Inference

Infer a healthcare network from the top 30 billing providers in Medicaid claims data from before and after a legislative change.

Distance

For RA, network distance is the sum of absolute differences in calculated probability distributions.

$$\hat{\theta} = \sum |q_{before} - q_{after}|$$

Permutation

Shuffle observations between datasets, infer networks from permuted datasets, calculate distances where the null is true.

Before data

“Before” when null is true

After data

“After” when null is true

Testing

Test null hypothesis by comparing the observed distance $\hat{\theta}$ against the reference distribution generated through permutation.

Conclusion

Networks can be tested for significant change over time when inferred from datasets of independent observations. PLS can measure change among pairs of providers, while RA identifies the higher-way relationships among them.
Statistical Analysis of Network Change

Teresa D. Schmidt and Martin Zwick

Inference

Infer a healthcare network from the top 30 billing providers in Medicaid claims data from before and after a legislative change.

Distance

For PLS, network distance is the average absolute difference in the networks’ connectivity matrices.

\[ \tilde{\theta} = \frac{1}{g(g-1)} \sum_{i \neq j} |\hat{s}_{ij} - \hat{s}_{ij}^2| \]

Permutation

Shuffle observations between datasets, infer networks from permuted datasets, calculate distances where the null is true.

Testing

Test null hypothesis by comparing the observed distance \( \tilde{\theta} \) against the reference distribution generated through permutation.

**Top 30 Provider Network Inference by Partial Least Squares Regression (PLS),**

defines a relationship by how much one provider contributes to latent terms of covariance with other provider’s billing patterns.

**PLS Results**

The billing provider network changed significantly over time \((p < .001)\). The structure of the network changed dramatically, and many strong connections appeared between healthcare organizations.

**Conclusion**

Networks can be tested for significant change over time when inferred from datasets of independent observations. PLS can measure change among pairs of providers, while RA identifies the higher-way relationships among them.
Statistical Analysis of Network Change

Teresa D. Schmidt and Martin Zwick

Healthcare Sector Network Inference by Reconstructability Analysis (RA), a data mining methodology that uses information theory to derive relationships among variables in the form of a best model (hypergraph).

Networks can be tested for significant change over time when inferred from datasets of independent observations. PLS can measure change among pairs of providers, while RA identifies the higher-way relationships among them.
Statistical Analysis of Network Change

Teresa D. Schmidt and Martin Zwick

Inference

Infer a network of healthcare sectors* from Medicaid claims data from before and after a legislative change.

Distance

For PLS, network distance is the average absolute difference in the networks’ connectivity matrices.

\[ \hat{\theta} = \frac{1}{g(g-1)} \sum_{i \neq j} |\hat{s}_{ij}^1 - \hat{s}_{ij}^2| \]

Permutation

Shuffle observations between datasets, infer networks from permuted datasets, calculate distances where the null is true.

Testing

Test null hypothesis by comparing the observed distance \( \hat{\theta} \) against the reference distribution generated through permutation.

PLS Results

The network of healthcare sectors changed significantly (\( p < .001 \)) over time. The facility sector became more strongly connected with primary, and primary became more connected with behavioral and less with specialty care.

Conclusion

Networks can be tested for significant change over time when inferred from datasets of independent observations. PLS can measure change among pairs of providers, while RA identifies the higher-way relationships among them.

---

* Sectors defined by taxonomy groups from the National Uniform Claims Committee

Healthcare Sector Network Inference by Partial Least Squares Regression (PLS), defines a relationship by how much one healthcare sector contributes to latent terms of covariance with other sector’s billing patterns.
Statistical Analysis of Network Change
Teresa D. Schmidt and Martin Zwick

Data Requirements by Reconstructability Analysis (RA)
Data must be formatted so that variables represent network ‘nodes’, and cases reflect observations of those nodes’ behavior. RA uses nominal data, but has no distributional assumptions.

Example Claims Data during Before Period

<table>
<thead>
<tr>
<th></th>
<th>Prov 1 (x₁)</th>
<th>Prov 2 (x₂)</th>
<th>Prov 3 (x₃)</th>
<th>Prov 4 (x₄)</th>
<th>Prov 5 (x₅)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient 1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Patient 2</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Patient 3</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Patient 4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Patient 5</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Patient 6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Patient 7</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Patient 8</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
</tbody>
</table>

Network Inference by Reconstructability Analysis
RA produces a best model of associations (e.g., by BIC), with calculated probabilities for all combinations of variable states. This best model can be interpreted as a hypergraph with three-way and higher-way associations between variables (nodes).

Example Calculated Probability Distribution from Before Period

<table>
<thead>
<tr>
<th></th>
<th>x₁</th>
<th>x₂</th>
<th>x₃</th>
<th>x₄</th>
<th>x₅</th>
<th>Probability for a Patient</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 0 0 0 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>q₀₀₀₀=0.0012</td>
</tr>
<tr>
<td>0 0 0 0 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>q₀₀₀₁=0.0004</td>
</tr>
<tr>
<td>0 0 0 1 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>q₀₀₁₀=0.0070</td>
</tr>
<tr>
<td>0 0 0 1 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>q₀₀₁₁=0.0100</td>
</tr>
<tr>
<td>0 0 1 0 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>q₀₁₀₀=0.0016</td>
</tr>
<tr>
<td>0 0 1 0 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>q₀₀₁₁=0.0009</td>
</tr>
</tbody>
</table>

Example Best Model during Before Period
\{x₁x₂ : x₂x₄ : x₃x₄ : x₄x₅\}

In the billing provider network, there is a variable for each billing provider, and a record for each distinct patient. Values indicate whether a provider billed for a patient 0, 1, or 2+ times during the time period. (Other levels could also be chosen.)

Statistical Analysis of Network Change

Teresa D. Schmidt and Martin Zwick

Data Requirements by Partial Least Squares (PLS)
Data must be formatted so that variables represent network ‘nodes’, and cases reflect observations of those nodes’ behavior. PLS uses continuous data, but has distribution assumptions.

Network Inference by Partial Least Squares
PLS determines the connectivity to any given node by the contributions of each other node to latent terms that vary with the given node’s behavior.

Claims Data during Before Period

<table>
<thead>
<tr>
<th></th>
<th>Prov 1 ((x_1))</th>
<th>Prov 2 ((x_2))</th>
<th>Prov 3 ((x_3))</th>
<th>Prov 4 ((x_4))</th>
<th>Prov 5 ((x_5))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient 1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Patient 2</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Patient 3</td>
<td>7</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Patient 4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Patient 5</td>
<td>2</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Patient 6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Patient 7</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Patient 8</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
</tbody>
</table>

In the billing provider network, there is a variable for each billing provider, and a record for each distinct patient during a time period. Values indicate the number of claims that each provider billed for each patient during a time period.