Secondary Analysis of Concussion Data

Martin Zwick
& S. Kolakowsky-Hayner, N. Carney, M. Balamane, T. Nettleton, D. Wright

Systems Science Program
Portland State University

zwick@pdx.edu
http://www.pdx.edu/sysc/research_dmm.html

Systems Science seminar, Nov 18, 2016
• ABSTRACT: Clinical studies are expensive & time-consuming. Typically in these studies specific hypotheses are subjected to confirmatory test. Yet the data may harbor evidence of unanticipated relations between variables. It is thus desirable to subject the data to secondary analyses in the hope of discovering novel & valuable associations. Exploratory analysis, however, is tentative: findings should be replicated in new data.

• This presentation reports some secondary analyses on concussion data. Data mining on 2 datasets will be discussed, & some unexpected findings reported. The analyses use reconstructability analysis (RA), a probabilistic graphical modeling method implemented in the Occam software package developed in the SySc Program, which is first briefly described.

1. Exploratory modeling with RA (Occam)
2. Sample results on Preece, Wright data sets
• **Stephanie Kolakowsky-Hayner**, Brain Trauma Foundation: Brain Trauma Evidence-Based Consortium (BTEC) project head
  – **Maya Balamane**, Assistant Program Manager

• **Nancy Carney**, OHSU, SySc-Psychology PhD: BTEC founder & previous head;
  – **Tracie Nettleton**, Research assistant

• BTEC funded by DoD via BTF & Stanford

• **PSU BTEC Project**
  – **Wayne Wakeland**, PI of overall project
  – & PI of **Dynamic Model Initiative Subproject**

• **Data Analytics (Occam) Subproject**
  – **Martin Zwick**, co-PI; Forrest Alexander, Peter Olson, Programmers
1. Exploratory modeling with RA (Occam)

- Exploratory modeling (data mining) with Reconstructability Analysis (RA):
  - to contribute to a clinically-useful TBI classification system & other BTEC projects
  - to extract additional information from past studies
  - to enhance RA methodology & Occam implementation for future data sets
Rationale for exploratory modeling

• Most studies are confirmatory, testing only specific hypotheses. Since studies are expensive & time-consuming, it is useful to explore what else might be discovered in the data.

• Exploratory studies can find unexpected non-linear & many-variable interaction effects (which should then be tested in confirmatory mode with new data).

• Exploratory studies (by data analysts) are unbiased.
Why RA & Occam software

• Explicitly designed for exploratory modeling
  – Analyzes both nominal & continuous (binned) variables
  – Easily interpretable; standard text input; web-accessible, emails results to user; available for research use

• Other statistical & machine-learning methods (log-linear, logistic regression, Bayesian networks, classification trees, support vector machines, neural nets) not well designed for exploration, or have limited model types, or have difficulty with nominal variables or with stochasticity
PAST/PRESENT RA APPLICATIONS

• **BIOMEDICAL**
  Gene-disease association, disease risk factors, gene expression, health care use & outcomes, dementia, diabetes, heart disease, prostate cancer, brain injury, primate health, surgery

• **FINANCE-ECONOMICS-BUSINESS**
  Stock market, bank loans, credit decisions, apparel analyses, market segmentation

• **SOCIAL-POLITICAL-ENVIRONMENTAL**
  Socio-ecological interactions, wars, urban water use, rainfall, forest attributes

• **MATH-ENGINEERING**
  Logic circuits, automata dynamics, genetic algorithm & neural network preprocessing, chip manufacturing, pattern recognition, decision analysis

• **OTHER**
  Textual analysis, language analysis
What RA is

- Reconstructability Analysis (RA) = Information theory + Graph theory, a probabilistic graphical modeling technique
- Model = structure applied to data
- RA structure = hypergraph (relations not only pairwise)
- RA model = a (joint or conditional) probability distribution simpler (fewer df) than the data, capturing much of the information in the data
Two types of RA explorations

• **Neutral search** (clustering): find relations among all variables

• **Directed search** (classification): predict DVs from IVs. Want:
  – High accuracy (information captured) (low error) measured by
    • $\%\Delta H = \%$ reduction of uncertainty (like variance)
    • $\%c = \%$ correct in prediction (a general measure)
  – High model simplicity (low complexity) = low $\Delta df$
  – Model selection criteria trade off these two objectives
Uncertainty reduction: the primary measure

• Reduction of uncertainty (Shannon entropy), a simple example

\[ \Delta H(Z) = \frac{T(A:Z)}{H(Z)} = 8\% \]

• \( p(Z_1)/p(Z_0) = 1:1 \), not knowing \( A \) → 2:1 or 1:2, knowing \( A \)

• 8% reduction in uncertainty (here) is large (unlike variance!)

\[
\begin{array}{c|cc|c}
& Z_0 & Z_1 & \text{df=3} \\
\hline
A_0 & .67*.5 & .33*.5 & .5 \\
A_1 & .33*.5 & .67*.5 & .5 \\
\end{array}
\]
Model selection criteria

Tradeoff between accuracy & simplicity (error & complexity)

- **Conservative**: Bayesian Information Criterion (BIC)
- **Aggressive**: Akaike Information Criterion (AIC)
  
  Incremental p-value (IncrP)

- AIC & BIC: linear combinations of error & complexity; BIC penalizes more for complexity: weights it by $\ln(N)$

- IncrP uses Chi-square p-values to pick models whose difference from -- & every incremental step from -- independence is statistically significant
Degrees of refinement of RA model search

Complexity
(degrees of freedom)

Variable-based

No loops
COARSE

With loops
FINE

State-based
ULTRA-FINE
4 variables, neutral systems: 114 models
Combinatorial explosion of possible structures

<table>
<thead>
<tr>
<th># variables</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>7 # neutral VB models (loops)</td>
<td>2</td>
<td>9</td>
<td>114</td>
<td>6,894</td>
<td>7,785,062</td>
<td>2.4 $10^{12}$</td>
</tr>
<tr>
<td>For 1 DV:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td># directed VB models (loops)</td>
<td>2</td>
<td>5</td>
<td>19</td>
<td>167</td>
<td>7,580</td>
<td>7.8 $10^6$</td>
</tr>
<tr>
<td># directed VB models (no loops)</td>
<td>2</td>
<td>4</td>
<td>8</td>
<td>16</td>
<td>32</td>
<td>64</td>
</tr>
<tr>
<td>For binary variables:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td># neutral SB models (loops)</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>even more severely exponential</td>
</tr>
</tbody>
</table>

NEED INTELLIGENT HEURISTICS

TO DO EXPLORATORY MODELING with 52 variables (Preece data) or 560 variables (Wright data)

Can now explore a few 100 variables; if parallelized could deal with more.
Searching the space of possible models
2. Sample results

2.1 Preece data: analysis complete
   auto accidents
   • Neutral coarse searches
   • Directed coarse, fine, & ultra-fine searches

2.2 Wright (PROTECT) data: analysis underway
   auto/motorcycle/bike accidents, hit pedestrians, falls
   • Directed coarse & fine searches

Other data sets to follow
2.1 Preece data

- 52 variables

- Variable types
  - P = patient characteristics (17 variables)
  - Y = symptoms (25): subjective reports
  - G = signs (4): objective indicators
  - C = cognitive deficits (5)
  - N = neurologic deficits (1)

- N = 337; reduces to 175 or less if exclude missing data
### Occam input file

**Partial, Preece** (note missing data)

```plaintext
# Occam input file (partial, Preece) (note missing data)

#### Subject id

<table>
<thead>
<tr>
<th>#</th>
<th>name</th>
<th>number</th>
<th>missing</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>subject_id</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>study</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

#### Symptoms

- `pain(16)`: Symptom (25)
- `sign(6)`: Neurologic (1) <= cognitive (5)

####-tage 1 study

<table>
<thead>
<tr>
<th>#</th>
<th>name</th>
<th>number</th>
<th>missing</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>patient</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>injury_group</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>sex</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

#### Psychological Measures

- `yapmscale(5)`: Standard pain scale used by hospitals
- `yemocscale(3)`: Yemoc scale
- `yemocscale(3)`: Yemoc scale
- `yemocscale(3)`: Yemoc scale
- `yemocscale(3)`: Yemoc scale
- `yemocscale(3)`: Yemoc scale

#### Additional Measures

- `glsleep`: GHSL
- `pul`: Pul
- `precentill`: Percent
- `ppadrug`: Previous medications
- `ppakiller`: Previous painkillers
- `ppprescription`: Previous prescriptions

#### Data

<table>
<thead>
<tr>
<th>#</th>
<th>variable</th>
<th>number</th>
<th>missing</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>id</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>age</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

---

**Note:**

- Different format for the 2 studies
- Includes a study that is not described in detail
- Missing data in various fields

---

**Source:**

- Preece et al., 2015
Neutral coarse search results

- A neutral search model is a set of associations (relations)
- Variables here with original (high) cardinalities & missing data

- Best BIC (conservative) model has:

<table>
<thead>
<tr>
<th>51 components: red: $p &lt; .05$; purple: $.05 &lt; p &lt; .1$; <strong>C, N</strong> variables in bold</th>
</tr>
</thead>
<tbody>
<tr>
<td>GpcYfg: GpcYcn: GpcYtk: GpcYbr: GpcYls: GpcYdv: GpcYrs: GpcYaz: GpcYrm: PlgPac:</td>
</tr>
<tr>
<td><strong>CnrCsr</strong></td>
</tr>
</tbody>
</table>
Neutral coarse search network

- Association network = hypergraph (but below is a graph)
- $23 \ p \leq 0.1$ (15 $p \leq 0.05$) associations in BIC model
## Neutral coarse search associations

- **Predictive success** (%$\Delta H$, $\Delta\%c$ relative to independence) ($p \leq 0.05$)

| v1   | v2   | $\%\Delta H(2|1)$ | $\%\Delta H(1|2)$ | p-value | N   | $\%\Delta c(2|1)$ | $\%\Delta c(1|2)$ | v1                          | v2                          |
|------|------|-------------------|-------------------|---------|-----|-------------------|-------------------|--------------------------|--------------------------|
| Ggc  | Pij  | 34.5              | 86.5              | 0.000   | 196 | 9.7               | 7.7               | glasgow coma scale       | Injury patient/control   |
| Gxc  | Pij  | 32.9              | 12.6              | 0.000   | 280 | 20.4              | 14.3              | external cause           | Injury patient/control   |
| Ped  | Pye  | 41.3              | 34.8              | 0.000   | 248 | 32.3              | 27.4              | highest educ level years |                           |
| Yem  | Ypn  | 6.4               | 6.1               | 0.000   | 218 | 5.0               | 2.3               | emotional problems painscale |                           |
| Yds  | Yem  | 6.0               | 27.8              | 0.000   | 210 | 3.8               | 0.0               | stress                   |                           |
| Ped  | Yem  | 43.6              | 26.0              | 0.000   | 210 | 1.4               | 1.9               | depression stress        |                           |
| Yds  | Yds  | 54.7              | 32.6              | 0.000   | 210 | 0.0               | 2.9               | anxiety stress           |                           |
| Ydd  | Yds  | 50.7              | 57.6              | 0.000   | 230 | 28.3              | 15.7              | current medications painkillers |                           |
| Gpc  | Pnp  | 57.0              | 100.0             | 0.000   | 52  | 11.5              | 30.8              | previous concussion # previous concussion |                           |
| Pac  | Plg  | 26.5              | 12.3              | 0.000   | 201 | 0.0               | 12.4              | caused accident case litigated |                           |
| Cnr  | Csr  | 48.6              | 48.3              | 0.000   | 210 | 34.3              | 31.0              | reaction time norm reaction time |                           |
| Psx  | Ycv  | 6.5               | 8.8               | 0.000   | 197 | 2.0               | 0.0               | sex corrected vision     |                           |
| Gpc  | Ydz  | 13.7              | 21.9              | 0.003   | 52  | 0                 | 9.6               | previous concussion dizzy |                           |
| Csr  | Pph  | 5.3               | 2.3               | 0.010   | 187 | 5.3               | 4.8               | reaction time previous head injury |                           |
| Gpc  | Yfr  | 9.1               | 17.3              | 0.011   | 52  | 1.9               | 9.6               | previous concussion frustrated |                           |
## Directed searches

- DVs (cognitive, neurological deficit variables)
  - #bins excludes missing values

<table>
<thead>
<tr>
<th>#bins</th>
<th>N</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>cdgtcorrect</td>
<td>6</td>
<td>Cdg</td>
</tr>
<tr>
<td>cnormsrt</td>
<td>6</td>
<td>Cnr</td>
</tr>
<tr>
<td>cspatialreac</td>
<td>6</td>
<td>csr</td>
</tr>
<tr>
<td>nlogmar</td>
<td>3</td>
<td>Nlr</td>
</tr>
</tbody>
</table>
Cnr coarse, fine, ultra-fine searches

Predict Cnr: reaction time, normalized by age, sex (rebin |Cnr| = 2: ~ 50-50)

<table>
<thead>
<tr>
<th>MODEL</th>
<th>Δdf</th>
<th>p</th>
<th>%ΔH</th>
<th>%c</th>
<th>N=175</th>
</tr>
</thead>
<tbody>
<tr>
<td>COARSE, single component predictors</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cdg Gpt Cnr</td>
<td>3</td>
<td>0.00</td>
<td>10.6</td>
<td>64.6</td>
<td>BIC, AIC</td>
</tr>
<tr>
<td>Pph Cdg Gpt Cnr</td>
<td>7</td>
<td>0.00</td>
<td>13.1</td>
<td>66.9</td>
<td>IncrP</td>
</tr>
<tr>
<td>Cnr (independence=reference)</td>
<td>0</td>
<td>1.00</td>
<td>0.0</td>
<td>50.9</td>
<td></td>
</tr>
<tr>
<td>FINE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cdg Cnr : Gpt Cnr</td>
<td>2</td>
<td>0.00</td>
<td>8.8</td>
<td>64.6</td>
<td>BIC</td>
</tr>
<tr>
<td>Pri Cnr : Pph Cnr : Cdg Gpt Cnr</td>
<td>6</td>
<td>0.00</td>
<td>14.7</td>
<td>70.3</td>
<td>AIC</td>
</tr>
<tr>
<td>Pye Cnr : Pph Cnr : Cdg Gpt Cnr</td>
<td>5</td>
<td>0.00</td>
<td>12.9</td>
<td>67.4</td>
<td>IncrP</td>
</tr>
<tr>
<td>ULTRA-FINE (state-based model)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pph\textsubscript{1} Cdg\textsubscript{1} Cnr : Cdg\textsubscript{0} Gpt\textsubscript{1} Cnr</td>
<td>2</td>
<td>0.00</td>
<td>12.4</td>
<td>64.8</td>
<td>BIC</td>
</tr>
<tr>
<td>Cnr (independence=reference)</td>
<td>0</td>
<td>1.00</td>
<td>0.0</td>
<td>50.9</td>
<td></td>
</tr>
</tbody>
</table>
## Cnr ultra-fine model

**Model:** $P_{ph_1} \ C_{dg_1} \ C_{nr} : C_{dg_0} \ G_{pt_1} \ C_{nr}$

**Odds (high is good)** = $C_{nr_0}/C_{nr_1}$ (model) = $p(\text{fast} = \text{normal reaction})/p(\text{slow})$

$P_{ph_1}$ previous head injury, $C_{dg_1}$ high digit score; $G_{pt_1}$ amnesia

**conditional probabilities of DV**

<table>
<thead>
<tr>
<th>IV states</th>
<th>N</th>
<th>data</th>
<th></th>
<th></th>
<th>model</th>
<th>Odds</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>$C_{nr_0}$</td>
<td>$C_{nr_1}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$C_{nr_0}$</td>
<td>$C_{nr_1}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Odds</td>
<td>Odds</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$p$</td>
<td>$p$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 0 0</td>
<td>20</td>
<td>0.40</td>
<td>0.60</td>
<td>0.52</td>
<td>0.48</td>
<td>1.1</td>
<td>.92</td>
</tr>
<tr>
<td>0 0 1</td>
<td>19</td>
<td>0.16</td>
<td>0.84</td>
<td>0.16</td>
<td>0.84</td>
<td>0.2</td>
<td>.00</td>
</tr>
<tr>
<td>1 0 0</td>
<td>30</td>
<td>0.57</td>
<td>0.43</td>
<td>0.52</td>
<td>0.48</td>
<td>1.1</td>
<td>.90</td>
</tr>
<tr>
<td>1 0 1</td>
<td>18</td>
<td>0.17</td>
<td>0.83</td>
<td>0.16</td>
<td>0.84</td>
<td>0.2</td>
<td>.00</td>
</tr>
<tr>
<td>0 1 0</td>
<td>24</td>
<td>0.50</td>
<td>0.50</td>
<td>0.52</td>
<td>0.48</td>
<td>1.1</td>
<td>.91</td>
</tr>
<tr>
<td>0 1 1</td>
<td>13</td>
<td>0.61</td>
<td>0.39</td>
<td>0.52</td>
<td>0.48</td>
<td>1.1</td>
<td>.93</td>
</tr>
<tr>
<td>1 1 0</td>
<td>38</td>
<td>0.76</td>
<td>0.23</td>
<td>0.73</td>
<td>0.27</td>
<td>2.7</td>
<td>.01</td>
</tr>
<tr>
<td>1 1 1</td>
<td>14</td>
<td>0.64</td>
<td>0.36</td>
<td>0.73</td>
<td>0.27</td>
<td>2.7</td>
<td>.09</td>
</tr>
<tr>
<td></td>
<td>176</td>
<td>0.51</td>
<td>0.49</td>
<td>0.51</td>
<td>0.49</td>
<td>1.0</td>
<td></td>
</tr>
</tbody>
</table>
**Cnr decision tree from conditional probabilities**

**Reaction time Odds** (probability fast/ probability slow) & p-values relative to marginal prob. (odds = 1)

![Decision Tree Diagram]

- **Amnesia**
  - **Digit symbol score**
    - low
      - **Previous head injury**
        - no
          - 1.1 .91
        - yes
          - .2 .00
    - normal
      - yes
        - 2.7 .01,.09
      - no
        - 1.1 .92
For low performance on digit symbol test, amnesia predicts slow reaction time.

For normal performance on digit symbol test, previous head injury increases the probability of fast (normal) reaction time. THIS IS ANOMALOUS.

– Need to see if it would be replicated in another data set.
– Possible explanation: prior exposure to Reaction Time test introduces a practice effect.
– If Reaction Time is so vulnerable to a practice effect that it no longer discriminates concussed from non-concussed, then it’s probably not an appropriate measure for this purpose.
2.2 Wright (PROTECT) data

• 560 variables (302 variables within 1st two weeks)

• Variable types
  – A = admin (32 variables) #1-32
  – P = patient characteristics (134 variables) #405-538
  – Y = symptoms (8 variables): subjective reports #551-558
  – G = signs (13 variables): objective indicators #539-550, 560
  – C = cognitive deficits (6 variables) #33-38
  – N = neurologic deficits (367 variables) #39-404, 559

• N = 882 patients
Directed searches

- DVs = deficit variables

<table>
<thead>
<tr>
<th>#bins</th>
<th>N</th>
<th># IVs</th>
</tr>
</thead>
<tbody>
<tr>
<td>mort2</td>
<td>Gvn</td>
<td>764</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0=not dead; 1=dead</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>gose</th>
<th>Nvm</th>
<th>882</th>
<th>Total extended Glasgow Outcome Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>1=death; 2=vegetative; 3,4 lower, upper severe disability; 5,6 lower, upper moderate disability; 7,8 lower, upper good recovery</td>
</tr>
</tbody>
</table>

Two lines of current investigation:

1. *Predict mortality at 2 weeks*

2. *Investigate possible progesterone effect*
**Predict mortality at 2 weeks**

- No surprises: GCS scores, days 2, 4, 9, are best predictors.

```
GCS day 2
  └── severe
      ├── GCS day 4
      │     └── severe
      │        ├── moderate/mild
      │        │     └── Increased probability of alive
      │        └── vegetative/missing
      └── moderate/mild
          └── Increased probability of alive

GCS day 8-10 + status day 13
  └── severe
      ├── vegetative/missing
      └── Increased probability of alive
```
Investigate possible progesterone effect

- Earlier studies suggested value of progesterone treatment
- These effects **not found** in Wright project
- Project is regarded as an exemplar of ‘failed’ studies

- Wright didn’t systematically look for complex effects
- Progesterone *might* have had effect in some *sub*population

- RA detects a **possible** predictive interaction effect
- Likely to be an artifact, but **under investigation**
A possible progesterone effect

- **Ngw** = sedation (0 no, 1 yes)
- **Pup** = progesterone treatment (0 no, 1 yes)
- **Gvn** = status at 2 weeks (0 alive, 1 dead)

<table>
<thead>
<tr>
<th>IV states</th>
<th>DV cond.prob.</th>
<th>N</th>
<th>Gvn₀</th>
<th>Gvn₁</th>
<th>Gvn₀</th>
<th>Gvn₁</th>
<th>p(margin)</th>
</tr>
</thead>
<tbody>
<tr>
<td>. 0</td>
<td>49</td>
<td></td>
<td>0.41</td>
<td>0.59</td>
<td>20</td>
<td>29</td>
<td>0</td>
</tr>
<tr>
<td>0 0</td>
<td>190</td>
<td></td>
<td>0.92</td>
<td>0.08</td>
<td>175</td>
<td>15</td>
<td>0.01</td>
</tr>
<tr>
<td>. 1</td>
<td>49</td>
<td></td>
<td>0.33</td>
<td>0.67</td>
<td>16</td>
<td>33</td>
<td>0</td>
</tr>
<tr>
<td>0 1</td>
<td>206</td>
<td></td>
<td>0.97</td>
<td>0.03</td>
<td>200</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>1 0</td>
<td>141</td>
<td></td>
<td>0.94</td>
<td>0.06</td>
<td>133</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>1 1</td>
<td>129</td>
<td></td>
<td>0.85</td>
<td>0.15</td>
<td>110</td>
<td>19</td>
<td>0.92</td>
</tr>
</tbody>
</table>

- **Pup benefits** if no sedation; **harms** if sedation
Effect may be an artifact

- Effect depends on another variable, Nod, being *missing*
- Nod = ‘Was GCS collected in previous 24 hrs’
- Nod missing $N=297$, not missing $N=467$
- If Nod *not missing*, effect *disappears*
- Value of results depends on *what Nod missing means*
- This is *being explored* with Wright

- *Missing data is frequently a confounder*
- *Analysis always depends on quality of the data*
Summary

• Preece data a test bed for analysis protocol. Analysis complete, being written up.

• As an exploratory study, results are tentative, needing confirmation on other data sets.

• Wright analysis underway

• These studies are driving methodological RA innovations.

• Hope for additional data sets (accident, military, sports), with higher N, fewer missing data, new variable types (imaging, genomic, proteomic).

• Work is collaborative with investigators who share data.
RA (DMM) web page
http://pdx.edu/sysc/research-discrete-multivariate-modeling
zwick@pdx.edu

Research: Discrete Multivariate Modeling

The methods used are also known in the systems literature as "reconstructability analysis" (RA). RA overlaps significantly with the fields of logic design and machine learning and with logistic statistical modeling. The papers "Wholes and Parts in General Systems Methodology" and "An Overview of Reconstructability Analysis" listed below offer a concise review of RA methodology.

Projects

Theory/Methodology
OCCAM: RA software for data analysis & data mining

Occam3 (web accessible; try it out)
User manual (PDF)

EDA: Extended Dependency Analysis
Heuristic RA search for loopless models
Download executable, sample files, and documentation (for Windows)

RA utility programs
RA software (Occam)

Occam is a Discrete Multivariate Modeling (DMM) tool based on the methodology of Reconstructability Analysis (RA). Its typical usage is for analysis of problems involving large numbers of discrete variables. Models are developed which consist of one or more components, which are then evaluated for their fit and statistical significance. Occam can search the lattice of all possible models, or can do detailed analysis on a specific model.

In Variable-Based Modeling (VBM), model components are collections of variables. In State-Based Modeling (SBM), components identify one or more specific states or substrates.

Occam provides a web-based interface, which allows uploading a data file, performing analysis, and viewing or downloading results.

- Run Occam
- For basic operation instructions, please see the manual: PDF
- Sample data files. You can download these to local files on your computer, then upload them via the Occam Web interface.
  A Neutral System
  A Directed System
- Links:
  Dr. Zwick's DMM Research Page
  Systems Science Graduate Program
  Occam-users mailing list (discussion)
  Occam-news mailing list (announcements)
- Contacts:
  Occam feedback email address
  Dr. Martin Zwick, Systems Science
  Joe Fusion, Graduate Assistant, Systems Science
PSU COURSES

• Discrete Multivariate Modeling (DMM)
  *theory course (SySc 551)*
  
  *Fall 2016*

• Data Mining with Information Theory (DMIT)
  *data analysis project course* (DMM not a prerequisite)
  
  *Winter 2017*

THANK YOU