Data Mining using the Omaha System
Oral Health in Dakota Co.

Gyorgy J. Simon
Jeanne Pfeiffer
Laura L. Fitzsimmons
Karen A. Monsen
Oral Health

• Oral health is a window to overall health
  – Gum disease may allow bacteria to enter the bloodstream
  – Chronic periodontitis can contribute to cardiovascular disease
  – Gum disease have been linked to premature birth

• Few databases contain information about oral health
  – One of the 42 central concepts in the Omaha System
Oral Health in Dakota Co.

• Familial data for 6,425 clients
  – Problems, Signs and Symptoms, Interventions,
  – 1,781 (27.7%) have oral problems
  – Study period: 2009 – 2011

• Find patterns in clients that are predictive of oral health problems
  – Initially, patterns in oral health data
  – Patterns in mothers predictive of oral health problem in children
  – Successful interventions
## Data Mining

<table>
<thead>
<tr>
<th>Data Mining</th>
<th>Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Exploratory</strong></td>
<td><strong>Confirmatory</strong></td>
</tr>
<tr>
<td>• Find novel, interesting patterns</td>
<td>• Confirm hypothesis</td>
</tr>
<tr>
<td><strong>Hypothesis generation</strong></td>
<td><strong>Hypothesis testing</strong></td>
</tr>
<tr>
<td>• Large number of hypotheses</td>
<td>• Few hypotheses</td>
</tr>
<tr>
<td>• Filter down to a smaller set</td>
<td></td>
</tr>
<tr>
<td><strong>No guarantees about results</strong></td>
<td><strong>Rigorous</strong></td>
</tr>
<tr>
<td><strong>Large number of predictors</strong></td>
<td><strong>Few, very relevant predictors</strong></td>
</tr>
</tbody>
</table>
Association Rule Mining

- Origins from sales data
- **Items**: articles carried by a store
- **Transactions**: items in the same shopping cart
- **Itemsets**: sets of items
- **Goal**: find all itemsets that are **frequently** purchased together

<table>
<thead>
<tr>
<th>Cart #</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>...</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>001</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>002</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>003</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>004</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Y</td>
</tr>
<tr>
<td>005</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
</tr>
</tbody>
</table>

- **Support**: # of transactions the itemset \( I \) appeared in
  - Support(ABC)=3
- **Frequent**: an itemset \( I \) is frequent, if \( \text{support}(I) > \text{minsup} \)

**Application to Oral Health**
- Items are problems
- Rows are patients

X: infrequent
Method

• Create a binary matrix
  – columns are Problems
  – rows are clients
  – Entry indicates whether the particular client has had the particular Problem during the study period and received Intervention for it

• Extract all combinations of Problems that exist in at least 5 clients with oral health problem
  – 2,900 combinations were found
Method 2

• 2,900 patterns are difficult to interpret

• Filtering based on predictive capability
  – Patterns that are not significantly predictive of oral health are irrelevant

• Filtering based on independence of items
  – Items co-occur more frequently than expected under the assumption that they co-occur at random

• Filtering based on independence of sub-patterns

• Summarize the patterns
Filtered and Summarized Patterns

**supOH Pattern**

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>728</td>
<td>Mentalhealth Pregnancy Income Residence Abuse FamPlanning Substance</td>
</tr>
<tr>
<td>672</td>
<td>Mentalhealth Income Postpartum Residence Abuse FamPlanning Substance</td>
</tr>
<tr>
<td>484</td>
<td>Mentalhealth Pregnancy Income Residence HealthCareSuprv Abuse FamPlanning Substance</td>
</tr>
<tr>
<td>452</td>
<td>Mentalhealth Income Postpartum Residence Caretaking HealthCareSuprv Abuse FamPlanning Substance</td>
</tr>
<tr>
<td>316</td>
<td>Mentalhealth Pregnancy Income Postpartum Residence HealthCareSuprv Abus FamPlanning Substance</td>
</tr>
<tr>
<td>14</td>
<td>Nutrition Residence</td>
</tr>
<tr>
<td>13</td>
<td>Mentalhealth Antepartum Caretaking</td>
</tr>
<tr>
<td>13</td>
<td>Antepartum FamPlanning</td>
</tr>
<tr>
<td>13</td>
<td>Mentalhealth Antepartum FamPlanning</td>
</tr>
</tbody>
</table>
Focus of PHN Practice

- Mental health, Income, Residence, Abuse, Family planning, Substance Use

+ Pregnancy (728)

+ Health care supervision, Pregnancy, Postpartum (313)

+ Health care supervision, Caretaking/parenting, PostPartum (452)

+ Postpartum (672)

+ Health care supervision, Pregnancy (484)
Summary

• Considering just Problems for individuals (as opposed to families) findings are consistent with use of evidence-based care plans that are in place and are being used in Dakota County
  – Demonstrates quality of PHN care/documentation
  – Suggests possible client types
  – May indicate evidence-based changes in PHN practice over time
Next Steps

• Select patients who have interventions for the Oral health
  – Consider family a unit (not a client)
  – Take signs/symptoms into account
  – Assess the effects of interventions

• Deidentification blinds us to dates
  – Temporal sequence of events is unclear
Acknowledgements

Jeanne Pfeiffer
University of Minnesota

Laura L. Fitzsimmons
Dakota County Public Health

Karen A. Monsen
University of Minnesota

Thank you!