

- Lack of standardized codes
- Volume of data – what is essential?
- Duplication for the same assessments, interventions, & goals
  - Different people building screens
  - Software upgrades
  - Discipline/practice specific needs
  - Typing (spaces, capitalization, plural, etc)
  - Pre-coordination
- No information models exist across EHR
- Data driven information modeling required

UMN – Academic Health Center

Flowsheets constitute 34% of all data

- 14,564 measure types
- 2,972 groups
- 562 templates
- 1.2 billion observations
- 2,000 measures cover 95% of observations

2,401,377 patients as of 6/10/16

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

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

**Purpose:**

- Support for the use of recognized terminologies supporting nursing practice
- Promote integration of terminologies into information technology solutions
- Facilitate interoperability between different concepts, nomenclatures, and information systems


---

**Position Statement**

When exchanging a Consolidated Continuity of Care Document (C-CDA) with another setting for problems and care plans,

- SNOMED CT® and LOINC® should be used for exchange
- LOINC® should be used for coding nursing assessments and outcomes
- SNOMED CT® for problems, interventions, and observation findings


**Pressure Ulcer Model - Standardized**

[NDNQI - Skin/ Pressure Ulcer](http://www.nursingworld.org/MainMenuCategories/Policy-Advocacy/Positions-and-Resolutions/ANAPositionStatements/Position-Statements-Alphabetically/Inclusion-of-Recognized-Terminologies-within-EHRs.html)
Unique Flowsheet IDs

Actual Data Collected

Skin Inspection
- Site: WDC
- Inspection: Type
- Body site inspected
- Skin: Temperature
- Skin: Moisture
- Skin: Turgor
- Skin: Integrity
- Skin: Characteristic
- Skin: Signs/Symptoms/Conditions
- Pressure Ulcer: Assessment
- Pressure Ulcer: Present
- Pressure Ulcer: Problems Assessed
- Pressure Ulcer: Location
- Drainage: Amount
- Drainage: Color or Characteristics
- Site: Assessment Wound
- Wound: Size
- Wound: Skin Edges
- Wound: Tunneling Length (cm)
- Wound: Undermining Length (cm)
- Pressure Ulcer: Dressing

Skin Color – Choice Lists
- Ashen: no abnormal color
- Blanched: non-blanchable
- Blue: other (see comments)
- Color consistent with ethnicity: pale
- Dusk: red
- Dusky: ruddy
- Ecchymotic: shiny
- Flushed: waxen
- Gray: white
- Jaundice: yellow
- Mottled:
Skin Inspection Choices

- abrasion
- excoriation
- pressure ulcer
- blister
- fragile
- pressure ulcer(s)
- body piercing
- incision
- rash
- burn(s)
- incision
- rash
- cracked
- incision(s)
- rash(s)
- cut(s)
- intact
- scale
- cyst
- itchy
- scar
- drain/device
- mass
- skin tear
- erosion
- petechiae
- tattoo
- subcutaneous emphysema (specify)
- wound
- erosion
- petechiae
- tattoo
- subcutaneous emphysema (specify)
- wound

Requirements for Useful Data

- Common data (information) models
- Standardized coding of data
- Standardized queries

PCORnet CDM Domains, v3.0

CONDITION
DEATH
DEATH_CAUSE
DEMOGRAPHIC
DIAGNOSIS
DISPENSING
ENCOUNTER
LAB_RESULT_CDM
PCORnet_TRIAL
PRESCRIBING
PRO_CDM
PROCEDURES
VITAL

Continuum of Care
Vision
Inclusion of Nursing Interprofessional Data

Creating Information Modes
Flowsheet Data

Sample Data Source - Clinical Data Models
- Flowsheet Data from 10/20/2010 - 12/27/2013
- 66,660 patients
- 199,665 encounters

Information Model Development Process

Flowsheet Information Models

<table>
<thead>
<tr>
<th>Topic</th>
<th>System</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pain</td>
<td>Cardiovascular System</td>
</tr>
<tr>
<td>Falls/Safety</td>
<td>Gastrointestinal System</td>
</tr>
<tr>
<td>Peripheral Neurovascular (VTE)</td>
<td>Neuromusculoskeletal System</td>
</tr>
<tr>
<td>Genitourinary System/CAUTI</td>
<td>Respiratory System</td>
</tr>
<tr>
<td>Pressure Ulcers</td>
<td>Vital Signs, Height &amp; Weight</td>
</tr>
<tr>
<td>Information Model Name</td>
<td>Number Flowsheet IDs (Mapped to Observables)</td>
</tr>
<tr>
<td>------------------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Cardiovascular System</td>
<td>241</td>
</tr>
<tr>
<td>Falls</td>
<td>59*</td>
</tr>
<tr>
<td>Gastrointestinal System</td>
<td>60</td>
</tr>
<tr>
<td>GU/CAUTI</td>
<td>79</td>
</tr>
<tr>
<td>Musculoskeletal System</td>
<td>276</td>
</tr>
<tr>
<td>Pain</td>
<td>309</td>
</tr>
<tr>
<td>Pressure Ulcers</td>
<td>104</td>
</tr>
<tr>
<td>Respiratory System</td>
<td>272</td>
</tr>
<tr>
<td>VTE</td>
<td>67</td>
</tr>
<tr>
<td>Vital Signs/Anthropometrics</td>
<td>85</td>
</tr>
</tbody>
</table>

Example Mapping Pain

Example of Mapping GU

Example High Level IM for GU
Next Steps

• Validating information models with other organizations
  • Allina, Cedars Sinai, Duke, Kaiser Permanente, MediComp, NY VNS
  • Beginning with Pain, then other quality measures
  • You are welcome to join also!
• Extracting their EHR flowsheet data into specific Excel spreadsheets
• Searching for concepts in the Pain IM
• Finalize a generalizable model – comparing across organizations

• Resources
  • Pain Information Model for Validation.xls
  • High Level Clinical Information Models from Flowsheet Data

Purpose

• Demonstrate a methodology for using EHRs data to estimate the compliance with the Surviving Sepsis Campaign (SSC) guideline recommendations
• Estimate the effect of the SSC individual and combined recommendations on the prevention of in-hospital mortality and sepsis-related complications

A Data Mining Approach to Determine Sepsis Guideline Impact on Inpatient Mortality and Complications

Bonnie L. Westra, PhD, RN, FAAN, FACMI
Lisiane Pruinelli, MSN, RN, PhD-C; Pranjul Yadav, PhD-C;
Alexander Huff; Jakob Johnson; Maribet McCarty, PhD, RN;
Vipin Kumar, PhD; Connie W. Delaney, PhD, RN, FAAN, FACMI;
Michael Steinbach, PhD; György J. Simon, PhD

Data and Cohort Selection

• De-identified EHR data from a Midwest health system was transferred to a CDR/UMN through a CTSA
• Timeframe: 1/1/09 – 12/31/11
• Inclusion criteria:
  • ICD-9 diagnosis = 995.92 and 785.5*
• Exclusion criteria:
  • Patients with cardiogenic shock
  • Patients with no antibiotic therapy
Study Design

- Baseline
  - In patients with at least 2 indications of sepsis, the time of the first sepsis indication
  - Pre-existing complications assessed at admission
  - Patient’s labs and vitals at baseline
  - Outcomes
    - Complications assessed at discharge
    - Updated for 30 days after discharge for complications

Baseline

- Sociodemographics
  - Age
  - Gender
  - Race/ethnicity
  - Payer (Medicaid for low income)
  - Charlson Index of Comorbidity

- Vital signs
  - Heart rate (HR)
  - Respiratory Rate (RR)
  - Temperature (Temp)
  - Mean arterial pressure (MAP)

- Laboratory results
  - Lactate
  - White Blood Cell Count (WBC)
  - Level of severity for each outcome complication

Baseline/Outcomes

- Complication (in hospital and up to 30 days after discharge)
  - Cerebrovascular
  - Respiratory
  - Cardiovascular
  - Kidney
  - In-hospital mortality

Sepsis Time Zero

At least 2 of the following criteria:

- MAP < 65
- HR > 100
- RR > 20
- Temp < 95 or > 100.94
- WBC < 4 or > 12
- Lactate > 2.0
### 3 Hour Bundle

- Lactate checked between admission and 0+3 hrs
- Blood culture drawn between admission and 0+3 hrs
- Broad-spectrum ABx administered between admission and 0+3 hrs
- Crystalloid fluid administered if MAP < 65 or lactate > 4 between admission and 0+3 hrs

### Data Preparation/Transformation

- Mapping the SSC guidelines to EHR data elements
- Data quality assessment based on literature and domain knowledge
  - Baseline missing values
  - For each data element, we evaluated range and created rules for suitable range to be included
  - Compared with other values i.e. MAP and SBP/DBP and health conditions
  - Determine the use of one or more flowsheet measures for vital signs
- Created ICD grouping / severity for complications - respiratory, cardiovascular, cerebrovascular, renal

### Method: Design

**X**: SSC guidelines
  - Interventions

**Y**: Outcomes
  - Mortality
  - Complications

**Z**: Confounders
  - Demographics such as age, gender
  - Comorbidities
  - Laboratories test results

Our aim is to estimate the causal effect of X on Y, while simultaneously adjusting for the effects of Z (i.e. Confounders).

### Method: Estimation Metric

- Metric for estimating the causal effect of intervention on outcome was ATT
- ATT is defined as the **Average Treatment effect on the Treated**
  \[\text{ATT} = \mathbb{E}(Y_1 - Y_0) = \mathbb{E}(Y_1) - \mathbb{E}(Y_0)\]
  - \(\mathbb{E}\) is the Expectation operator
  - \(Y_1\) represents the outcome when intervention was prescribed
  - \(Y_0\) represents the outcome when no intervention was prescribed
Method: Bootstrapped PSM

- Propensity score matching (PSM) was used for estimating the causal metric.
- PSM is a statistical matching technique to estimate the effect of a treatment, policy, or other intervention by accounting for the covariates that predict receiving the treatment.
- Caliper Matching was used to match patients from treatment (intervention was prescribed) to that of control (intervention was not prescribed).
- Bootstrapping was used in conjunction with PSM to estimate the CI's.

Baseline Features (n=177)

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Patient Count</th>
<th>Condition (Severity Score)</th>
<th>Patient Count</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Count/Mean (IQR)</td>
<td>Count (range of score)2)</td>
<td></td>
</tr>
<tr>
<td>Age (years)</td>
<td>61 (51-71)</td>
<td>Cardiovascular</td>
<td>109 (0-2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0-12)</td>
<td>32 (3-4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>36 (5+)</td>
</tr>
<tr>
<td>Gender (Male)</td>
<td>102</td>
<td>Cerebrovascular</td>
<td>158</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0-4)</td>
<td>19 (3)</td>
</tr>
<tr>
<td>Race (Caucasian)</td>
<td>97</td>
<td>Respiratory</td>
<td>156</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0-6)</td>
<td>21 (3+)</td>
</tr>
<tr>
<td>Ethnicity (Latino)</td>
<td>11</td>
<td>Kidney</td>
<td>121</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0-12)</td>
<td>56 (2+)</td>
</tr>
<tr>
<td>Payer (Medicaid)</td>
<td>102</td>
<td>Charlson Index</td>
<td>92</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0-17)</td>
<td>58 (2-4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>27 (5+)</td>
</tr>
</tbody>
</table>

Results of 0+ 3Hrs - 3 Hour Rules

<table>
<thead>
<tr>
<th>Rules Description</th>
<th>Patient Count / %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Lactate checked?</td>
<td>100/ 77 57% 0</td>
</tr>
<tr>
<td>2. Blood culture done?</td>
<td>97/ 80 55% 0</td>
</tr>
<tr>
<td>3. Broad spectrum antibiotic?</td>
<td>151/ 26 85% 0</td>
</tr>
<tr>
<td>4. Crystalloid fluid resuscitation for hypotension (MAP&lt;65) or if Lactate &gt;4</td>
<td>97/ 41 70% 39</td>
</tr>
</tbody>
</table>

Effect of Elements on Outcomes

<table>
<thead>
<tr>
<th></th>
<th>Mortality</th>
<th>Cardiovascular</th>
<th>Respiratory</th>
<th>Cerebrovascular</th>
<th>Kidney</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lactate checked</td>
<td>-0.066</td>
<td>0.617</td>
<td>-0.004</td>
<td>0.192</td>
<td>-0.204</td>
</tr>
<tr>
<td>Blood Culture checked</td>
<td>-0.051</td>
<td>0.514</td>
<td>0.534</td>
<td>0.003</td>
<td>-0.115</td>
</tr>
<tr>
<td>Antibiotic (Broad Spectrum)</td>
<td>-0.024</td>
<td>0.167</td>
<td>0.278</td>
<td>0.119</td>
<td>0.159</td>
</tr>
<tr>
<td>Crystalloid</td>
<td>0.058</td>
<td>-0.136</td>
<td>-0.561</td>
<td>0.045</td>
<td>-0.354</td>
</tr>
</tbody>
</table>
**Limitations**

- Data from a single center
- Small sample size
- No comparison group other than patients with sepsis who received or did not receive recommendations
- We may misestimate “TimeZero”, as the majority of patients were in the ED when “TimeZero” occurred
- Timeframe of data - under vs over coding for sepsis

**Summary**

- Flowsheet data extends clinical data available for research
- Standardization is needed to make flowsheet data useful
- JAMIA feedback (revisions needed – but…)
  - “The models are intended to facilitate navigation and extraction of data to support cohort discovery and comparative effectiveness research. This is a very important and relevant objective.”
  - “This work is incredibly important to standardize patient data (vitals, assessments, etc) that will be vital to any number of research and quality improvement activities. A set of standardized information models can allow local variation but provide a standard to aggregate and share data.”
Questions?
Bonnie L. Westra, PhD, RN, FAAN, FACMI
westr006@umn.edu