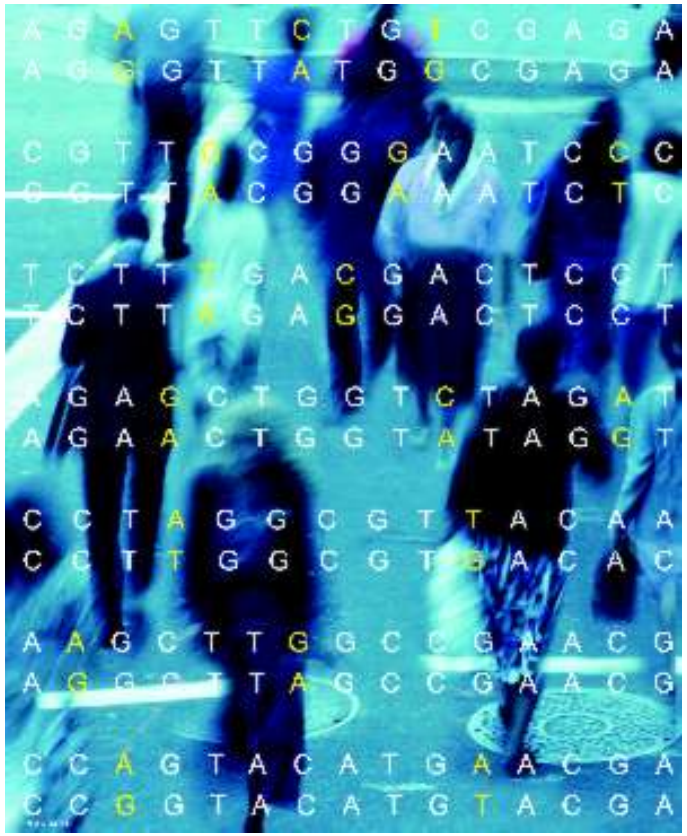
UNIVERSITY of CALIFORNIA, SAN DIEGO
SCHOOL OF MEDICINE

Computer Models for Medical Diagnosis and Prognostication

Lucila Ohno-Machado, MD, PhD
Division of Biomedical Informatics

- Clinical pattern recognition and predictive models
- Evaluation of binary classifiers (calibration)
- Ethical implications for clinical practice

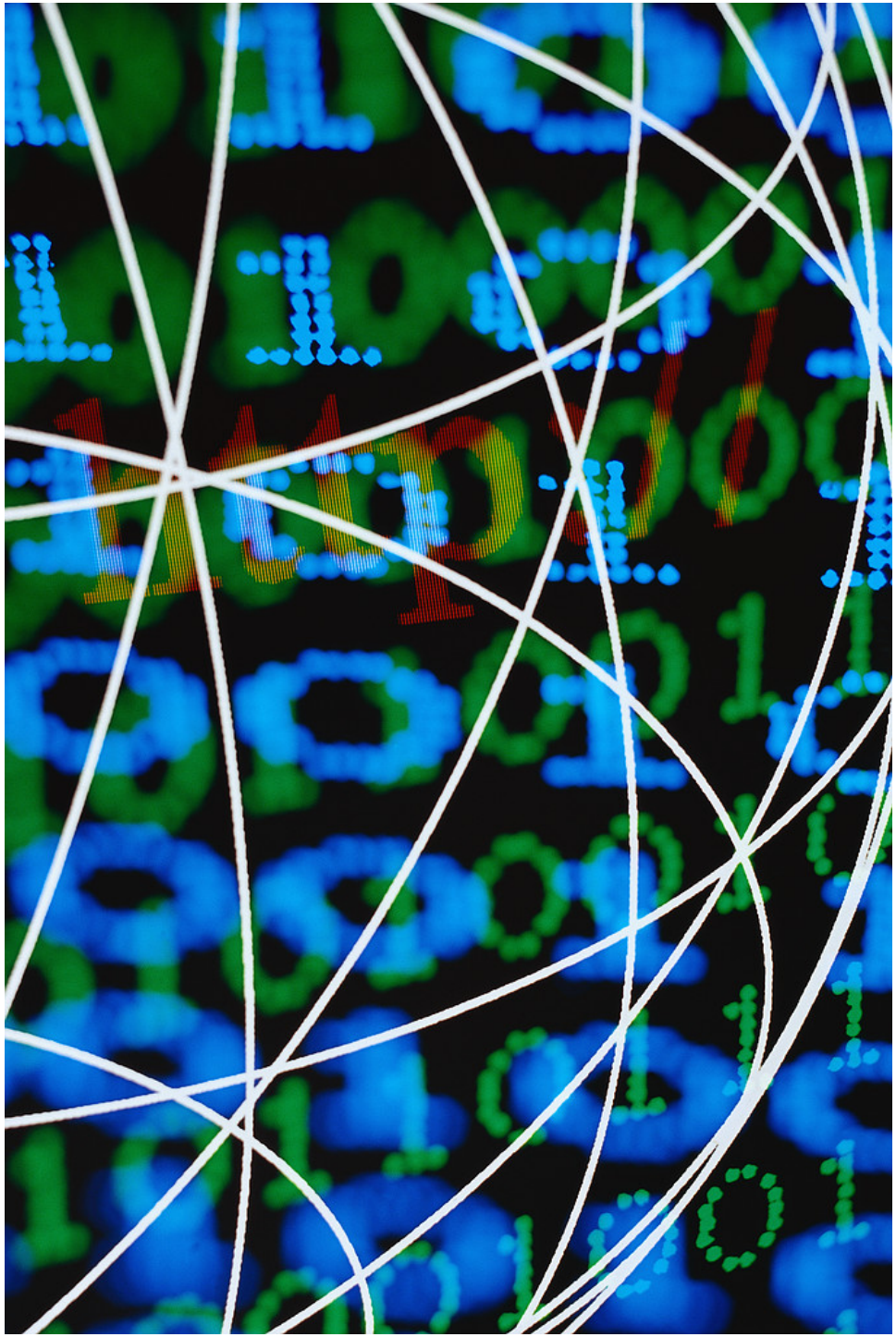


Source: DOE

Computer-Based Models for Medical Diagnosis and Prognostication

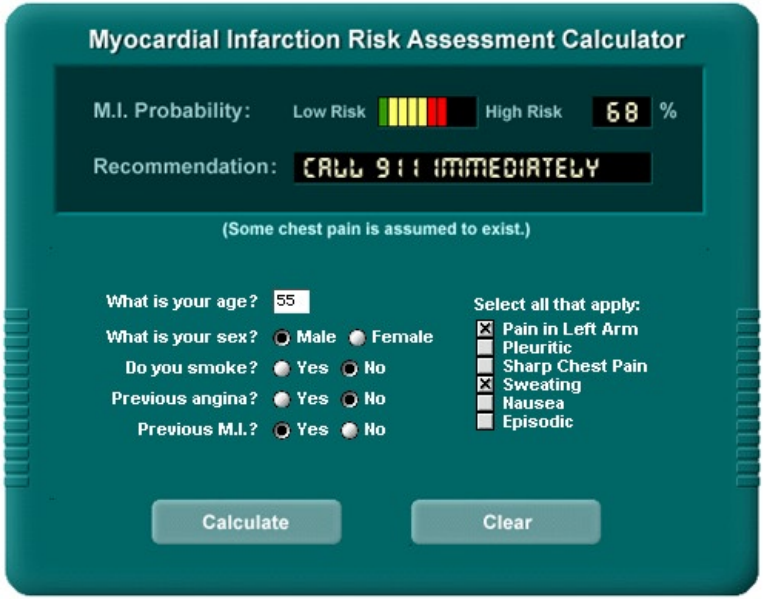
Lucila Ohno-Machado, MD, PhD

Division of Biomedical Informatics
University of California San Diego



Risk Assessment

- Popular risk calculators
  - Gail Model (Breast cancer)
  - Framingham Risk Calculator (CVD)
  - APACHE (ICU mortality)



- Use of individual estimates
  - Prophylaxis for breast cancer
  - Cholesterol management guidelines
  - *Continuation of life support*

Results - BCRA - Windows Internet Explorer

http://www.cancer.gov/bcrisktool/RiskAssessment.aspx?current_age=42&age_at_menarche=10&age_at_first_live_l

Live Search

Norton Safe Search | Search | Cards & Log-ins

Results - BCRA

Page

National Cancer Institute

U.S. National Institutes of Health | www.cancer.gov

Breast Cancer Risk Assessment Tool

An Interactive Tool For Measuring the Risk of Invasive Breast Cancer

- Risk Calculator
- About the Tool
- Breast Cancer Risk
- Mobile Access
- Download Source Code

Page Options

- Print Page
- Email Page

Quick Links

- Breast Cancer Home Page
- Breast Cancer: Prevention, Genetics, Causes
- Initial Results of STAR Released

The Breast Cancer Risk Assessment Tool is an interactive tool designed by scientists at the National Cancer Institute (NCI) and the National Surgical Adjuvant Breast and Bowel Project (NSABP) to estimate a woman's risk of developing invasive breast cancer. The tool has been updated for African American women based on the Contraceptive and Reproductive Experiences (CARE) Study. See About the Tool for more information.

Results (Breast Cancer Risk)

New Risk Calculation

Reminder: *The Breast Cancer Risk Assessment Tool was designed for use by health professionals. If you are not a health professional, you are encouraged to discuss these results and your personal risk of breast cancer with your doctor.*

Race/Ethnicity:

White

5 Year Risk

- This woman (age 42) 3.1%
- Average woman (age 42): 0.7%

Explanation

Based on the information provided (see below), the woman's estimated risk for

Initial Results of STAR Released

Current Clinical Trials: Breast Cancer *In Situ*: Treatment

Current Clinical Trials: Breast Cancer Prevention

Current Clinical Trials: Breast Cancer Screening

Estimating Breast Cancer: Q&A

Understanding Cancer Risk

National Cancer Institute

Need Help?
Contact us by phone, Web, and e-mail
1-800-4-CANCER

Explanation

Based on the information provided (see below), the woman's estimated risk for developing invasive breast cancer over the next 5 years is 3.1% compared to a risk of 0.7% for a woman of the same age and race/ethnicity from the general U.S. population. This calculation also means that the woman's risk of NOT getting breast cancer over the next 5 years is 96.9%.

Lifetime Risk

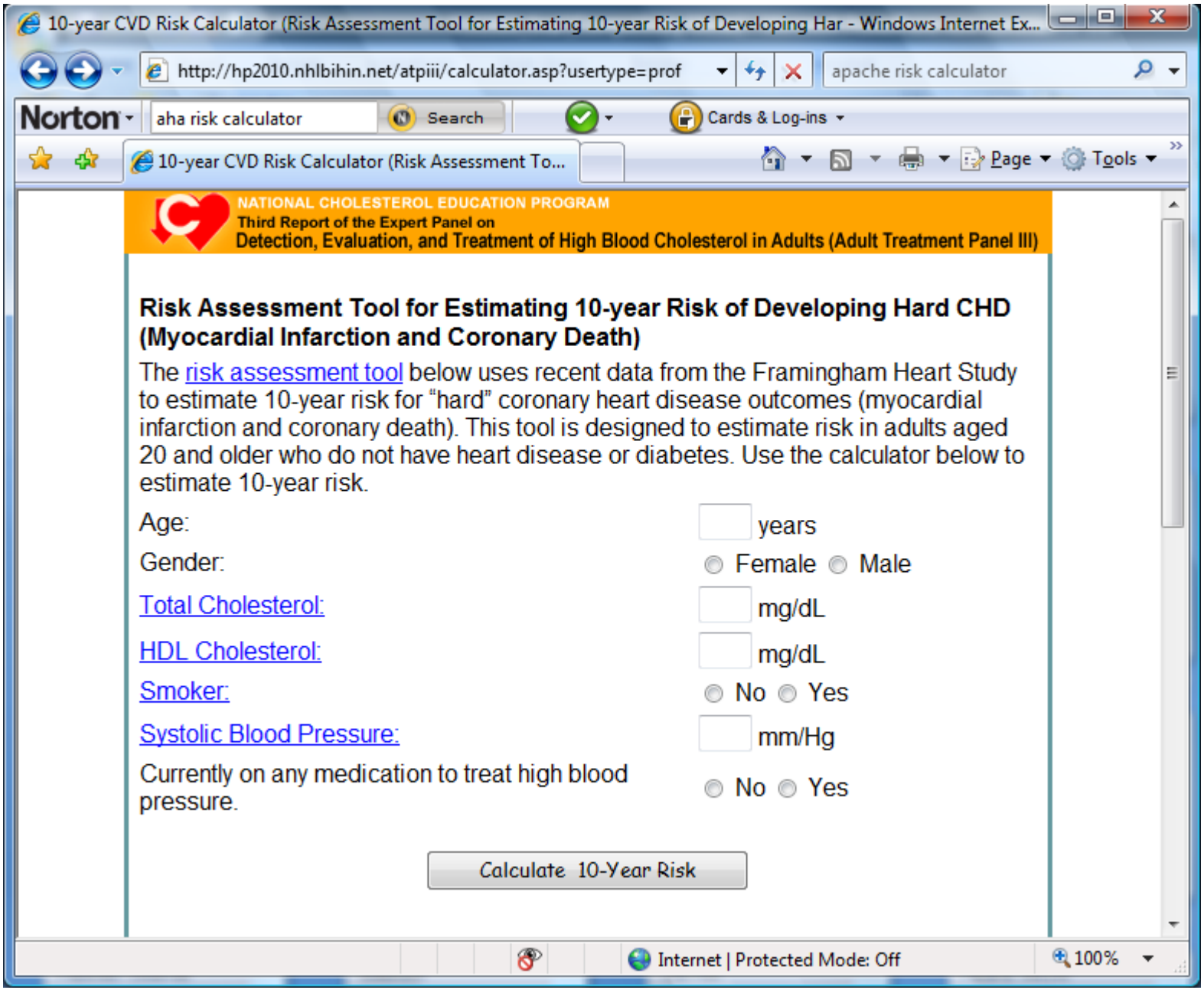
- This woman (to age 90): 31.1%
- Average woman (to age 90): 12.2%

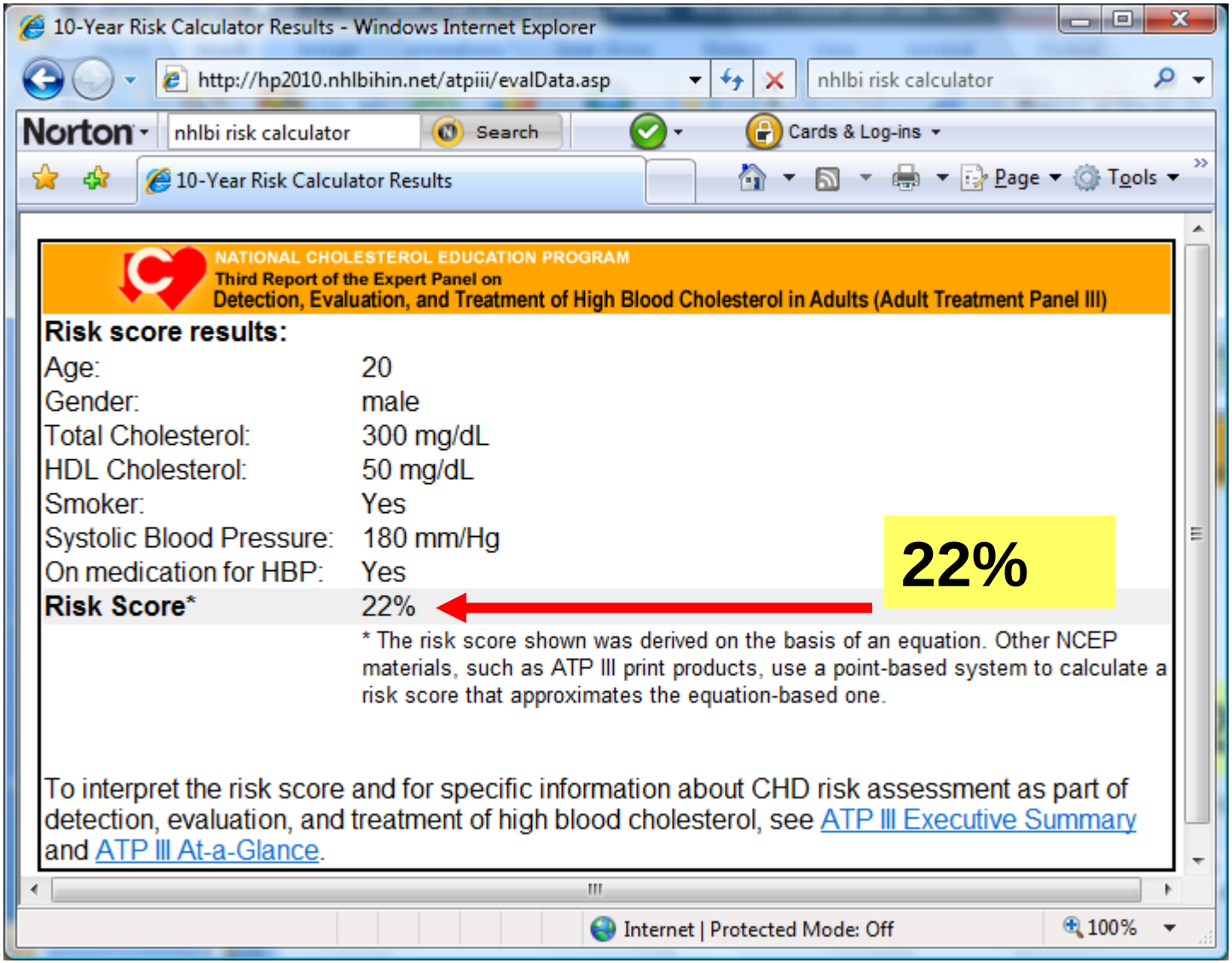
Explanation

Based on the information provided (see below), the woman's estimated risk for developing invasive breast cancer over her lifetime (to age 90) is 31.1% compared to a risk of 12.2% for a woman of the same age and race/ethnicity from the general U.S. population.

These results are based upon the following answers:

Does the woman have a medical history of any breast cancer or of ductal carcinoma in situ (DCIS) or lobular carcinoma in situ (LCIS)?	No
What is the woman's age?	42
What was the woman's age at the time of her first menstrual period?	7 to 11
What was the woman's age at the time of her first live birth of a child?	> =30


10-year CVD Risk Calculator (Risk Assessment Tool for Estimating 10-year Risk of Developing Har - Windows Internet Ex...
http://hp2010.nhlbihin.net/atpiii/calculator.asp?usertype=prof
apache risk calculator
Norton
aha risk calculator
Search
Cards & Log-ins
10-year CVD Risk Calculator (Risk Assessment To...
Page
Tools
NATIONAL CHOLESTEROL EDUCATION PROGRAM
Third Report of the Expert Panel on
Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III)
Risk Assessment Tool for Estimating 10-year Risk of Developing Hard CHD (Myocardial Infarction and Coronary Death)
The risk assessment tool below uses recent data from the Framingham Heart Study to estimate 10-year risk for "hard" coronary heart disease outcomes (myocardial infarction and coronary death). This tool is designed to estimate risk in adults aged 20 and older who do not have heart disease or diabetes. Use the calculator below to estimate 10-year risk.
Age: years
Gender: Female Male
Total Cholesterol: mg/dL
HDL Cholesterol: mg/dL
Smoker: No Yes
Systolic Blood Pressure: mm/Hg
Currently on any medication to treat high blood pressure. No Yes
Calculate 10-Year Risk
Internet | Protected Mode: Off
100%


10-Year Risk Calculator Results - Windows Internet Explorer
http://hp2010.nhlbihin.net/atpiii/evalData.asp
nhlbi risk calculator
Norton
nhlbi risk calculator
Search
Cards & Log-ins
10-Year Risk Calculator Results
Page
Tools
NATIONAL CHOLESTEROL EDUCATION PROGRAM
Third Report of the Expert Panel on
Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III)
Risk score results:
Age: 20
Gender: male
Total Cholesterol: 300 mg/dL
HDL Cholesterol: 50 mg/dL
Smoker: Yes
Systolic Blood Pressure: 180 mm/Hg
On medication for HBP: Yes
Risk Score* 22%
22%
* The risk score shown was derived on the basis of an equation. Other NCEP materials, such as ATP III print products, use a point-based system to calculate a risk score that approximates the equation-based one.
To interpret the risk score and for specific information about CHD risk assessment as part of detection, evaluation, and treatment of high blood cholesterol, see ATP III Executive Summary and ATP III At-a-Glance.
Internet | Protected Mode: Off
100%

Heart Attack Risk Calculator v2.1.0.75 - Beta - Windows Internet Explorer

https://www.americanheart.org/gglRisk/locale/en_US/index.html?gtype=health

Heart Attack Risk Calculator - Beta

American Heart Association. Learn and Live.

Results

Follow the directions below. Click the "?" buttons for more information.

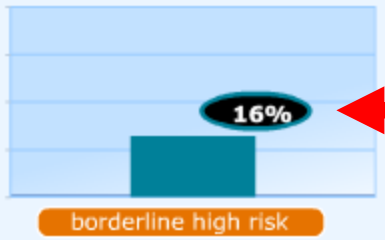
Your 10-Year Heart Attack Risk Estimate

Register | Log In

You can control your "modifiable" risk factors.

Click each tab to view your modifiable risk factors, including the "risk category" for each.

HEART ATTACK RISK (from current risk factors)



The graph shows your estimated risk of having a heart attack or dying of heart disease within the next 10 years.

Your risk is determined by y current *risk factors*, like age smoking.

16%

You can control some risk factors through lifestyle changes or medications.

These risk factors are called "modifiable" risk factors.

Cholesterol | Blood Pressure | Smoking

Modifiable Risk Factors

	CURRENT RISK FACTORS
Total cholesterol (mg/dL)	300 high risk
HDL (good) cholesterol (mg/dL)	50 low risk

In general, *lowering* your risk factor levels will *lower* your projected heart attack risk estimate. *However, the HDL (good) cholesterol risk factor is the exception.*

What should I do to lower my risk?

- [x] Lower blood pressure (systolic)
- [x] Lower total cholesterol
- [ ] Increase HDL ("good") cholesterol
- [x] Stop smoking

This tool uses a scoring system from the Framingham Heart Study to assess your risk for heart attack or dying of coronary heart disease in the next 10 years. It is important to discuss your situation with your doctor.

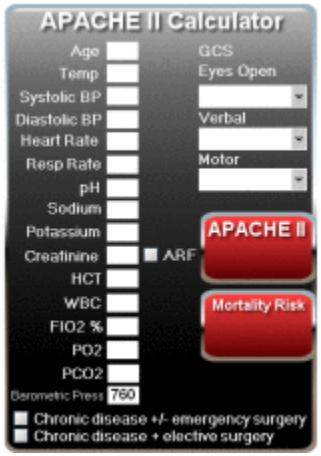
Exit | Informational Library | My History

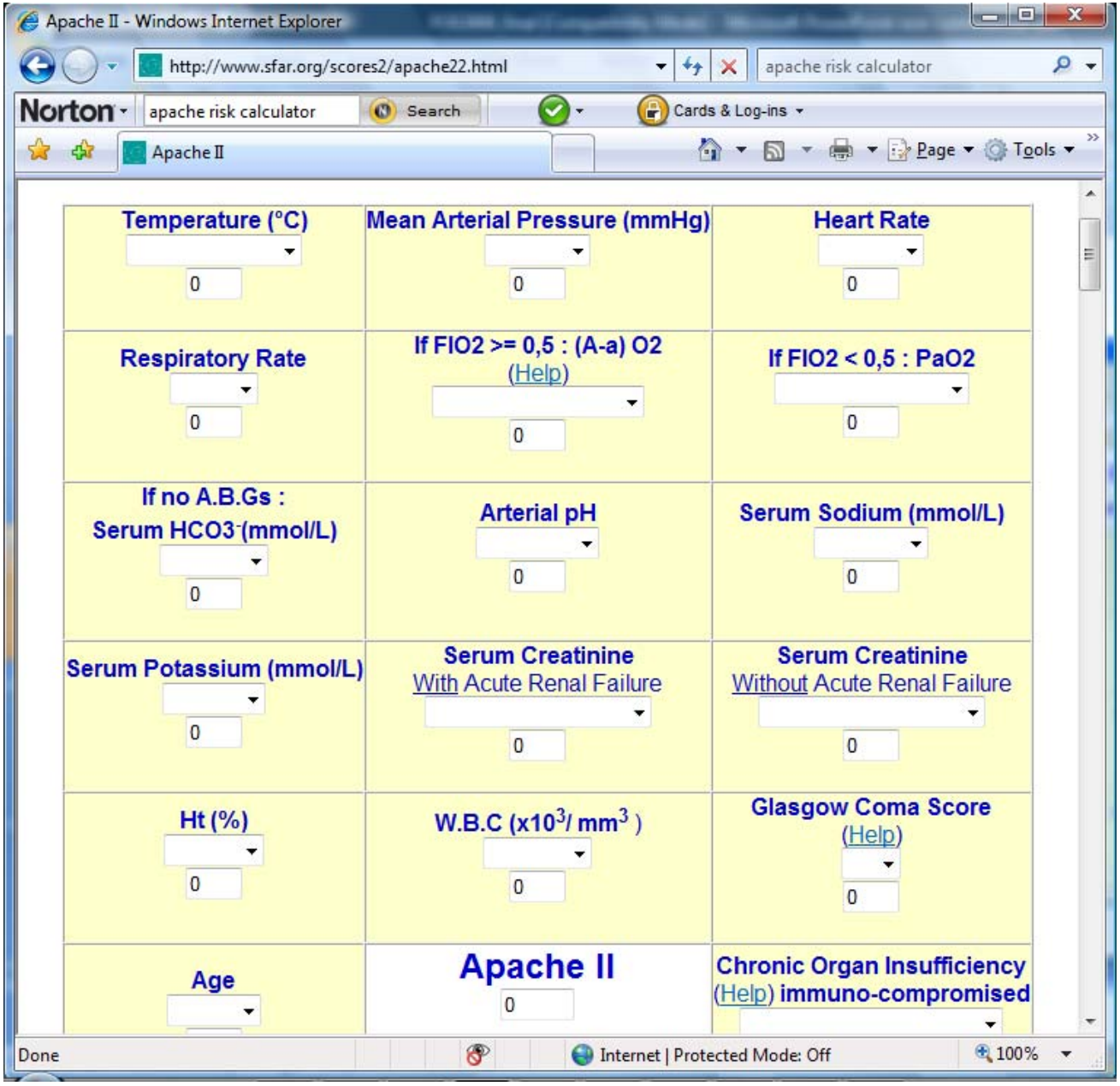
Previous 7/10 Next

Done | Internet | Protected Mode: Off | 100%

APACHE II

- Mortality in intensive care units (ICUs)
- 12 physiologic predictors





Every Genome - Personal Genome Sequencing Service FAQs - Illumina, Inc. - Windows Internet Explorer

http://www.everygenome.com/getting_started/faqs_genome_sequence.ilmn

Live Search

Norton Safe Search | Search | Cards & Log-ins

Every Genome - Personal Genome Sequencing Se...

Page

illumina

About Illumina | Personal Genomics 101 | Getting Started | Resources | Home

Getting Started

- Receiving a Genome Service
- What Can I Learn?
- Getting a Genotyping Service
- Getting a Genome Sequence
  - Our Process
  - Process FAQs
- Find a Doctor

Contact Us

Getting a personal genome sequencing service: FAQs.

Click topics below to expand or get all the answers.

How will I receive my personal genome sequence?

A computer loaded with your genome sequence data and a visual display program will be sent to your doctor, who will deliver it to you and go over the results with you. This program will allow you to view the sequence and to see the positions of important features, including genes, single nucleotide polymorphisms (SNPs) and other sources of genetic variation.

How often should I review my personal genome sequence data with my doctor to take account of ongoing scientific discoveries and medical recommendations?

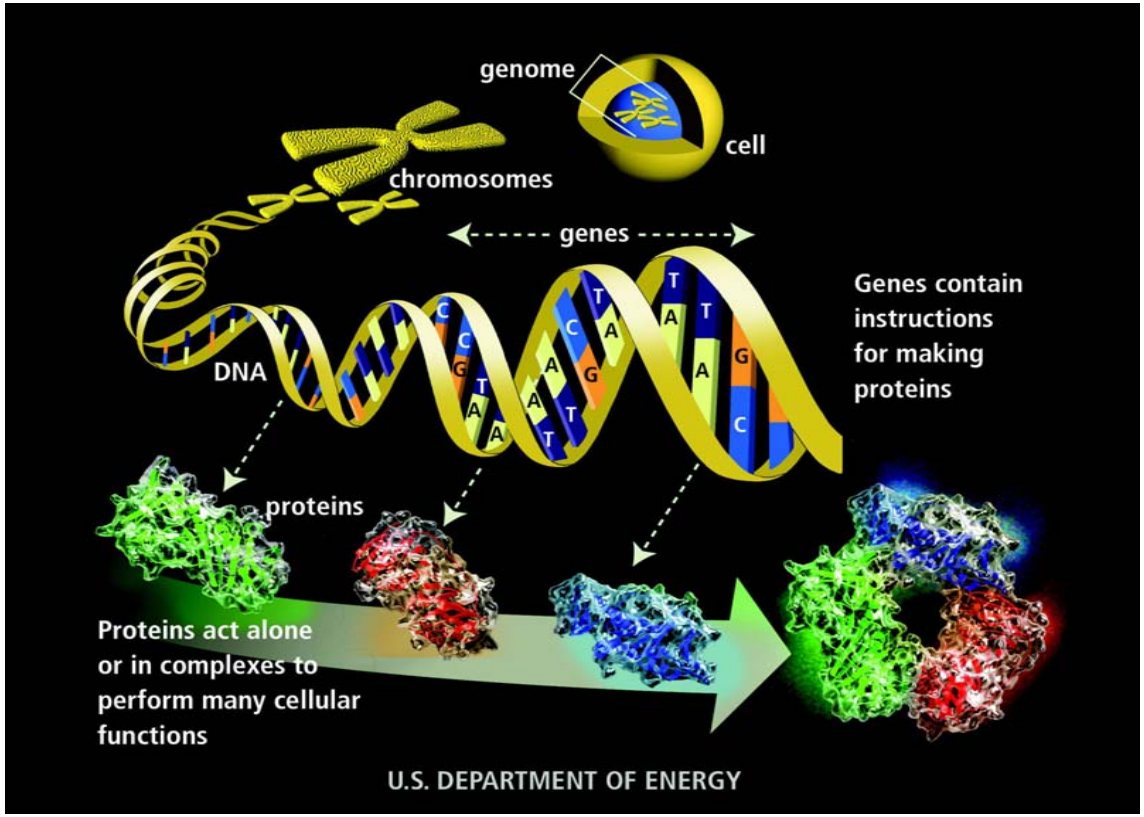
Plans for periodic review of personal genome sequence data should be discussed with your doctor as part of your consultation once the data are delivered. You may want to schedule an annual review as part of your overall care plan.

How will Illumina protect my privacy?

When you submit a sample, your sample will be labeled with a random ID number. Your de-identified sample will be sent to the Illumina laboratory generating your personal genome sequence information, and your signed consent form will be sent separately to specially designated individuals at Illumina who are carefully screened off from the rest of the organization and subject to the strictest of confidentiality obligations. These individuals will have access only to certain personal information (such as your name, address, and account number), but will not have access to your sample and personal genome sequence. Similarly, the Illumina laboratory generating your personal genome sequence information will not have

Individualized Genome

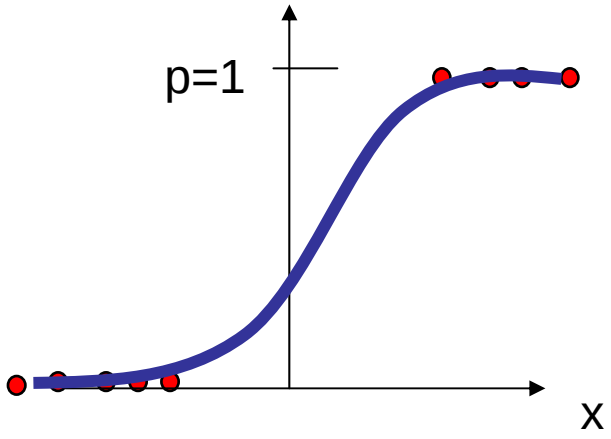
- How many individual genotypes are needed to predict disease?



Logistic Regression

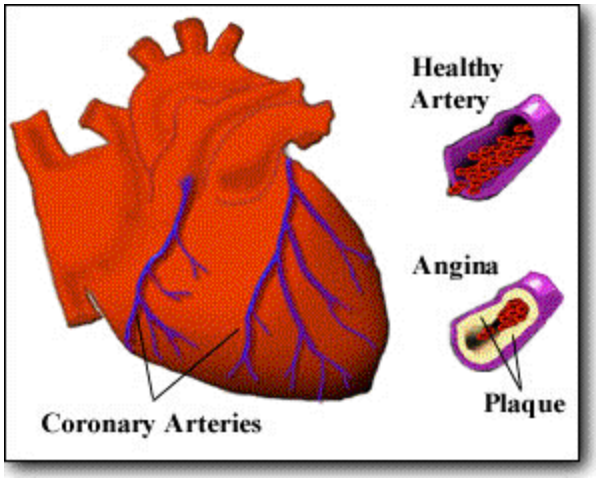
$$p_i = \frac{1}{1+e^{-(\beta_0+\beta_i x_i)}}$$

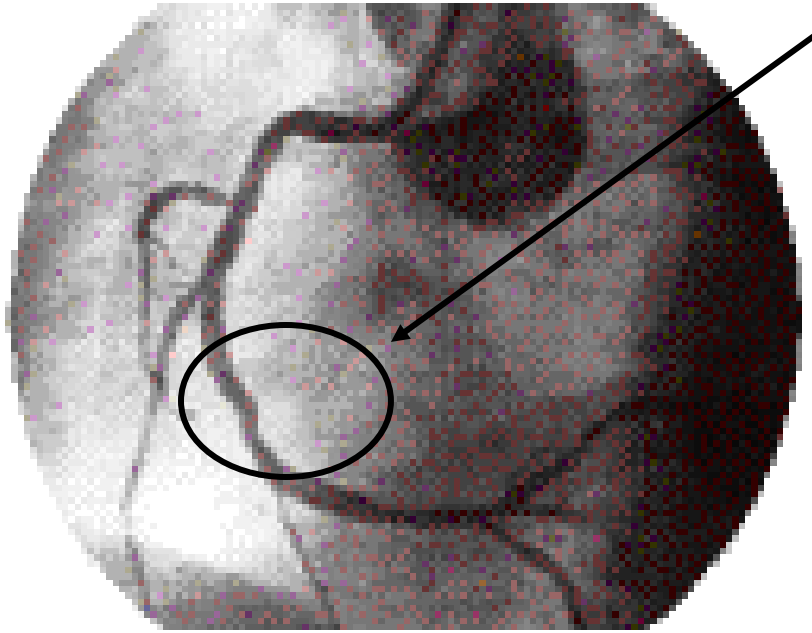
$$\log\left[\frac{p_i}{1-p_i}\right] = \beta_0 + \beta_i x_i$$



x

Coronary Angioplasty and Stenting





Risk of death in angioplasty

National average of deaths after angioplasty is 2%, which is stated in the informed consent.

"Informed consent and good clinical practice require a discussion of risks and benefits…"

Alexander et al, 52th ACC meeting

Less than 10% of the patients have an estimated risk of death around 2%.

Are we lying to the other 90%?

Dataset: Attributes Collected

History	Presentation	Angiographic	Procedural	Operator/Lab
age	acute MI	occluded	number lesions	annual volume
gender	primary	lesion type (A,B1,B2,C)	multivessel	device experience
diabetes	rescue	graft lesion	number stents	daily volume
iddm	CHF class	vessel treated	stent types (8)	lab device experience
history CABG	angina class	ostial	closure device	unscheduled case
baseline creatinine	cardiogenic shock		gp 2b3a antagonists	
CRI	failed CABG		dissection post	
ESRD			rotablator	
hyperlipidemia			atherectomy	
			angiojet	
			max pre stenosis	
			max post stenosis	
			no reflow	

Data Source:
- Medical Record
- Clinician Derived
- Other

Resnic et al, J Am Col Card 2001; Matheny et al, J Biomed Inf 2005

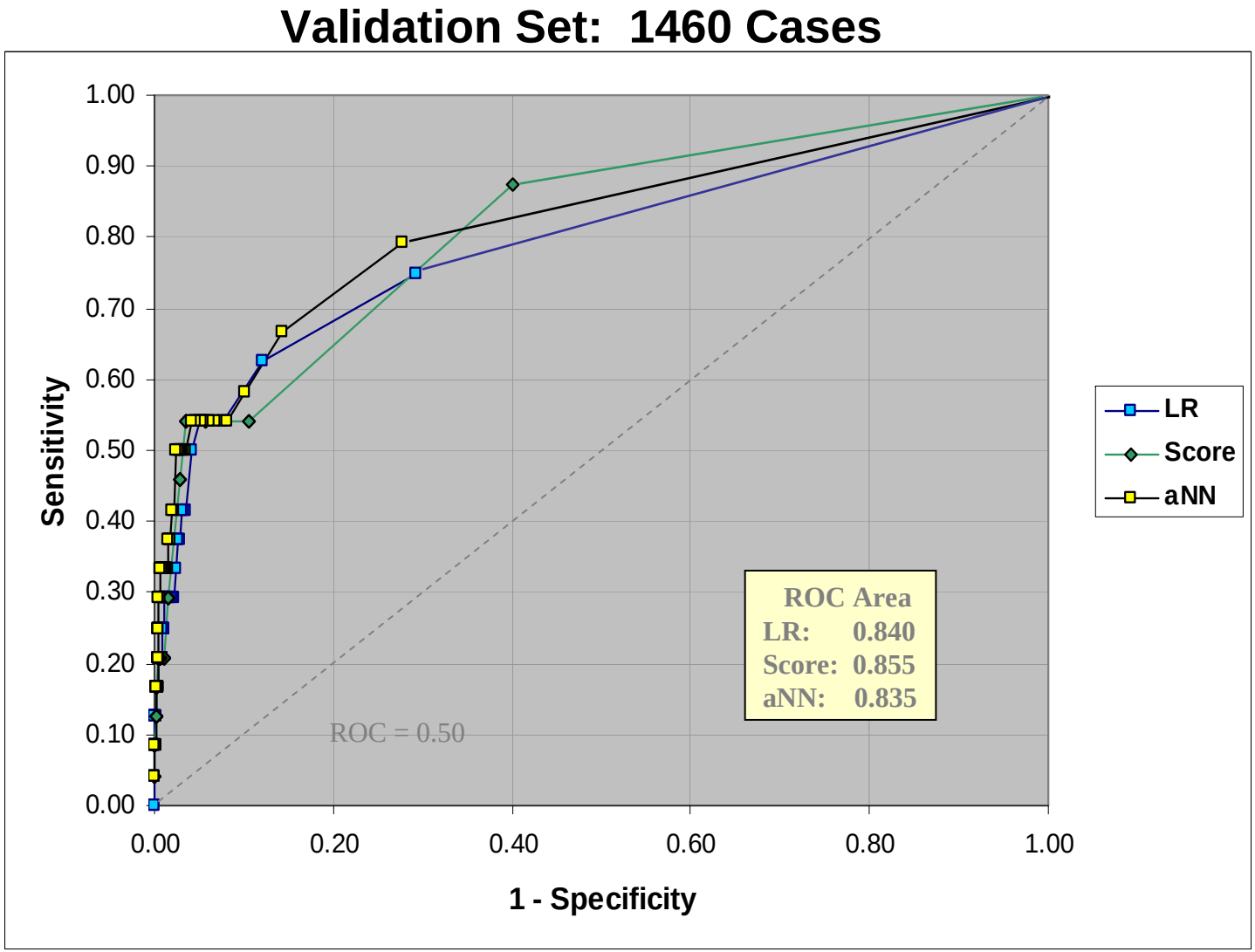
Study Population: Descriptive Statistics

	Development Set	Validation Set	
Cases	2,804	1,460	
Women	909 (32.4%)	433 (29.7%)	p=.066
Age > 74yrs	595 (21.2%)	308 (22.5%)	p=.340
Acute MI	250 (8.9%)	144 (9.9%)	p=.311
Primary	156 (5.6%)	95 (6.5%)	p=.214
Shock	62 (2.2%)	20 (1.4%)	p=.058
Class 3/4 CHF	176 (6.3%)	80 (5.5%)	p=.298
gp IIb/IIIa antagonist	1,005 (35.8%)	777 (53.2%)	p<.001
Death	67 (2.4%)	24 (1.6%)	p=.110
Death, MI, CABG (MACE)	177 (6.3%)	96 (6.6%)	p=.739

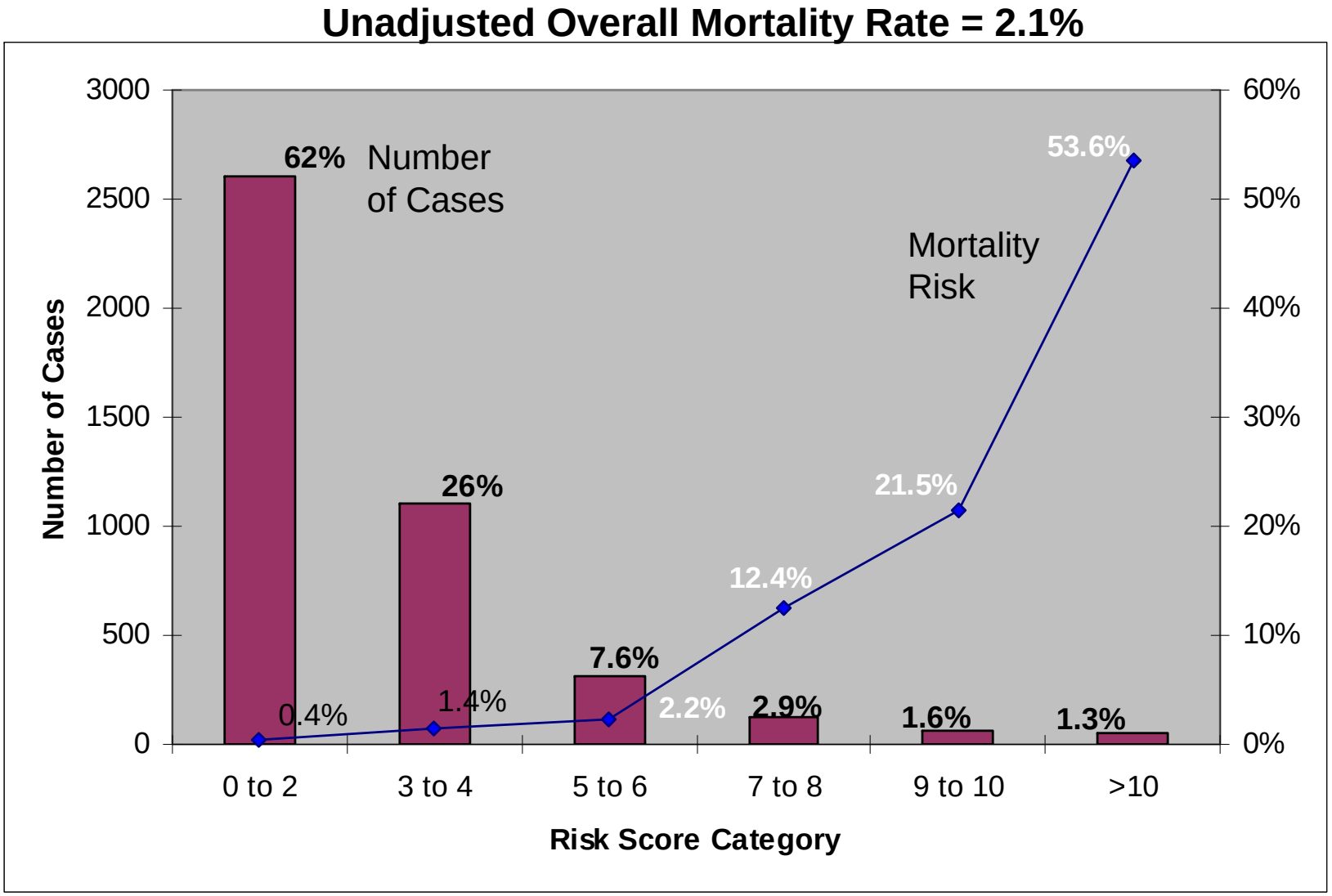
Multivariate Models

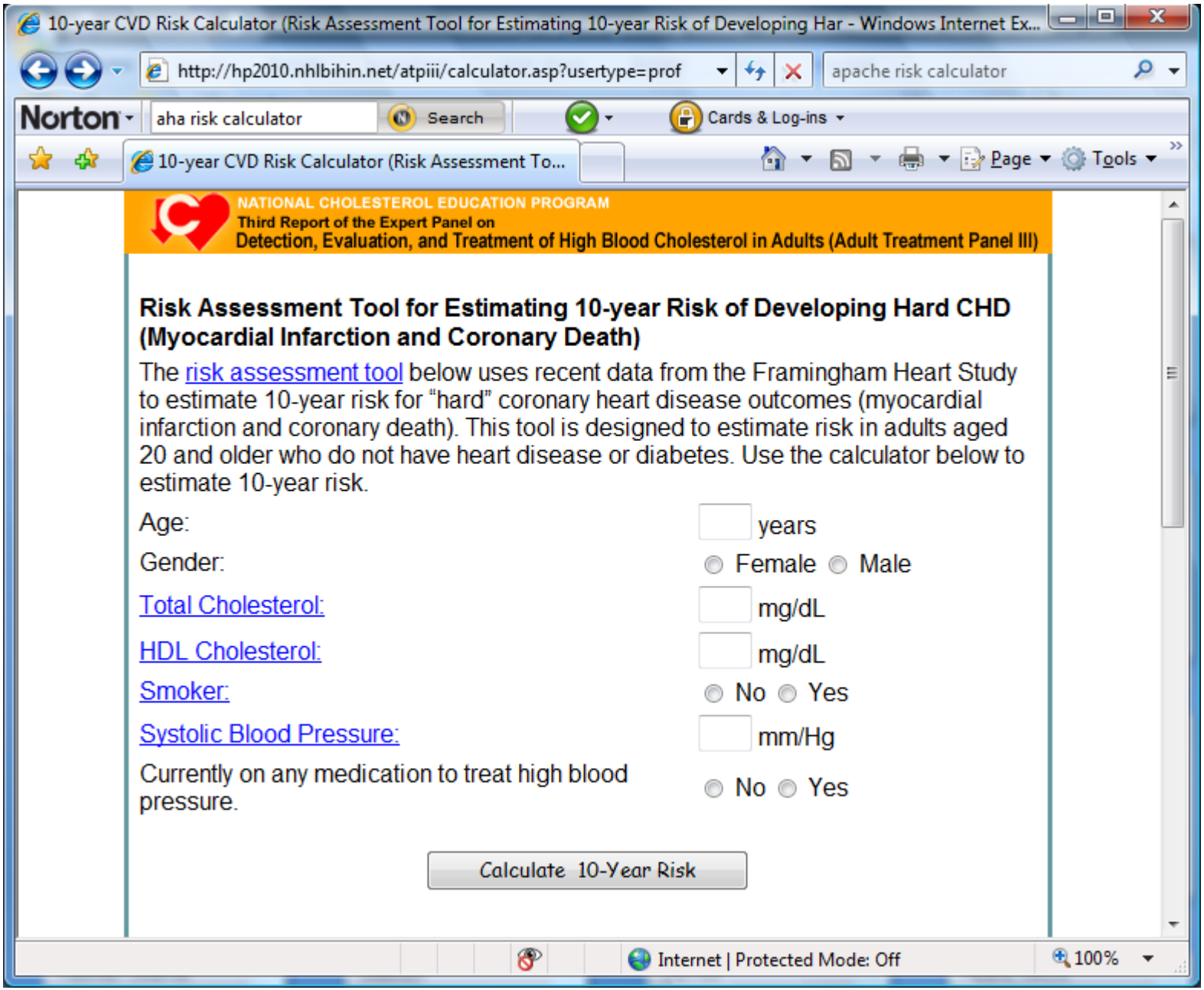
	Logistic Regression Model		Prognostic Risk Score Model		Artificial Neural Network
	Odds Ratio	p-value	beta coefficient	Risk Value	
Age > 74yrs	2.51	0.02	0.921	2	
B2/C Lesion	2.12	0.05	0.752	1	
Acute MI	2.06	0.13	0.724	1	
Class 3/4 CHF	8.41	0.00	2.129	4	
Left main PCI	5.93	0.03	1.779	3	
IIb/IIIa Use	0.57	0.20	-0.554	-1	
Stent Use	0.53	0.12	-0.626	-1	
Cardiogenic Shock	7.53	0.00	2.019	4	
Unstable Angina	1.70	0.17	0.531	1	
Tachycardic	2.78	0.04	1.022	2	
Chronic Renal Insuf.	2.58	0.06	0.948	2	

Logistic Regression, Score, and Neural Networks



Risk Score of Death

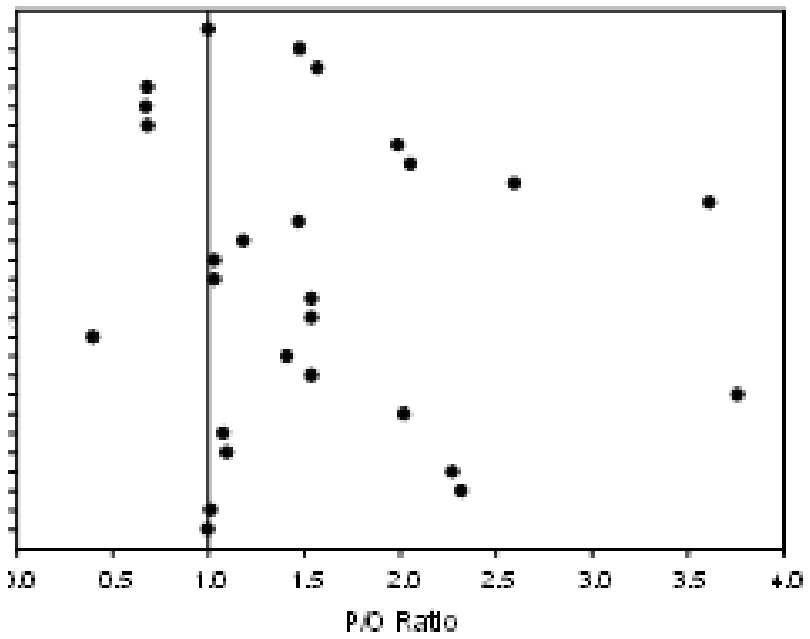



10-year CVD Risk Calculator (Risk Assessment Tool for Estimating 10-year Risk of Developing Har - Windows Internet Ex...
http://hp2010.nhlbihin.net/atpiii/calculator.asp?usertype=prof
apache risk calculator
Norton
aha risk calculator
Search
Cards & Log-ins
10-year CVD Risk Calculator (Risk Assessment To...
Page
Tools
NATIONAL CHOLESTEROL EDUCATION PROGRAM
Third Report of the Expert Panel on
Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III)
Risk Assessment Tool for Estimating 10-year Risk of Developing Hard CHD (Myocardial Infarction and Coronary Death)
The risk assessment tool below uses recent data from the Framingham Heart Study to estimate 10-year risk for "hard" coronary heart disease outcomes (myocardial infarction and coronary death). This tool is designed to estimate risk in adults aged 20 and older who do not have heart disease or diabetes. Use the calculator below to estimate 10-year risk.
Age: years
Gender: Female Male
Total Cholesterol: mg/dL
HDL Cholesterol: mg/dL
Smoker: No Yes
Systolic Blood Pressure: mm/Hg
Currently on any medication to treat high blood pressure. No Yes
Calculate 10-Year Risk
Internet | Protected Mode: Off
100%

External Validations for CVD Models

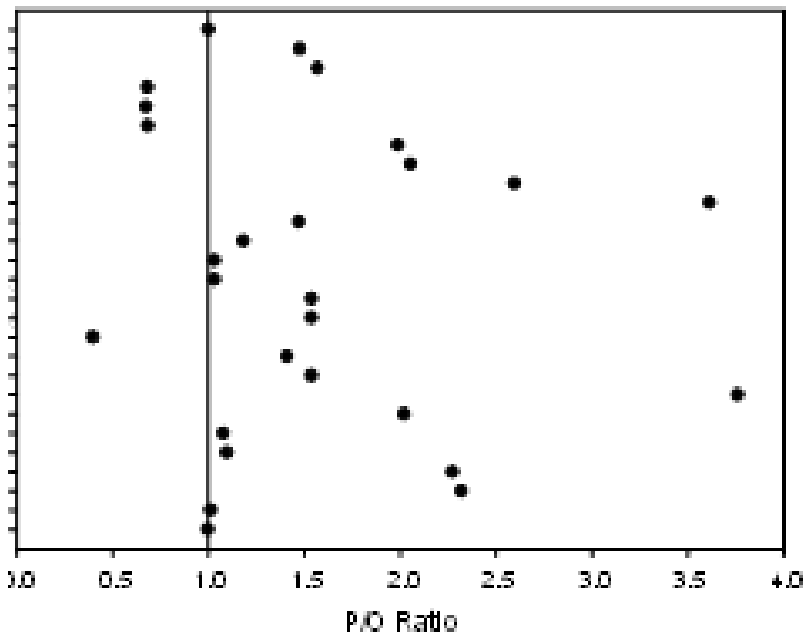
Model/Cohort	AKA	Year Published	External Validations
Framingham Risk Score	FRS	Dawber et al, 1951[8]	1
Framingham Risk Score	FRS	Kannel et al, 1976[9]	4
Framingham Risk Score	FRS	Anderson et al, 1991[2]	29
Glostrup	Glostrup	Schroll et al, 1992[10]	1
European Society of Cardiology	ESC	Pyorala et al, 1994[11]	1
Framingham Risk Score	FRS	Wilson et al, 1998[12]	32
Framingham Risk Score for ATP III	FRS ATP III	ATP III, 2001[13]	5
Framingham Risk Score	FRS	D'Agostino et al, 2001[14]	9
UK Prospective Diabetes Study	UKPDS	Stevens et al, 2001[15]	1
Framingham Point System	FPS	ATP III, 2002[1]	2
Prospective Cardiovascular Munster Study	PROCAM	Assman et al, 2002[16]	6
Finnish Diabetes Risk Score	FINDRISC	Lindstrom et al, 2003[17]	6
Systematic Coronary Risk Evaluation	SCORE	Conroy et al, 2003[18]	8
Diabetes Epidemiology: Collaborative Analysis of Diagnostic Criteria in Europe	DECODE	Balkau et al, 2004[19]	1
ASSessing cardiovascular risk using SIGN guidelines	ASSIGN	SIGN, 2007[20]	2
		Total	108

Predicted / Observed



Framingham models tested on
European populations

Predicted / Observed



Framingham models tested on European populations

European models tested on North American populations

Questions

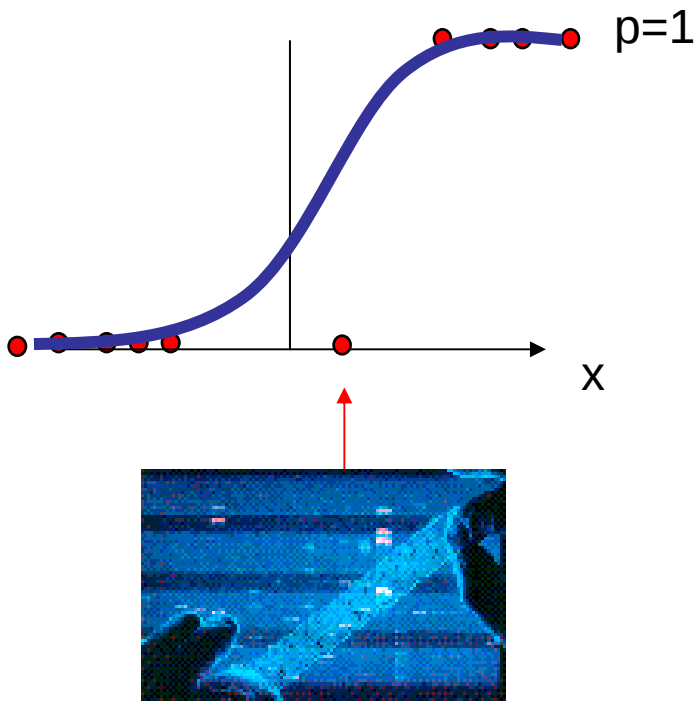
- Which model is right?

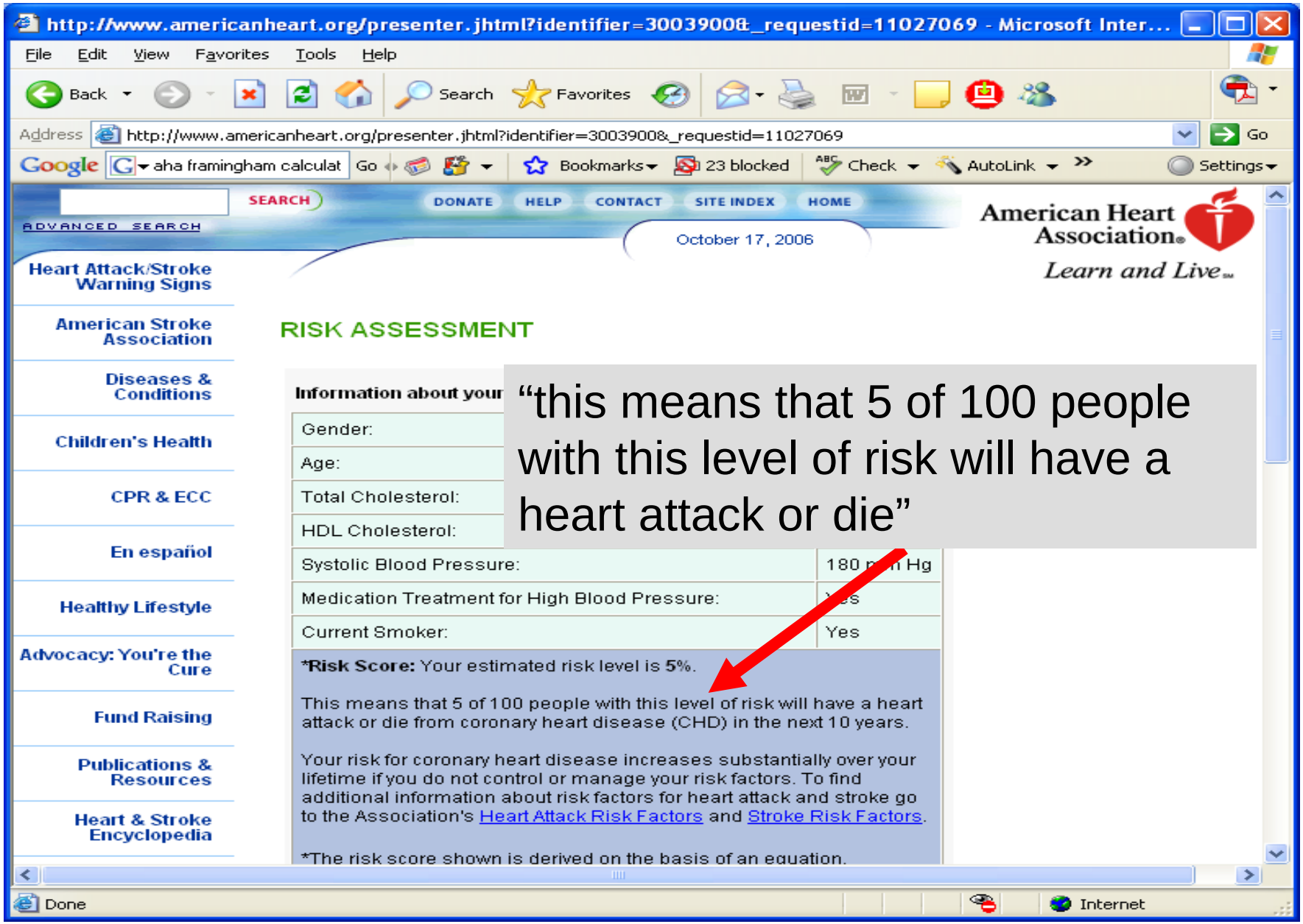
- “True” probability would be the gold-standard
  - What is the true probability?

- Are the models adequate in discrimination and calibration?

Your Risk

"this program shows the estimated health risks of people with your *same* age, gender, and risk factor levels"



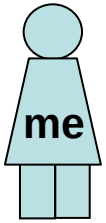


"this means that 5 of 100 people with this level of risk will have a heart attack or die"

Patients “like you”

Input space

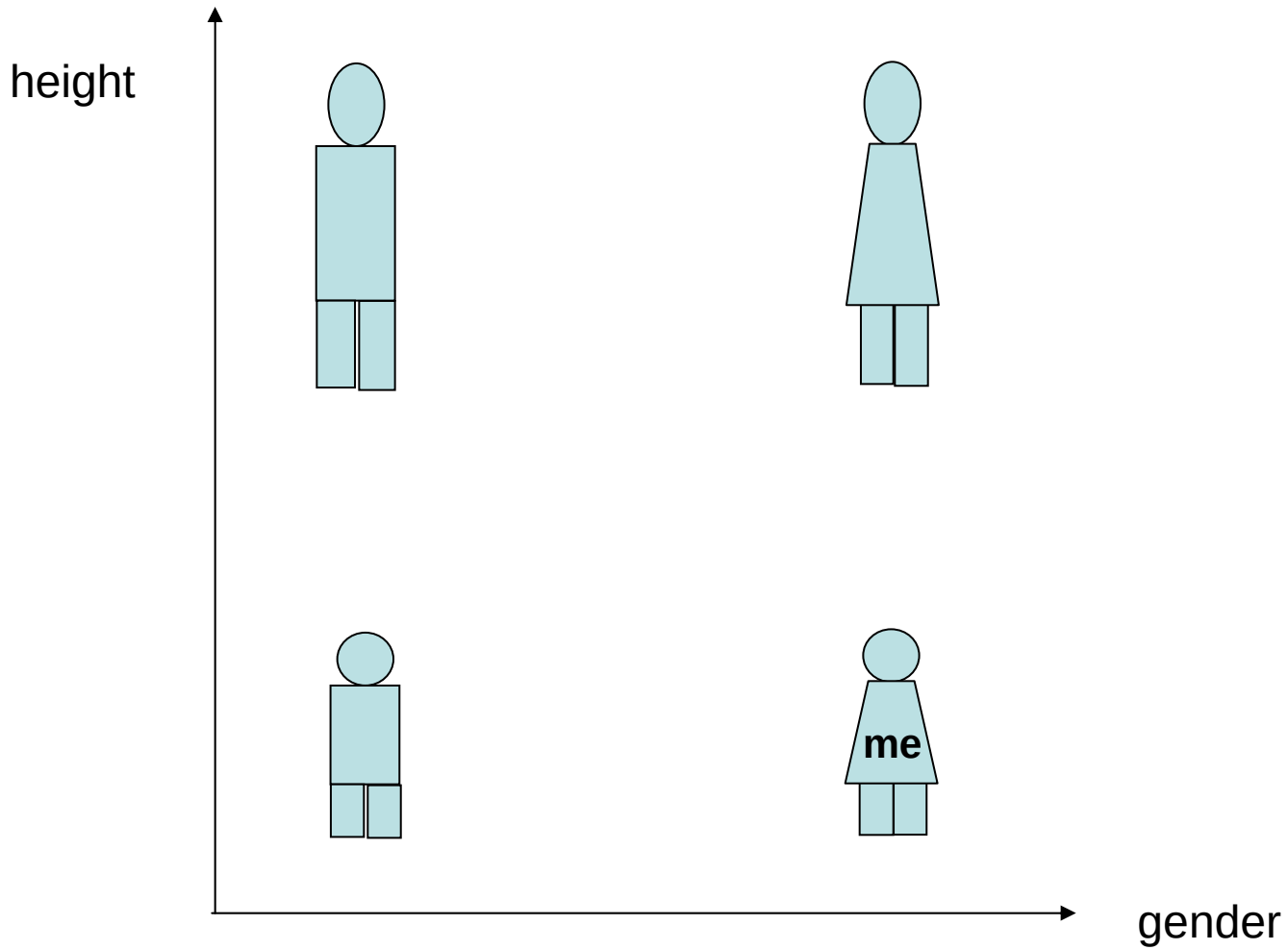
“people with your *same* age, gender, and risk factor levels”



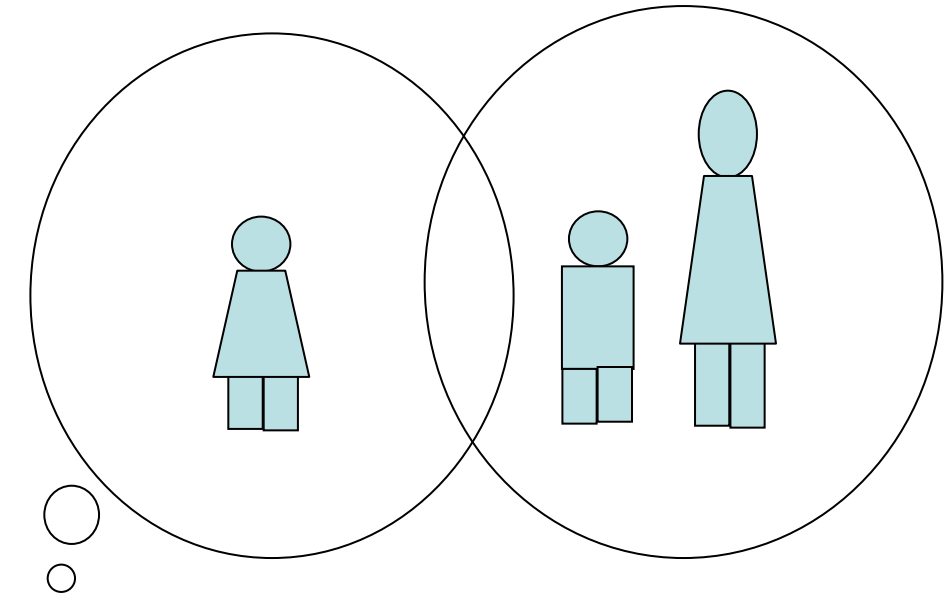
Output space

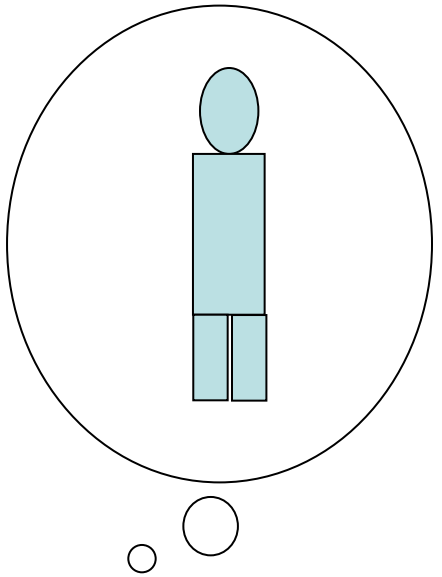
“people with this level of risk”

Patients “like you”

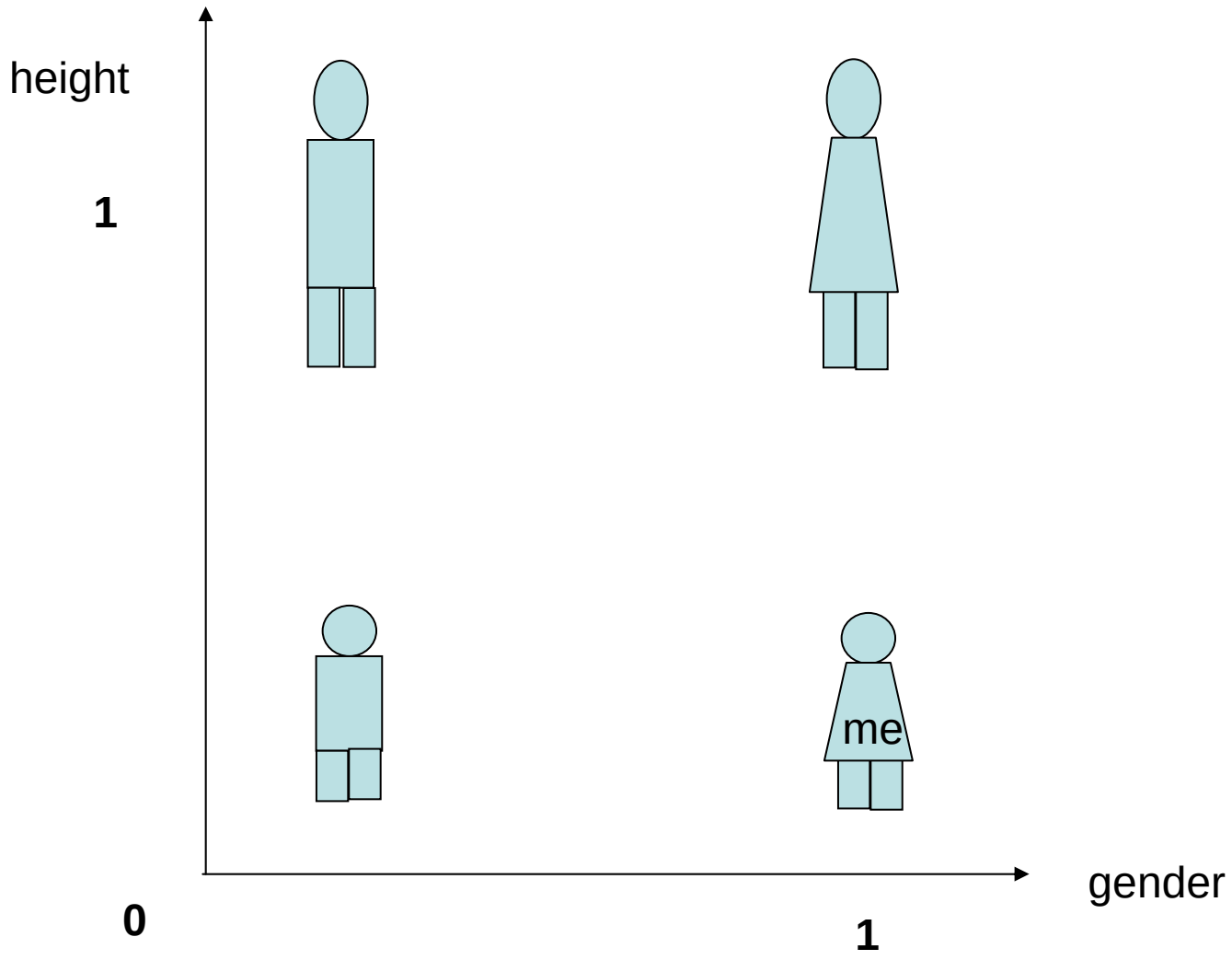


Patients “like you”

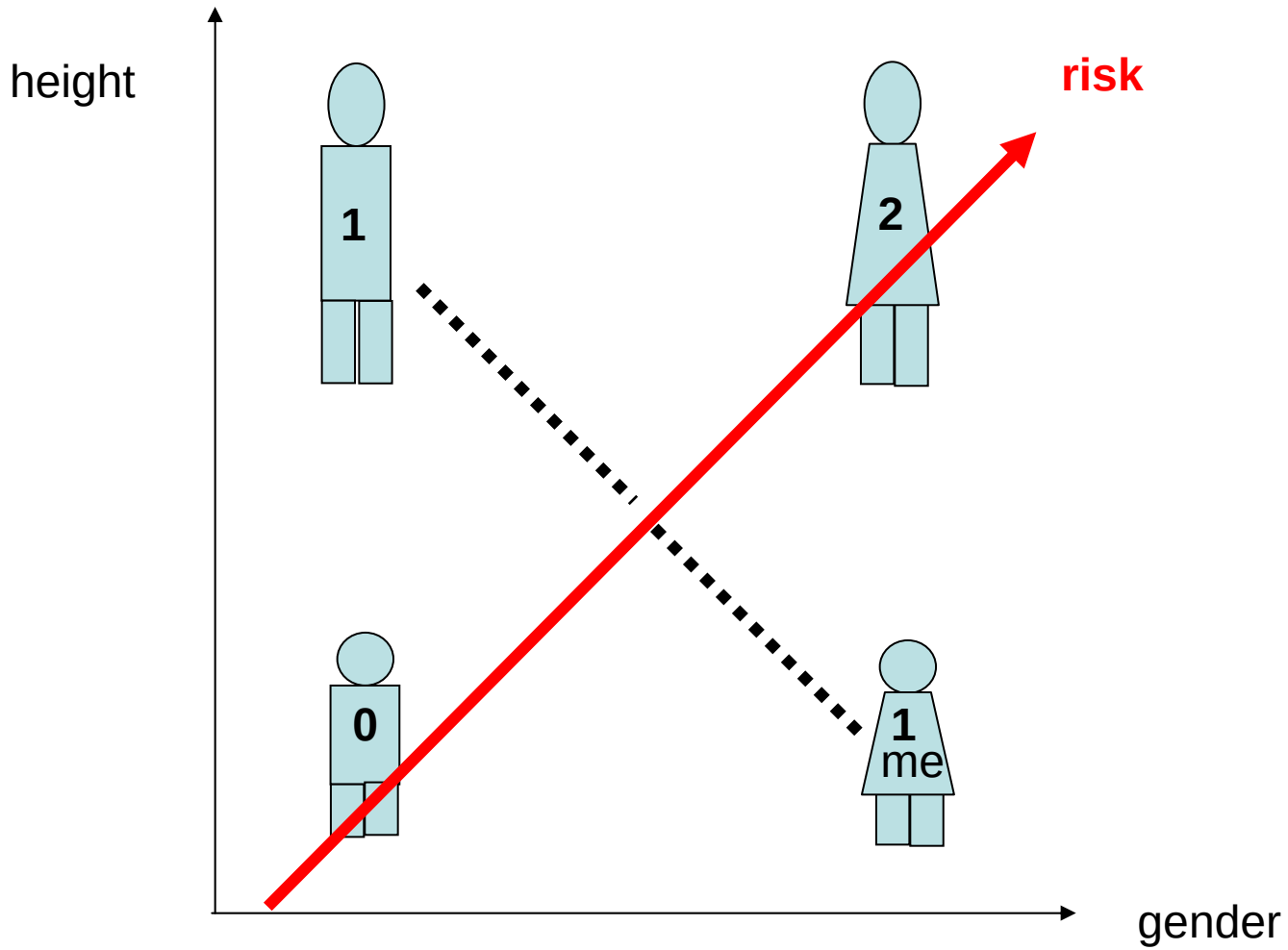




Patients “like you”



Patients “like you”



Evaluation of Predictive Models

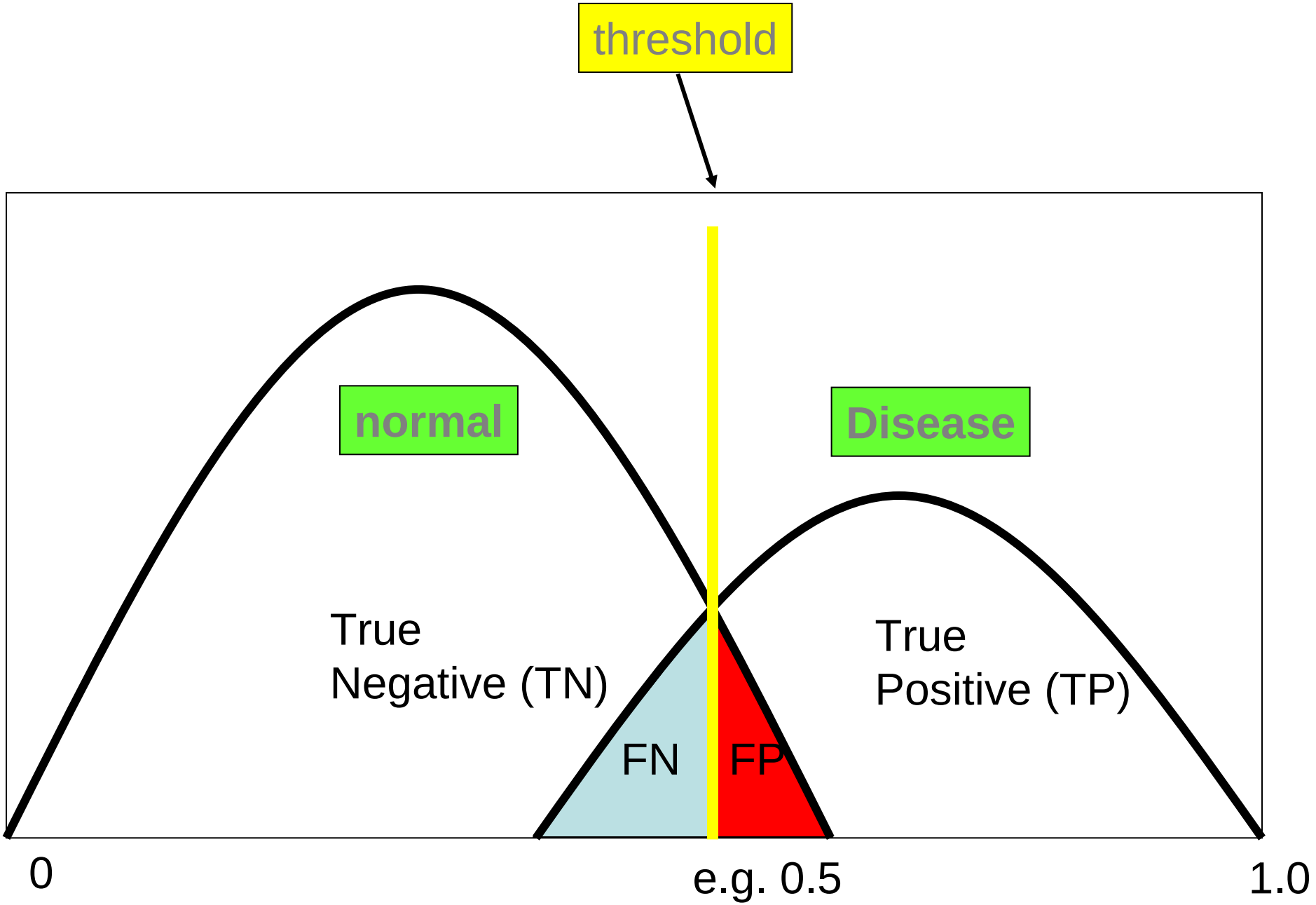
- Error
- Discrimination
  - Area under ROC
- Calibration
  - Plot of groups: observed vs expected
  - Hosmer-Lemeshow statistic

Discrimination of Binary Outcomes

- **Estimate** and **Observed** outcome (“gold standard”, “true”)

Estimate	True
0.3	0
0.2	0
0.5	1
0.1	0

- Classification into category 0 or 1 is based on thresholded estimates (e.g., if estimate > 0.5 then consider “positive”)


threshold
normal
Disease
True
Negative (TN)
True
Positive (TP)
FN
FP
0
e.g. 0.5
1.0

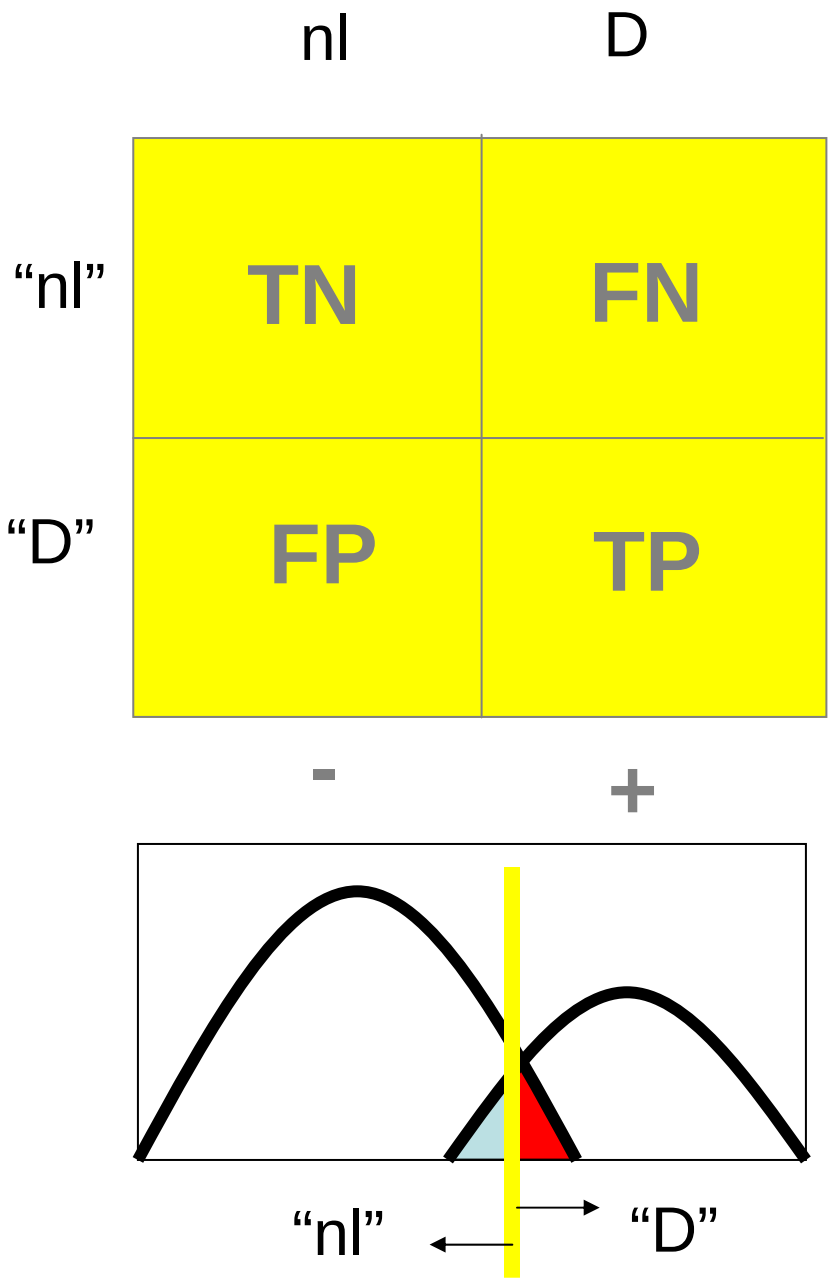
Sens = TP/TP+FN

Spec = TN/TN+FP

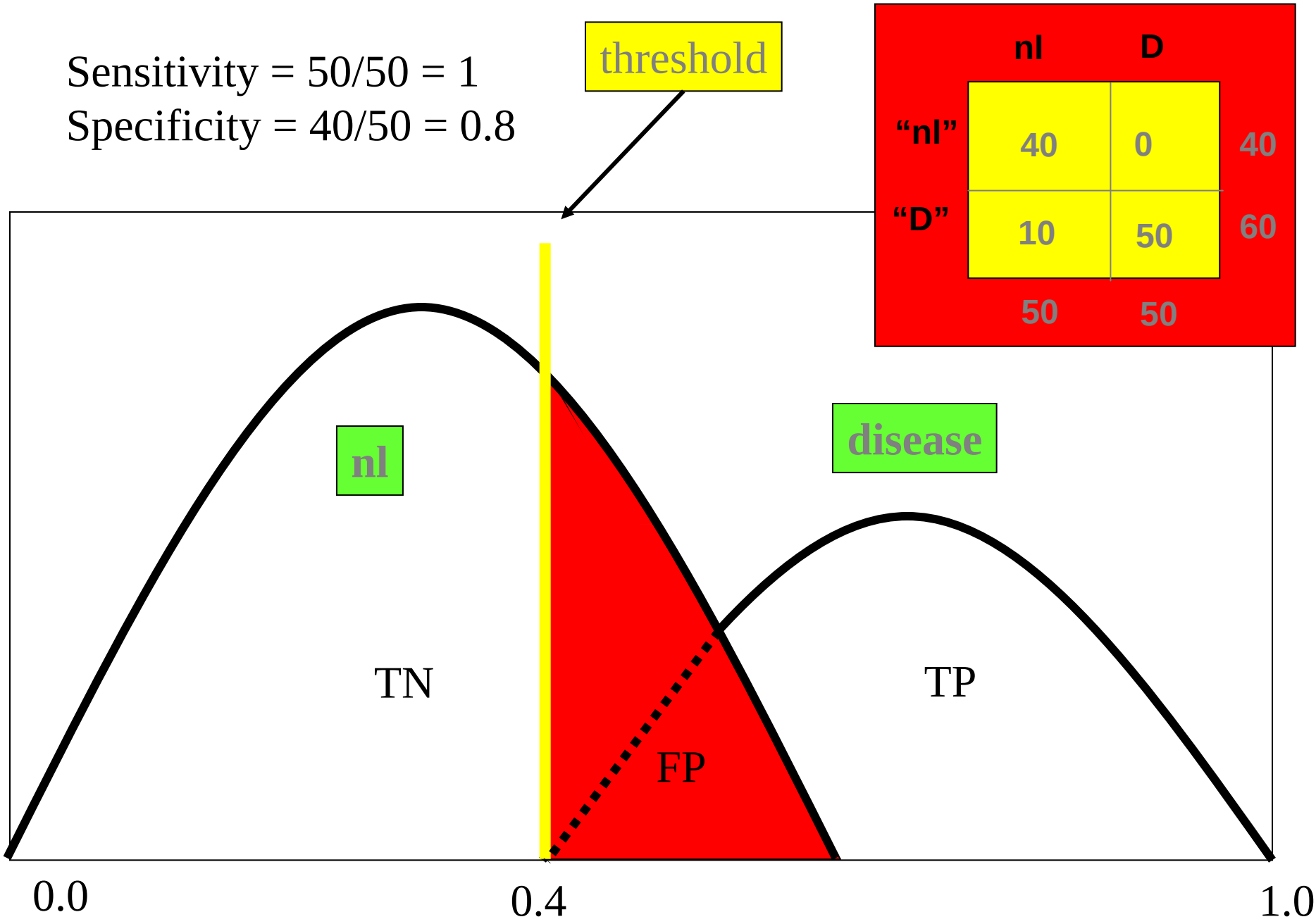
PPV = TP/TP+FP

NPV = TN/TN+FN

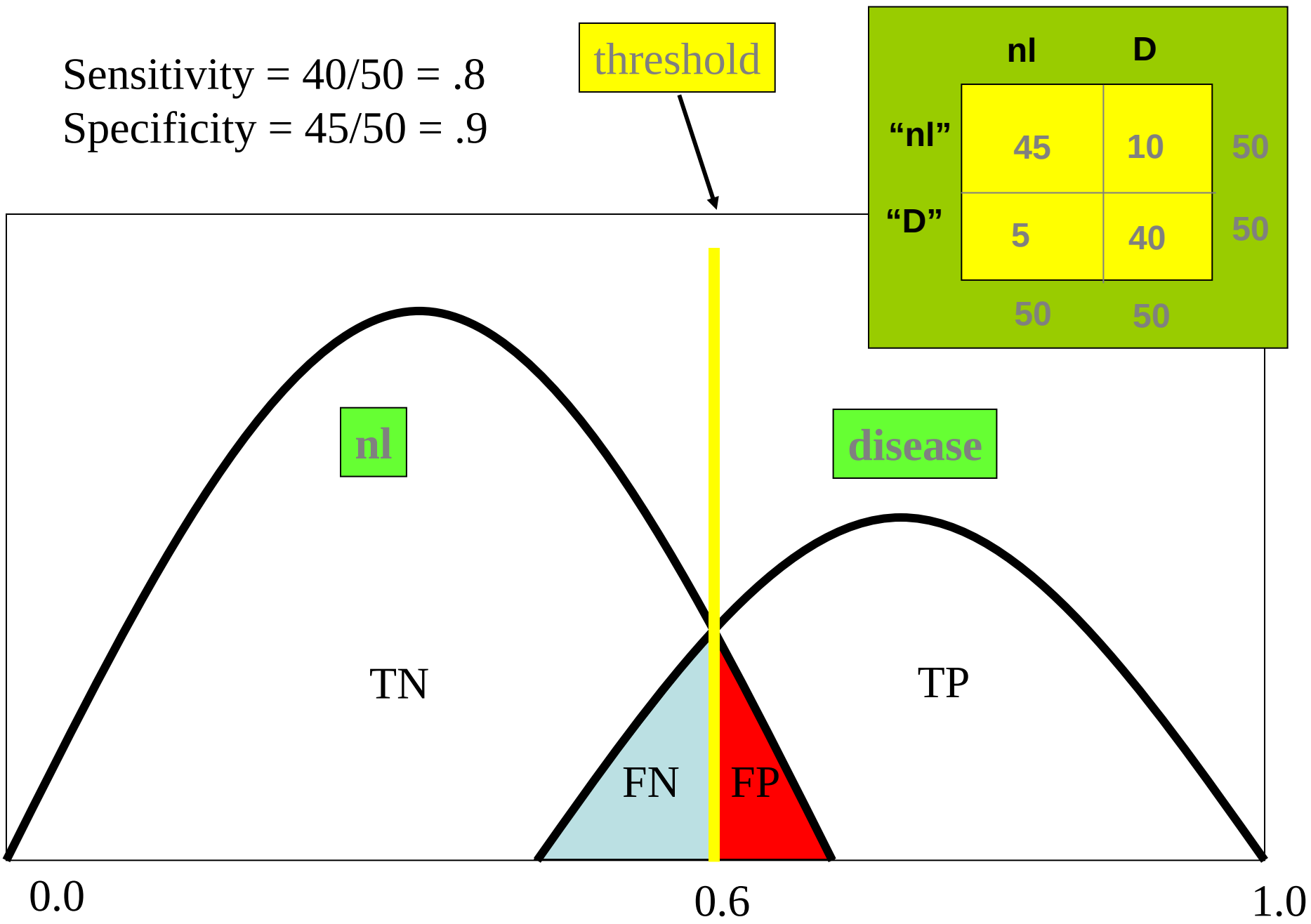
Accuracy = TN +TP



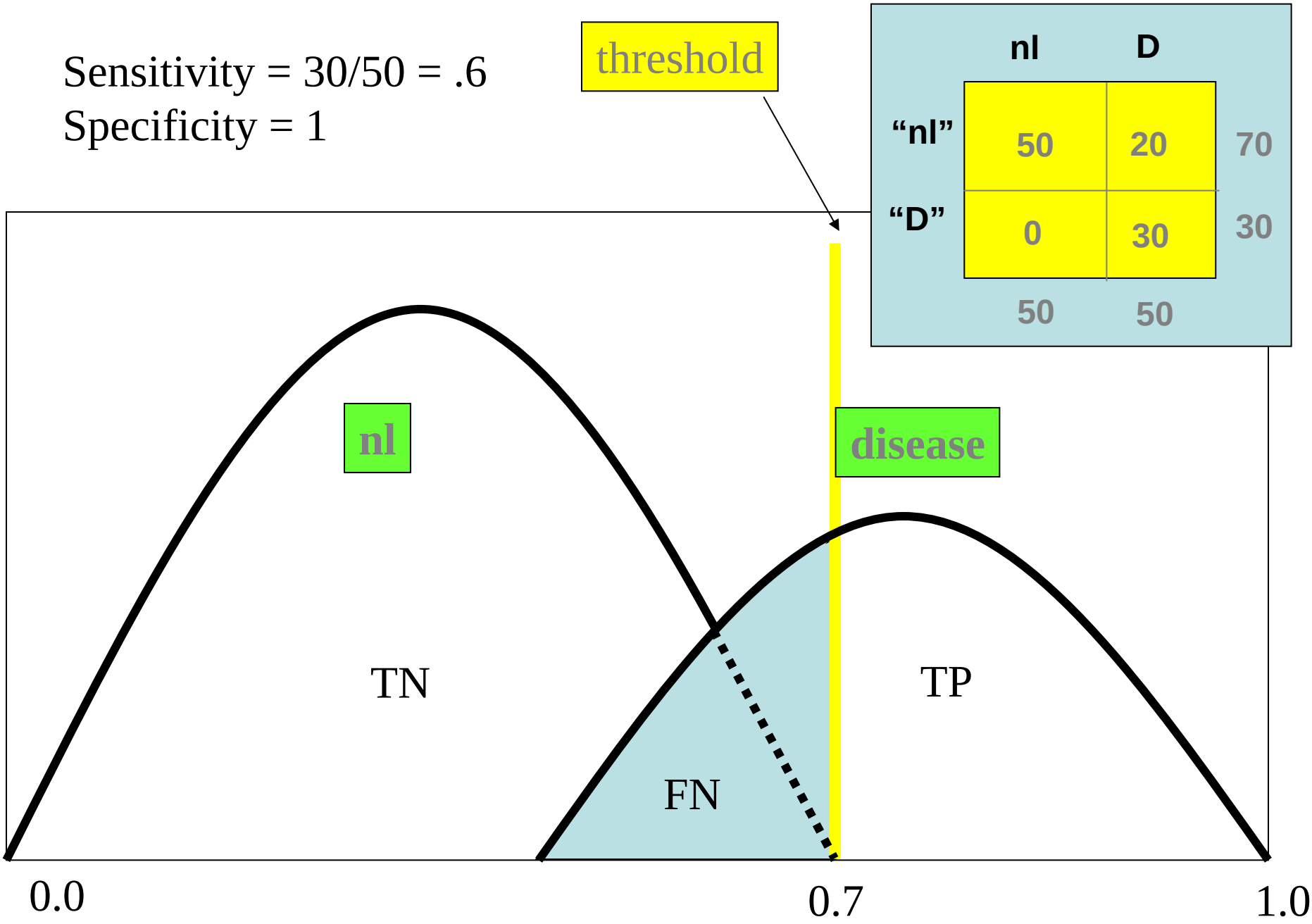
Sensitivity = 50/50 = 1
Specificity = 40/50 = 0.8

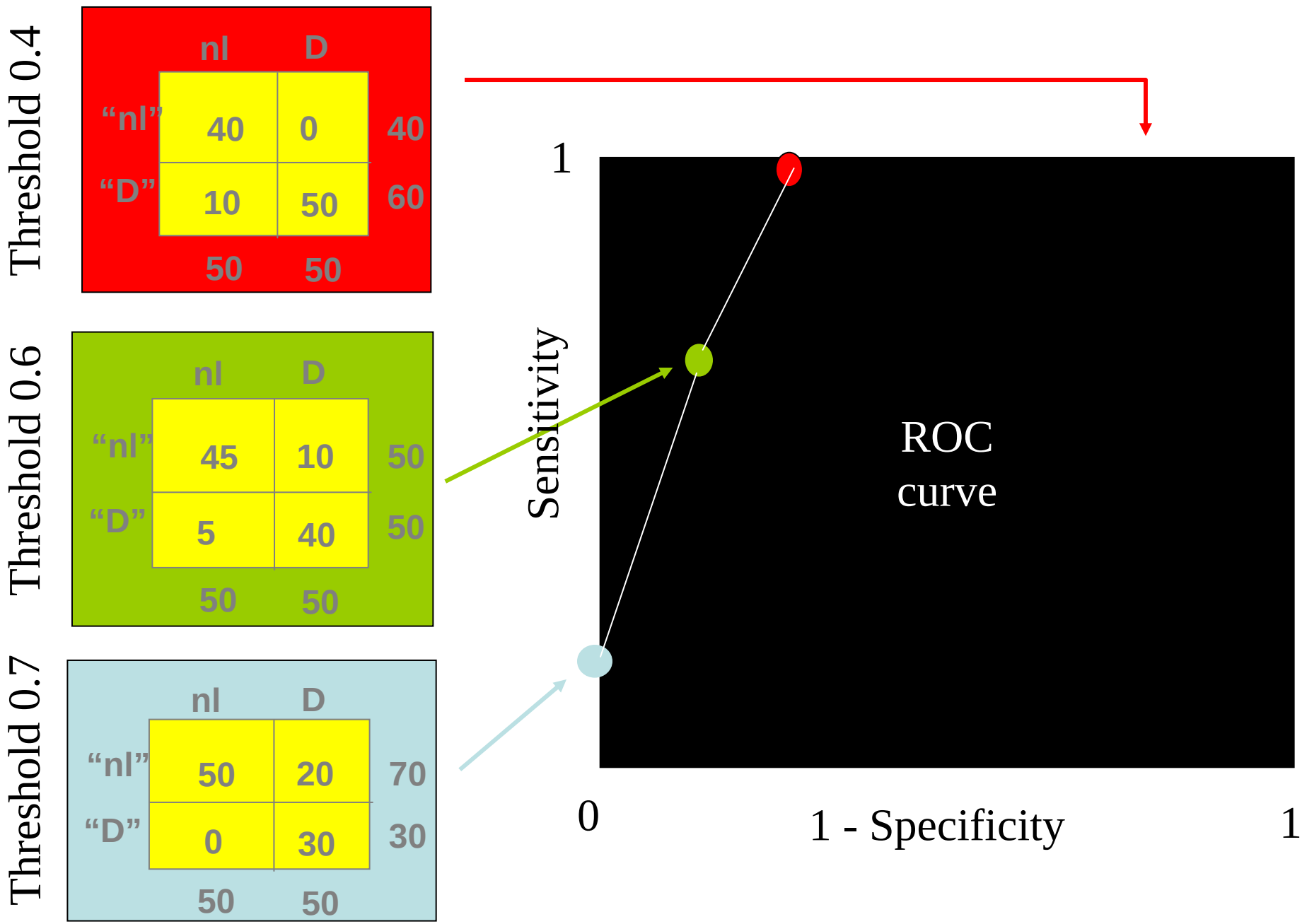


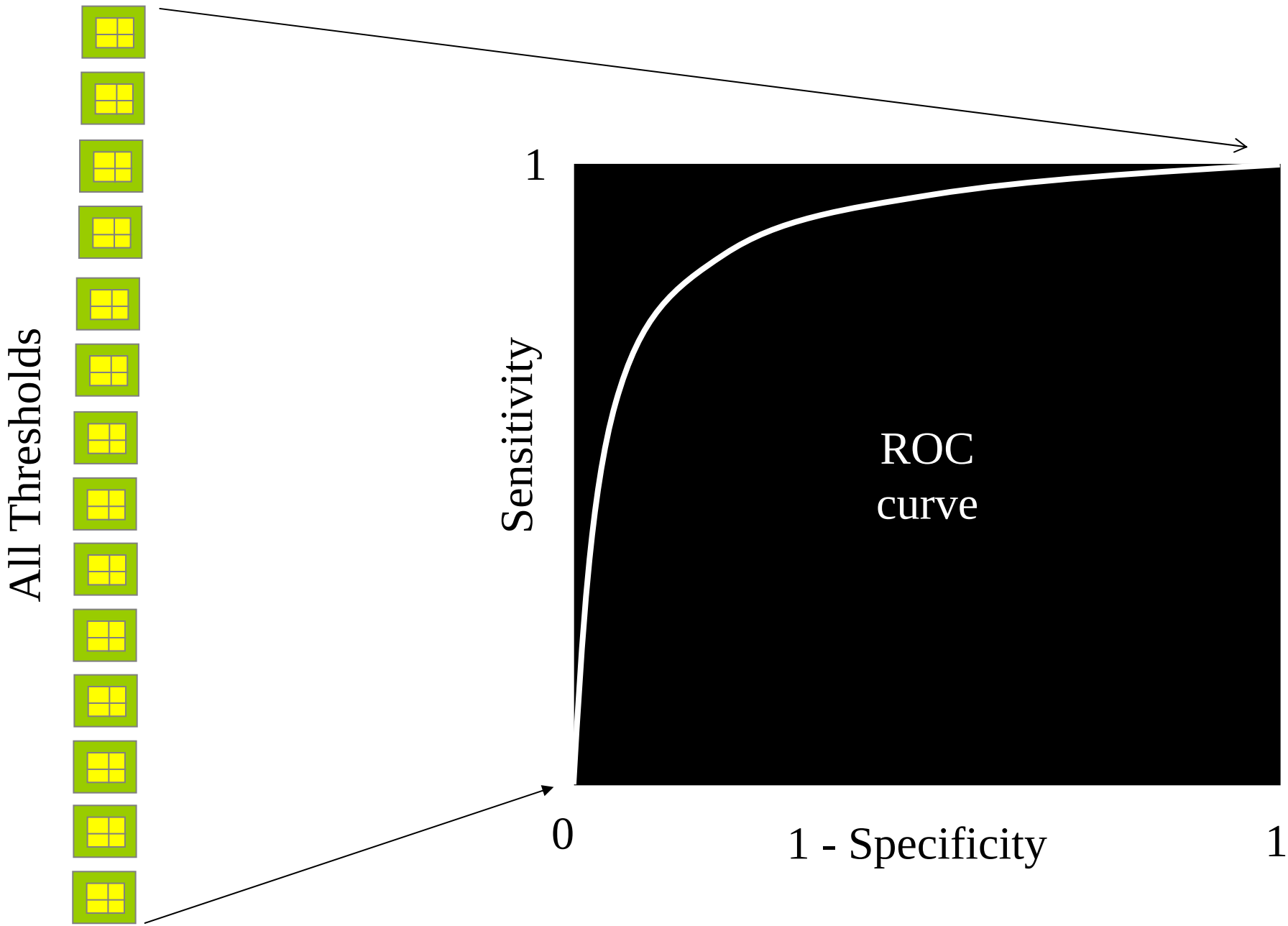
Sensitivity = 40/50 = .8
Specificity = 45/50 = .9



Sensitivity = 30/50 = .6
Specificity = 1




Threshold 0.4
nl
D
"nl"
"D"
40
0
10
50
40
60
50
50
Threshold 0.6
nl
D
"nl"
"D"
45
10
5
40
50
50
50
50
Threshold 0.7
nl
D
"nl"
"D"
50
20
0
30
70
30
50
50
Sensitivity
1
0
1 - Specificity
1
ROC
curve


All Thresholds
1
Sensitivity
ROC
curve
0
1 - Specificity
1

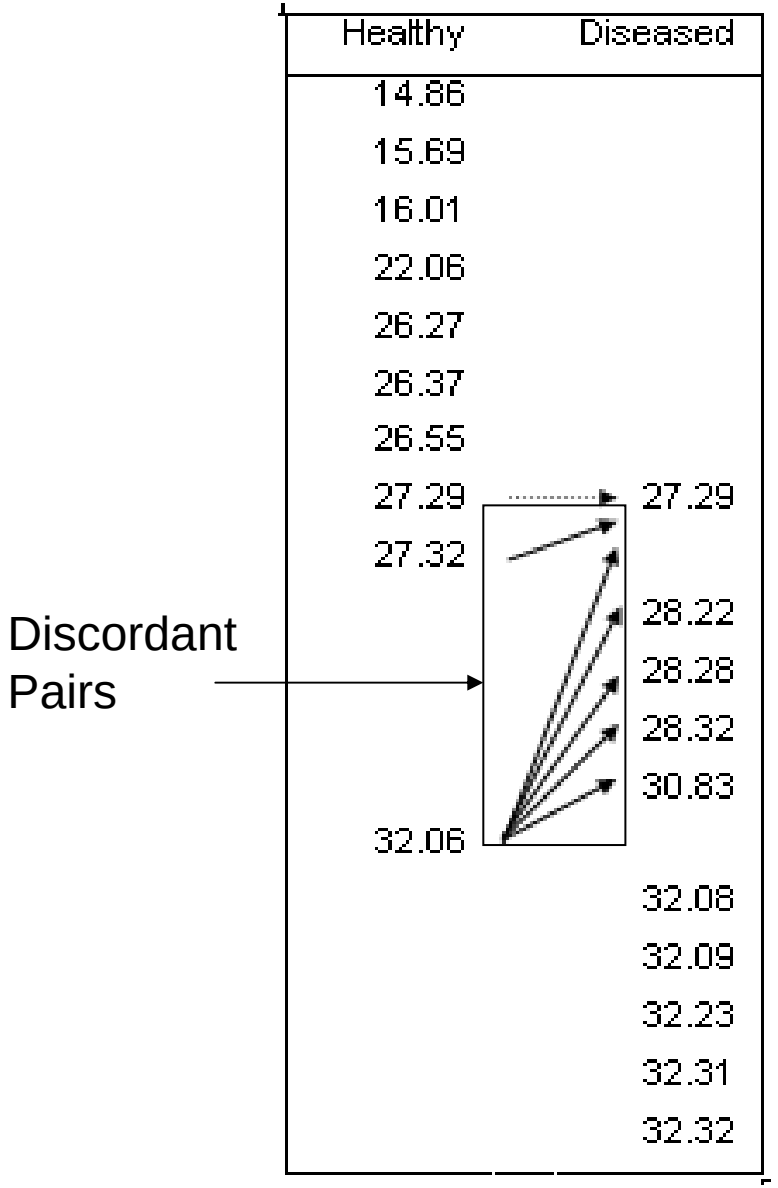
Areas Under the ROC curve

concordance index

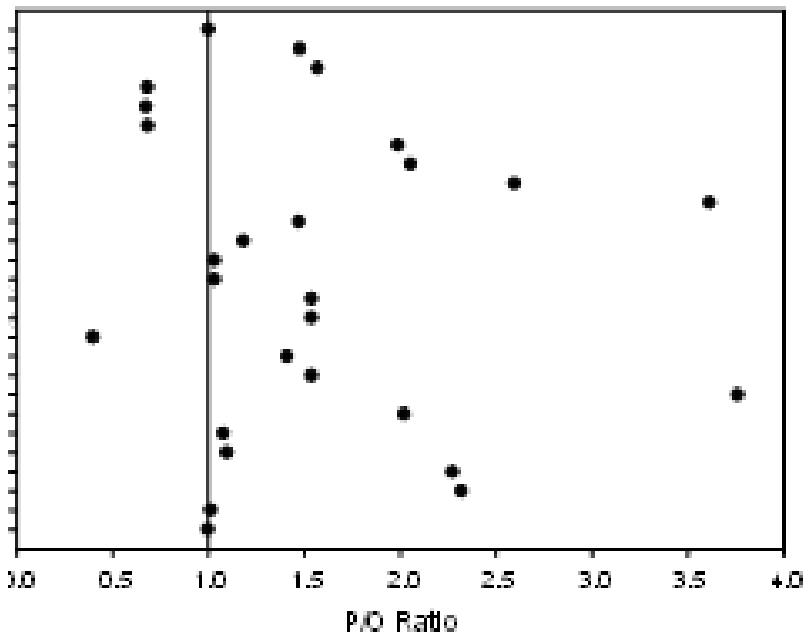
- measure adequacy of risk ranking

(#concordant pairs + ½ #ties) /all pairs

- do not measure adequacy of risk estimates (collective nor individual)



Predicted / Observed



Framingham models tested on
European populations

Calibration

- Measures how close the average estimate is to the observed proportion
- Goodness-of-fit
  - Hosmer-Lemeshow statistics

D = 0

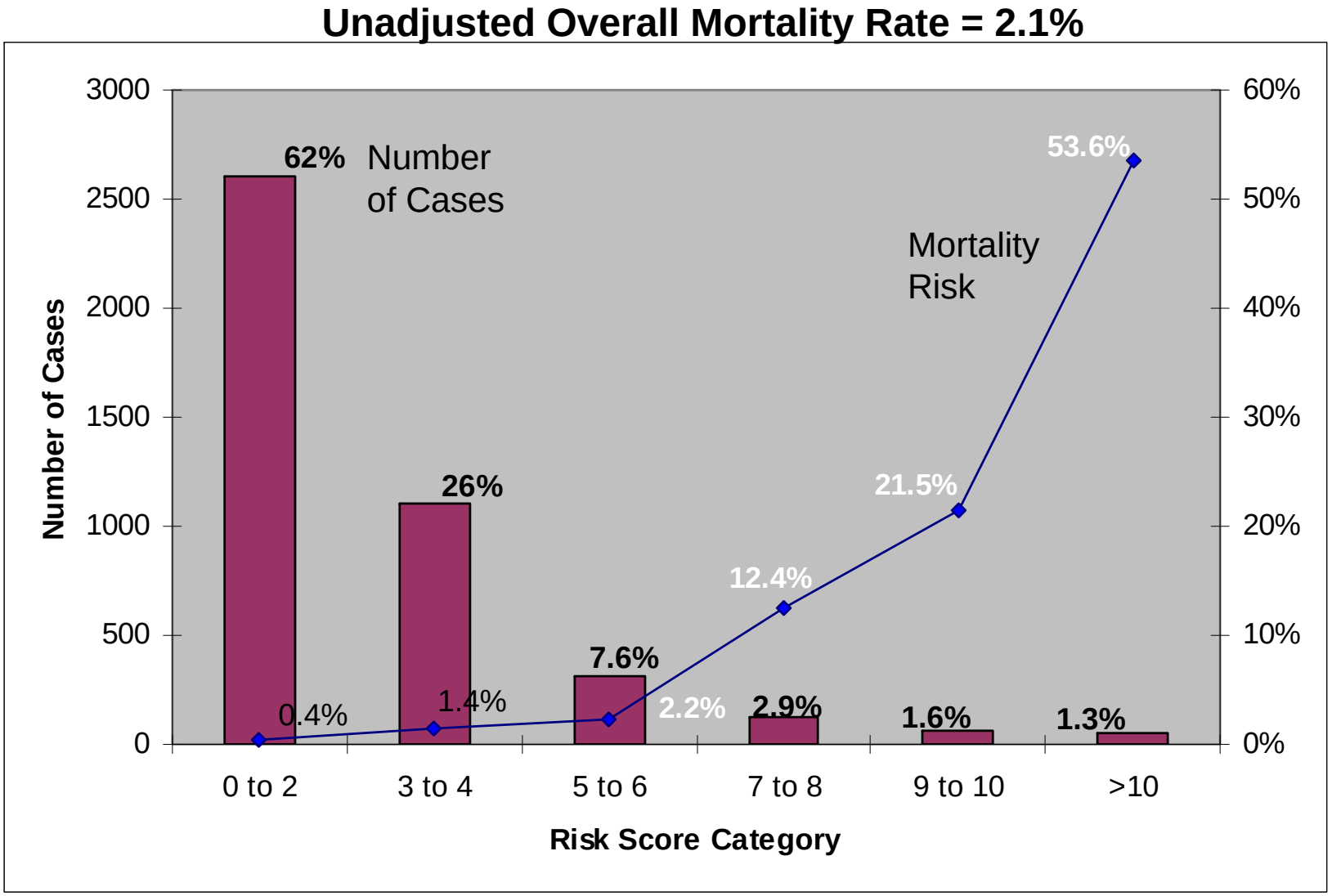
Estim. π	Obs. o	Tercile Sums	
.1486	0		
.1569	0		
.1601	0		
.2206	0		
.2627	0		
.2637	0		
.2655	0	1.478	0
.2729	0		
.2729	1		
.2732	0		
.2822	1		
.2828	1		
.2832	1		
.3083	1	1.976	5
.3206	0		
.3208	1		
.3209	1		
.3223	1		
.3231	1		
.3232	1	1.931	5

Calibration

$$HL = \sum_{D=0}^{1} \sum_{l=1}^{3} \left[\frac{\left(\pi_{Dl} - o_{Dl} \right)^2}{\pi_{Dl}} \right]$$

D = 0				D = 1			
Estim. π	Obs. o	Tercile Sums		Estim. π	Obs. o	Tercile Sums	
.1486	0			.8514	1		
.1569	0			.8431	1		
.1601	0			.8399	1		
.2206	0			.7794	1		
.2627	0			.7373	1		
.2637	0			.7363	1		
.2655	0	1.478	0	.7345	1	5.522	7
.2729	0			.7271	1		
.2729	1			.7271	0		
.2732	0			.7268	1		
.2822	1			.7178	0		
.2828	1			.7178	0		
.2832	1			.7172	0		
.3083	1	1.976	5	.7168	0	5.051	2
.3206	0			.6794	1		
.3208	1			.6792	0		
.3209	1			.6791	0		
.3223	1			.6777	0		
.3231	1			.6769	0		
.3232	1	1.931	5	.6768	0	4.069	1
				HL C		15.54	
				p		0.049	

Risk Score of Death



Interventional Cardiology Models

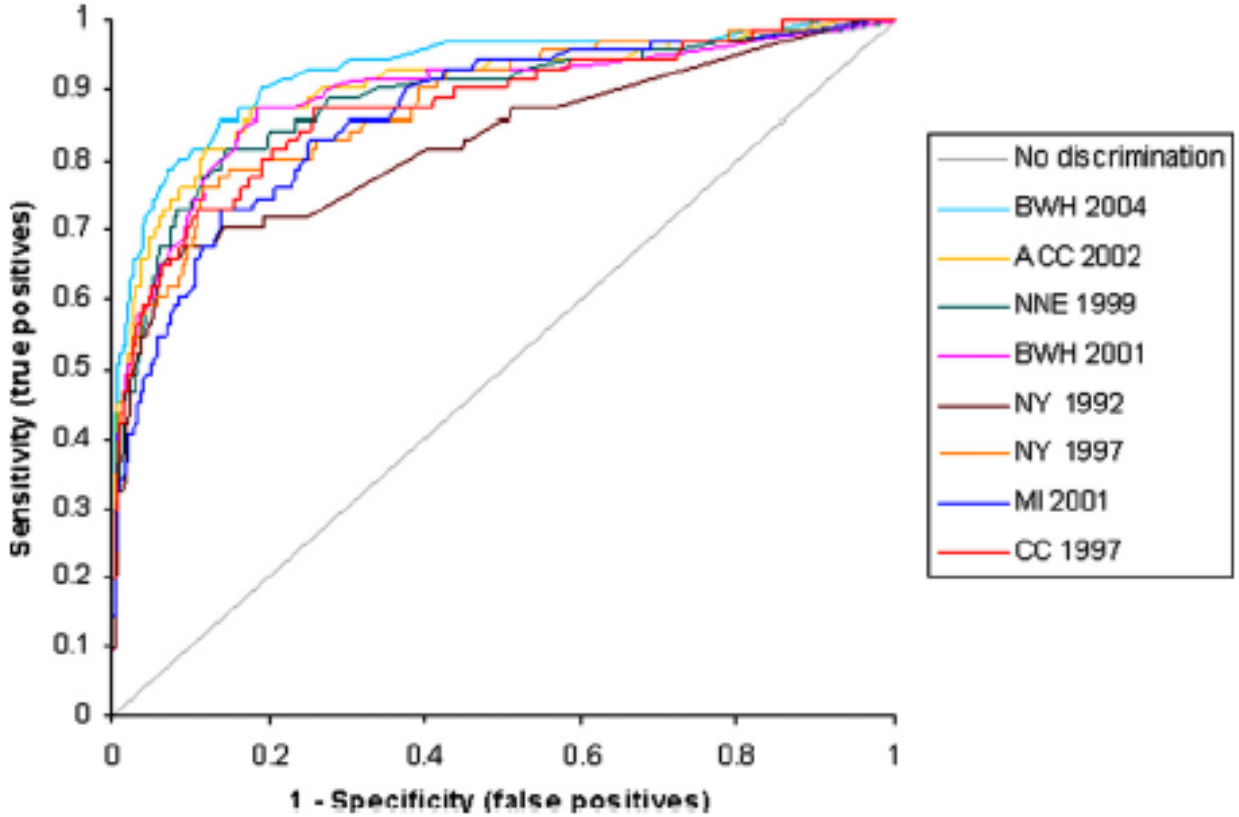
Summary of the training datasets for the models used in this study

Model	Dates		Location	Sample	AUC	HL(p)	Validation type
NNE [22] 1999	1/1/1994	12/31/1996	NH, ME, MA, VT (7)	15331	0.88	0.09	Bootstrap resampling
NY [23] 1992	1/1/1991	6/30/1991	NY	5827	0.884	NA	Subset significance
NY [24] 1997	1/1/1991	12/31/1994	NY	62670	0.892	0.11	Subset significance
MI [25] 2001	10/1/1999	8/30/2000	Detroit, MI	10796	0.90	0.5	Training/test
ACC [26] 2002	1/1/1998	9/30/2000	National	100253	0.89	0.133	Training/test
BWH [28] 2001	1/1/1997	12/31/1999	Boston, MA	2804	0.86	0.11	Training/test
CC [27] 1997	1/11993	12/31/1994	Cleveland, OH (5)	12985	0.846	NA	Bootstrap resampling

Sample, sample size. AUC, area under the receiver operating characteristic. HL(p), Hosmer–Lemeshow p value.

Validation

- 5278 patients from BWH (2001-2004) (external validation set)
- Comparisons use Areas under the ROC curve (AUC) and the Hosmer-Lemeshow goodness-of-fit statistic (deciles)



Summary of discrimination and calibration performance for each model

Curve	Deaths	AUC	95% CI	HL χ^2	95% CI	HL(p)
NY 1992	96.7	0.82	0.76–0.88	31.1	13.9–50.0	<0.001
NY 1997	61.6	0.88	0.81–0.92	32.2	16.4–45.5	<0.001
CC 1997	78.8	0.88	0.82–0.93	27.8	19.6–38.7	<0.001
NNE 1999	56.2	0.89	0.84–0.94	45.9	31.9–67.4	<0.001
MI 2001	61.8	0.86	0.81–0.90	30.4	16.7–43.1	<0.001
BWH 2001	136.1	0.89	0.84–0.93	39.7	23.2–73.3	<0.001
ACC 2002	49.9	0.90	0.84–0.95	42.0	24.9–63.3	<0.001
BWH 2004	70.5	0.93	0.89–0.96	7.61	1.5–14.2	0.473

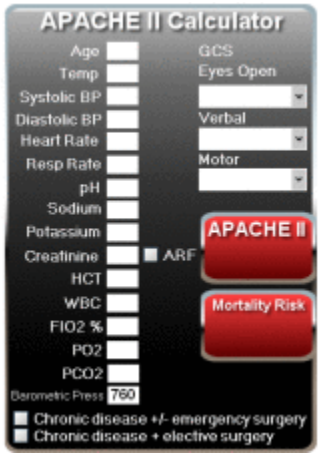
Calibration

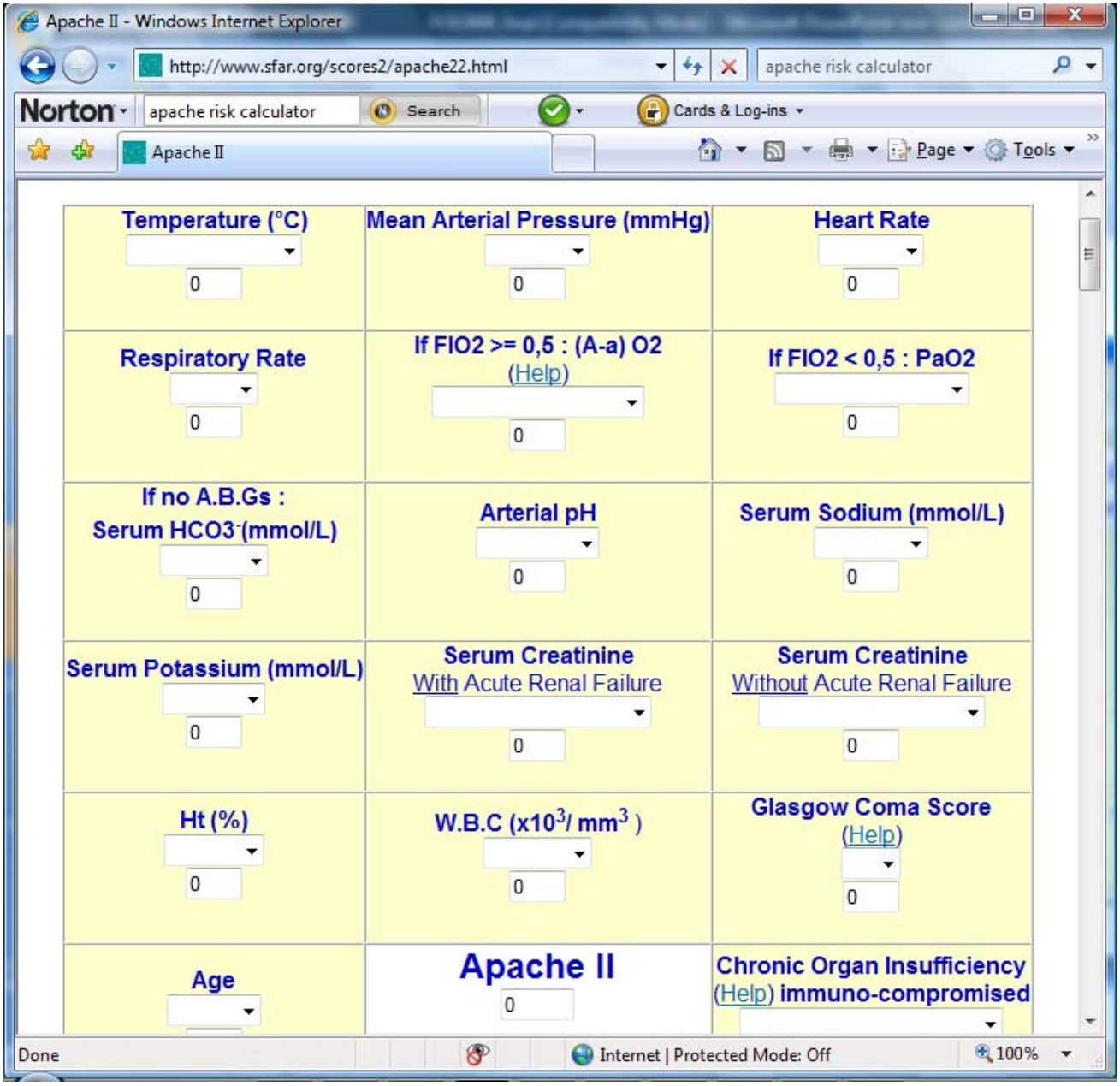
Are predictions obtained from external models good for individual counseling?

Curve	HL χ^2	95% CI	HL(p)	95% CI
NY 1992	31.1	13.9–50.0	<0.001	<0.001–0.003
NY 1997	32.2	16.4–45.5	<0.001	<0.001–0.004
CC 1997	27.8	19.6–38.7	<0.001	<0.001–0.013
NNE 1999	45.9	31.9–67.4	<0.001	<0.001–<0.001
MI 2001	30.4	16.7–43.1	<0.001	<0.001–0.011
BWH 2001	39.7	23.2–73.3	<0.001	<0.001–0.001
ACC 2002	42.0	24.9–63.3	<0.001	<0.001–0.002
BWH 2004	7.61	1.5–14.2	0.473	0.073–0.992

APACHE II

- Mortality in intensive care units (ICUs)
- 12 physiologic predictors





Summary of all comparison studies in terms of discrimination (AUC)

Author	AP-II	MPM_0	MPM_{24}	SAPS	AP-III	SAPS-II	$MPM\text{-}II_0$	$MPM\text{-}II_{24}$
Castella (86)	0.867	0.865						
Rowan (87)	0.830	0.740						
Wilairatana (88)	0.723			0.710	0.694			
Del Bufalo (89)	0.808					0.735		
Castella (90)[a]	0.852	0.773	0.825	0.798	0.866			
Castella (90)[b]	0.857	0.778	0.815	0.799		0.855	0.815	0.833
Moreno (91)						0.822	0.785	
Nouira (92)	0.820					0.840	0.850	0.882
Tan (93)	0.880					0.870		
Patel (94)	0.702					0.672		0.695
Vassar (95)	0.870				0.890			
Katsaragakis (96)	0.839					0.870		
Livingston (97)	0.763				0.795	0.784	0.741	0.791
Capuzzo (98)	0.805					0.816		
Markgraf (99)	0.832				0.846	0.846		
Beck (100)	0.835				0.867	0.852		

*a = Full sample, b = validation sample, AP = APACHE.

Summary of HL-GOF H and C statistics. χ^2 values and degrees of freedom are listed where available, p values are listed otherwise.

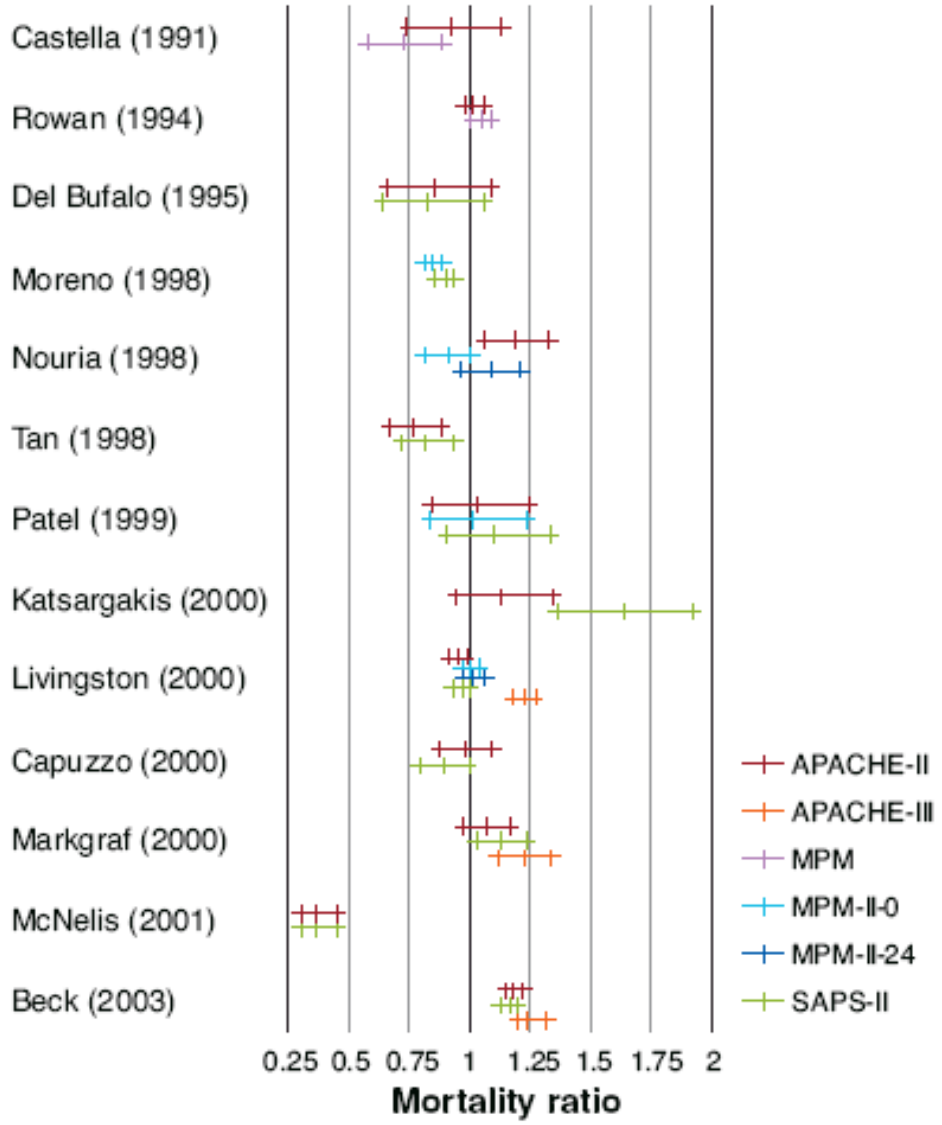
Author	AP-II	MPM_0	MPM_{24}	SAPS	AP-III	SAPS-II	$MPM\text{-}II_0$	$MPM\text{-}II_{24}$
HL-H								
Castella (86)	**9.42**	59.25						
Rowan (87)	80.66 (8df)						2514.86 (8df)	
Del Bufalo (89)	**7.73 (7df)**					14.84 (7df)		
Castella (90)[a]	$p < 0.001$	$p < 0.001$	$p < 0.001$					
Castella (90)[b]	$p = 0.005$	$p < 0.001$	$p < 0.001$			$p = 0.024$	$p = 0.072$	$p = 0.093$
Moreno (91)						218.2 (10df)	437.1 (10df)	
Nouira (92)	32.15 (10df)					76.89 (10df)	38 (10df)	19.9 (10df)
Patel (94)	**14.33**					22.58		20.7
Katsaragakis (96)	16.56 (8df)					77.54 (9df)		
Capuzzo (98)	**3.87 (10df)**					**7.62 (10df)**		
Markgraf (99)	**11.8 (8df)**				48.4 (8df)	20.5 (8df)		
HL-C								
Rowan (87)	57.25 (8df)						1737.41 (8df)	
Castella (90)[a]	$p < 0.001$	$p < 0.001$	$p < 0.001$					
Castella (90)[b]	$p = 0.025$	$p < 0.001$	$p < 0.001$			$p = 0.102$	$p = 0.015$	$p = 0.026$
Moreno (91)						208.4 (10df)	368.2 (10df)	
Nouira (92)	25.95 (10df)					73.78 (10df)	36.66 (10df)	29.59 (10df)
Tan (93)	43.96					49.06		
Katsaragakis (96)	18.14 (8df)					60.48 (9df)		
Livingston (97)	67.41					142.03	451.85	100.77
Capuzzo (98)	**5.05 (10df)**					**9.32 (10df)**		
Beck (100)	232.1				443.3	287.5		

[a]Full sample.

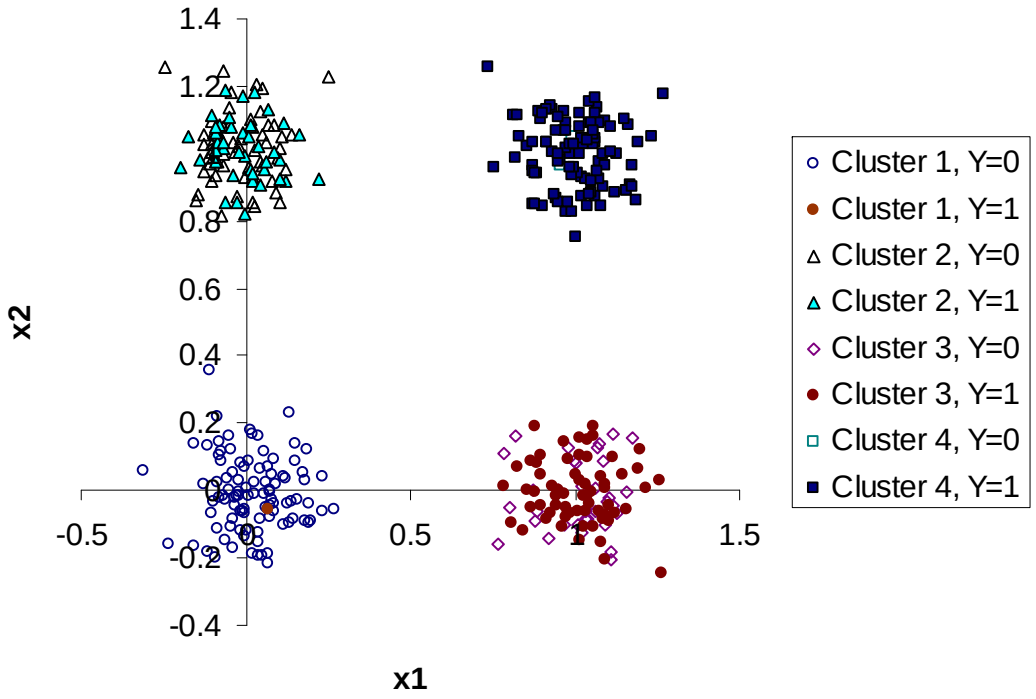
[b]Validation sample.

Bold entries indicate adequate calibration.

Standardized mortality ratio in different study comparisons



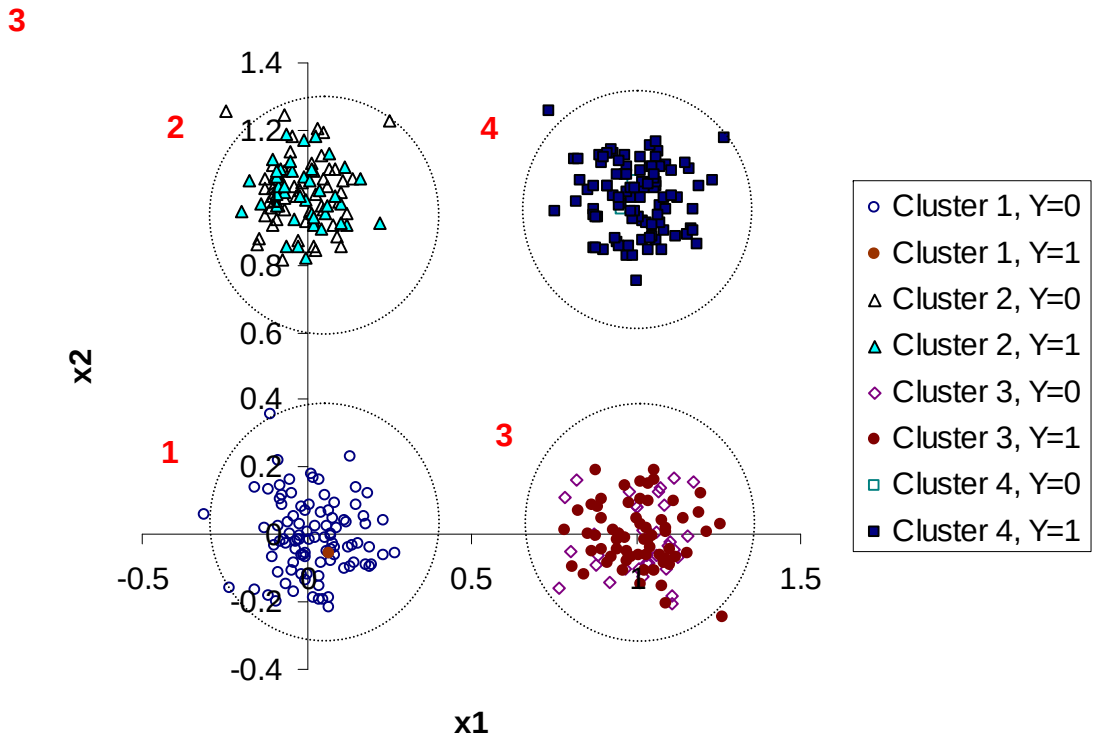
3

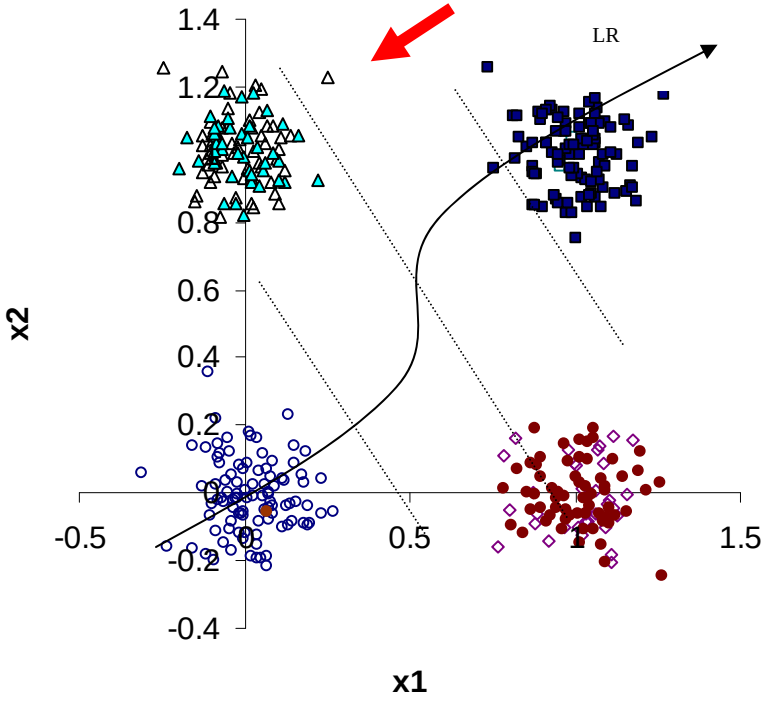


Simulated data set

Clusters 1 to 4 centered on (0,0), (0,1), (1,0), and (1,1), with Gaussian noise

True probability for clusters 1 to 4: 0.01, 0.40, 0.60, and 0.99

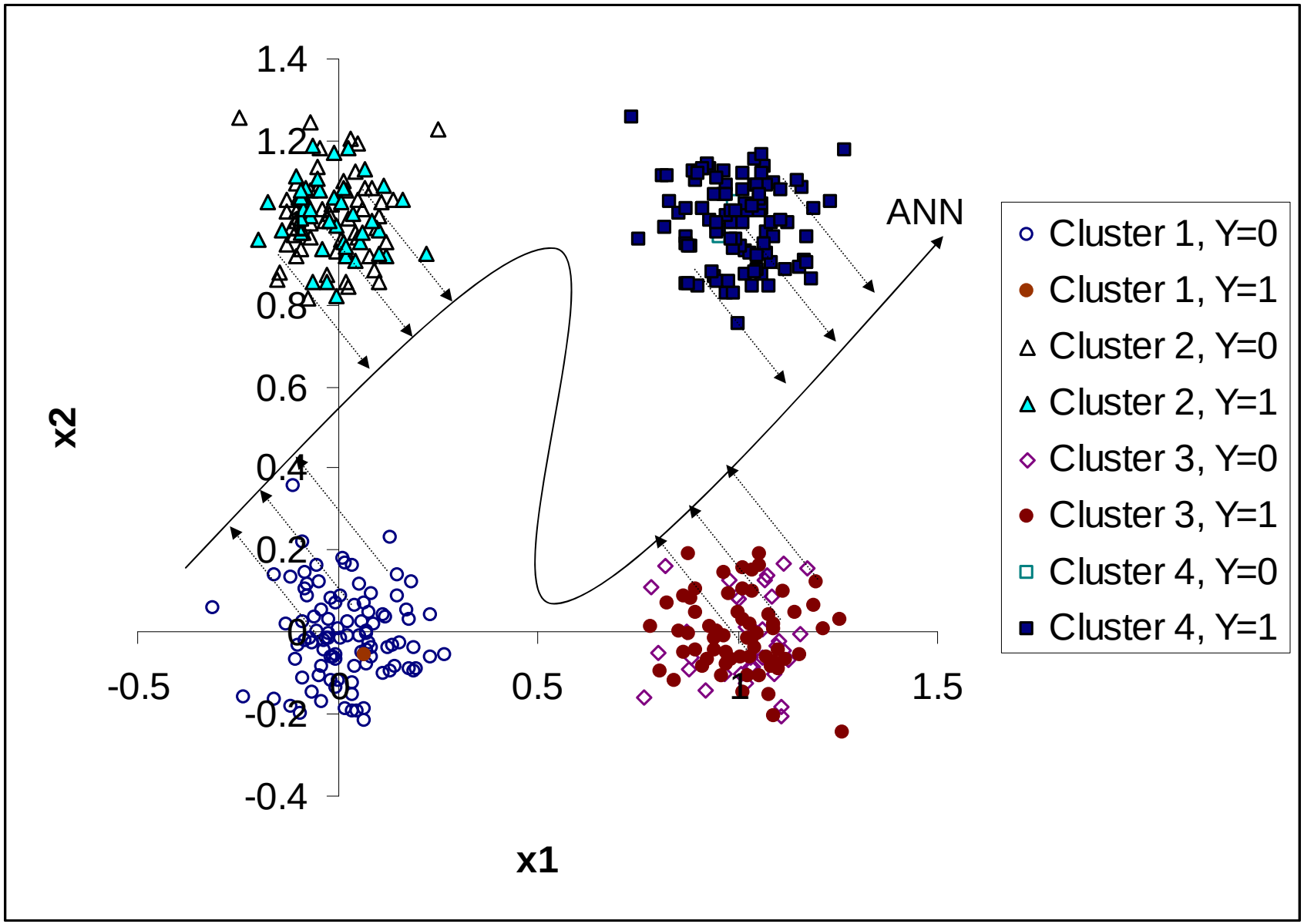


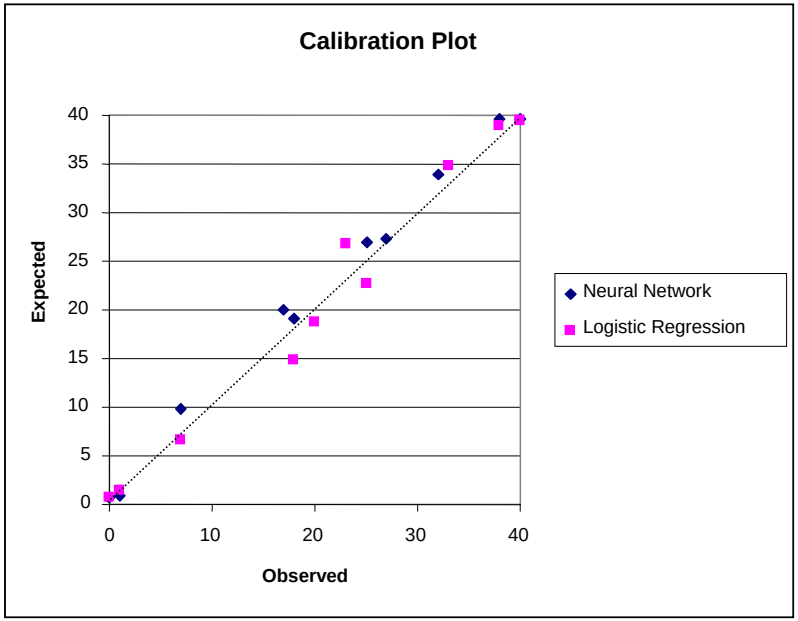


Simulated data set

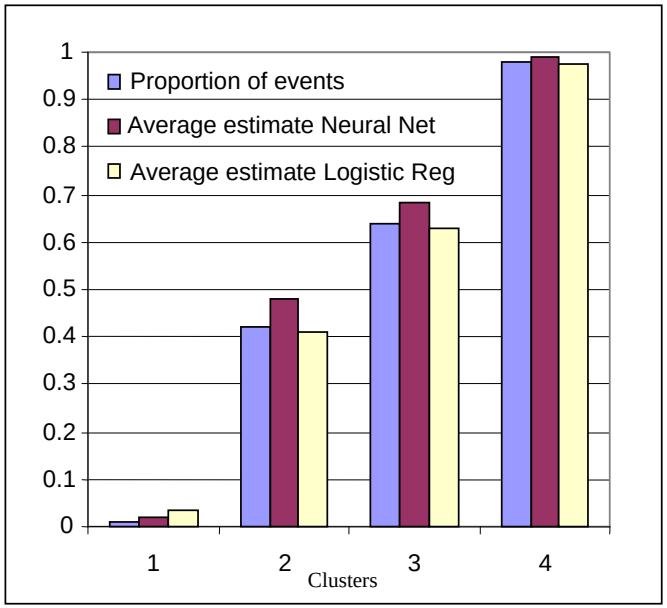
Clusters 1 to 4 centered on (0,0), (0,1), (1,0), and (1,1), with Gaussian noise

True probability for clusters 1 to 4: 0.01, 0.40, 0.60, and 0.99

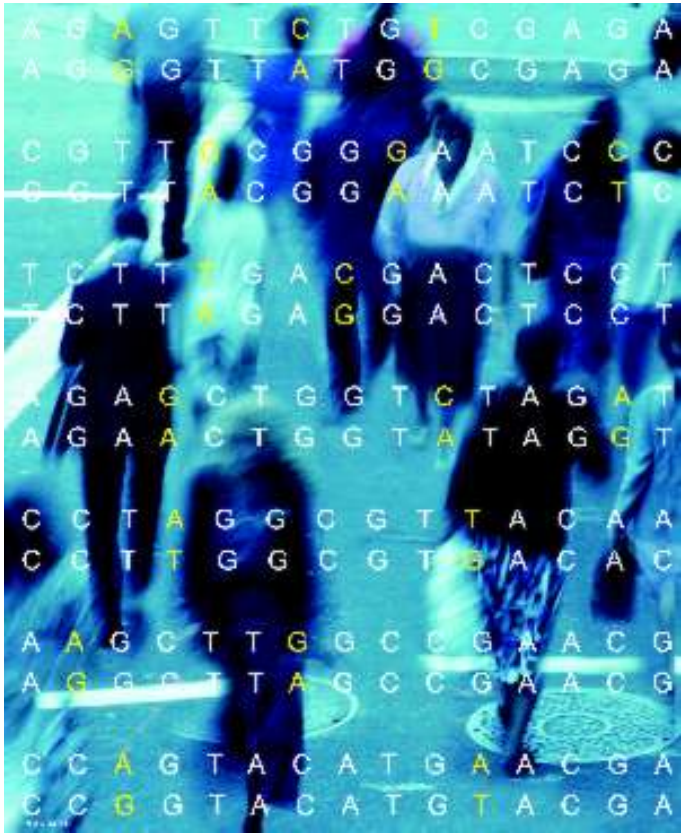

1.4
1.2
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.5
0
0.5
1
1.5
x2
x1
ANN
Cluster 1, Y=0
Cluster 1, Y=1
Cluster 2, Y=0
Cluster 2, Y=1
Cluster 3, Y=0
Cluster 3, Y=1
Cluster 4, Y=0
Cluster 4, Y=1



	LR	Neural Network
Sum of squared errors	52.363	50.894
Mean squared error	0.130	0.127
Cross-entropy error	154.543	150.838
Mean cross-entropy error	0.386	0.377
Sum of residuals	103.226	100.412
Mean residual	0.2580	0.251
AUC	0.889	0.895
HL-C	6.437	11.773
p	0.598	0.161



	LR	Neural Network
Cluster 2 min (GS: 0.4)	.20	.43
Cluster 2 max	.80	.58
Cluster 3 min (GS: 0.6)	.29	.65
Cluster 3 max	.85	.73



Source: DOE

Genotype — genome

transcription

RNA — transcriptome

translation

Protein — proteome

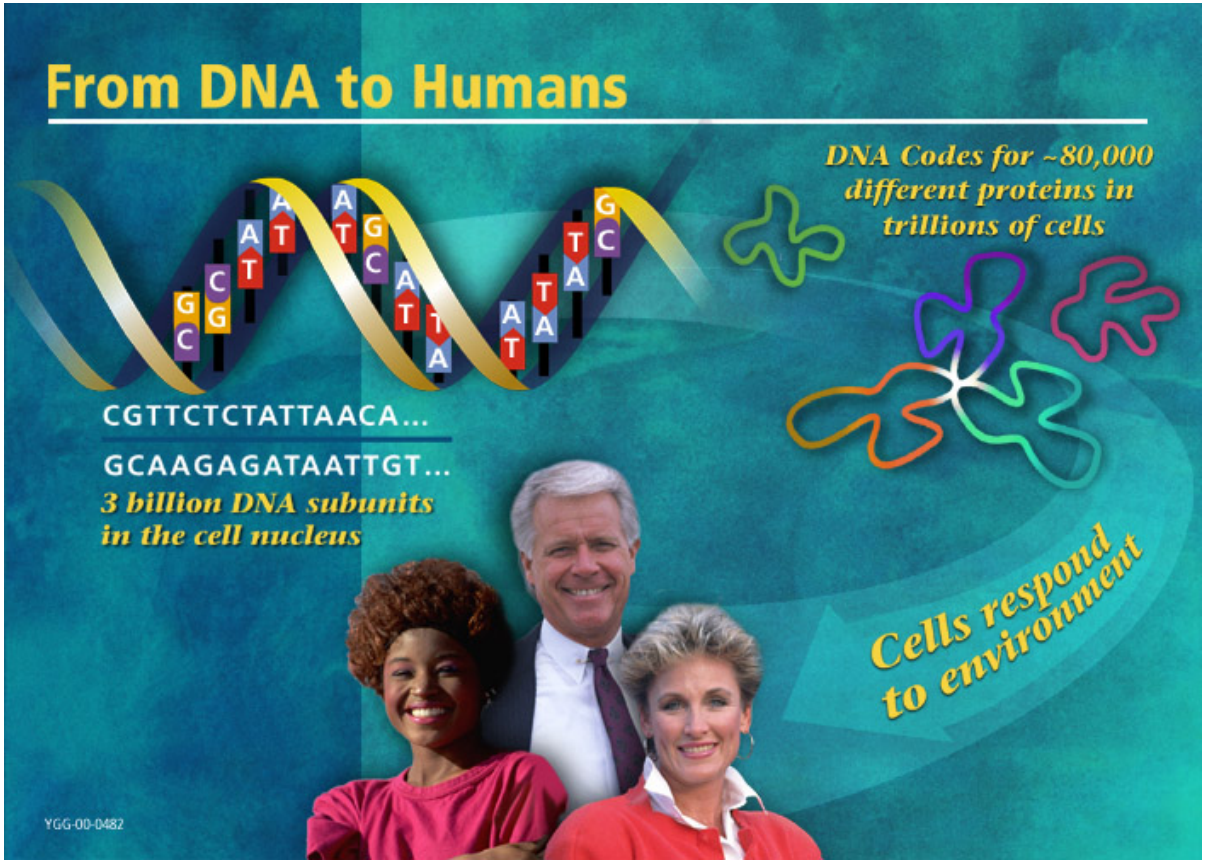
Phenotype — physical exam, imaging

Physiology — tests

Metabolites — laboratory

Will we ever achieve “individualized” risk assessment?

If so, how can we evaluate it?



Acknowledgments

- NIH, Komen Foundation
- Fred Resnic, Michael Matheny