Data Acquisition Collaboration for Nursing Cost Study Using “Big Data”

Peggy Jenkins PhD, RN
Assistant Professor, University of Colorado College of Nursing

Esther Chipps PhD, RN, NEA-BC
Clinical Nurse Scientist, Associate Professor of Clinical Nursing

The Ohio State University Wexner Medical Center, The Ohio State University College of Nursing
Disclosure Slide

The authors have nothing to disclose.
Definitions of “Big Data”

• High volume or complex data pulled from multiple sources (Westra et al., 2015)
• Media term to describe data from huge electronic repositories (Clancy, 2016)
• Volume, Variety, Velocity, Veracity (Linnen, 2016)
• Disruptive innovation changing science (Kitchin, 2014)
• Sociological revolution (Brennan & Bakken, 2015)
Background

- 1980’s- Prospective Payment/ Diagnostic Related Groups (DRG’s)
- 1990’s Managed Care
- 2000’s Continued emphasis on cost containment
- 2012- Affordable Care Act/ Value Based Healthcare
  - Need to measure nursing cost per patient and link to patient outcomes to derive value of nursing (Caspers & Pickard, 2010; Pappas, 2013; Welton & Harper, 2015)
Science of Nursing Cost

- Nurse scientists have documented the need to measure variability in nursing cost at the patient level for decades (Chiang, 2009; Diers, Bozzo, & RIMS Acuity Project Group, 1997; Edwardson & Giovannetti, 1987; Naylor, Munro, & Brooten, 1991; Pappas, 2007; Sovie, 1988; Thompson & Diers, 1985; Wilson, Prescott, & Aleksandrowicz, 1988).
- Wide variation exists in nursing literature on definition of nursing cost (Wilson, Prescott, & Aleksandrowicz, 1988; Chiang, 2009; Witzell, Ingersoll, Schultz, & Ryan, 1996; Pappas, 2007; Welton, Zone-Smith, and Bandyopadhyay, 2009).
- Various methods used to derive nursing cost in nursing studies because patient level nursing cost data was not available (Chiang, 2009; Dowless, 2007; Knauf, Ballard, Mossman, & Lichtig, 2006; Pappas, 2007).
- Big Data provides contemporary source for advancing science of nursing cost.
Collaborating to Obtain Data

- PhD student working with Nurse Scientist
- Negotiations among vendor and hospital staff
- De-identified data issues
- Query created collaboratively
- IRB and legal issues
OSU Wexner Medical Center purchased patient assignment software in 2008.

The goal was an “end to end” workforce management solution which included patient acuity, staffing and scheduling, workload management and productivity/performance.

Workforce Performance Management System designed to help us to get the right nurse for the right patient at the right price at the right time.

At the time of purchase and installation, there was no specific research objective.
The Modules

- Schedule Manager (Client & Web)
- Demand Manager (Web)
- Outcomes Driven Acuity (Web)
- Patient Assignment (Web)
Workforce Performance Management System

- Workforce performance management system makes it possible to view the quantitative impact of Acuity, CHURN, DPE on staffing requirements
- User since June 2008
- Scheduler nursing (48 units) and non nursing units (51 units)
  - Demand Manager - 2009
  - Acuity - 2009
  - Interface Acuity to EMR - 2013
  - Patient Assignment – Summer 2015
**Demand Acuity Report**

### 1100 - 1500 in FTE

**Wed 1/8/2014**

<table>
<thead>
<tr>
<th>Global Skill</th>
<th>Acuity Target</th>
<th>Actual</th>
<th>Var</th>
<th>Util</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>RN</td>
<td>9.42</td>
<td>9.00</td>
<td>-0.42</td>
<td>105%</td>
<td></td>
</tr>
<tr>
<td>PCA</td>
<td>3.99</td>
<td>4.00</td>
<td>0.01</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>CHG RN</td>
<td>1.00</td>
<td>1.00</td>
<td>0.00</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>Diff</td>
<td>0.00</td>
<td>0.81</td>
<td>0.81</td>
<td>0%</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>14.41</td>
<td>14.81</td>
<td>0.40</td>
<td>97%</td>
<td></td>
</tr>
</tbody>
</table>

**Volume:** 36 A: 0 D: 4 T: 0 TO: 1

### 1500 - 1900 in FTE

**Wed 1/8/2014**

<table>
<thead>
<tr>
<th>Global Skill</th>
<th>Acuity Target</th>
<th>Actual</th>
<th>Var</th>
<th>Util</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>RN</td>
<td>8.99</td>
<td>9.00</td>
<td>0.01</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>PCA</td>
<td>3.79</td>
<td>4.00</td>
<td>0.21</td>
<td>95%</td>
<td></td>
</tr>
<tr>
<td>CHG RN</td>
<td>1.00</td>
<td>1.00</td>
<td>0.00</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>13.78</td>
<td>14.00</td>
<td>0.22</td>
<td>98%</td>
<td></td>
</tr>
</tbody>
</table>

**Volume:** 31 A: 1 D: 6 T: 3 TO: 0

### 2100 - 2500 in FTE

**Wed 1/8/2014**

<table>
<thead>
<tr>
<th>Global Skill</th>
<th>Acuity Target</th>
<th>Actual</th>
<th>Var</th>
<th>Util</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>RN</td>
<td>89.67</td>
<td>9.00</td>
<td>-0.67</td>
<td>107%</td>
<td></td>
</tr>
<tr>
<td>PCA</td>
<td>3.99</td>
<td>4.00</td>
<td>0.01</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>CHG RN</td>
<td>1.00</td>
<td>1.00</td>
<td>0.00</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>Diff</td>
<td>0.00</td>
<td>0.81</td>
<td>0.81</td>
<td>0%</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>84.66</td>
<td>14.81</td>
<td>0.15</td>
<td>99%</td>
<td></td>
</tr>
</tbody>
</table>

**Volume:** 36 A: 0 D: 4 T: 0 TO: 1

### 2500 - 2900 in FTE

<table>
<thead>
<tr>
<th>Global Skill</th>
<th>Acuity Target</th>
<th>Actual</th>
<th>Var</th>
<th>Util</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>RN</td>
<td>9.11</td>
<td>9.00</td>
<td>-0.11</td>
<td>101%</td>
<td></td>
</tr>
<tr>
<td>PCA</td>
<td>3.79</td>
<td>4.00</td>
<td>0.21</td>
<td>95%</td>
<td></td>
</tr>
<tr>
<td>CHG RN</td>
<td>1.00</td>
<td>1.00</td>
<td>0.00</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>13.89</td>
<td>14.00</td>
<td>0.10</td>
<td>99%</td>
<td></td>
</tr>
</tbody>
</table>

**Volume:** 31 A: 1 D: 6 T: 3 TO: 0

### Notes

- **RN:** Registered Nurse
- **PCA:** Patient Care Assistant
- **CHG RN:** Critical Care Graduate Nurse
Data Acquisition

Data source was repository collected from three existing databases

1) Workforce Performance Management
2) Medical Information Management System/Information Warehouse
3) Human Resources Database
Resources to Coordinate Data Acquisition

- Software company Chief Nursing Officer
- Software company Data Manager
- OSU Legal staff
- OSU Information Warehouse
- OSU Nurse Informaticists
- OSU Nurse Manager
- OSU Human Resource Data Manager
Necessary Approvals

- Nursing Leadership
- IRB
  - OSU Exempt IRB
  - University of Colorado Exempt IRB
- Approval through OSU Legal Department
Patient Variables Extracted from Databases

- Medical Record Number
- Patient Encounter Number
- Age/Sex
- Patient’s acuity score
- Total number of hours that patient was on the unit
- Total hospital LOS
- Patients day of Admission
- ED admit
- Time in OR
Non-patient Variables Extracted from Databases

- Number of patients admitted per shift
- Number of patients discharged per shift
- Calculated nursing care hours per shift
- Date and number of patients assigned to each RN per shift
- Type of RN personnel (Skill level)
Variables from Medical Information Management System

- Patient Medical Record Number (MRN)
- Patient’s severity adjusted DRG (MSDRG)
- Patient Length Of Stay
- Patient age/gender
Variables from Human Resources Database

- Employee ID
- Age/Gender
- Nurse’s Wage
- Nurse’s education (BSN or non-BSN)
- Clinical Ladder status
- Certification
- Years in organization
- Total years as RN
De-identification Process

- Patient data from assignment software and IW was linked at OSU via patient medical ID number.
- Once linked the MRN was removed and each patient randomly assigned a number that cannot be linked to each identifier.
- Hospitalization dates removed and LOS calculated using “random secret referencing”.
- Human resources data regarding nursing staff was de-identified.
### ID Maps

<table>
<thead>
<tr>
<th>N</th>
<th>FAKE_MRN</th>
</tr>
</thead>
<tbody>
<tr>
<td>111111</td>
<td>1</td>
</tr>
<tr>
<td>222222</td>
<td>2</td>
</tr>
<tr>
<td>333333</td>
<td>3</td>
</tr>
<tr>
<td>444444</td>
<td>4</td>
</tr>
<tr>
<td>555555</td>
<td>5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>EMPID</th>
<th>FAKE_EMP</th>
</tr>
</thead>
<tbody>
<tr>
<td>8183838</td>
<td>1</td>
</tr>
<tr>
<td>4141140</td>
<td>2</td>
</tr>
<tr>
<td>6162310</td>
<td>3</td>
</tr>
<tr>
<td>7172538</td>
<td>4</td>
</tr>
<tr>
<td>200087256</td>
<td>5</td>
</tr>
<tr>
<td>200210277</td>
<td>6</td>
</tr>
<tr>
<td>5147082</td>
<td>7</td>
</tr>
<tr>
<td>222222211</td>
<td>8</td>
</tr>
<tr>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>-----</td>
<td>----</td>
</tr>
<tr>
<td>PT3557</td>
<td>2582</td>
</tr>
<tr>
<td>PT4359</td>
<td>3046</td>
</tr>
<tr>
<td>PT4231</td>
<td>1392</td>
</tr>
<tr>
<td>PT3023</td>
<td>764</td>
</tr>
<tr>
<td>PT1904</td>
<td>2851</td>
</tr>
<tr>
<td>PT3557</td>
<td>2582</td>
</tr>
<tr>
<td>PT415</td>
<td>1316</td>
</tr>
<tr>
<td>PT1272</td>
<td>1942</td>
</tr>
<tr>
<td>PT2902</td>
<td>3133</td>
</tr>
</tbody>
</table>
Data Management/Analysis

Organization (Long, 2009)
- Merged five excel files into Stata using patient and nurse common identifiers
- Stata do-files and Master do-files written to aid replication
- Variables named, cleansed, constructed
- Patients with missing data dropped

Analysis
- Descriptive analysis of all variables
- ANOVA to measure difference in 3 groups
- Regression analysis
- STATA v.12
Data Merger Architecture

“Patient Age & LOS” \( N = 4589 \)

Merge with “Location & LOS” \( N = 4589 \)

“Patient Age & LOS2” \( N = 4589 \)
Generate Inclusion Criteria

Drop 1051 not on study unit entire LOS

“Inclusion LOS” \( N = 3528 \)

Merge with “Assign Care & RN Intensity” \( n = 4200 \)

“Intensity” \( N = 3538 \)

Drop 185 patients with missing intensity

“Intensity C” \( N = 3353 \)

Drop 242 patients with missing days intensity

“Main Patient File” \( N = 3111 \)
44,842 shift obs

Merge Drop 71 Missing Employee observations

“Main File” \( N = 3111 \) patients
\( N = 150 \) nurses
44,771 shift obs

“HR Unit” \( N = 67 \)

“HR Float” \( N = 91 \)

Append

“Main HR File” \( N = 154 \)
Drop 4 duplicates
Results (Jenkins & Welton, 2014)

Patient Sample
- 3111 patients
- 54% female (n = 1688)
- Mean age 55 years (range 18-89)
- Mean Total days hospital 3.64 (range 0-25)
- Mean hours on unit 84.37 (range 1.12-600.18)

Nurse Sample

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Mean (SD)</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RN (n = 101)</td>
<td>Non-RN (n = 49)</td>
</tr>
<tr>
<td>Age</td>
<td>41.1 (11.2)</td>
<td>39.6 (13.1)</td>
</tr>
<tr>
<td></td>
<td>22-66</td>
<td>21-69</td>
</tr>
<tr>
<td>Years Service Organization</td>
<td>8.1 (4.6)</td>
<td>6.2 (4.3)</td>
</tr>
<tr>
<td></td>
<td>1-30</td>
<td>2-13</td>
</tr>
<tr>
<td>Years Service Unit</td>
<td>6.1 (4.2)</td>
<td>5.7 (4.2)</td>
</tr>
<tr>
<td></td>
<td>.05-13.9</td>
<td>.11-13.9</td>
</tr>
<tr>
<td>Wage</td>
<td>$31.21 ($7.16)</td>
<td>$14.19 ($4.81)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>92 (91%)</td>
<td>41 (84%)</td>
</tr>
<tr>
<td>Male</td>
<td>9 (9%)</td>
<td>8 (16%)</td>
</tr>
<tr>
<td>BSN</td>
<td>41 (41%)</td>
<td>NA</td>
</tr>
<tr>
<td>Clinical Ladder</td>
<td>2 (2%)</td>
<td>NA</td>
</tr>
<tr>
<td>Certification</td>
<td>10 (10%)</td>
<td>NA</td>
</tr>
<tr>
<td>Float</td>
<td>62 (41%)</td>
<td>27 (18%)</td>
</tr>
</tbody>
</table>
Ongoing Nursing Value Work

- National Nursing Value Workgroup (Welton & Harper, 2015)
- Defining standards for measuring nursing value (Garcia et al., 2015)
- Grant for multiple site study using nursing cost methodology
- Continue to generate nursing science on value of nursing using big data (Brennan & Bakken, 2015; Harper & Parkerson, 2015; Keenan, 2014; Westra et al., 2016)
References


References


