9-2016

Exploratory Modeling of TBI Data

Martin Zwick
Portland State University, zwick@pdx.edu

Stephanie Kolakowsky-Hayner
Mt. Sinai

Sadie Carney
Portland State University

Maya Balamane
Brain Trauma Foundation

Tracie Nettleton
Oregon Health & Science University

See next page for additional authors

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

Exploratory Data Modeling of Traumatic Brain Injury

Martin Zwick

Systems Science Graduate Program
Portland State University

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

Bethesda, June 9, 2015

- RaDaR (raw data analysis)-Occam subproject
  - Martin Zwick, Tracie Nettleton
  - Hugo duCoudray Forrest Alexander, Naghmeh Daneshi, Peter Olson
- Brain Trauma Evidence-Based Consortium (BTEC)
- Dr. Nancy Carney, OHSU, head
- Funded by DoD via Brain Trauma Foundation, Stanford

1. Objectives, exploratory modeling

2. Preece data; approach

3. Some results on Preece dataset

1. Objectives, exploratory modeling

- Exploratory modeling (data mining) using Reconstructability Analysis (RA) on multiple data sets to contribute to:
  - a clinically useful TBI classification system
  - other BTEC subprojects, e.g., dynamic modeling

- now Preece data on auto accidents
- other data sets to follow

What RA is

- Reconstructability Analysis (RA) = Information theory + Graph theory
- RA: a probabilistic graphical modeling technique
  – Graph = model: node = variable; link = relationship
  – Hypergraphs = associations between >2 variables
- RA can detect many-variable or non-linear interactions *not hypothesized* *<u>in advance</u>*
- RA model = a (conditional) probability distribution simpler (fewer df) than data, capturing much of the information in the data

***Why RA** & Occam software*

- Explicitly designed for exploratory modeling
  - Analyzes both nominal & continuous (binned) variables
  - Easily interpretable method & output
  - Standard text input; Occam emails results to user
  - Occam web-accessible; available for research use

- Related 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* difficulty with nominal variables or stochasticity

2. Preece data; approach

- 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

Variables (1/3)

- Patient (P) variables (17)

pinjgrp,5,	pij	Injury Group patient or control
page,7,	pag	age
psex,2,	psx	sex
pyred,6,	pye	years of education
pedlevel,8,	ped	highest level of education
puhrsleep,5,	pul	usual # of hrs of sleep: less than or greater than normal (8 hr)
precentill,3,	pri	recent illness 0 no 1 yes
pmedication,3,	pmd	current medications 0 no 1 yes
ppainkller,3,	ppk	currently on painkillers 0 no 1 yes
ppreheadinj,3,	pph	have they had previous head injury 0 no 1 yes
pprecon,3,	ppc	previous concussion 0 no 1 yes
pnumprecon,8,	pnp	how many previous concussions
pdbqerror,13,	pqe	Driver Behavior Questionaire self reported driving errors/violatior
pdbqviol,14,	pqv	Driver Behavior Questionaire violations
plitigat,4,	plg	was the case litigated
prespacc,6,	pac	who was responsible for the accident
pfsiq,5,	piq	full scale IQ calculated from national adult reading test

Variables *(2/3)*

- Symptom (Y) variables (25)

ypainscale,5,	ypn	standard painscale used by hospitals	
yemoscale,5,	yem	sacle defining emotional state(0 no problems 1 few 2 moderate 3 many problems)	
ydassd,5,	ydd	Depression Anxiety Stress Scales: depression	
ydassa,6,	yda	Depression Anxiety Stress Scales: anxiety	
ydasss,4,	yds	Depression Anxiety Stress Scales: stress	
yheadache,6,	yhs	Rivermead	headache
ydizz,5,	ydz	Rivermead	dizzy
ynausea,5,	yna	Rivermead	nausea
ynoisesens,6,	yns	Rivermead	noise sensitivity
yslpdis,6,	ysd	Rivermead	sleep disorder
yfatigue,6,	yfa	Rivermead	fatigue
yirritable,6,	yir	Rivermead	irritable
ydepressed,5,	ydp	Rivermead	depressed
yanxious,6,	yax	Rivermead	anxious
yfrustrated,5,	yfr	Rivermead	frustrated
yforgtful,6,	yfg	Rivermead	forgetful
ypoorconc,6,	ycn	Rivermead	poor concentration
ylongthink,6,	ytk	Rivermead	long time to think
yblurredvis,6,	ybr	Rivermead	blurred vision
ylightsens,5,	yls	Rivermead	light sensitivity
ydoublevis,6,	ydv	Rivermead	double vision
yrestless,6,	yrs	Rivermead	restless
ydazed,5,	yaz	Rivermead	dazed
yrivmead,5,	yrm	summation of Rivermead post concussion symptom questionaire	
ycrrectedvis,3,	ycv	corrected vision	

Variables *(3/3)*

- Sign (G) & Deficit (C, N) variables (4, 5, 1)

ghrssleep,5,	ghl	number of hours of sleep, divided in less than normal normal=8hr and greater than normal
ggcs,4,	ggc	Glasgow coma scale a measure of level of unconsciousness; lower = deeper unconsciousness
gextcause,8,	gxc	external cause of the injury
gpta,3,	gpt	post traumatic amnesia
chazpt,10,	chp	hazard perception test measures how quickly potential driving hazards are predicted
cnormsrt,6,	cnr	Spatial Reaction Time normalized for age and sex
cspatialreac,6,	csr	Spatial Reaction Time tests how quickly patient responds to a visual stimuli
cdgtcorrect,7,	cdg	Digit Symbol Substitution neuropsychological test
cstarcan,4,	csc	Star Cancelation Test a test of spatial neglect
nlogmar,4,	nlr	LogMAR Logarithm of the Minimum Angle of Resolution: a visual acuity test

Approach *(1/3)*

2 types of model searches

- *Neutral*: find relationships among all variables ('clustering')
- *Directed*: predict C, N variables from P, Y, G ('classification')
  - reference = independence model
  - predictive success (information captured) measured by
    - %ΔH = %reduction of uncertainty: (information-theoretic measure)

      Uncertainty is *like* variance

      Rule of thumb: %ΔH = 8% *can be* a sizeable effect
    - %c = %correct (general measure)
  - want low model complexity = Δdf

Approach *(2/3)*

3 degrees of search refinement (IVs: A,B,C…; DV: Z)

- *Coarse search*: variable-based models w/o loops, e.g., A B z , ...

  Fast, can handle *many* variables

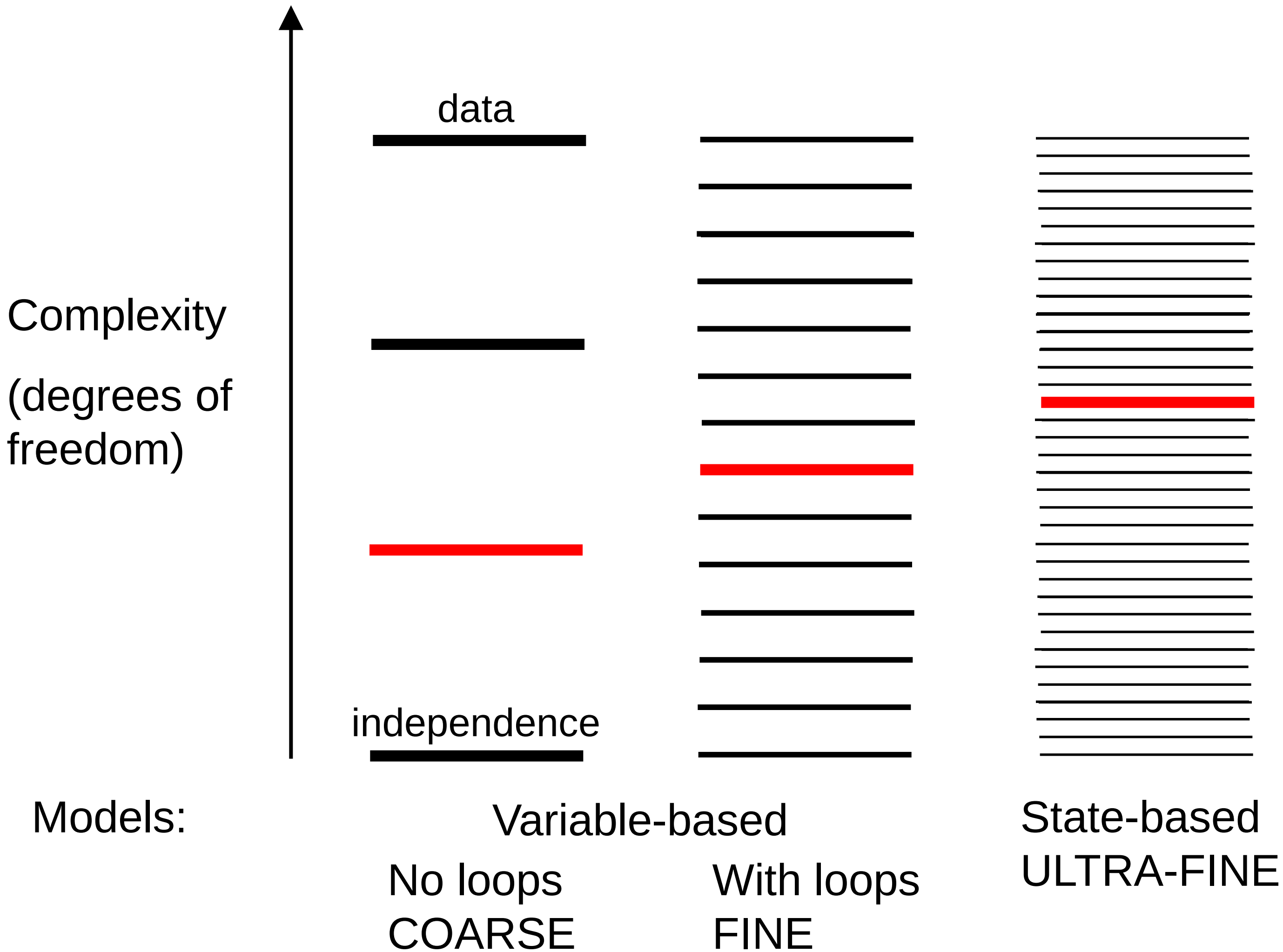
- *Fine search*: variable-based models w' loops, e.g., A B z : B C z

  Slow, can handle 100s of variables

- *Ultra-fine search*: state-based models, e.g., A_2 B_1 z : B_0 z

  Very slow, less than 10 variables

Three degrees of search refinement



Approach *(3/3)*

3 model selection criteria (information-complexity tradeoff)

- *Conservative:* Bayesian Information Criterion (BIC)
- *Aggressive:* Akaike Information Criterion (AIC)

  Incremental p-value (IncrP)

- AIC & BIC: linear combinations of error (opposite of information) & 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

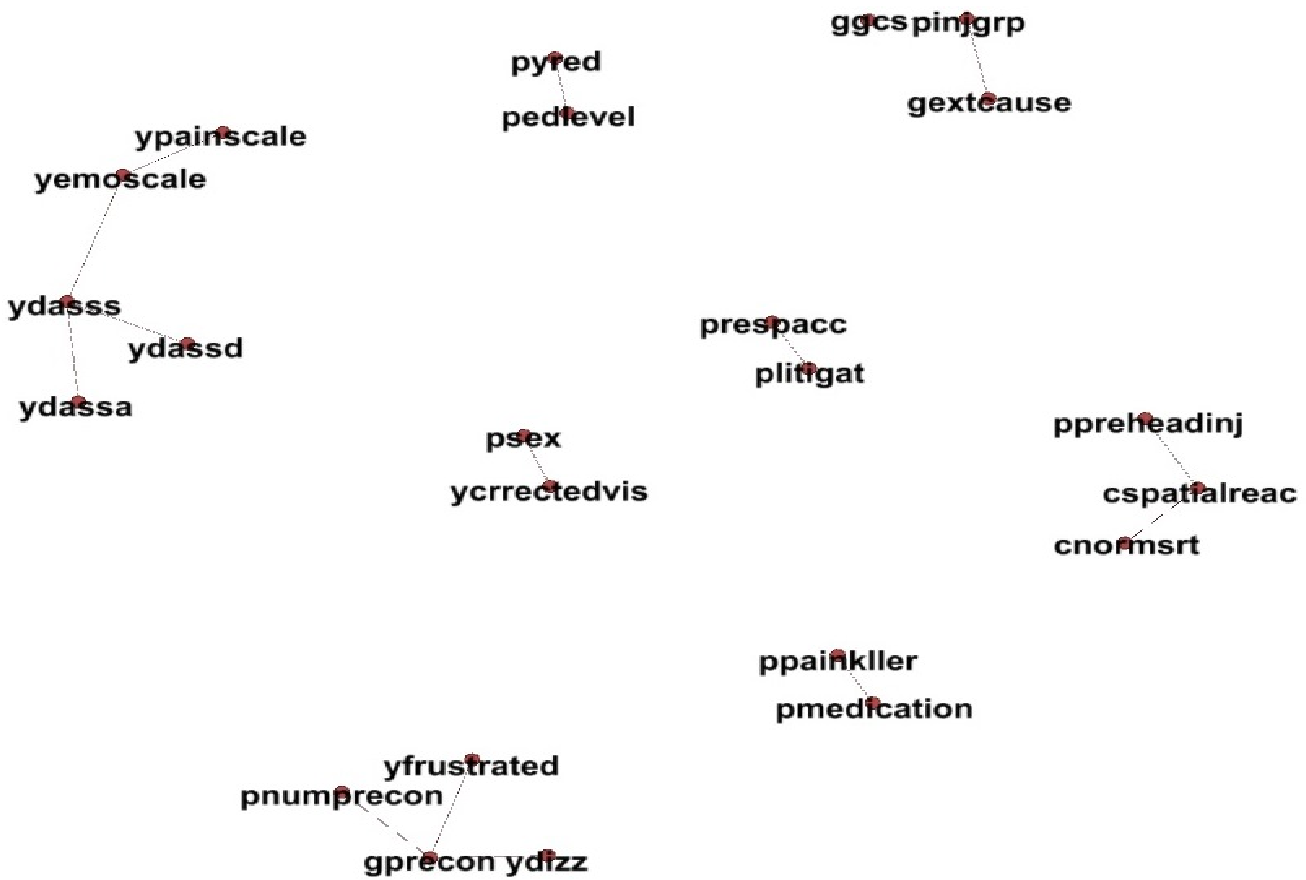
Some issues: binning, missing data, small N, validation

3. Results on Preece dataset

- Neutral *coarse searches*
  - find associations among all P, Y, G, C, N variables
- Directed *coarse, fine, ultra-fine searches*
  - predict C, N from P, Y, G & from *other* C, N variables

Neutral coarse search *(graph of associations)*

- 15 **p ≤ 0.05** associations in BIC model (2 involve C)



Neutral coarse search (15 associations)

- Predictive success (%ΔH, Δ%c relative to independence) (p ≤ 0.05)

v1	v2	%ΔH(2|1)	%ΔH(1|2)	p-value	N	Δ%c(2|1)	Δ%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 = deficit variables

- Priorities (Dr. Carney): focus here on predicting Cdg, Cnr
- #bins excludes missing values ; will often aggregate states into fewer bins

	#bins		N	
cdgtcorrect	6	**Cdg**	255	**Digit Symbol Substitution neuropsychological test**
				Most important/reliable 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)
				Less important/reliable
cstarcan	3	csc	50	Star Cancelation Test a test of spatial neglect
chazpt	9	chp	282	Hazard perception test: how quickly potential driving hazards are predicted

Cdg ***directed coarse, fine, ultra-fine searches***

Predict Cdg: digit symbol substitution test (rebin |Cdg| = 2: ~ 50-50)

MODEL (IV component omitted)	Δdf	p	%ΔH	%c				
COARSE, single predictors					**ΔBIC**	**N=240**		
Pij Cdg	3	0.00	11.9	68.3	**47.6**	**patient injury type**		
Ped Cdg	7	0.00	11.7	65.0	5.9	education level		
Ggc Cdg	3	0.00	5.6	65.0	18.3	Glasgow coma scale		
Cnr Cdg	5	0.00	3.5	60.8	6.1	**spatial reaction, normalized**		
Pye Cdg	1	0.00	3.0	68.3	**27.9**	**years education**		
Csr Cdg	5	0.00	2.5	63.3	0.4	spatial reaction		
Cdg (independence=reference)	0	1.00	0.0	***50.8***	0			
FINE					**Criterion**	**N=240**	|Cnr|=6, incl missing	
Pij Cdg : Pye Cdg	4	0.00	25.5	72.9	**BIC**			
Pij Cdg : Pye Cdg : Cnr Cdg	9	0.00	32.8	76.7	**AIC**			
Pij Cdg : Psx Cdg : Pye Cdg : Cnr Cdg	10	0.00	32.9	76.3	**IncrP**	**sex**		
ULTRA-FINE (state-based model)						**N=175**	|Cnr|=2, no missing	
Pij_2 Cnr_1 Cdg : Pye_0 Cdg	2	0.00	13.5	68.6	**BIC**			
Cdg (independence=reference)	0	1.00	0.0	***50.9***				

Cdg ***ultra-fine (state-based)*** ***model*** *3/3*

Model: $Pij_2\ Cnr_1\ {}_{Cdg}$: $Pye_0\ {}_{Cdg}$

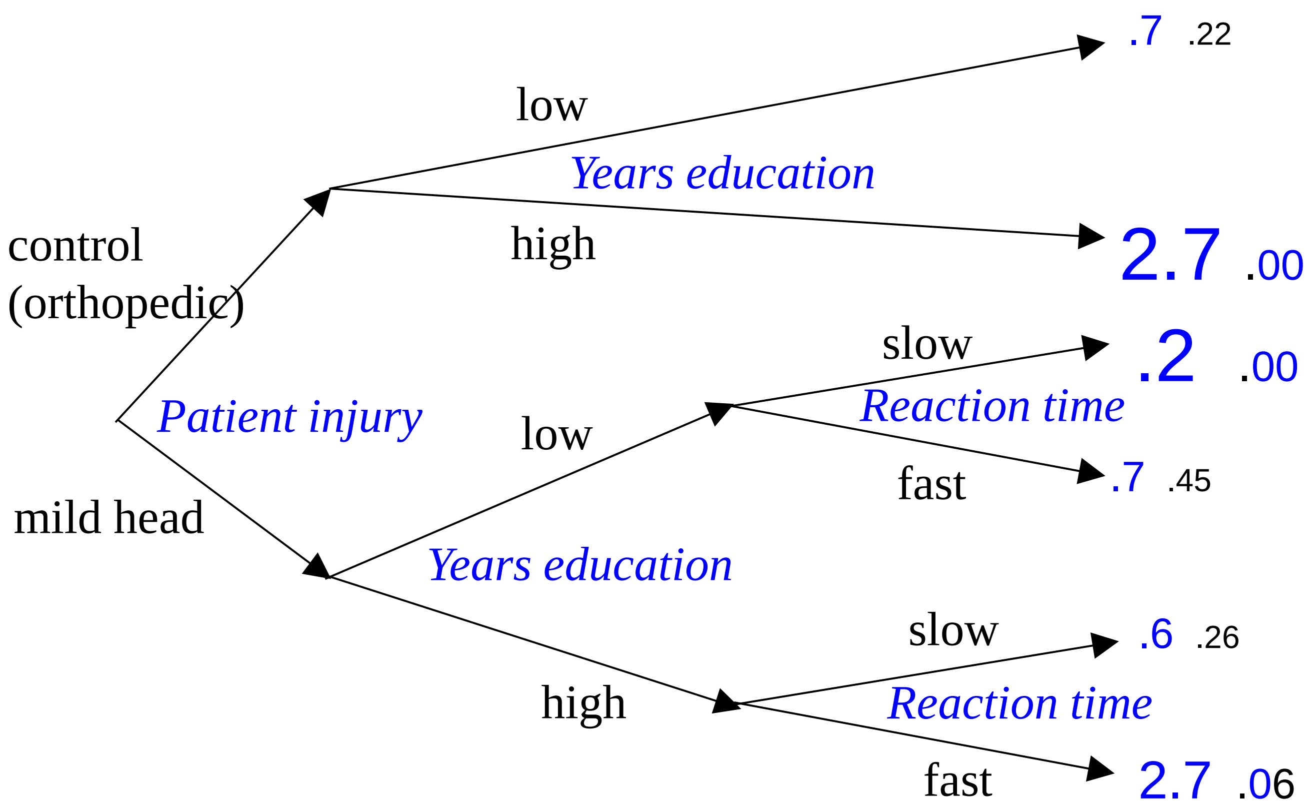
Odds (high is good) = $\mathbf{Cdg_1}/Cdg_0$(model) = p(**high** digit score)/p(low score)

Pij_1 control, Pij_2 mild head injury; Pye_0 low years educ.; Cnr_0 = fast reaction

				conditional probabilities of DV					
IV states				data		model			
Pij	Pye	Cnr	N	Cdg_0	Cdg_1	Cdg_0	$\mathbf{Cdg_1}$	Odds	p
1	0	0	18	0.50	0.50	0.59	**0.41**	0.7	.41
1	0	1	22	0.68	0.32	0.59	**0.41**	0.7	.36
1	1	0	38	0.21	0.79	0.27	**0.73**	2.7	.01
1	1	1	20	0.35	0.65	0.27	**0.73**	2.7	.05
2	0	0	15	0.53	0.47	0.59	**0.41**	0.7	.45
2	0	1	24	0.88	0.13	0.86	**0.14**	0.2	.00
2	1	0	18	0.33	0.67	0.27	**0.73**	2.7	.06
2	1	1	20	0.60	0.40	0.62	**0.38**	0.6	.26
			175	0.49	0.51	0.49	**0.51**	1.0	

Cdg decision tree ***from conditional probabilities***

Digit Symbol score odds (prob. high performance/ prob. low performance) & p-values relative to marginal prob. (odds = 1):



Cdg decision tree**, **verbally

- For all patients, education predicts performance on digit symbol test: more education predicts better performance.
  - Education is a confounding variable for digit symbol test in discriminating concussion, & must be controlled for
- For controls (orthopedic), reaction time does not predict digit symbol score.
- For TBI patients, fast reaction time predicts better digit symbol performance beyond influence of education.

Cnr directed 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 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)						
Pph_1 Cdg_1 Cnr : Cdg_0 Gpt_1 Cnr	2	0.00	12.4	64.8	**BIC**	
Cnr (independence=reference)	0	1.00	0.0	***50.9***		

Cnr *ultra-fine (state-based)* *model*

Model: $Pph_1\ Cdg_1\ _{Cnr}$: $Cdg_0\ Gpt_1\ _{Cnr}$

Odds (high is good) = $\mathbf{Cnr_0}/Cnr_1$(model) = p(**fast** = normal reaction)/p(slow)

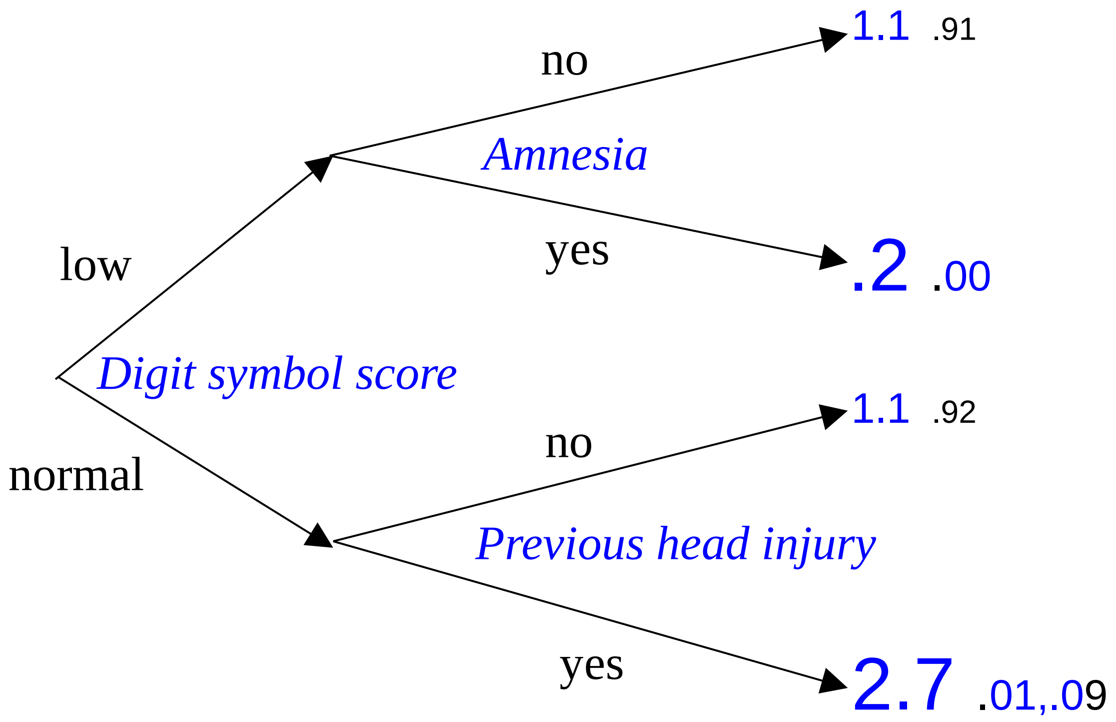
Pph_1 previous head injury, Cdg_1 high digit score; Gpt_1 amnesia

				conditional probabilities of DV					
IV states				data		model			
Pph	Cdg	Gpt	N	Cnr_0	Cnr_1	$\mathbf{Cnr_0}$	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 score odds (probability normal/ 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.
  - We need to see if it would be replicated in another data set.
  - One 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.

Future ***(1/2)***

- Preece data a test bed for future analyses.

- Results are preliminary & tentative, *illustrative* of *type* of results from exploratory analysis.

- Need to *confirm* results with other data sets or future studies.

Future ***(2/2)***

- Hoping for *more* data sets (accident, military, sports), *higher* N, *fewer* missing data, *additional* types of variables (imaging, genomic, proteomic).

- Work to be fully collaborative with investigators sharing data.

- THANK YOU

RA (DMM) web page

http://pdx.edu/sysc/research-discrete-multivariate-modeling

zwick@pdx.edu