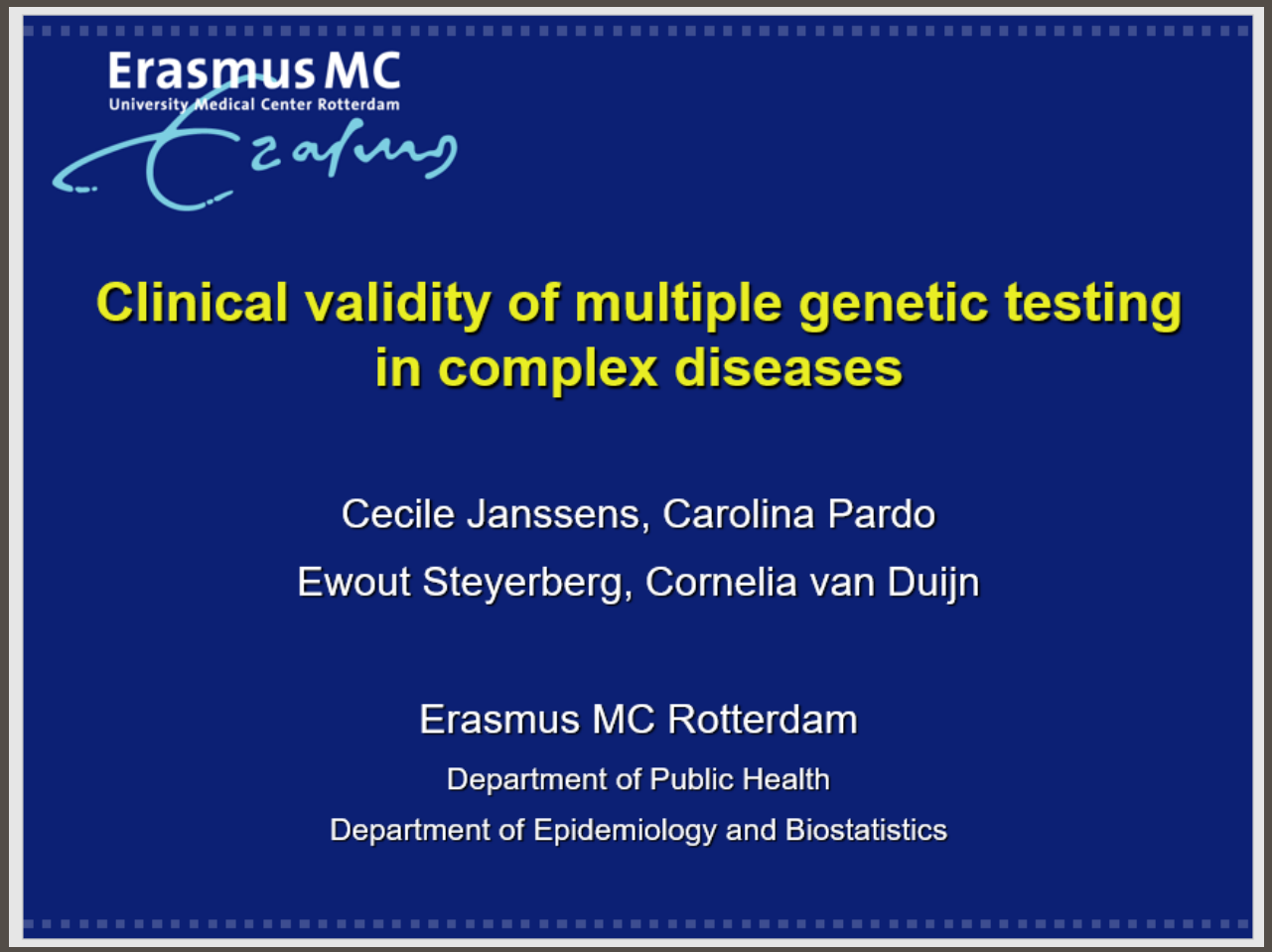
EMORY | ROLLINS SCHOOL OF PUBLIC HEALTH

Polygenic Risk Research

Lessons Learned From the Pre-GWAS days

A. Cecile J.W. Janssens, PhD
Research professor of epidemiology
Department of Epidemiology
@cecilejanssens

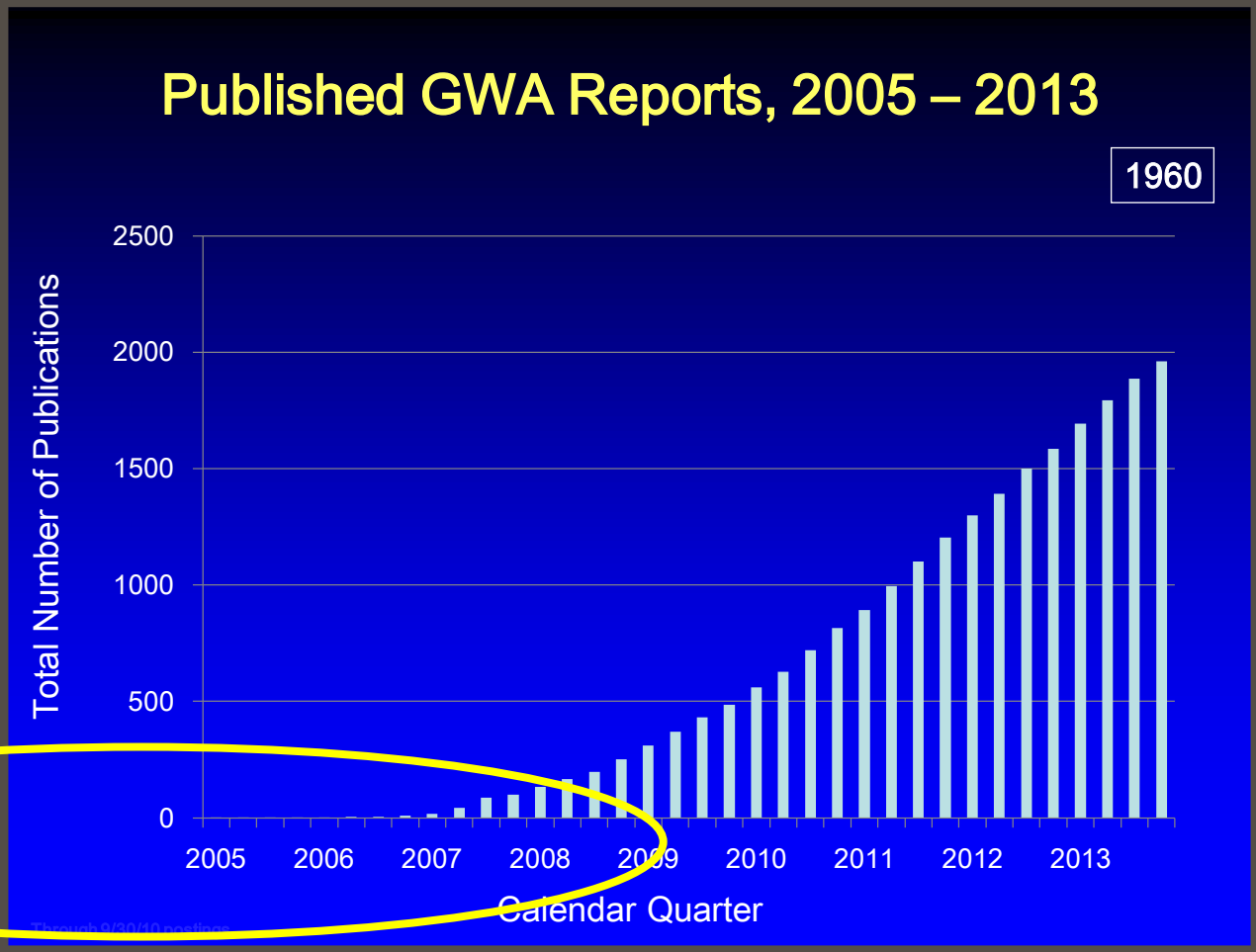
My first presentation polygenic risk



October 2003, UCSF Seminar

EMORY

Many current issues in prediction look like what was discussed in pre-GWAS days



GWAS Catalog, downloaded June 2, 2015

EMORY

1998

The new genetics

The new genetics in clinical practice

John Bell

1999

1998 ASHG PRESIDENTIAL ADDRESS
Making Genomic Medicine a Reality

Arthur L. Beaudet

2000

The New England Journal of Medicine

WILL GENETICS REVOLUTIONIZE MEDICINE?

NEIL A. HOLTZMAN, M.D., M.P.H.
THERESA M. MARTEAU, PH.D.

2001

Misconceptions about the use of genetic tests in populations

Paolo Vineis, Paul Schulte, Anthony J McMichael

First mentions of genetic information, susceptibility for common diseases, not yet polygenic models

EMORY

2002

article

Polygenic susceptibility to breast cancer and implications for prevention

Paul D.P. Pharoah[1,2], Antonis Antoniou[3], Martin Bobrow[4], Ron L. Zimmern[2], Douglas F. Easton[3] & Bruce A.J. Ponder[1]

Published online: 4 March 2002, DOI: 10.1038/ng853

- First mention of risk distributions
- Fitted on cancer data from relatives of BC patients
- Concluded that polygenic model fitted well
- No mention of individual variants or how to build polygenic risk models

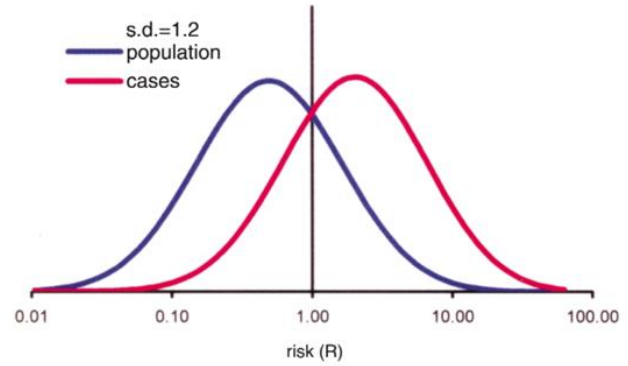


Fig. 1 Distribution of breast cancer risk in the population and in individual cases. Risks are shown on a log scale; the arithmetical average risk for the entire population has been set at 1.0 (see Methods). The risk distribution in individuals who will develop breast cancer (cases) is shifted to the right. The standard deviation describes the spread of risk between high and low values within the population, and thus the potential to discriminate different levels in different individuals.

EMORY

2003

Am. J. Hum. Genet. 72:636–649, 2003

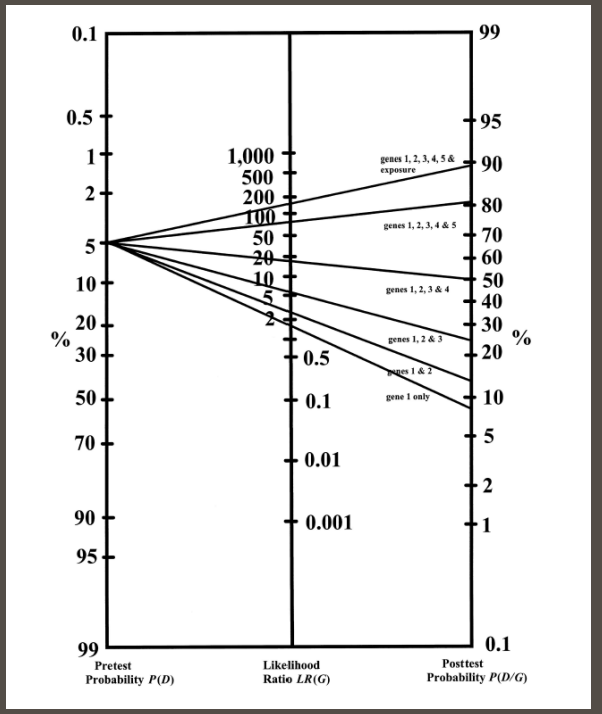
Improving the Prediction of Complex Diseases by Testing for Multiple Disease-Susceptibility Genes

Quanhe Yang,[1] Muin J. Khoury,[2] Lorenzo Botto,[1] J. M. Friedman,[4] and W. Dana Flanders[3]

[1]National Center on Birth Defects and Developmental Disabilities and [2]Office of Genomics and Disease Prevention, Centers for Disease Control and Prevention, and [3]Department of Epidemiology, School of Public Health, Emory University, Atlanta; and [4]Department of Medical Genetics, University of British Columbia, Vancouver

- First study to show how multiple genes can be combined to predict risk, using regression analysis
- Focused on posterior risk for carriers of one or more multiple risk alleles

- (very strong per-allele effects by today's standards (RR 1.5-3.5))



EMORY

2004

Am. J. Hum. Genet. 74:585–588, 2004

Revisiting the Clinical Validity of Multiplex Genetic Testing in Complex Diseases

To the Editor:

The usefulness of genetic testing to identify high-risk patients for common multifactorial diseases is subject to debate. Optimism about the public health opportunities is counterbalanced with skepticism, since genetic factors appear to play a role in only a minority of patients with complex diseases, the number of genes involved is large, and their penetrance is incomplete (Holtzman and Marteau 2000; Vineis et al. 2001).

A. CECILE J. W. JANSSENS,[1] M. CAROLINA PARDO,[2] EWOUT W. STEYERBERG,[1] AND CORNELIA M. VAN DUIJN[2]

Am. J. Hum. Genet. 74:588–589, 2004

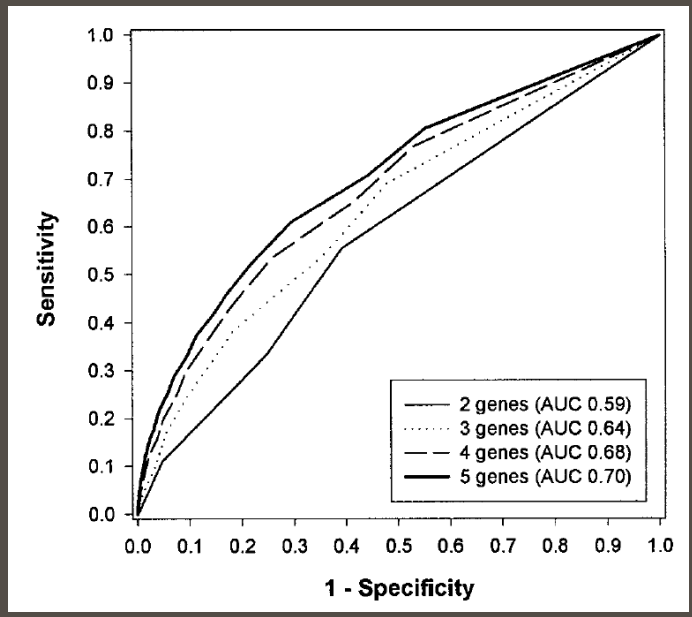
Revisiting the Clinical Validity of Multiplex Genetic Testing in Complex Diseases: Reply to Janssens et al.

To the Editor:

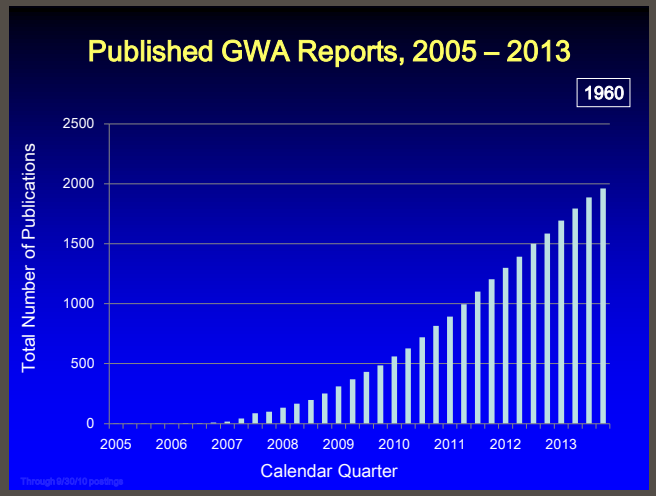
We appreciate the comments by Janssens and her associates (2004 [in this issue]) regarding our study on the use of likelihood ratios to improve the prediction of complex diseases by testing for multiple-susceptibility genes (Yang et al. 2003). As Janssens et al. correctly point out, our study considers only the predicted probability of disease for subjects who have all positive testing results, and this is likely to be an infrequent occurrence. We think that the suggestion made by Janssens et

QUANHE YANG,[1] MUIN J. KHOURY,[2] LORENZO BOTTO,[1] J. M. FRIEDMAN,[4] AND W. DANA FLANDERS[3]

- Evaluation of test performance should include all people, also noncarriers of risk alleles
- Proposed using Area under the Receiver Operating Curve (AUC)



EMORY



Pre-GWAS → no SNP data to work with

Two major advantages:

- Had to use simulated data: all parameters (# SNPs, ORs, allele freqs, population risk) can be varied to investigate and help understand impact on predictive performance of polygenic risk
  → If simulation is valid, then its observations apply to real data too
- Were not in a hurry: there was time to think about how to evaluate polygenic risk

EMORY

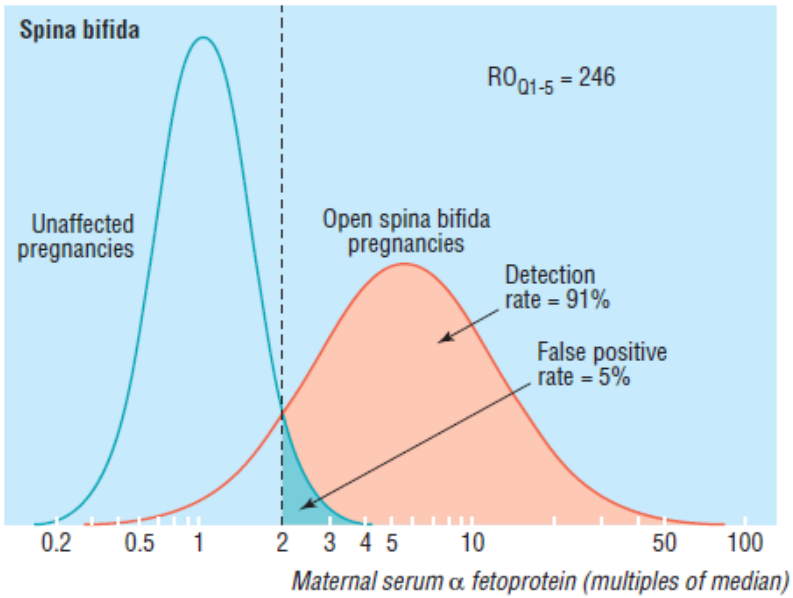
1999

When can a risk factor be used as a worthwhile screening test?

N J Wald, A K Hackshaw, C D Frost

Summary points

To be a worthwhile screening test, a risk factor must be strongly associated with a disorder



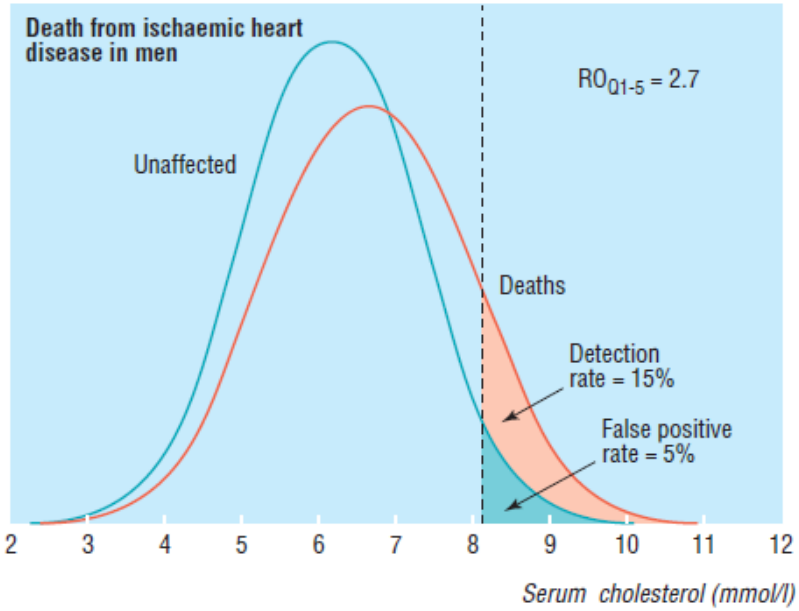
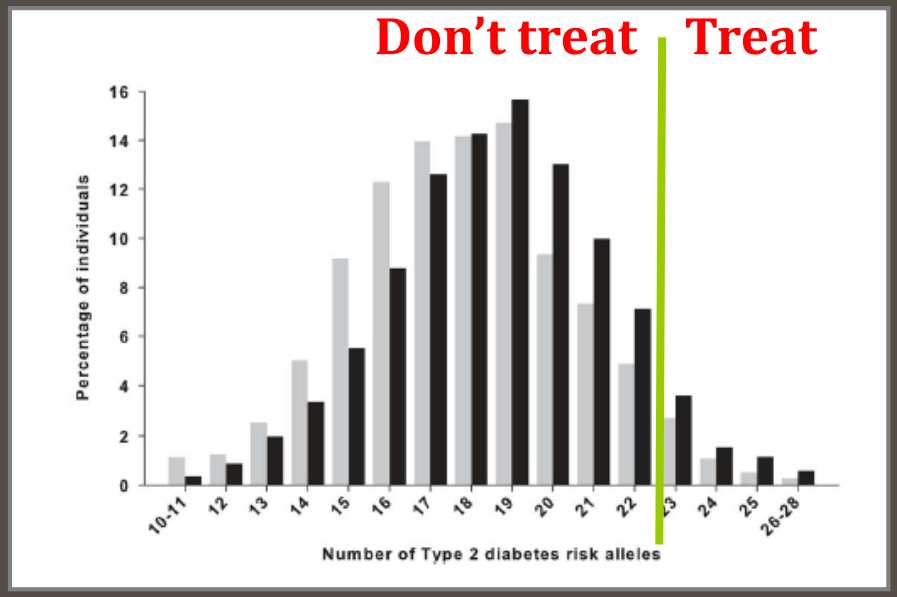


Fig 4 Distribution of maternal serum α fetoprotein in pregnancies affected and unaffected by open spina bifida (derived from Wald et al[2]) and distribution of serum cholesterol in men who did and did not die of ischaemic heart disease (derived from Wald et al[1])

BMJ VOLUME 319 11 DECEMBER 1999 www.bmj.com

EMORY

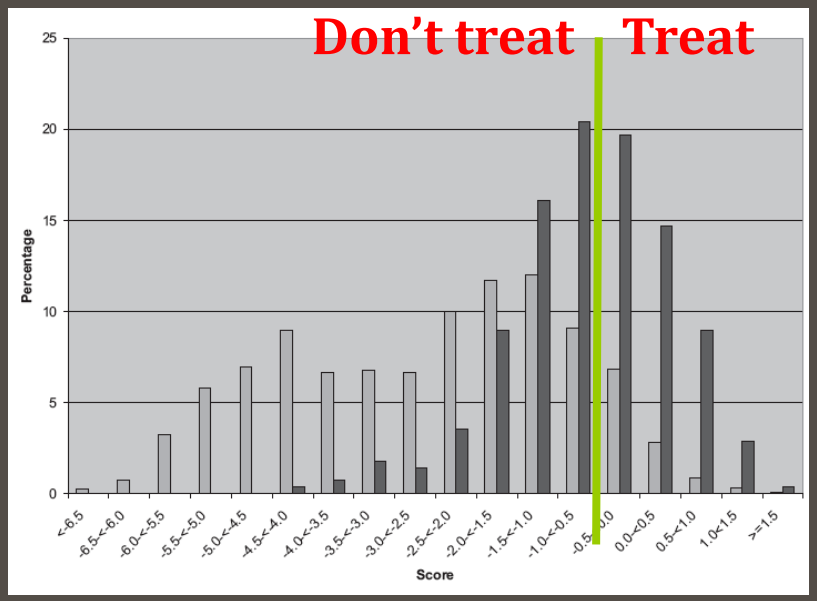
Type 2 diabetes



Lango et al *Diabetes* 2008

AUC = 0.60

AMD



Seddon et al. *IOVS* 2009

AUC = 0.76

AUC = degree of separation between risk distributions of affected and unaffected individuals—nothing more, nothing less

0.50: complete overlap ~ random prediction
1.0: complete separation ~ perfect prediction

EMORY

How to get high AUC: common variants with strong effects

Type 2 diabetes
AUC = 0.60

TCF7L2	1.36	SLC30A8	1.10
KCNJ11	1.25	TSPAN8	1.09
CDKN2A/2B	1.21	CDC123	1.10
PPARG	1.21	WFS1	1.07
ADAM30	1.15	TCF2	1.07
CDNK2A/2B	1.13	ADAMTS9	1.05
IGF2BP2	1.12	HHEX-IDE	1.02
FTO	1.11	THADA	1.04
CDKAL1	1.11	JAZF1	1.00

Hypertriglyceridemia
AUC = 0.80

APOA5 19WW	7.36
APOA5 -1131CC	5.57
APOE non-e3	2.14
GCKR TT	2.11
TRIB1 AA	2.02
TBL2 CC	2.81
GALNT2 GG	2.10

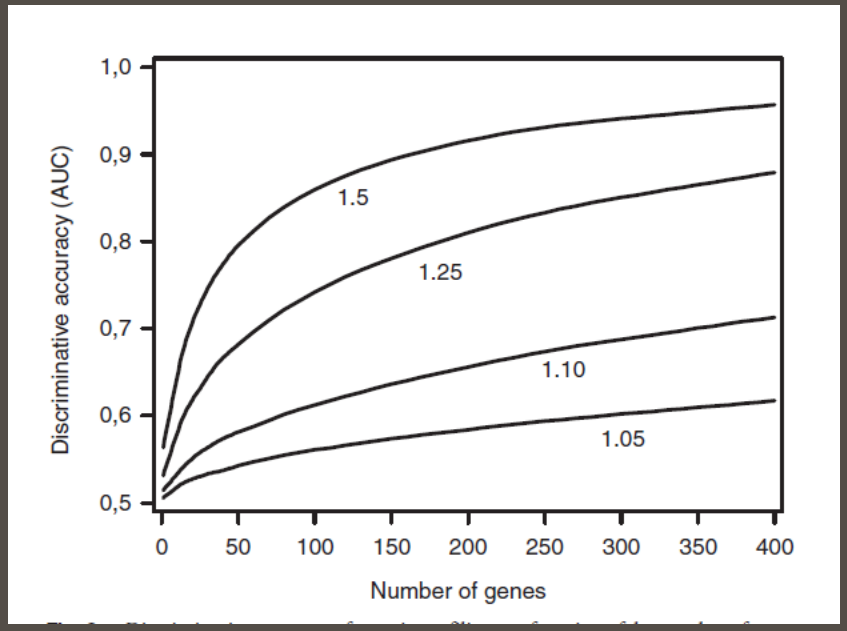
Lango et al *Diabetes* 2008; Wang et al. Hum Mol Genet 2008

EMORY

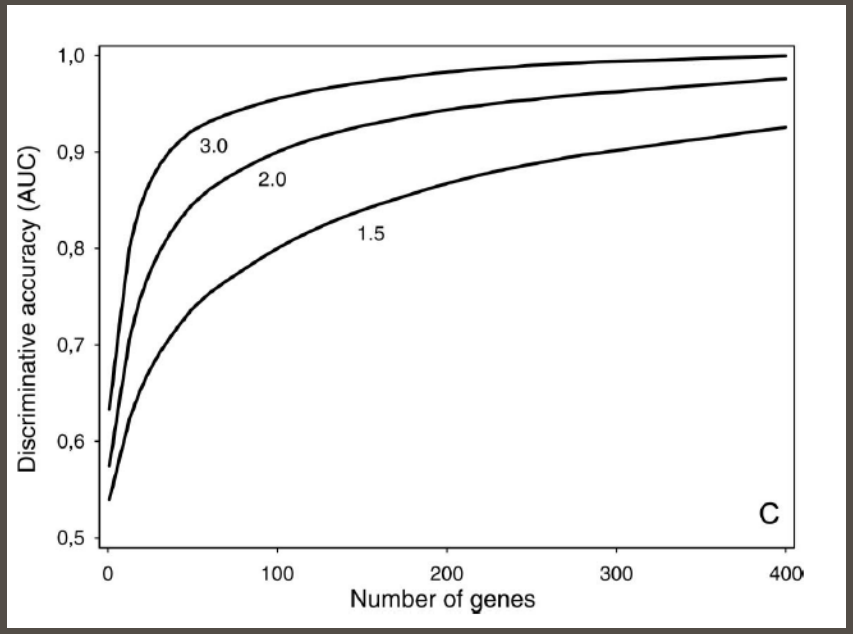
No exception: only strong variants lead to higher AUC

(higher AUC = more separation risk distributions)

Simulation study: impact of number of genes and OR on AUC



All variants same OR



First 20 variants: ORs from <max> to 1.15
Allele freq from 0.05 to 0.30
Next 380 variants: OR from 1.15 to 1.05
Allele freq from 0.30 to 0.50

Janssens et al. Genet Med 2006

EMORY

Predictive performance polygenic risk scores

- Mostly modest: AUC up to ~0.65
- AUC generally (much) lower than clinical prediction models
- Modest improvement beyond clinical models
- Exceptions when some SNPs have stronger effects, e.g., age-related macular degeneration, Crohn disease

- Can we do better?

Quality of Prediction = quality of data & quality of prediction model

Data	Model	Prediction
Excellent	Excellent	Excellent
Excellent	Poor	Poor
Poor	Excellent	Poor
Poor	Poor	Poor

Much focus on data these days:
how are we doing on modeling risk?

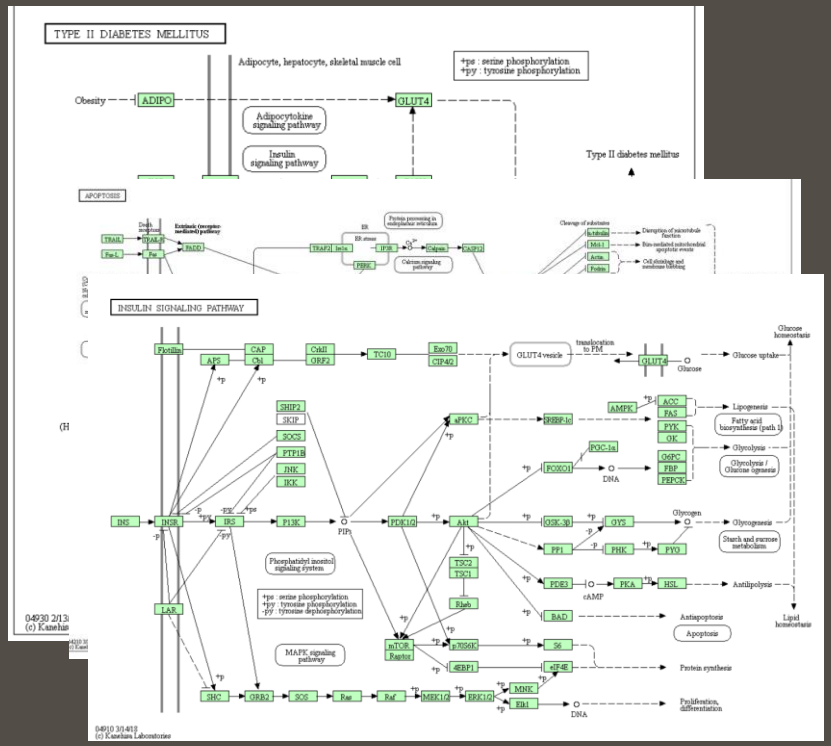
EMORY

Polygenic risk score poor reflection of pathways

Polygenic risk score

$$score = \beta_1 * snp_1 + \beta_2 * snp_2 + \cdots \beta_n * snp_n$$

Molecular pathways

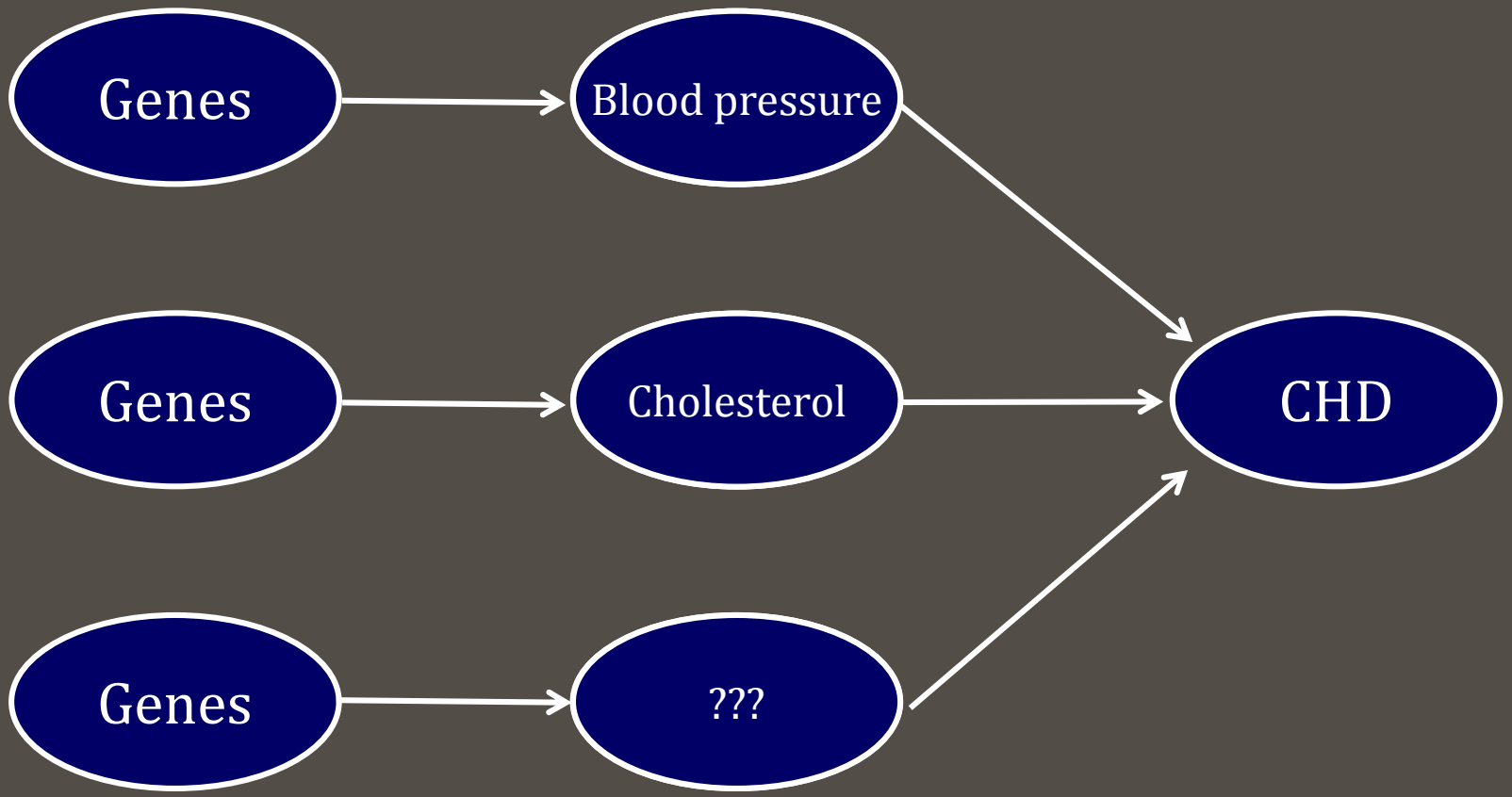


EMORY

How about modeling clinical + genetic models?

EMORY

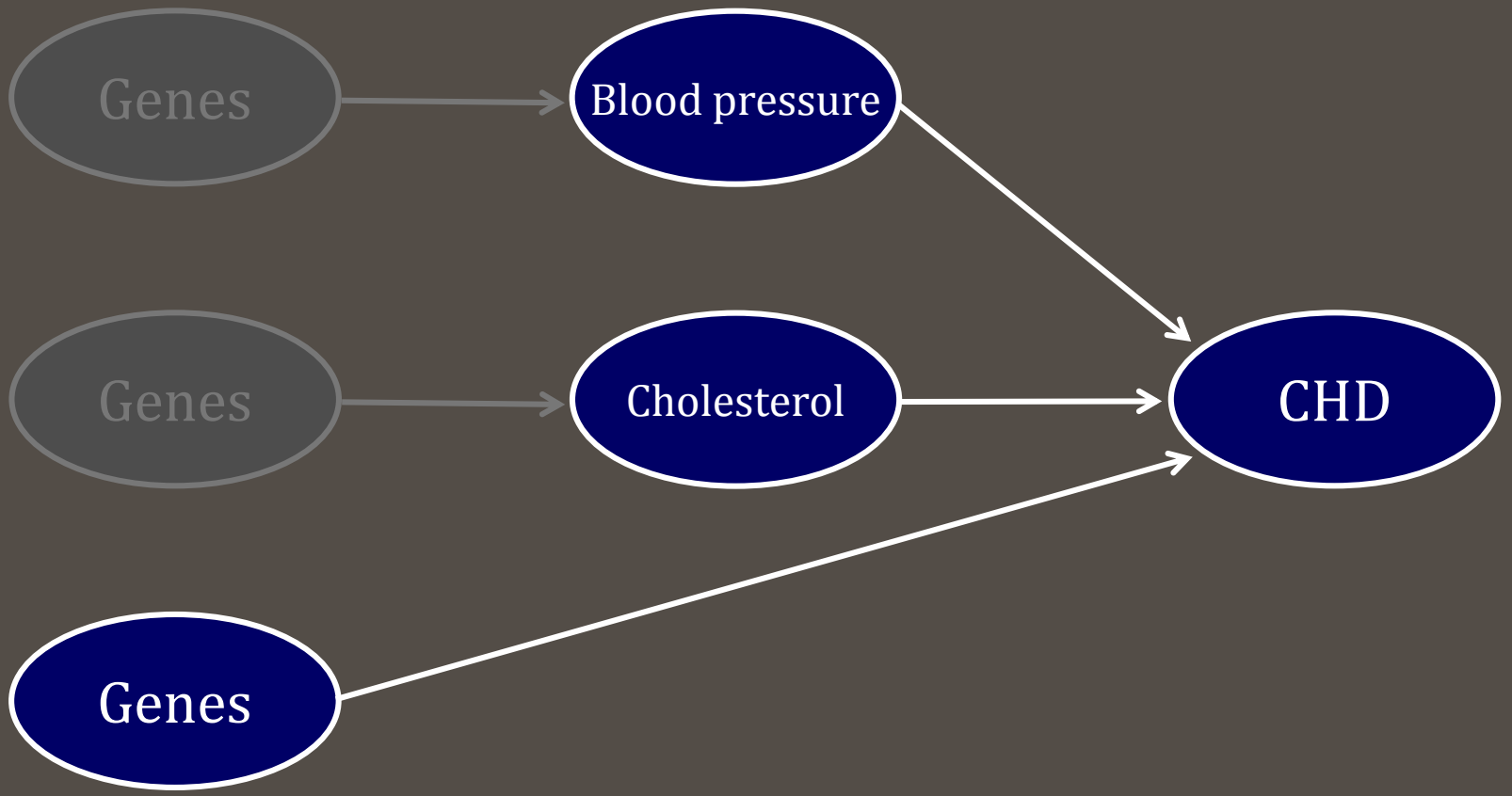
From pre-GWAS:
genes 'only' improve prediction if not mediated



Adapted from: Janssens & van Duijn, Hum Mol Genet 2008

EMORY

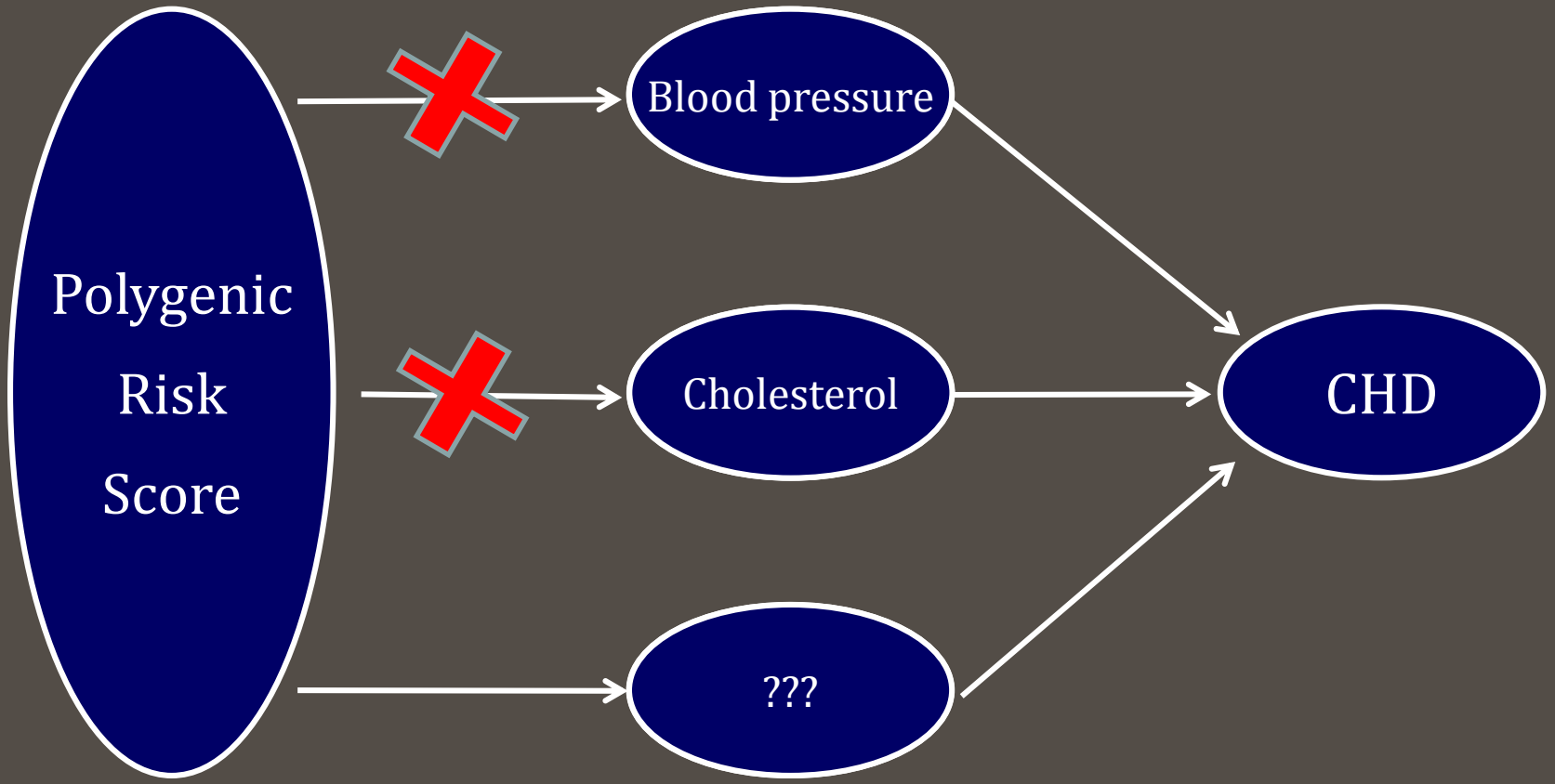
From pre-GWAS:
genes ‘only’ improve prediction if not mediated



Adapted from: Janssens & van Duijn, Hum Mol Genet 2008

EMORY

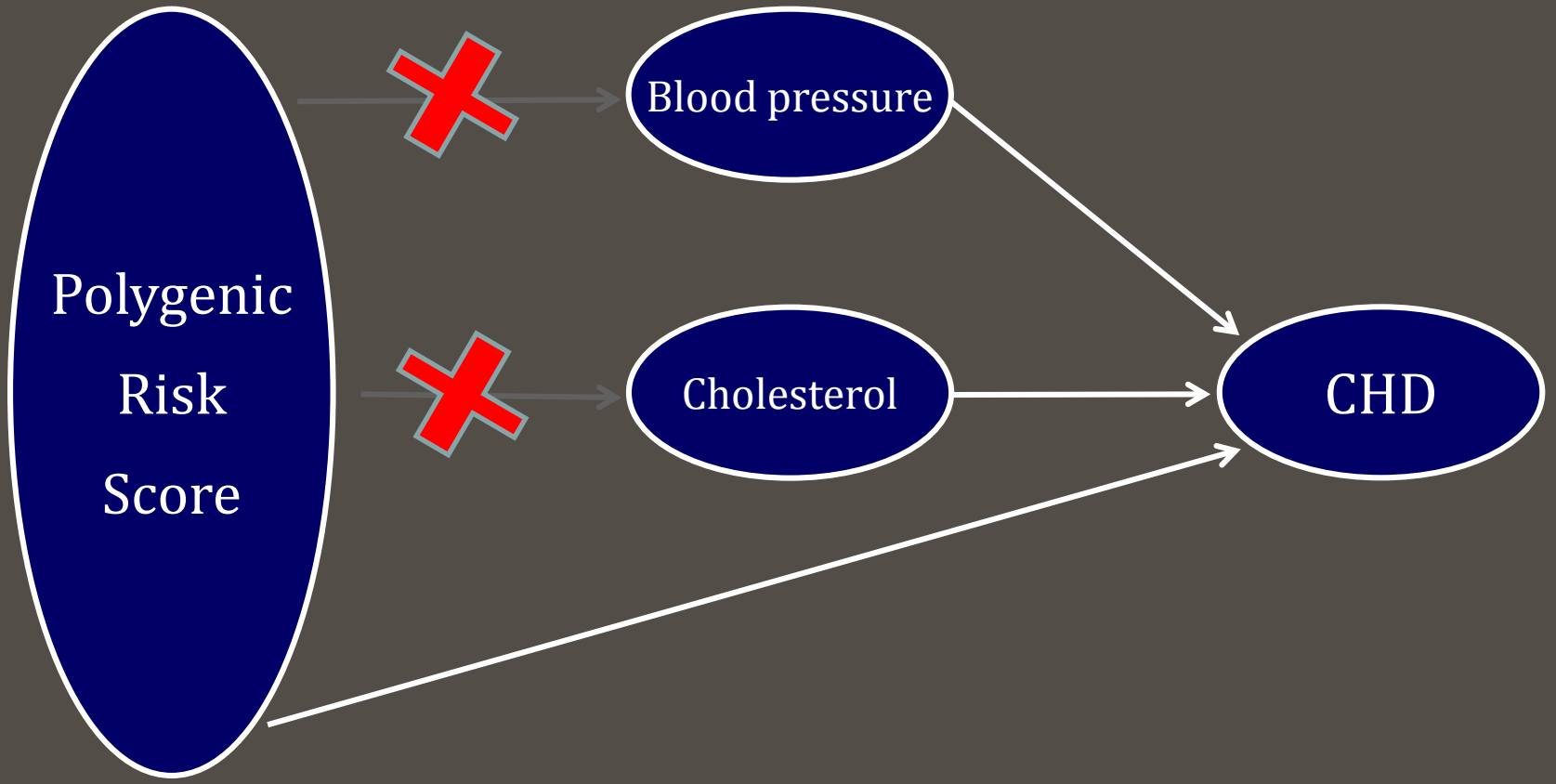
When predisposing genes are *combined* in polygenic risk score, the resulting score is no longer related to each clinical risk factor



Janssens et al. *Submitted*

EMORY

Because polygenic risk score is no longer associated to clinical risk factors, score seems independent risk factor

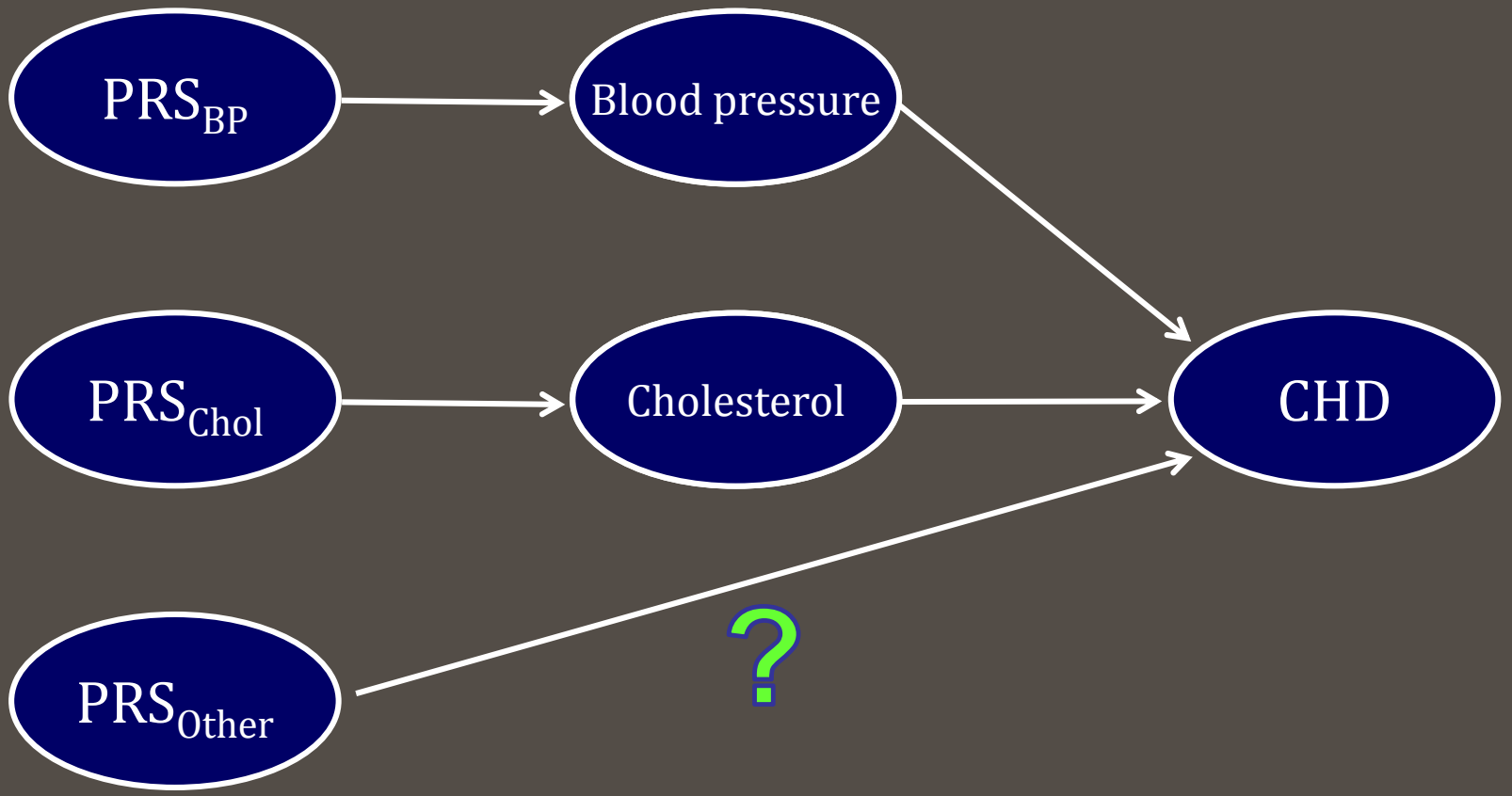


Janssens et al. *Submitted*

EMORY

Modeling polygenic risk scores should be improved so that clinical risk factors get opportunity to mediate

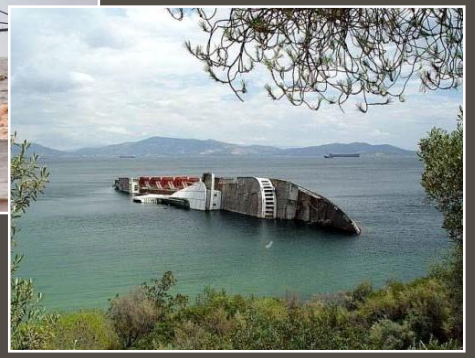
e.g.:



Janssens et al. *Submitted*

EMORY

Can causal mechanisms of complex outcomes be modeled or are their causes too complex?



EMORY

Herald of Free Enterprise

Capsized on March 6 1987, killing 193 people

Capsized because multiple factors happened simultaneously, among which:

- Bow doors open: responsible employee had fallen asleep and there was no double checking of doors
- Full ballast tanks → ship lower on water
- Delayed departure → higher speed → higher waves
- Open car compartment, cars not secured → adding imbalance

Janssens & van Duijn *Hum Mol Genet 2008*

EMORY

Why-Because Graph

Herald of Free Enterprise, March 6th, 1987

Braband et al. 2003

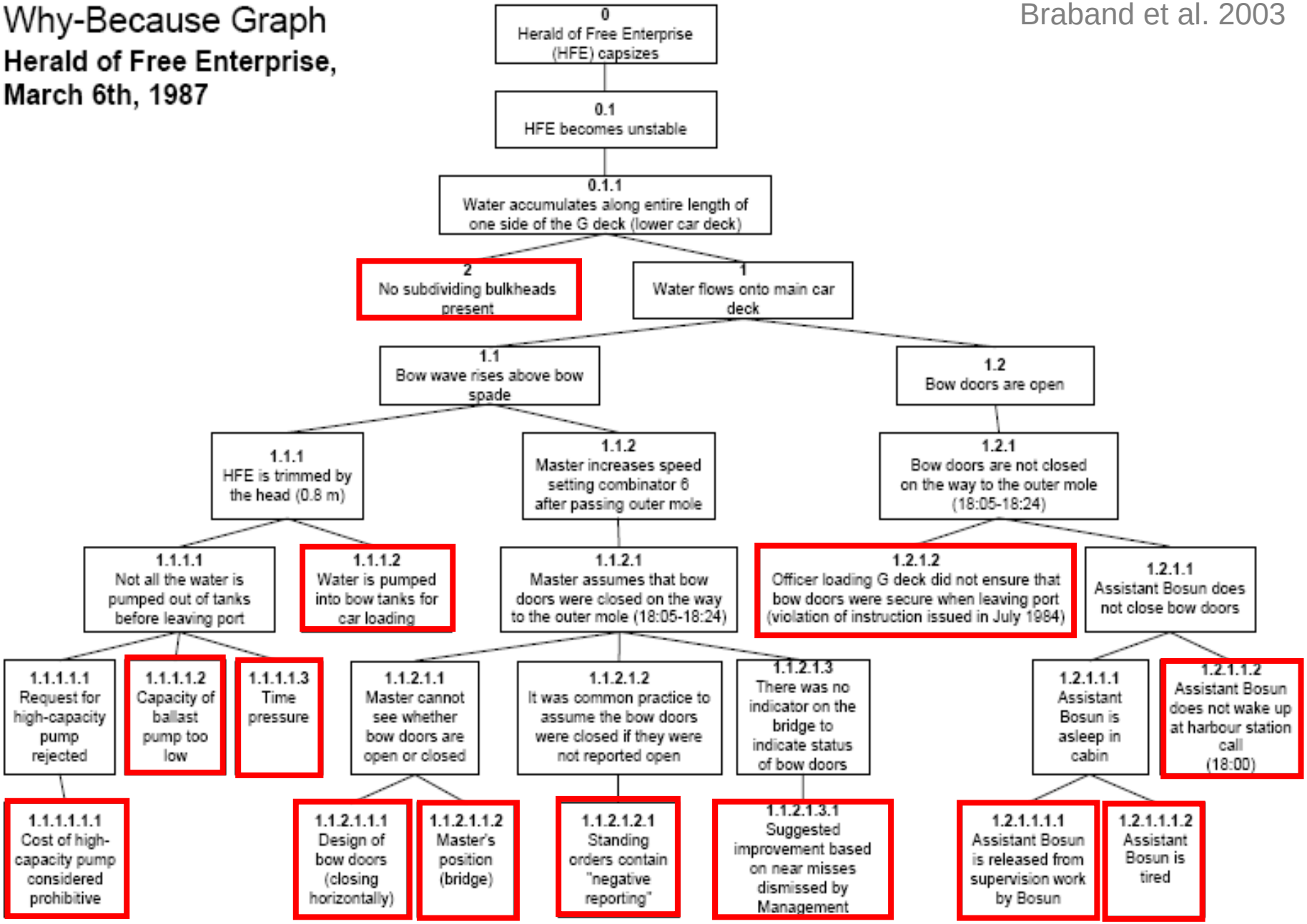


Figure 2 - Why-Because Graph

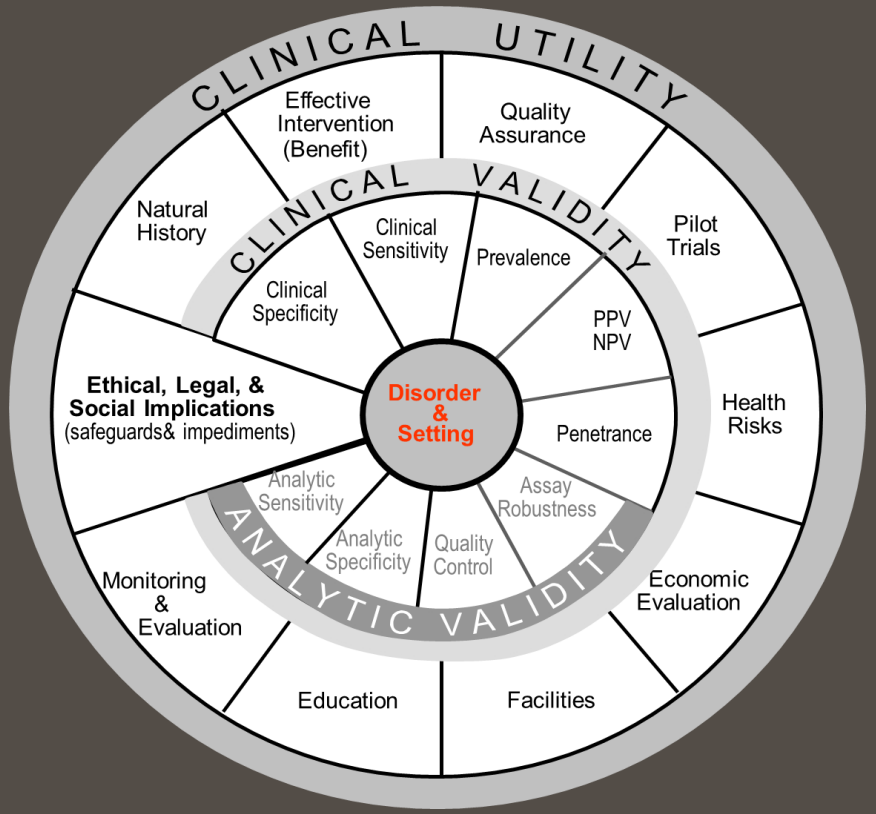
ACCE model: evaluating genetic tests

2003

- Comprehensive framework
- Key: Disorder & Setting: What is predicted in whom, for what purpose?
- Assessment changes if setting changes (different population or purpose)

- Claims often based on statistical significance of PRS association

  → Association determines clinical validity but itself is not part of evaluation



Haddow & Palomaki, Human Genome Epidemiology, 2003

EMORY

2008

Purpose: Increasing efficiency of healthcare

Polygenes, Risk Prediction, and Targeted Prevention of Breast Cancer

Paul D.P. Pharoah, Ph.D., Antonis C. Antoniou, Ph.D., Douglas F. Easton, Ph.D., and Bruce A.J. Ponder, F.R.S.

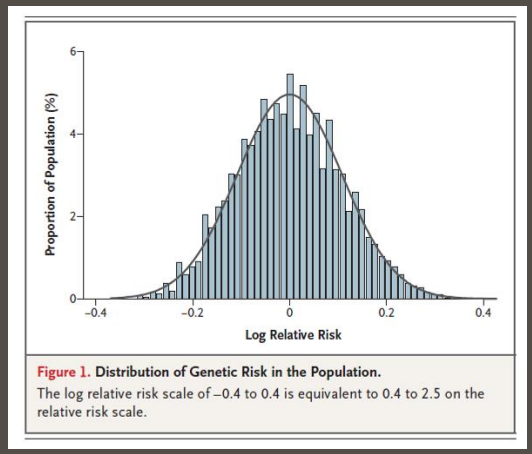


Figure 1. Distribution of Genetic Risk in the Population.
The log relative risk scale of −0.4 to 0.4 is equivalent to 0.4 to 2.5 on the relative risk scale.

Table 2. Absolute Risks of Breast Cancer According to Percentile of Population.*

Percentile of Population	Relative Risk	Lifetime Risk†	10-Yr Risk at 50 Yr of Age†	Age at Which 10-Yr Risk ≥2.3%
		%		yr
5	0.63	6.1	1.5	NA‡
10	0.69	6.7	1.6	NA‡
20	0.77	7.4	1.8	NA‡
40	0.90	8.6	2.1	53
60	1.03	9.7	2.4	49
80	1.20	11.0	2.7	45
90	1.35	12.0	3.0	43
95	1.49	14.0	3.4	41

Then: keeping healthcare costs the same, but redistribute efforts
Now: often proposing ‘new’ care to high-risk groups, but is more care affordable?

EMORY

Purpose: Changing health behavior

- Little (no?) evidence of long-term impact on health behavior
- Limitation: mostly simple tests or simple risk scores; impact unknown when polygenic risk scores are really predictive
- Future: not one PRS, but for every disease

Genetic testing or reduction of exposures

The major diseases in western societies are multifactorial. Thus, lung cancer is not wholly attributable to smoking, but to many linked factors of which smoking is one.

A: Exposure	Disease	Proportion attributable to exposure*
Tobacco smoke	Lung cancer	90%
	Bladder cancer	70% (men)/30% (women)
	Larynx cancer	90%
	Coronary heart disease	12·5%
	Chronic bronchitis	80%
B: Disease	**Low-penetrant genes**	**Odds ratio†**
Lung cancer	*CYP1A1 Msp* I (Asian)	1·73
	CYP1A1 Msp I (white)	1·04
	CYP1A1 exon 7 (Asian)	2·25
	CYP1A1 exon 7 (white)	1·30
	CYP2D6	1·26
	GSTM1	1·34
Bladder cancer	*NAT-2* slow	1·37
	GSTM1	1·57
Colon cancer	*NAT-2* rapid	1·19

*Ref 18. †Ref 5.

Table 2: **An example of one exposure resulting in many diseases (A) and one disease resulting from low-penetrant genes (B)**

Vineis et al., *Lancet,* 2001

What is behavioral response when:

PRS report

High : CVD

Average: type 2 diabetes
dementia

Low: obesity
asthma
depression

EMORY

Moving forward

- Improve modeling to better reflect underlying mechanisms
  - May increase predictive performance of polygenic risk scores
  - May reduce their value added to clinical factors

- Improve assessment of potential utility of polygenic risk scores
  - Assess scores in target population
  - Apply appropriate performance metrics
  - Interpret in appropriate context: predictive enough? Actionable/informative? Affordable?
  - Compare with existing (nongenetic) risk models

2000

The New England Journal of Medicine

Will Genetics Revolutionize Medicine?

In our rush to fit medicine with the genetic mantle, we are losing sight of other possibilities for improving the public health. Differences in social structure, lifestyle, and environment account for much larger proportions of disease[42,43] than genetic differences. Although we do not contend that the genetic mantle is as imperceptible as the emperor's new clothes were, it is not made of the silks and ermines that some claim it to be. Those who make medical and science policies in the next decade would do well to see beyond the hype.

Neil A. Holtzman, M.D., M.P.H.
Theresa M. Marteau, Ph.D.

EMORY