Evidence-Based Assessment: From simple clinical judgments to statistical learning

Eric A. Youngstrom, PhD

University of North Carolina at Chapel Hill

Empirical Article

Clinical Psychological Science
1–23

DOI: 10.1177/2167702617741845
www.psychologicalscience.org/CPS
SAGE

Evidence-Based Assessment From Simple Clinical Judgments to Statistical Learning: Evaluating a Range of Options Using Pediatric Bipolar Disorder as a Diagnostic Challenge

Eric A. Youngstrom[1], Tate F. Halverson[1], Jennifer K. Youngstrom[1], Oliver Lindhiem[2], and Robert L. Findling[3]
[1]University of North Carolina at Chapel Hill, [2]University of Pittsburgh, and [3]Johns Hopkins University

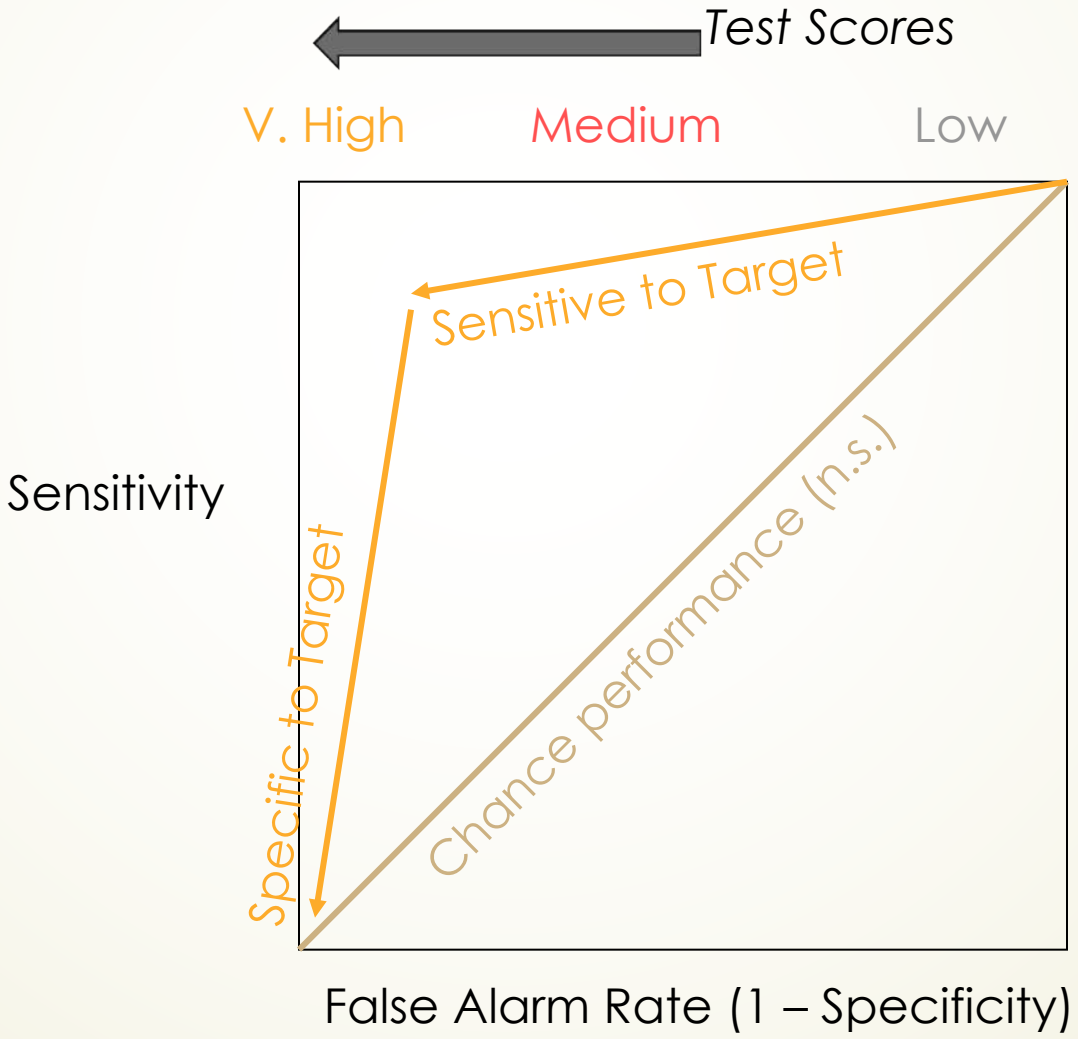
Looking at clinical decision-making from many angles

- My journey from age 18
- Researcher/Teacher/Clinician/Parent
- Revisioning my research
- How to teach it better? Or transfer the flag?
- How to incorporate new research, more information?

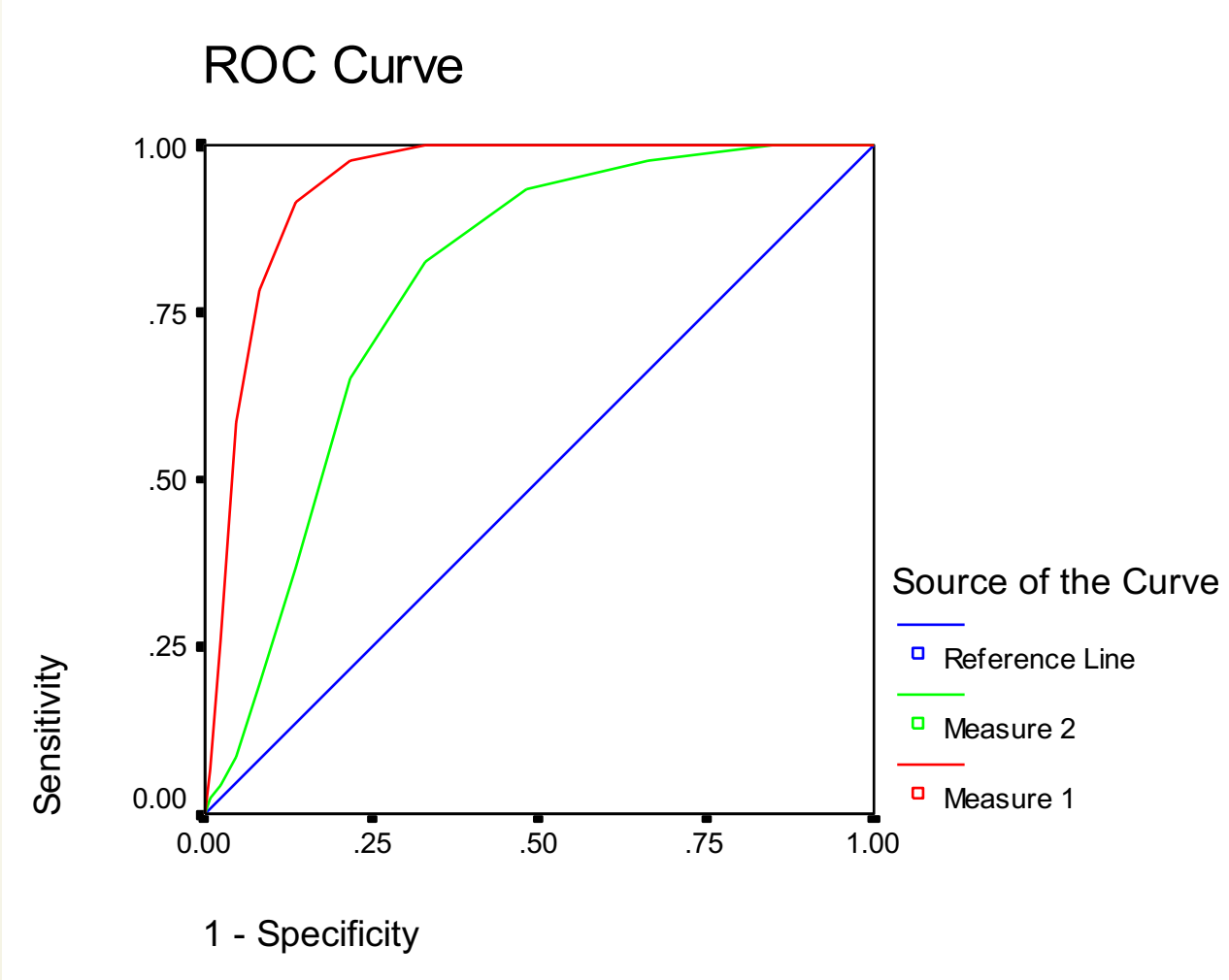
Early 1990s

- Where were you?
- (Working on master's thesis)
- First modern "sightings" of pediatric bipolar
  - Geller 1993 Depression Trial
  - Wozniak 1995 JAACAP paper (ADHD sample)
  - 1999 Papolos book

Evaluating Diagnostic Efficiency:
Receiver Operating Characteristics (ROC)



A Visual Comparison of Diagnostic Efficiency



Measure 1 clearly performs better across a wide range of scores.

AUC #1 = .94
AUC #2 = .79
(Measure 1 is better, $p < .0005$)

Areas Under the Curve (AUC)

- ~~Excellent: .90 +~~ → Be suspicious!
- Good: .70 to .89
- Fair: .60 to .69
- Poor: < .60
- Chance: .50
  (If you get a number significantly below .50, you are using a good test backwards!)

A Primer on Receiver Operating Characteristic Analysis and Diagnostic Efficiency Statistics for Pediatric Psychology: We Are Ready to ROC

Eric A. Youngstrom,[1,2] PhD

[1]*Department of Psychology and* [2]*Department of Psychiatry, University of North Carolina at Chapel Hill*

All correspondence concerning this article should be addressed to Eric A. Youngstrom, PhD, Department of Psychology, University of North Carolina at Chapel Hill, Davie Hall CB 3270, Chapel Hill, NC 27599-3270, USA. E-mail: eay@unc.edu

Received March 6, 2013; revisions received May 27, 2013; accepted July 9, 2013

(come to the workshops!)

Bringing Bayes to clinicians…

Comparing the Diagnostic Accuracy of Six Potential Screening Instruments for Bipolar Disorder in Youths Aged 5 to 17 Years

ERIC A. YOUNGSTROM, PH.D., ROBERT L. FINDLING, M.D., JOSEPH R. CALABRESE, M.D., BARBARA L. GRACIOUS, M.D., CHRISTINE DEMETER, B.A., DENISE DELPORTO BEDOYA, M.A., AND MEGAN PRICE, M.A.

TABLE 4

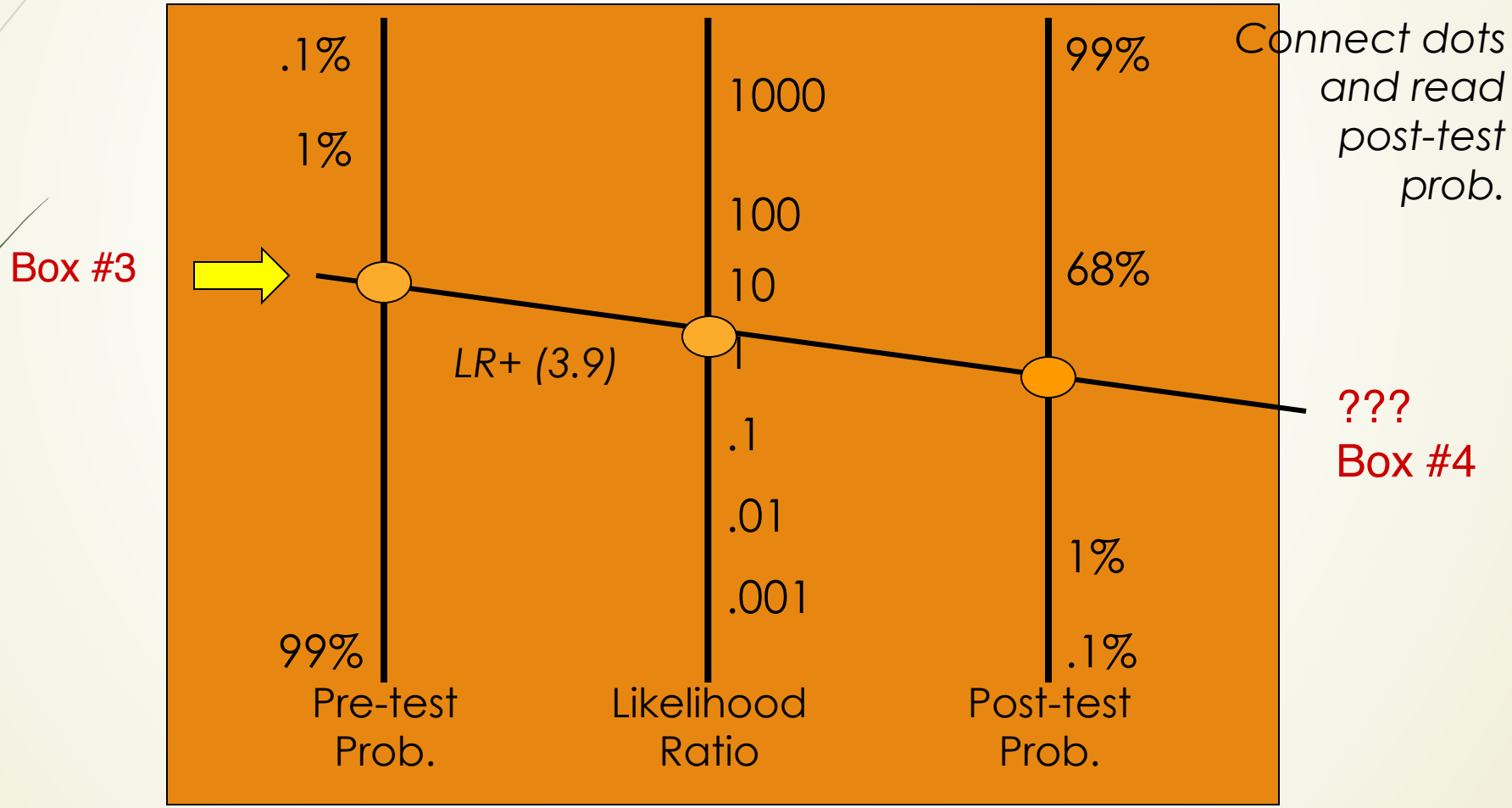
Change in Odds of Bipolar Diagnosis (Likelihood Ratios) for Index Test Scores

Ages 5–10 LR: 50.3% Prevalence of Bipolar Disorders

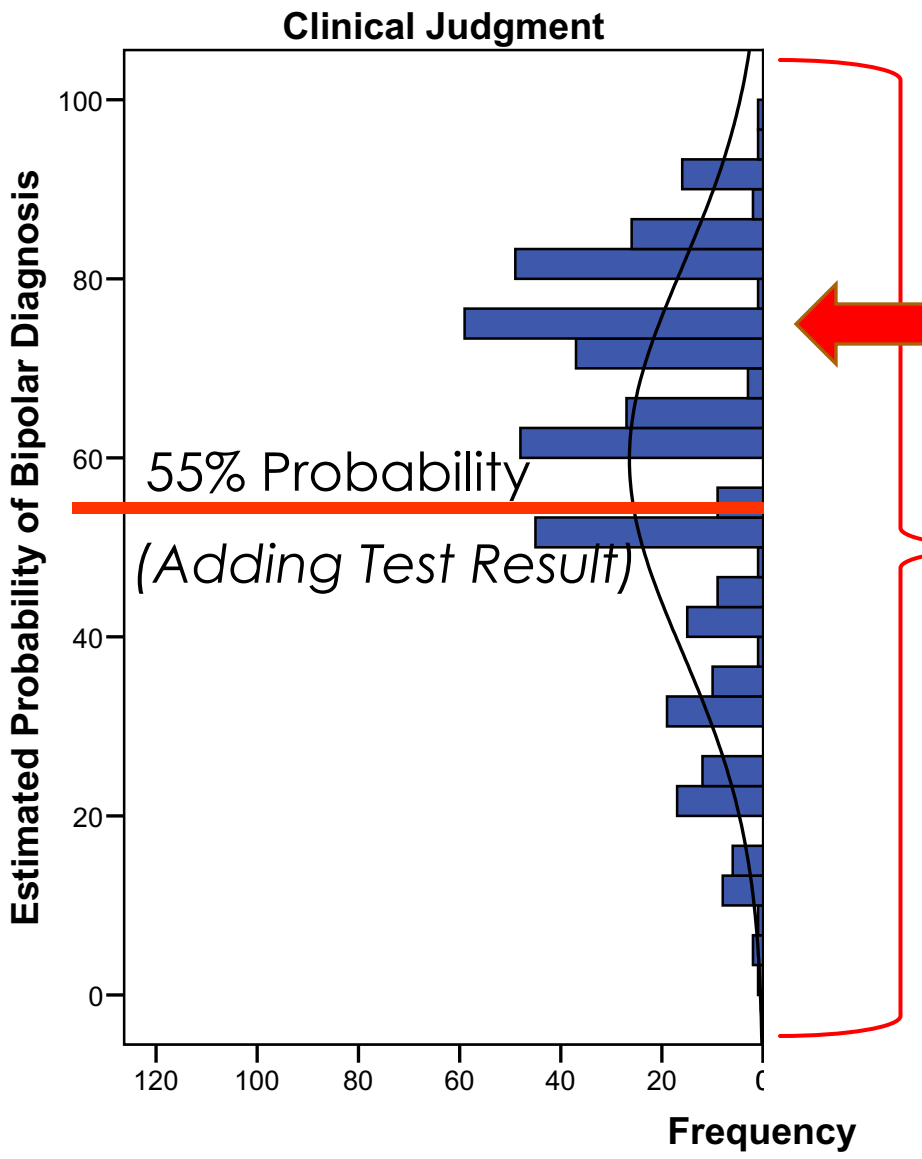
	Range					
Summary	Low	Mod. Low	Neutral	Mod. High	High	Very High
CBCL						
Score	<58	58–67	68–72		**73+**	
LR	0.07	0.47	1.50		**3.91**	

LR+ = 3.9

Using a Nomogram
Add a CBCL Test Result



Is the Nomogram Worth Using?

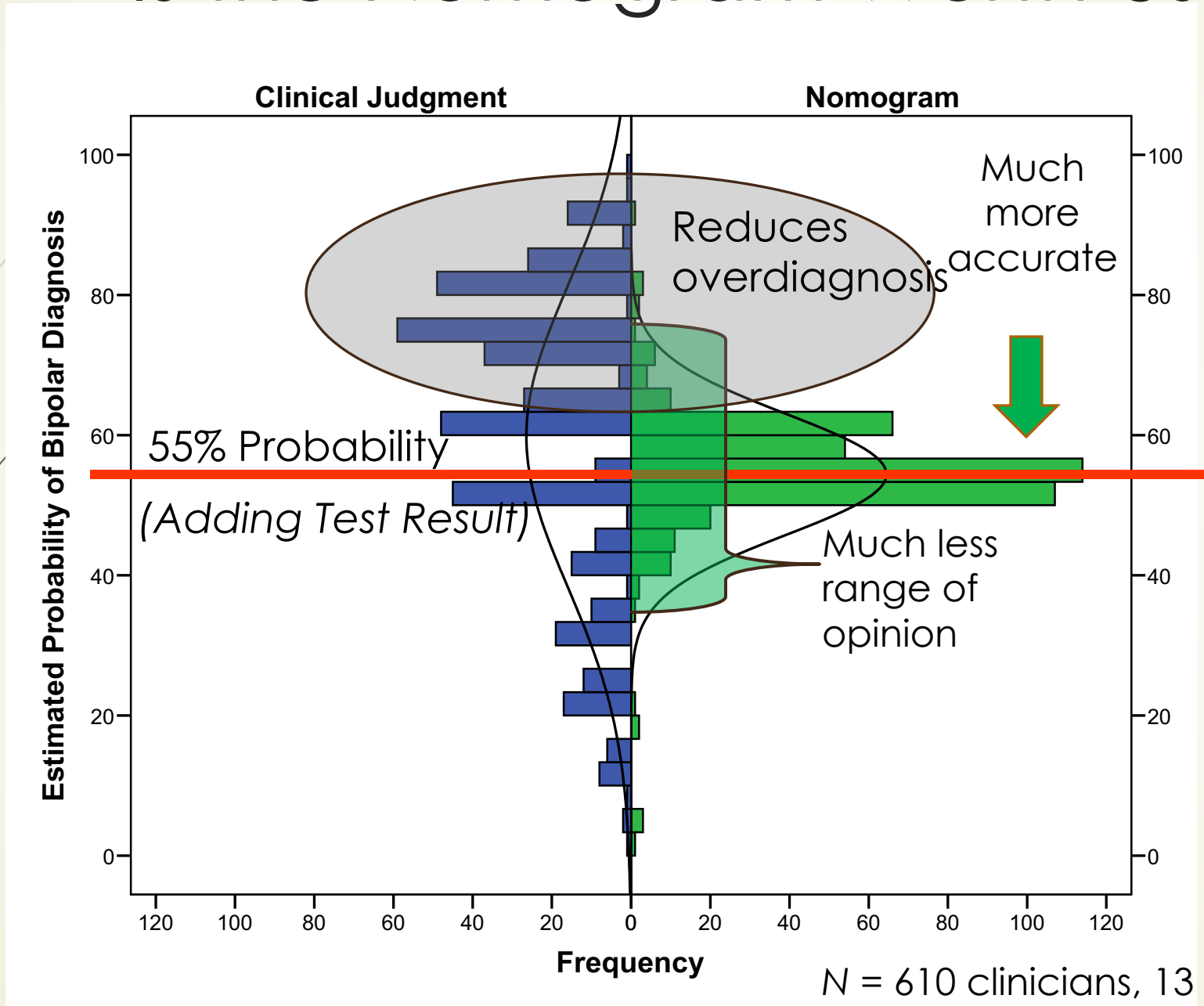


Most tend to overdiagnose

Still extreme range of opinion

N = 610 clinicians, 13 sites

Is the Nomogram Worth Using?



Statistical Learning Models

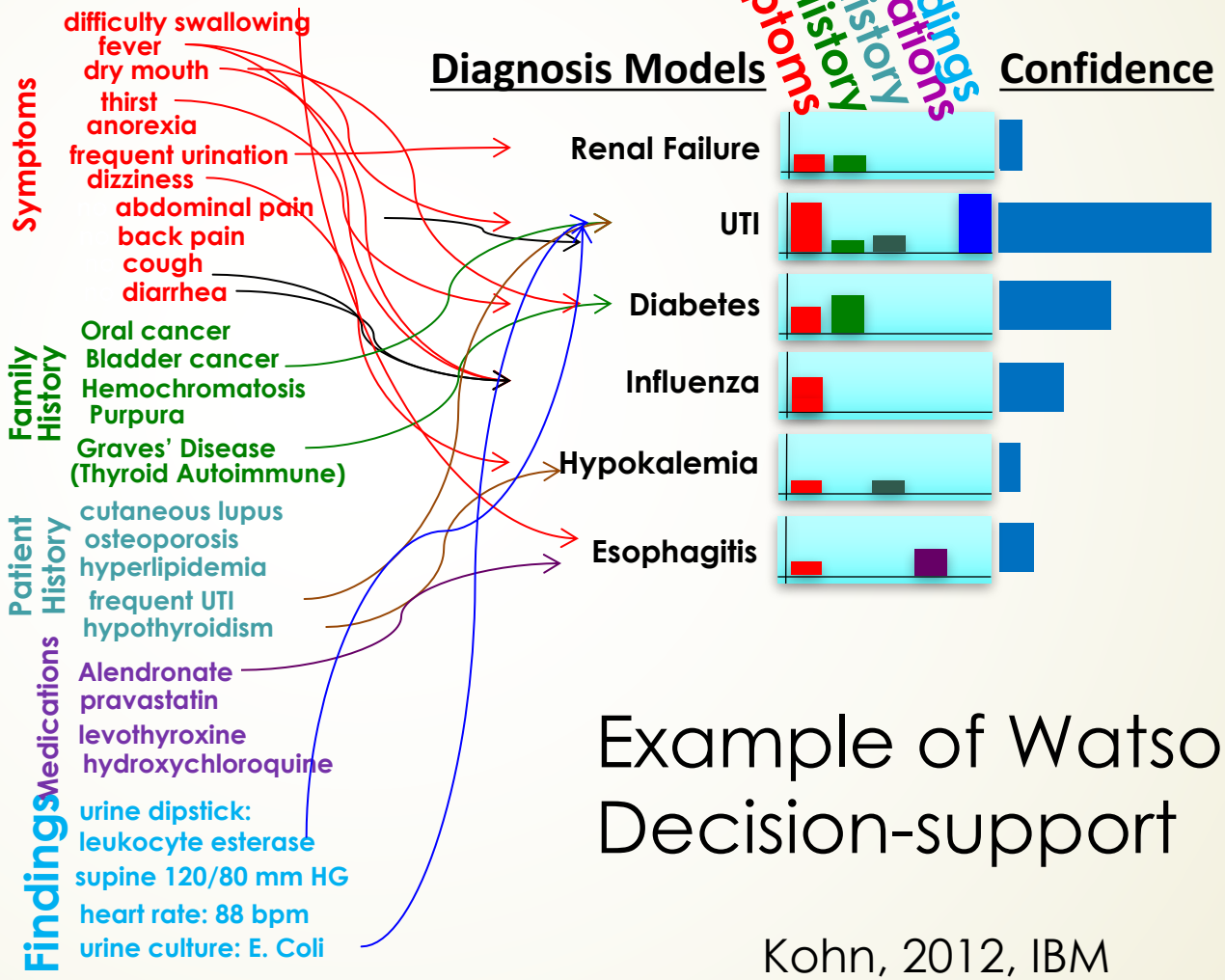
- Count how many buzzwords you have heard:
  - ❑ Data mining, ❑ Machine learning, ❑ Watson,
    ❑ Statistical learning… ❑ "big data," ❑ Internet of Things…
  - It's not just for psychology: Netflix, Amazon, IBM, Google
- Turns out that most of the methods are things that we learned in grad school!
- Key is to have computer do the heavy work:
  - Automate the model building and testing
  - Bias-Variance Trade-off (~Type I versus Type II error)
  - Use internal cross-validation to pick a model that is likely to generalize

IBM Watson wins on *Jeopardy!*

- **Natural language, unlike chess**
- **Better approximates clinical interview**
- **Medical decision-making**

14 February, 2011

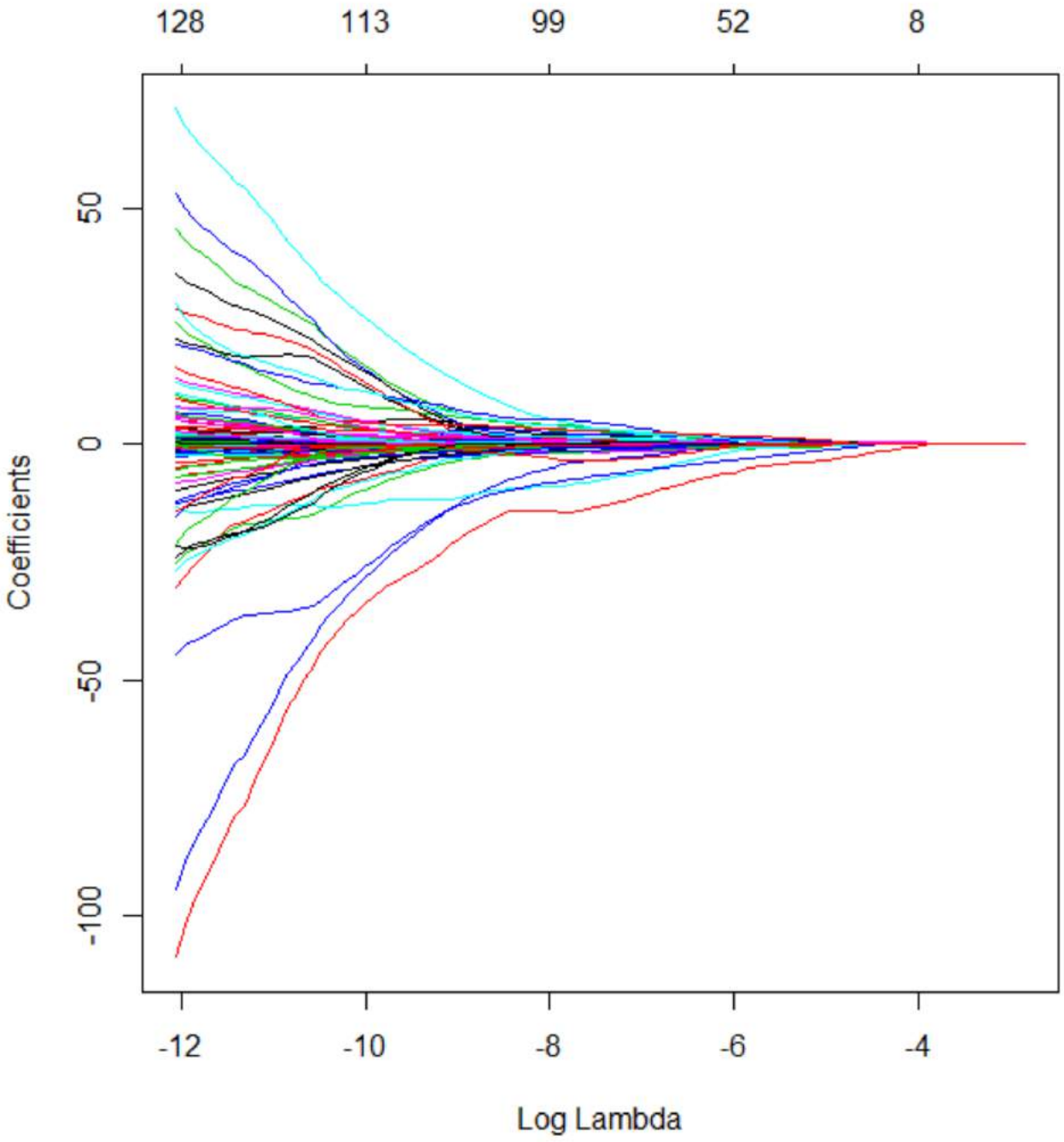
Putting the proper pieces together at the point of impact can be life changing

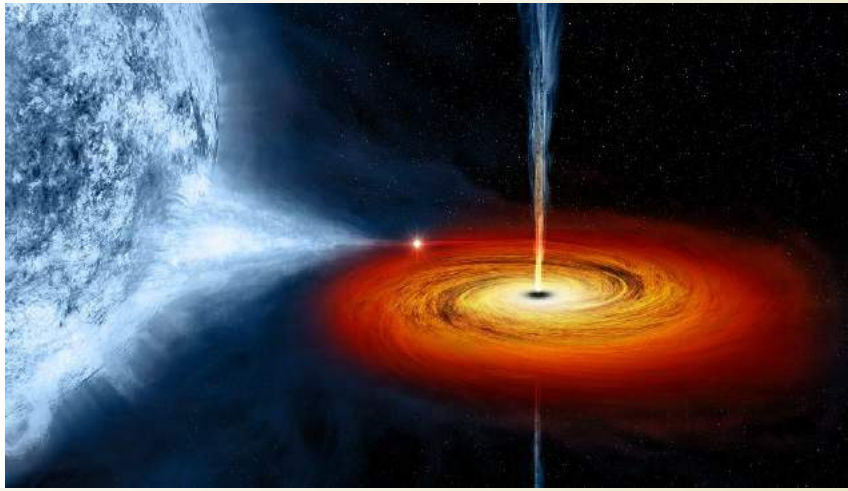


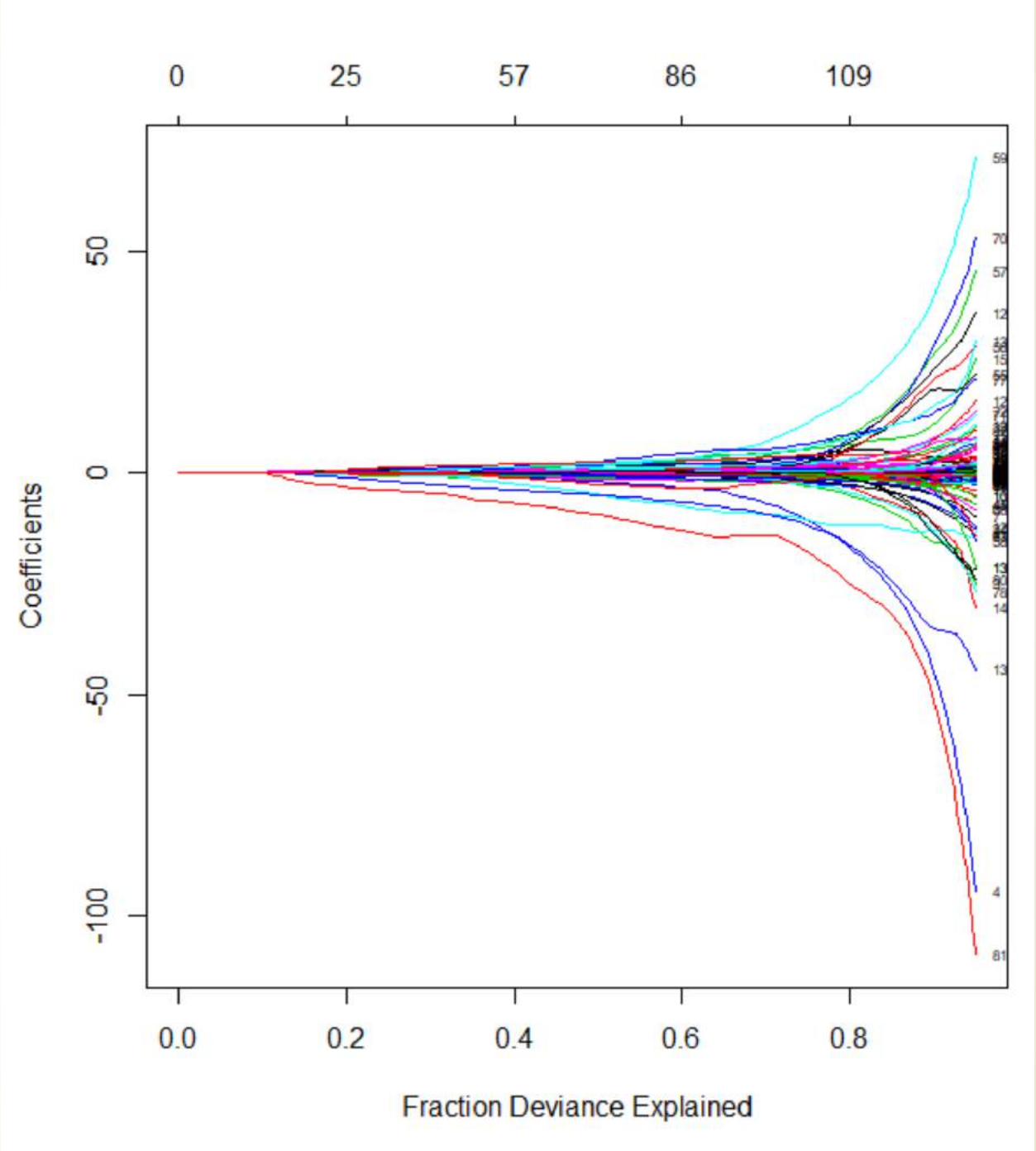
Example of Watson Decision-support

Kohn, 2012, IBM

LASSO regression

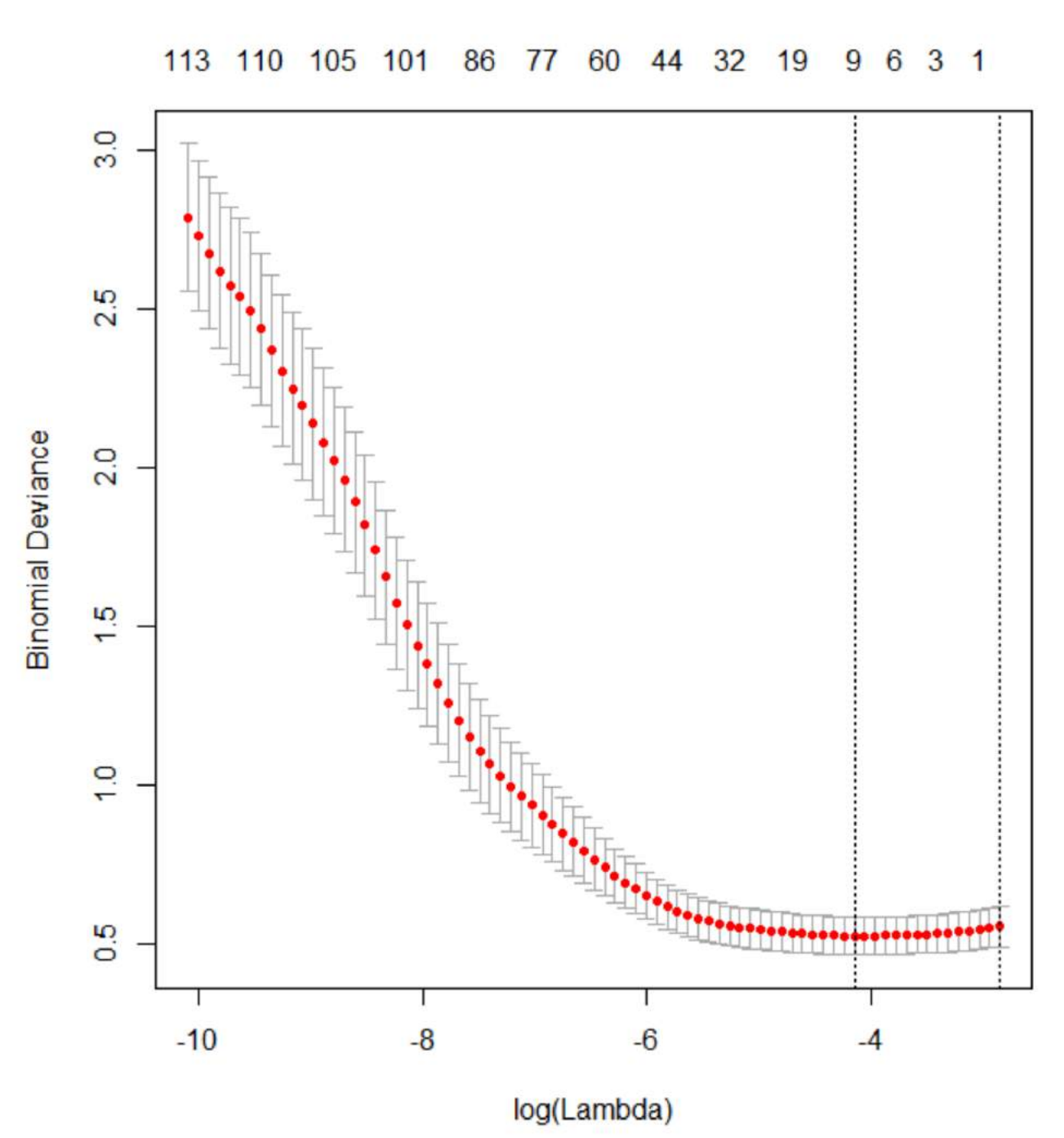





0
25
57
86
109
50
0
-50
-100
Coefficients
0.0
0.2
0.4
0.6
0.8
Fraction Deviance Explained

How to pick?

10-fold cross-validation



Contestants

- ~~There ain't no such thing!~~
- ~~Bet the base rate~~
- Take the best screener – test positive or negative?
- ~~Bayes Theorem – too hard~~
  - Nomogram: Just connect the dots!
- Multilevel Likelihoods, two predictors
- Logistic regression
  - 1 predictor (every score gets its own prediction)
  - Multiple predictors
- LASSO

(but could also do quadratic discrim, random forests….)

Increasing model complexity

Table 1. Candidate Variables Included in Each Prediction Model

Variable	Take the best screener	Probability nomogram	Multilevel and multipredictor nomogram	Logistic regression (1 *df*)	Augmented logistic regression (5 *df*)	LASSO (136 candidate variables)
PGBI10M	X	X	X	X	X	X
Family bipolar history			X		X	X
Sex (female)					X	X
Youth age (years)					X	X
Race (White yes/no)					X	X
PGBI–depression						X
PGBI–hypo/biphasic						X
PGBI–sleep						X
PGBI 7 Up						X
PGBI 7 Down						X
Diagnosis count						X
Other diagnoses[a]						X
Two-way interactions						X

Unfair advantages:

LASSO = least absolute shrinkage and selection operation; PGBI = Parent General Behavior Inventory; PGBI10M = PGBI 10-item mania scale.
[a]Dummy codes for attention-deficit/hyperactivity disorder, oppositional defiant disorder, conduct disorder, anxiety, and posttraumatic stress disorder.

The challenge:
Identify cases with bipolar disorder…

…under clinically realistic conditions

Place your bets – human versus LASSO?

Criterion	Bet
Statistical significance?	Both
Clinical significance?	Both
Best accuracy?	
Usability?	

Not much of a contest

- We know that regression will produce optimized weights
- LASSO is getting extra variables that clinician wouldn't

Next questions:

- How much better is the statistical model?
- And would it work at your clinic? ←the external validation question

Plot twist:
There's a second clinic

Academic

Community

Academic and Community Samples: Different on almost every variable

Table 2. Demographics and Clinical Characteristics by Clinic Setting

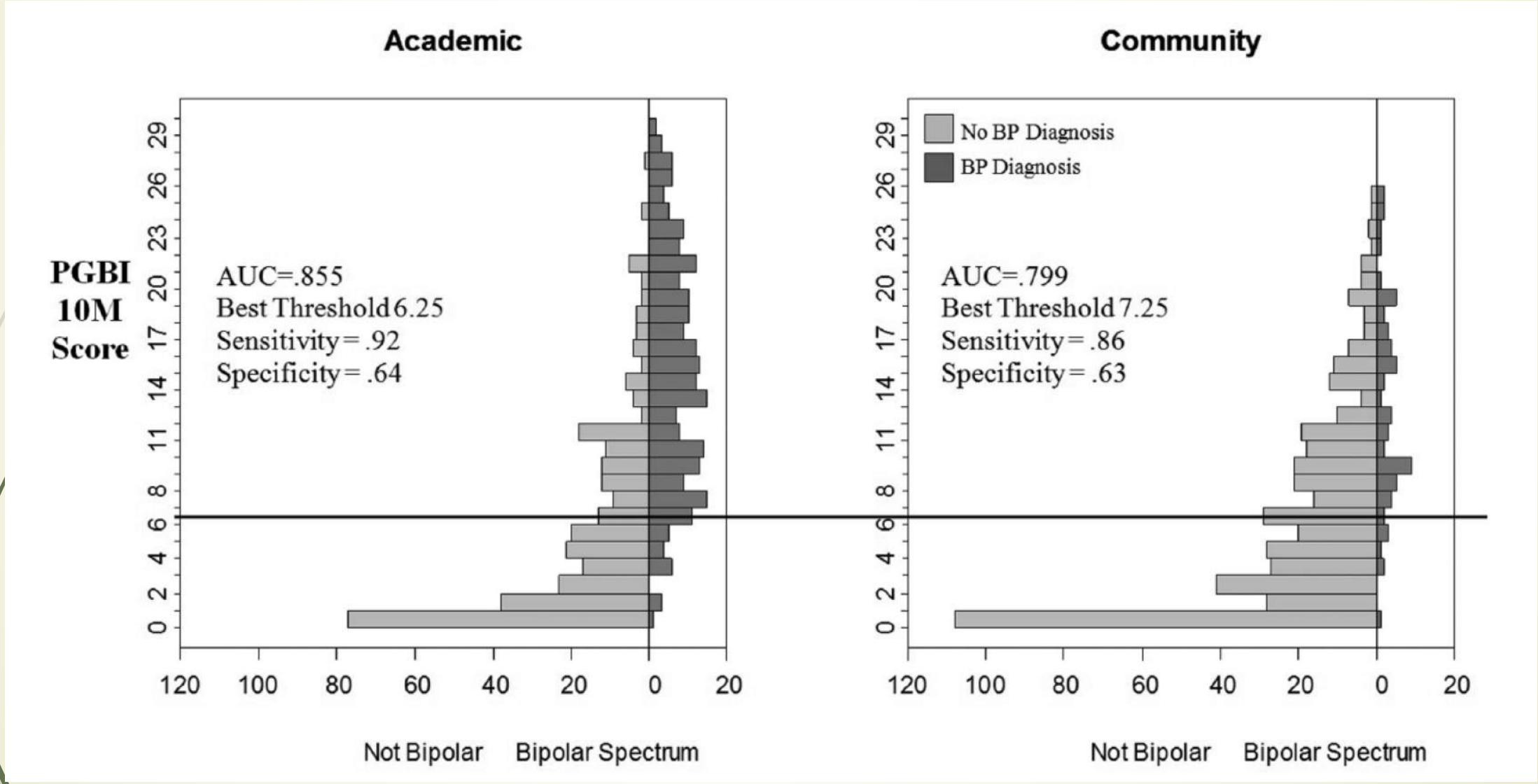
	Academic clinic (N = 550)	Community clinic (N = 511)	Effect size[a]
Youth demographics			
Male, % (n)	60% (217)	60% (205)	.01[n.s.]
Age, M (SD)	11.40 (3.23)	10.53 (3.41)	.26***
White, % (n)	79% (433)	6% (31)	.74***
Family income[b]	2.45 (1.21)	1.28 (0.64)	1.20***
Clinical characteristics			
Family history of bipolar	35% (194)	32% (165)	.03[n.s.]
YMRS	11.65 (11.86)	6.05 (8.41)	.54***
CDRS-R	35.49 (16.08)	29.95 (13.20)	.38***
PGBI10M	10.13 (7.88)	7.47 (6.35)	.37***
PGBI–hypo/biphasic	24.66 (16.84)	19.70 (14.22)	.32***
PGBI–depression	36.19 (25.67)	24.48 (21.49)	.49***
7 Up	5.16 (4.61)	4.11 (3.83)	.25***
7 Down	6.24 (5.28)	3.21 (4.04)	.64***
PGBI–sleep scale	5.87 (4.74)	4.06 (4.18)	.41***

Academic and Community Samples: Big differences in diagnoses

Table 2. Demographics and Clinical Characteristics by Clinic Setting

	Academic clinic (N = 550)	Community clinic (N = 511)	Effect size[a]
Number Axis I diagnoses	2.15 (1.34)	2.69 (1.38)	–.39***
Bipolar spectrum diagnosis	44% (241)	13% (65)	.34***
Any attention-deficit/hyperactivity	54% (295)	66% (338)	–.13***
Any oppositional defiant disorder	30% (167)	38% (196)	–.08**
Any conduct disorder	8% (44)	12% (61)	–.07*
Any anxiety disorder	8% (45)	27% (138)	–.25***
Any posttraumatic stress disorder	2% (11)	11% (54)	–.18***

Score distributions on PGBI-10M



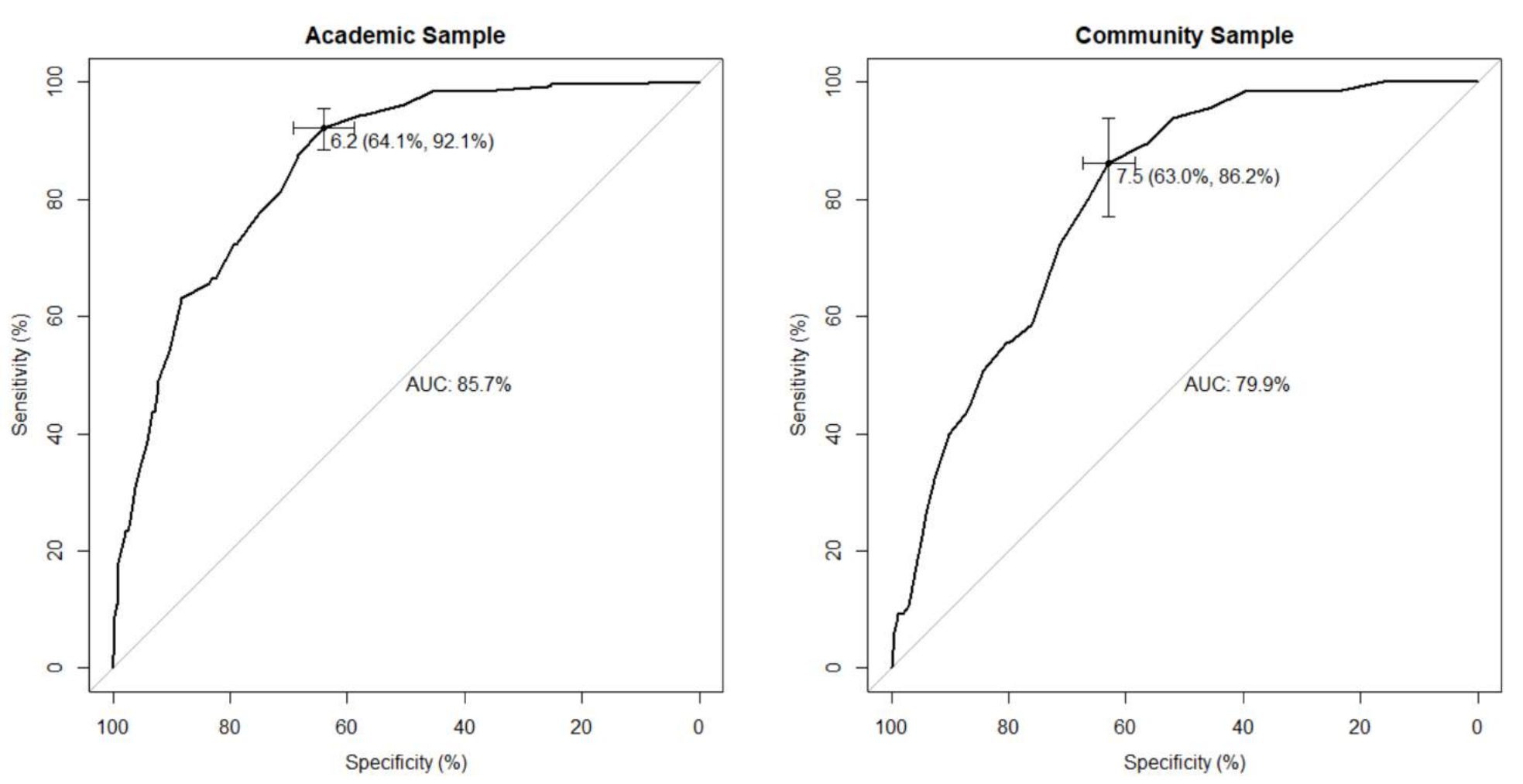
Round 1 results: Academic Clinic

Table 3. Accuracy Statistics for Discrimination (AUC) and Calibration (Spiegelhalter's z) for Internal Validation and Cross-Validation in an Academic Sample and External Cross-Validation in the Community Sample

	Academic sample (N = 550)	
Model	AUC	Spiegelhalter's z
Bet the base rate	.500 (.025)	$0.01^{n.s.}$
Take the best (dichotomize PGBI10M)	.781 (.020)	$-0.01^{n.s.}$
Nomogram	.781 (.020)	
Multilevel and two-variable nomogram	.882 (.014)	$0.19^{n.s.}$
Logistic regression (1 *df*)	.857 (.016)	$0.13^{n.s.}$
Logistic regression (5 *df*)	.890 (.014)	$-0.06^{n.s.}$
LASSO (136 candidates)	.902 (.013)	-3.72^{***}
Diagnosis upper limit	.925[a]	—

[a]The KSADS diagnosis kappa of .85 imposes an upper bound on the AUC (Kraemer, 1992).

PGBI-10M works in both clinics



Round 2 results: Community Clinic

Table 3. Accuracy Statistics for Discrimination (AUC) and Calibration (Spiegelhalter's z) for Internal Validation and Cross-Validation in an Academic Sample and External Cross-Validation in the Community Sample

	Academic sample (N = 550)	External cross-validation: Academic weights in community sample (N = 511)
Model	AUC	AUC
Bet the base rate	.500 (.025)	.500 (.038)
Take the best (dichotomize PGBI10M)	.781 (.020)	.729 (.029)
Nomogram	.781 (.020)	.729 (.029)
Multilevel and two-variable nomogram	.882 (.014)	.775 (.025)
Logistic regression (1 *df*)	.857 (.016)	.799 (.024)
Logistic regression (5 *df*)	.890 (.014)	.775 (.026)
LASSO (136 candidates)	.902 (.013)	.801 (.024)
Reversed LASSO (community weights)	.864 (.015)	.830 (.023)
Diagnosis upper limit	.925[a]	.925[a]

[a]The KSADS diagnosis kappa of .85 imposes an upper bound on the AUC (Kraemer, 1992).

Supplemental Table 1
LASSO models built in the Academic sample (*N*=550), in the Community sample predicting KSADS diagnoses (*N*=511), and in the Community sample predicting chart diagnoses (*N*=511).

	Academic		Community		Chart Diagnoses	
Variable	Min	1SE	Min	1SE	Min	1SE
Intercept	-3.28	-2.52	-3.32	-2.17	-3.40	-2.47
PGBI10M	**0.18**	**0.13**	**0.04**			
Family Bipolar History	0.25	0.79				
Number of Diagnoses	**0.38**	**0.19**	**0.13**			
Youth Age x Family Bipolar History	0.06					
Youth Age x PGBI Depression	0.00					
Youth Age x PTSD	0.07					
PGBI10M x Female	0.01					
PGBI10M x White	0.04	0.01				
Female x Number of Diagnoses	0.02					
Female x CD	-0.55					
Family Bipolar History x White	0.03					
Family Bipolar History x PGBI Hypo/Biphasic	0.01	0.00				
Family Bipolar History x ADHD	0.92	0.48				
Family Bipolar History x Anxiety	-1.10					
Family Bipolar History x PTSD	-0.85					
White x Number of Diagnoses	0.01					
White x Anxiety	**-0.06**		**-0.26**			
PGBI Sleep x PGBI Depression	0.00					
PGBI Sleep x ADHD	-0.01					
PGBI Depression x Anxiety	0.00					
ADHD x ODD	-0.41					
ADHD x CD	0.20					
ADHD x Anxiety	-0.23					
CD x Anxiety	0.03					
PGBI Sleep			0.02			
Youth Age x PGBI10M			0.00			
PGBI10M x Number of Diagnoses			0.01	0.01		
PGBI10M x CD			0.01			
PGBI Hypo/Biphasic x Number of Diagnoses			0.00			
Number of Diagnoses x PTSD			-0.06			
Youth Age x PGBI10M					5.00E-03	
Youth Age x Number of Diagnoses					9.47E-03	
PGBI10M x ODD					6.63E-03	
Female x PTSD					1.73E-01	
Family Bipolar History x ODD					3.22E-01	
White x PGBI Depression					2.35E-02	
White x PTSD					-6.93E-01	
PGBI Sleep x Number of Diagnoses					5.87E-06	
CD x Anxiety					5.41E-01	

Good news!
PGBI & Family History

Discovery!
White x Anxiety

But:
many more predictors in Academic than Community?

Just when you thought it was over… **ROUND 3!**

- What if we used billing diagnoses to train the model?

Genetic Testing for Health, Disease & Ancestry; DNA Test - 23andMe - Windows Internet Explorer

https://www.23andme.com/

Odum Qualtrics

File Edit View Favorites Tools Help

Favorites | Suggested Sites | Free Hotmail | Web Slice Gallery

Genetic Testing for Healt... | The Odum Institute for Rese...

Page | Safety | Tools

23andMe genetics just got personal.

Search 23andMe Go | Log in | Claim Codes | Blog | Help | Cart

welcome | ancestry | health | how it works | store

Choose the DNA test that's right for you.

Fill in your family tree.
Ancestry Edition, $399 Learn more
Buy Now

Take charge of your health.
Health Edition, $429 Learn more
Buy Now

Choose to have it all.
23andMe Complete, $499
Buy Now

Now only $299 8/23/12

Find a disease or trait that we cover:

News and Press

Scientific Resources and Principles

(Rapidly changing ethics and guidelines!

he University o | ROCR: Classifier Visualizat | How to Use Your Evernote | hBARSCI - Checkout | Welcome to 23andMe - 23

https://www.23andme.com/ancestry-only-notice/?redirect=aKPL41CQYzwY3JIV4c10RZIbRYDa1HDQzdIKBCI8yH4nFqI8biPQiI4rM

23andMe

welcome | health | ancestry | how it works | buy | search

Please read the following statement below and click the "I understand" button to enter 23andMe.com.

Welcome to 23andMe.

At this time, we have suspended our health-related genetic tests to comply with the U.S. Food and Drug Administration's directive to discontinue new consumer access during our regulatory review process.

We are continuing to provide you with both ancestry-related genetic tests and raw genetic data, without 23andMe's interpretation.

If you are an existing customer please click the button below and then go to the health page for additional information, including information about refunds.

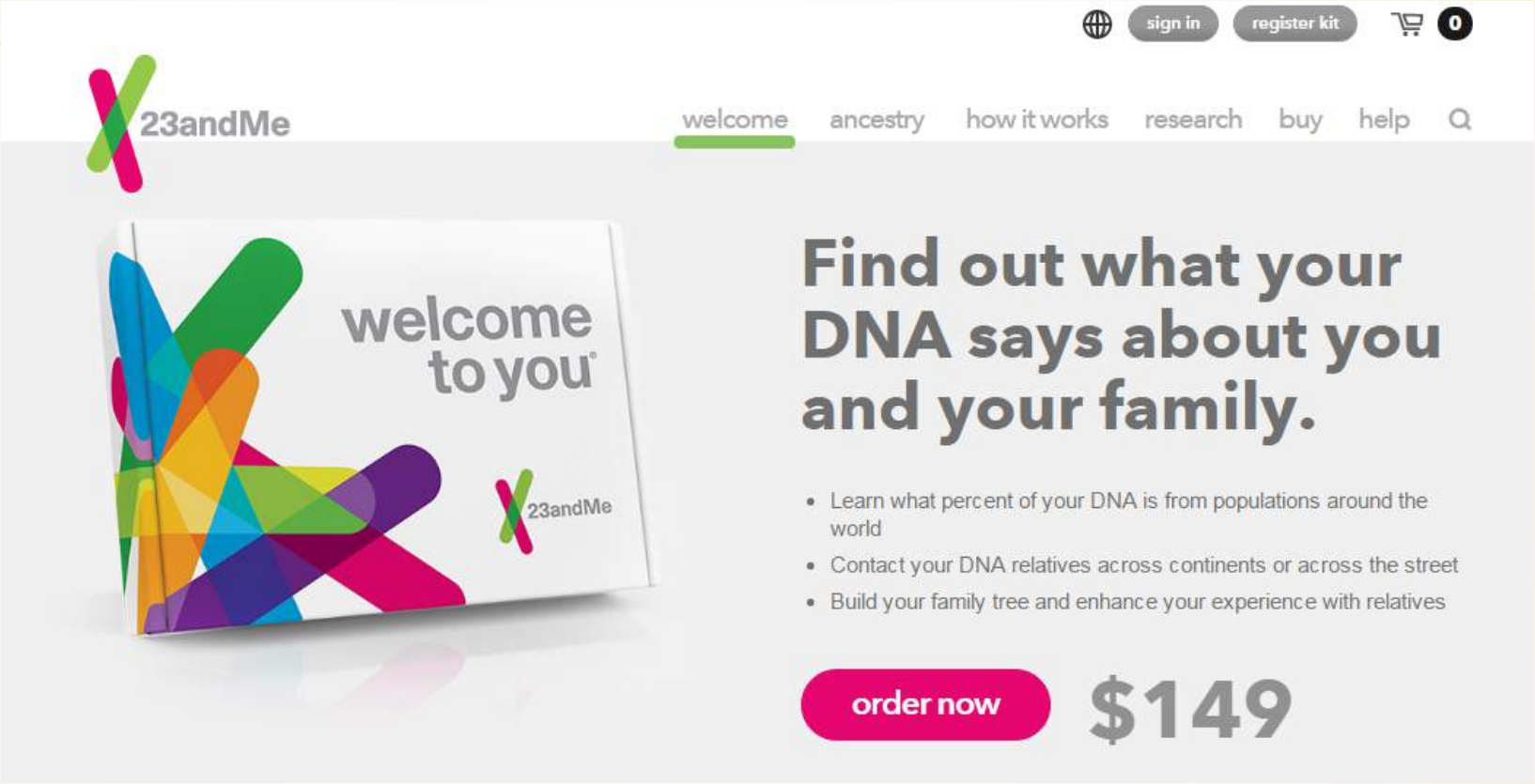
We remain firmly committed to fulfilling our long-term mission to help people everywhere have access to their own genetic data and have the ability to use that information to improve their lives.

Upon entering the site, please confirm you understand the new changes in our services.

I understand that 23andMe only sells ancestry reports and raw genetic data at this time. I understand 23andMe will not provide health-related reports. However, 23andMe may provide health-related results in the future, dependent upon FDA marketing authorization.

I UNDERSTAND

June 2016 Version
(with 20% discount for additional people in 2017)



Google's NIH steal Tom Insel on the 'major paradigm shift' of digitizing mental health care

By MEGHANA KESHAVAN

1 Comment / 213 Shares / Sep 18, 2015 at 3:01 PM

Supplemental Table 1

LASSO models built in the Academic sample (*N*=550), in the Community sample predicting KSADS diagnoses (*N*=511), and in the Community sample predicting chart diagnoses (*N*=511).

	Academic		Community		Chart Diagnoses	
Variable	Min	1SE	Min	1SE	Min	1SE
Intercept	-3.28	-2.52	-3.32	-2.17	-3.40	-2.47
PGBI10M	**0.18**	**0.13**	**0.04**			
Family Bipolar History	0.25	0.79				
Number of Diagnoses	**0.38**	**0.19**	**0.13**			
Youth Age x Family Bipolar History	0.06					
Youth Age x PGBI Depression	0.00					
Youth Age x PTSD	0.07					
PGBI10M x Female	0.01					
PGBI10M x White	0.04	0.01				
Female x Number of Diagnoses	0.02					
Female x CD	-0.55					
Family Bipolar History x White	0.03					
Family Bipolar History x PGBI Hypo/Biphasic	0.01	0.00				
Family Bipolar History x ADHD	0.92	0.48				
Family Bipolar History x Anxiety	-1.10					
Family Bipolar History x PTSD	-0.85					
White x Number of Diagnoses	0.01					
White x Anxiety	**-0.06**		**-0.26**			
PGBI Sleep x PGBI Depression	0.00					
PGBI Sleep x ADHD	-0.01					
PGBI Depression x Anxiety	0.00					
ADHD x ODD	-0.41					
ADHD x CD	0.20					
ADHD x Anxiety	-0.23					
CD x Anxiety	0.03					
PGBI Sleep			0.02			
Youth Age x PGBI10M			0.00			
PGBI10M x Number of Diagnoses			0.01	0.01		
PGBI10M x CD			0.01			
PGBI Hypo/Biphasic x Number of Diagnoses			0.00			
Number of Diagnoses x PTSD			-0.06			
Youth Age x PGBI10M					5.00E-03	
Youth Age x Number of Diagnoses					9.47E-03	
PGBI10M x ODD					6.63E-03	
Female x PTSD					1.73E-01	
Family Bipolar History x ODD					3.22E-01	
White x PGBI Depression					2.35E-02	
White x PTSD					-6.93E-01	
PGBI Sleep x Number of Diagnoses					5.87E-06	
CD x Anxiety					5.41E-01	

Conclusions

- Naïve Bayesian approaches (even nomogram) would be a big step forward
- They generalize better than expected
  - Can include local rates, information
- LASSO, etc.
  - More accurate in training sample
  - External validity is a big hurdle
  - need more implementation support

Wikipedia: "Best of the Free" Assessments

- Write pages for free use tools that have good score psychometrics across samples
  - Link to copies of measures
  - Solves Awareness and Access issues
  - Supported by grants from SCCAP, APS, SSCP, APA CODAPAR & D12
- http://hgaps.org

Free Evidence-based Assessments (and embed the interpretation)

DBSA
Depression and Bipolar Support Alliance

Crisis | Donate | Newsletter Sign-up | Search... | Go

We've been there. We can help.

EDUCATION info, training, events | WELLNESS OPTIONS treatment, tools, research | PEER SUPPORT peer groups, inspiration | HELP OTHERS family, friends, peers | ABOUT DBSA who we are

Like 56 | Tweet | Pin it | G+1 1 | Share

Mental Health Screening Center

These online screening tools are not a substitute for consultation with a health professional. Regardless of the results of a screen, if you have any concerns, see your doctor or mental health professional.

Depression

This screening form was developed from the Quick Inventory of Depressive Symptomatology Self-Report (QIDS-SR) which is in the public domain.
Take a confidential online depression screening

Anxiety

The Hamilton Anxiety Scale was authored by Max Hamilton. It is in the public domain.
Take a confidential online anxiety screening

Mania

Used with permission. This screening form was developed from the Mood Disorder Questionnaire (MDQ)
Take a confidential online mania screening

Childhood Mania

The Child Mania Rating Scale (CMRS) is a parent screening instrument for mania based on DSM-IV

EDUCATION

Mood Disorders
- Depression
- Bipolar Disorder
- Anxiety
- Screening Center
- Co-occurring Illnesses/Disorders
- Related Concerns

Educational Materials
- Brochures
- Podcasts
- Publications
- Videos
- Living Successfully Course

- Also on Wikipedia & Wikiversity
- Can help us frame the feedback & suggest resources:
- http://tinyurl.com/ebafeedback

Thank You!