

11-18-2016

Secondary Analysis of Concussion Data

Martin Zwick

Portland State University, zwick@pdx.edu

Stephanie Kolakowsky-Hayner

Mt. Sinai

Nancy Carney

Oregon Health and Science University

Maya Balamane

Brain Trauma Foundation

Tracie Nettleton

Oregon Health and Science University

See next page for additional authors

Let us know how access to this document benefits you.

Follow this and additional works at: https://pdxscholar.library.pdx.edu/sysc_fac



Part of the [Logic and Foundations Commons](#)

Citation Details

Zwick, Martin; Kolakowsky-Hayner, Stephanie; Carney, Nancy; Balamane, Maya; Nettleton, Tracie; and Wright, D., "Secondary Analysis of Concussion Data" (2016). *Systems Science Faculty Publications and Presentations*. 126.

https://pdxscholar.library.pdx.edu/sysc_fac/126

This Presentation is brought to you for free and open access. It has been accepted for inclusion in Systems Science Faculty Publications and Presentations by an authorized administrator of PDXScholar. For more information, please contact pdxscholar@pdx.edu.

Authors

Martin Zwick, Stephanie Kolakowsky-Hayner, Nancy Carney, Maya Balamane, Tracie Nettleton, and D. Wright

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

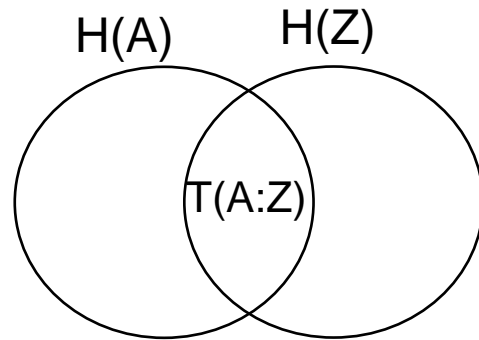
- **Reconstructability Analysis (RA)** = Information theory + Graph theory, a **probabilistic graphical modeling** technique
- Model = **structure** applied to **data**
- RA structure = hypergraph (relations not only pairwise)
- **RA model** = a (joint or conditional) probability distribution **simpler** (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 Δdf
 - Model selection criteria trade off these two objectives

Uncertainty reduction: the primary measure

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



	Z_0	Z_1	
A_0	$.67 * .5$	$.33 * .5$	$.5$
A_1	$.33 * .5$	$.67 * .5$	$.5$
df=3	$.5$	$.5$	

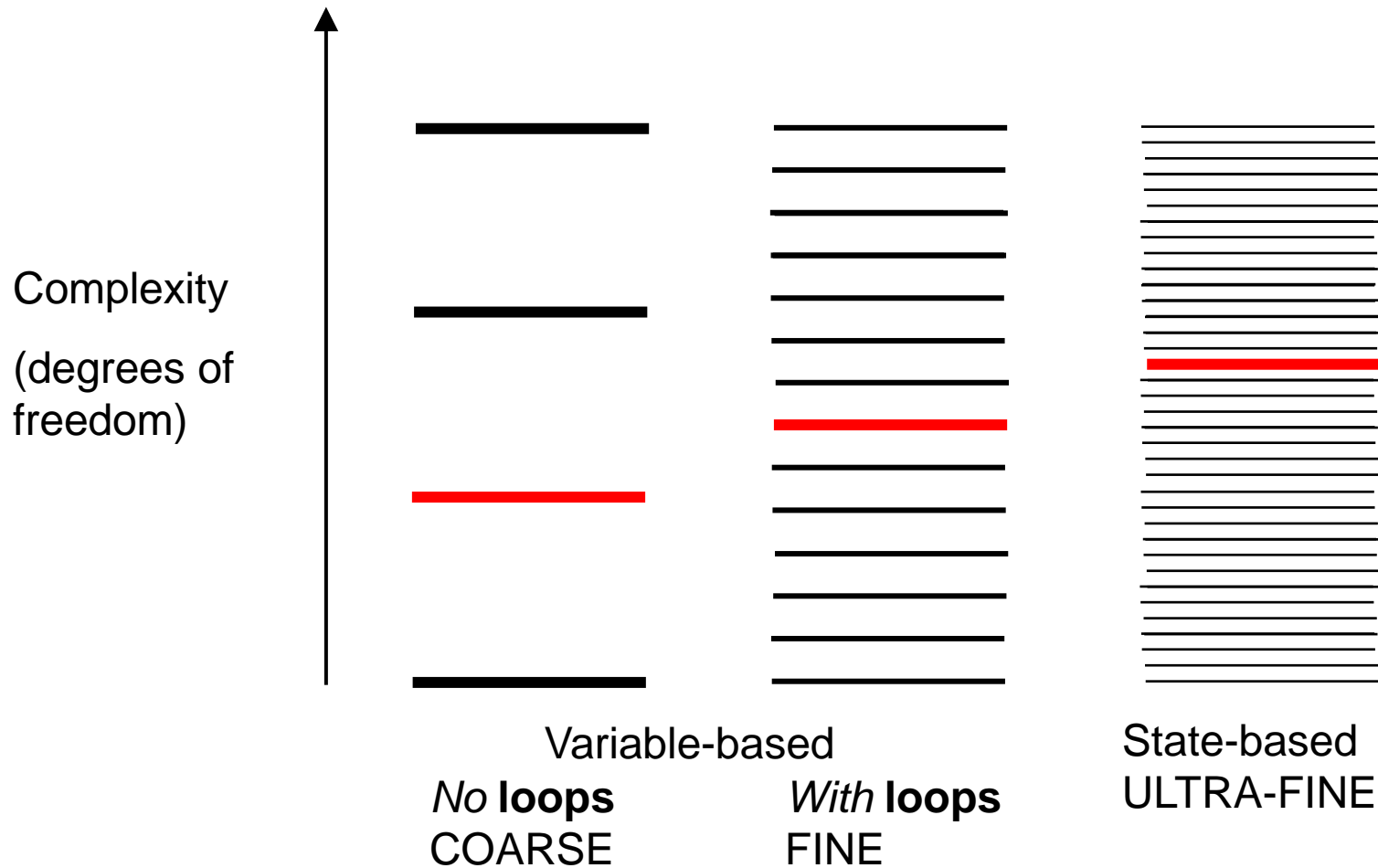
- $p(Z_1)/p(Z_0) = 1:1$, not knowing A \rightarrow $2:1$ or $1:2$, knowing A
- $\Delta H(Z) = T(A:Z) / H(Z) = 8\%$
- 8% reduction in **uncertainty** (here) is *large* (unlike **variance!**)

Model selection criteria

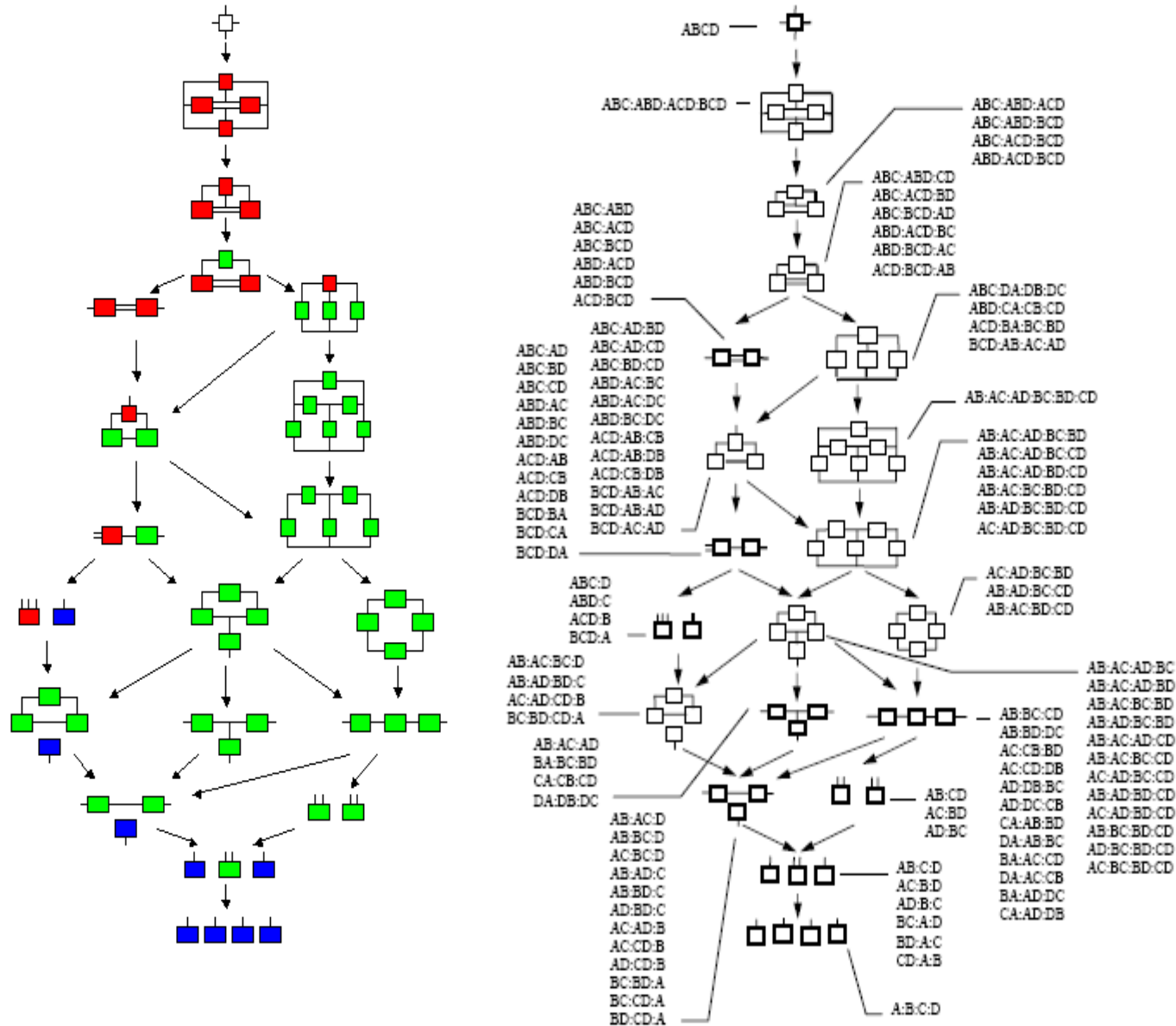
Tradeoff between accuracy & simplicity (error & complexity)

- *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



4 variables, neutral systems: 114 models



Combinatorial explosion of possible structures

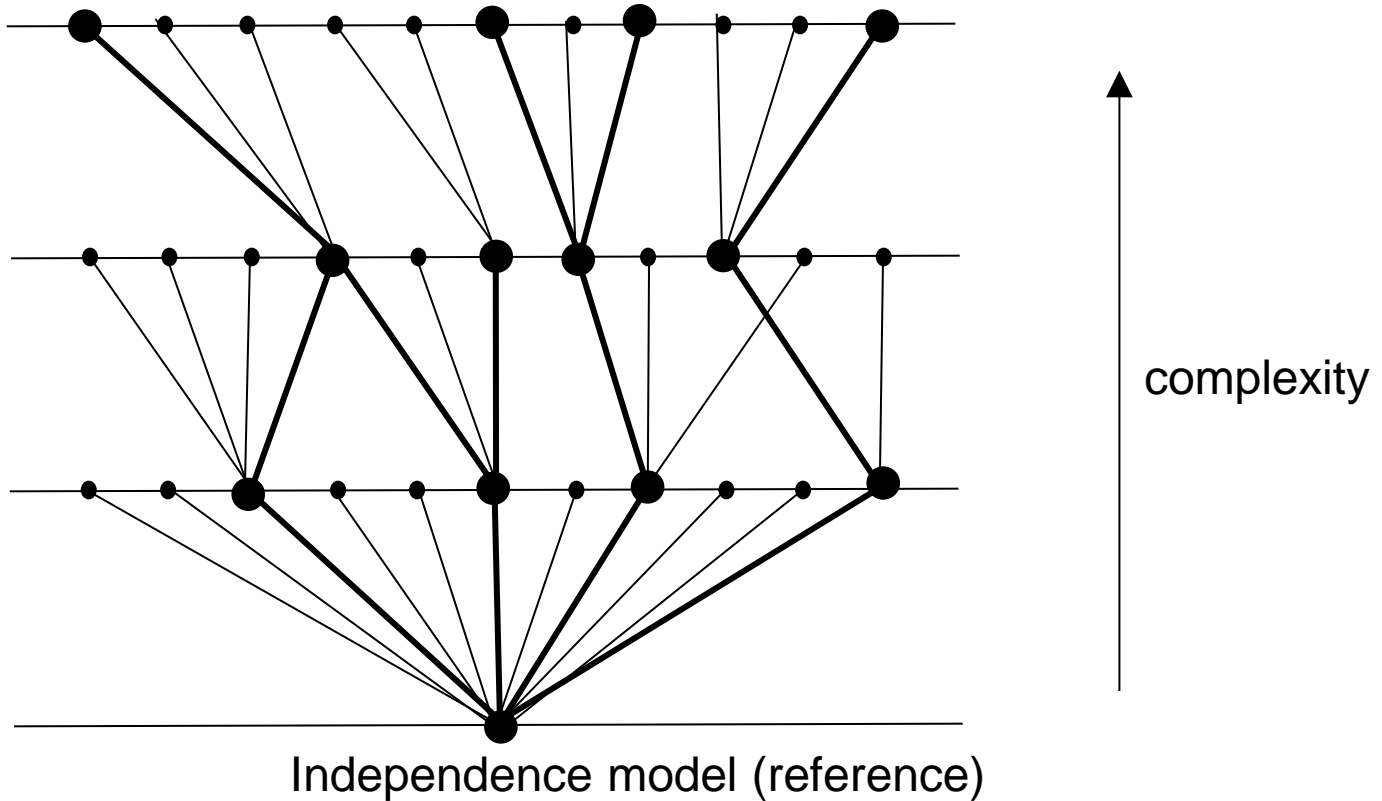
# variables	2	3	4	5	6	7
# neutral VB models (loops)	2	9	114	6,894	7,785,062	$2.4 \cdot 10^{12}$
For 1 DV:						
# directed VB models (loops)	2	5	19	167	7,580	$7.8 \cdot 10^6$
# directed VB models (no loops)	2	4	8	16	32	64
For binary variables:						
# neutral SB models (loops)	14	<i>even more severely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

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)

```

preece04slide - Notepad
File Edit Format View Help

:action
search

:nominal
#
subjectid,1, 0, id #1 P= patient(16) Y=symptom (25) G=sign (6) N=neurologic (1) c= cognitive (5)
study,2, 0, st #2 different Format for the 2 studies
pinjgrp,5, 1, pij #3 which study the data is from (1)PAH N=55 or(2) RBWH N=282
page,7, 1, pag #4 p Injury Group patient or control
psex,2, 1, psx #5 p sex
pyred,6, 1, pye #6 p years if education
pedlevel,8, 1, ped #7 p highest level of education
ypainscale,5, 1, ypn #8 y standard painscale used by hospitals
yemoscale,5, 1, yem #9 y sacle defining emotional state(0 no problems 1 few 2 moderate 3 many problems)
ydassd,5, 1, ydd #10 y Depression Anxiety Stress Scales measure of DEPRESSION(subjective experience questionnaire)
ydassa,6, 1, yda #11 y Depression Anxiety Stress Scales measure of ANXIETY(subjective experience questionnaire)
ydass,4, 1, yds #12 y Depression Anxiety Stress Scales measure of STRESS (Subjective experience questionnaire)
ghrsleep,5, 1, ghl #13 g number of hours of sleep, divided in less than normal normal=8hr and greater than normal
puhrsleep,5, 1, pul #14 p usual number of hours of sleep, divided in less than normal normal=8hr and greater than normal
precentill,3, 1, pri #15 p recent illness 0 no 1 yes
pmedication,3, 1, pmd #16 p current medications 0 no 1 yes
ppainkiller,3, 1, ppk #17 p currently on painkillers 0 no 1 yes
ppreheadinj,3, 1, pph #18 p have they had previous head injury 0 no 1 yes
pprelocigt,7, 1, gpl #19 g how long unconscious
gprecon,3, 1, gpc #20 g previous concussion 0 no 1 yes
pnumprecon,8, 1, pnp #21 p how many previous concussions. N = 16 there were only 7 different values so each code defines a raw value
ggcs,4, 1, ggc #22 g glasgow coma scale a measure of the level of unconsciousness lower score = deeper level of unconsciousness
chazpt,10, 1, chp #23 c hazard perception test measures how quickly potential driving hazards are predicted
pdbqerror,13, 1, pqe #24 p Driver Behavior Questionnaire errors self reported driving errors and violations
pdbqviol,14, 1, pqv #25 p Driver Behavior Questionnaire violations

:data
# variable number, short name, number of missing
#1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
# ID st ij ag sx ye ed pn em dd da ds dl hl ul r1 md pk ph pl pc 285 285 141 55 161 150 57
# 0 1 1 1 1 2 3 116 119 127 127 127 86 85 83 47 107 137 154 285 285 141 55 161 150 57
1 0 3 1 1 1 2 3 1 1 1 1 0 1 0 1 0 1 1 1 5 1 4 0 1 1
2 0 3 1 0 0 0 1 1 1 1 1 0 0 0 0 0 0 0 0 1 4 0 0 1 1
4 0 3 3 0 0 2 2 1 1 1 1 3 3 0 1 1 1 1 0 0 0 0 0 0 0
7 0 3 0 1 1 2 1 1 1 1 1 0 0 1 1 1 1 1 0 0 0 0 0 0 0
8 0 3 2 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
9 0 3 1 1 3 5 1 1 1 1 1 2 2 0 1 1 1 1 1 1 2 0 0 1 1 1
11 0 2 3 1 0 0 1 1 1 1 1 2 0 1 1 1 1 1 1 0 0 2 1 1 1 1
13 0 3 2 1 4 6 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0
15 0 2 3 1 0 3 1 1 1 1 1 1 0 0 1 1 1 1 1 1 6 2 1 1 1 1
16 0 3 1 1 1 1 1 1 1 1 1 2 0 1 2 0 1 1 1 1 0 0 1 1 1 1
17 0 0 0 1 2 4 1 1 1 1 1 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0
18 0 0 0 1 3 2 1 1 1 1 1 1 0 0 1 0 0 0 0 0 1 3 1 1 1 1
19 0 0 0 1 4 5 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0
20 0 0 0 1 3 4 1 1 1 1 1 1 0 0 0 0 0 0 0 0 1 1 1 1 1 1
21 0 0 0 0 3 4 1 1 1 1 1 1 0 0 0 0 0 0 0 0 1 1 1 1 1 1
22 0 0 0 1 4 5 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0
23 0 0 3 1 4 5 1 1 1 1 1 1 0 0 0 0 0 0 0 0 1 1 1 1 1 1
24 0 0 0 1 1 2 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 1
26 0 2 1 1 0 1 1 1 1 1 1 2 0 0 0 1 1 1 1 1 0 0 2 1 1 1
27 0 0 1 0 3 5 1 1 1 1 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0
28 0 2 0 1 0 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 2 1 1 1 1
29 0 3 4 1 3 5 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0
30 0 3 3 1 0 3 1 1 1 1 1 0 2 1 1 1 1 1 1 0 0 1 1 1 1 1
3]

```

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:

51 components: red: $p < .05$; purple: $.05 < p < .1$; C, N variables in bold									
PijGpc:	PijGgc:	PijGxc:	Pag:	PsxYcv:	PyePed:	PyePri:	YpnYem:	YemYds:	YddYds:
YdaYds:	YdsPph:	GhlPri:	PulPri:	PriPph:	PriCdg:	PriNlr:	PmdPpk	Gpc:PpkPph:	PphGpl:
PphPqe:	PphPqv:	PphPlg:	PphCsr:	PphYcv:	PphPiq:	PphGpt:	GpcPnp:	GpcChp:	GpcCsc:
GpcYhs:	GpcYdz:	GpcYna:	GpcYns:	GpcYsd:	GpcYfa:	GpcYir:	GpcYdp:	GpcYax:	GpcYfr:
GpcYfg:	GpcYcn:	GpcYtk:	GpcYbr:	GpcYls:	GpcYdv:	GpcYrs:	GpcYaz:	GpcYrm:	PlgPac:
CnrCsr									

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 of education
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	emotional problems
Ydd	Yds	43.6	26.0	0.000	210	1.4	1.9	depression	stress
Yda	Yds	54.7	32.6	0.000	210	0.0	2.9	anxiety	stress
Pmd	Ppk	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

	#bins		N							
cdgtcorrect	6	Cdg	255	Digit Symbol Substitution neuropsychological test						
cnormsrt	6	Cnr	210	Spatial Reaction Time normalized for age and sex						
cspatialreac	6	csr	214	Spatial Reaction Time test: how quickly patient responds to visual stimuli						
nlogmar	3	Nlr	209	LogMAR Log of Minimum Angle of Resolution (visual acuity)						

Cnr coarse, fine, ultra-fine searches

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

MODEL	Δdf	p	% ΔH	%c		N=175		
COARSE, single component predictors								
Cdg Gpt Cnr	3	0.00	10.6	64.6	BIC, AIC	Cdg = digit symbol test		
Pph Cdg Gpt Cnr	7	0.00	13.1	66.9	IncrP		Gpt = amnesia	
Cnr (independence=reference)	0	1.00	0.0	50.9			Pph = previous head injury	
FINE								
Cdg Cnr : Gpt Cnr	2	0.00	8.8	64.6	BIC			
Pri Cnr : Pph Cnr : Cdg Gpt Cnr	6	0.00	14.7	70.3	AIC	Pri = recent illness		
Pye Cnr : Pph Cnr : Cdg Gpt Cnr	5	0.00	12.9	67.4	IncrP	Pye = years education		
ULTRA-FINE (state-based model)								
Pph₁ Cdg₁ Cnr : Cdg₀ Gpt₁ Cnr	2	0.00	12.4	64.8	BIC			
Cnr (independence=reference)	0	1.00	0.0	50.9				

Cnr ultra-fine model

Model: Pph₁ Cdg₁ Cnr : Cdg₀ Gpt₁ Cnr

Odds (high is good) = Cnr₀/Cnr₁(model) = p(fast = normal reaction)/p(slow)

Pph₁ previous head injury, Cdg₁ high digit score; Gpt₁ amnesia

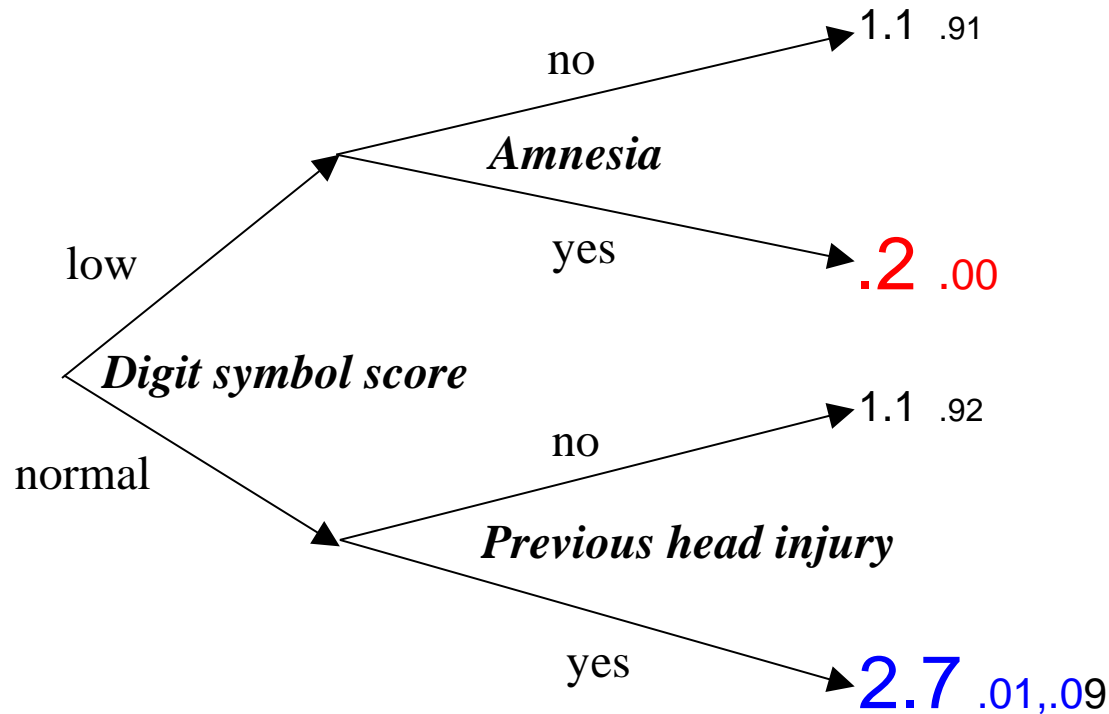
conditional probabilities of DV

IV states				data		model			
Pph	Cdg	Gpt	N	Cnr ₀	Cnr ₁	Cnr ₀	Cnr ₁	Odds	p
0	0	0	20	0.40	0.60	0.52	0.48	1.1	.92
0	0	1	19	0.16	0.84	0.16	0.84	0.2	.00
1	0	0	30	0.57	0.43	0.52	0.48	1.1	.90
1	0	1	18	0.17	0.83	0.16	0.84	0.2	.00
0	1	0	24	0.50	0.50	0.52	0.48	1.1	.91
0	1	1	13	0.61	0.39	0.52	0.48	1.1	.93
1	1	0	38	0.76	0.23	0.73	0.27	2.7	.01
1	1	1	14	0.64	0.36	0.73	0.27	2.7	.09
			176	0.51	0.49	0.51	0.49	1.0	

Cnr decision tree from conditional probabilities

Reaction time **Odds** (probability **fast**/ probability **slow**)

& p-values relative to marginal prob. (odds = 1)



Cnr decision tree, verbally

- 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

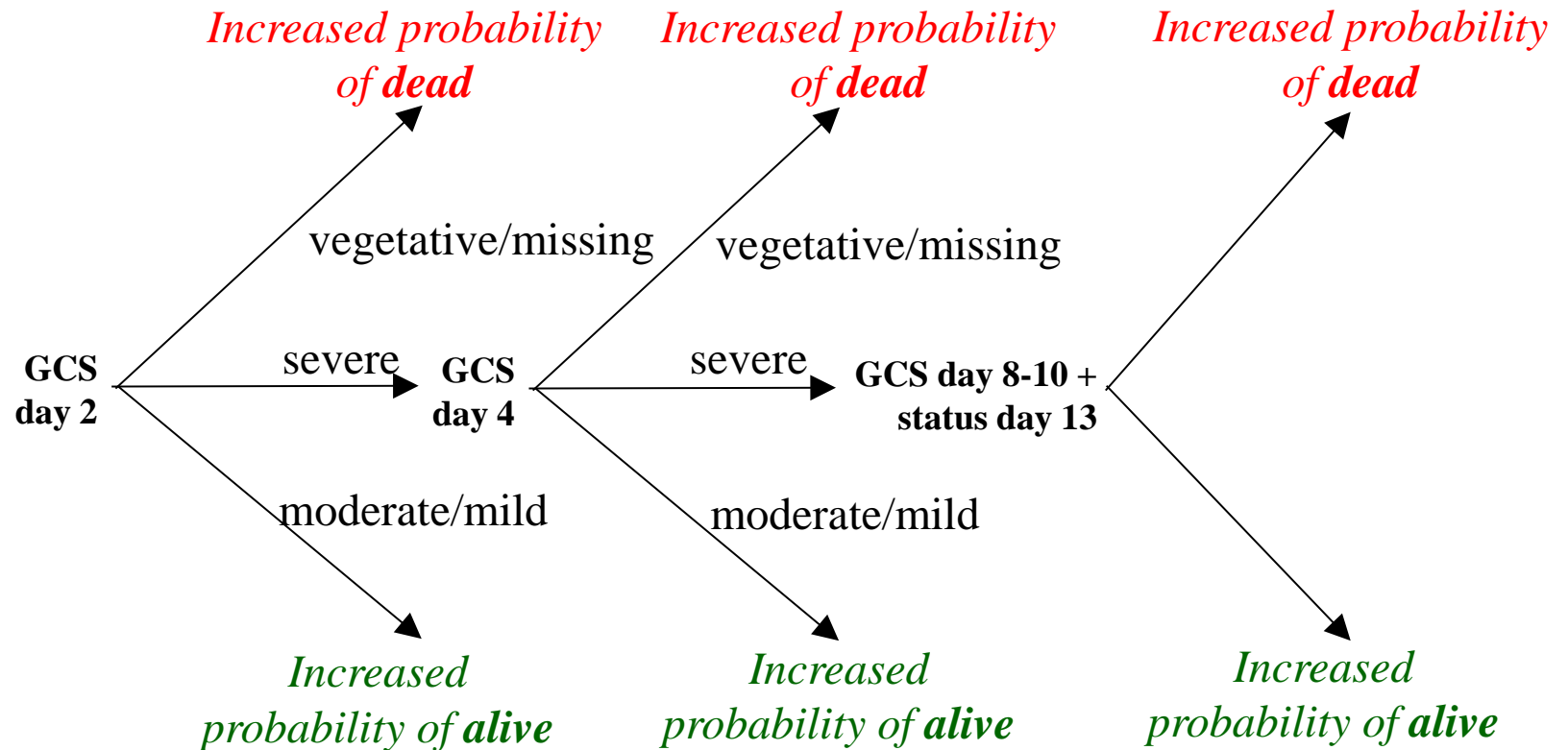
Directed searches

- DVs = deficit variables

	#bins		N		# IVs				
mort2	2	Gvn	764	Mortality at 2 weeks	302				
	0=not dead; 1=dead								
gose	8	Nvm	882	Total extended Glasgow Outcome Scale					
	1=death; 2=vegetative; 3,4 lower, upper severe disability;								
	5,6 lower, upper moderate disability; 7,8 lower, upper good recovery								
Two lines of current investigation:									
	<i>1 Predict mortality at 2 weeks</i>								
	<i>2 Investigate possible progesterone effect</i>								

Predict mortality at 2 weeks

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



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)

IV states		N	DV cond.prob.		N		p(margin)
Ngw	Pup		Gvn ₀	Gvn ₁	Gvn ₀	Gvn ₁	
.	0	49	0.41	0.59	20	29	0
0	0	190	0.92	0.08	175	15	0.01
.	1	49	0.33	0.67	16	33	0
0	1	206	0.97	0.03	200	6	0
1	0	141	0.94	0.06	133	8	0
1	1	129	0.85	0.15	110	19	0.92
		764	0.86	0.14	654	110	

- 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

Portland State Systems Science Graduate Program | Research: Discrete Multivariate Modeling - Mozilla Firefox

File Edit View History Bookmarks Tools Help Google

Portland State Systems Science Graduate Pr... +

www.pdx.edu/sysc/research-discrete-multivariate-modeling

Portland State UNIVERSITY Systems Science Graduate Program myPSU Contact SYSC Quick Menu

Courses Program Faculty Students Research Resources

PSU » System Science » Research » Research: Discrete Multivariate Modeling

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 log-linear 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.

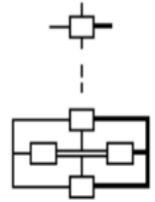
Projects

Theory/Methodology

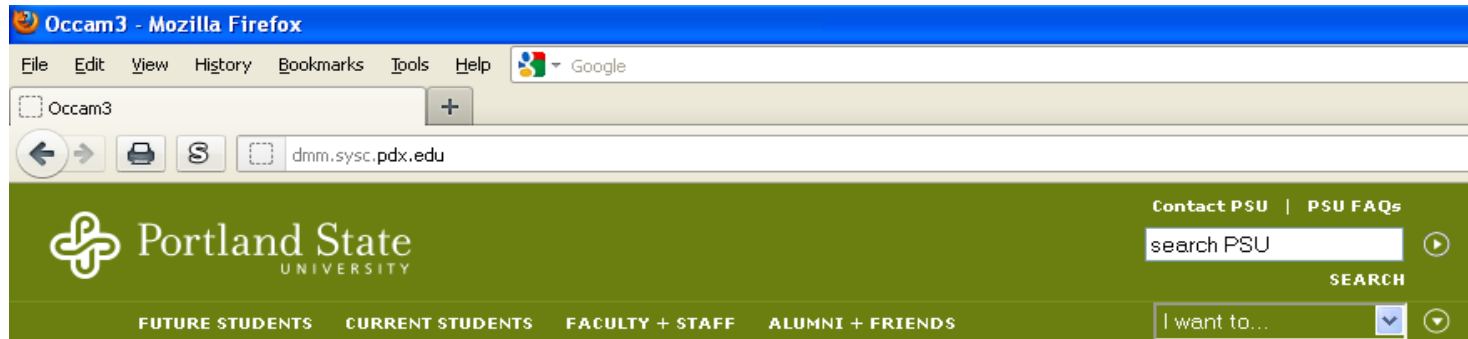
- OCCAM: RA software for data analysis & data mining**
 - [Occam3](#) (web accessible; try it out)
 - [User manual \(PDF\)](#)
- EDA: Extended Dependency Analysis**
 - Heuristic RA search for loopless models.
 - [Download](#) executable, sample files, and documentation (for Windows)

RA utility programs

Below is the lattice of structures for a 4-variable *directed* system with 1 dependent variable (output). Boxes = relations; lines = variables; bold lines = the dependent variable.



RA software (Occam)



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 substates.

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