



**KnowYourNumber™**  
The Key To Proactive Good Health



## Development of Disease Specific and Morbidity Based Health Risk Assessments

BioSignia, Inc.

Guizhou Hu, Ph.D.

Martin Root, Ph.D.

# Presentation Outline

---

- HRA vs. disease prediction models
- Developing disease prediction models - evidence-based approach
- Model validation
- Application of the models

# Health Risk Appraisal (HRA)

---

- HRA: a technique for estimating the odds that a person with certain characteristics will die within a given time span.
  - Mortality-based and non-disease specific
  - Lacks a link with health care cost and life quality
  - “Debit / credit” methodology has no basis in statistical probability theory
  - Simple, cheap and have been widely used

# Disease Onset Predication Model

---

- A technique for estimating the odds that a person with certain characteristics will have a specific disease within a given time span.

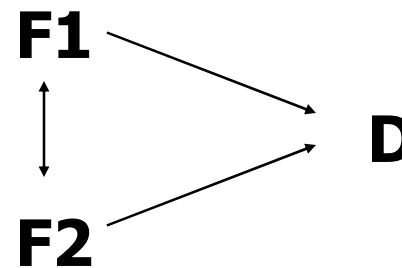
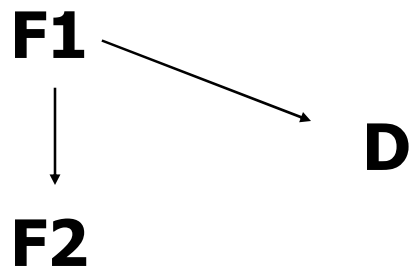
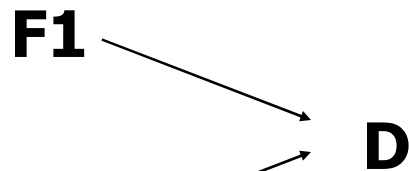
Compared to HRA's it is more informative for

- Patients
- Health economists
- Decision makers of health promotion

# The Challenge for Disease Risk Assessment

The complicated inter-correlation among risk factors

- Four possible ways that two risk factors are associated with disease risk



# Developing Disease Risk Assessment

---

## Empirical study

- Established traditional methodology
- Possible selection bias, small N
- Limited risk factors considered

## Evidence-based approach

- No selection bias, big N
- Comprehensive
- Lack of established methodology

# Evidence-Based Medicine

---

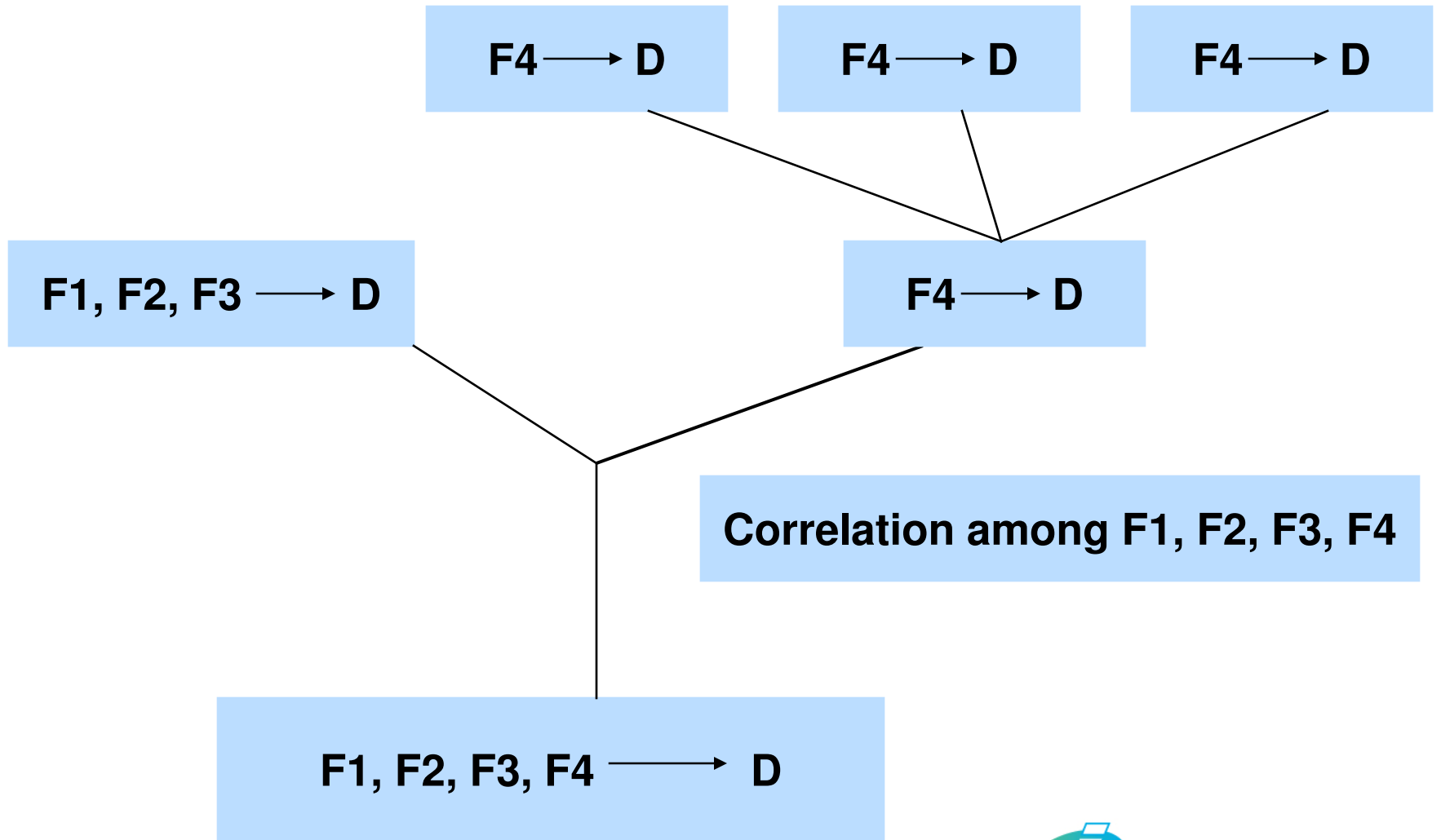
- Conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients.
- Integrating the best available external evidence from systematic research with individual clinical expertise and the patient's values.

# Development of BioSignia's Evidence-Based Disease Prediction Models

---

Develop a methodology that can integrate results from disparate studies:  
“synthesis analysis”

# Synthesis Analysis



# Synthesis Analysis

---

**disease risk = age , sex**

**disease risk = BMI**

**correlation  
among, age,  
sex, BMI**

**disease risk = age , sex , BMI**

# Development of Coronary Heart Disease (CHD) Prediction Model

---

## Identify all risk factors related to CHD

- age, sex, BP, cholesterol, smoking, diabetes, obesity, physical activity, family history, HRT, fibrinogen, albumin, lipoprotein(a), c-reactive protein, homocysteine, etc.

## Search for published prediction model

- Framingham CHD prediction model

# Framingham CHD Prediction Equation

- Male:  $m = 11.1122 - 0.9119 \cdot \log(\text{sbp}) - 0.2767 \cdot \text{smoke} - 0.7181 \cdot \log(\text{total}/\text{HDL}) - 0.5865 \cdot \text{LVH} - 1.4792 \cdot \log(\text{age}) - 0.1759 \cdot \text{diabetes}$
- Female:  $m = 11.1122 - 0.9119 \cdot \log(\text{sbp}) - 0.2767 \cdot \text{smoke} - 0.7181 \cdot \log(\text{total}/\text{HDL}) - 0.5865 \cdot \text{LVH} - 5.8549 + 1.8515 \cdot (\log(\text{age}/74))^2 - 0.3758 \cdot \text{diabetes}$
- $u = (\log(t) - 0.4181 - m) / \exp(-0.3155 - 0.2784 \cdot m)$
- $p = 1 - \exp(-\exp(u))$

Keaven Anderson et al.

***“An updated coronary risk profile--a statement for health professionals”***

**AHA statement Circulation 83:356-62, 1991**

# Risk Factors Not Considered in the Framingham Model

---

- Fibrinogen
- Lipoprotein(a)
- HRT
- Exercise
- Family history
- Albumin
- Homocysteine
- C-reactive protein

## Effect of Other Risk Factors on CHD

---

Physical exercise RR = 1.9, 36 cohort studies

HRT RR = 0.56, 16 prospective studies

Fibrinogen RR = 1.8 every 0.1g/dL, 18 studies

Albumin RR = 0.67 every 4g/L, 8 studies

Lipoprotein(a) RR = 1.9 every 50 mg/dL, 2 studies

# Synthesis Analysis

**CHD = gender, age, cholesterol, smoking, LVH, diabetes, BP**

**CHD = family history**

**correlation among gender, age, cholesterol, smoking, LVH, diabetes, BP, family history**

**CHD = gender, age, cholesterol, smoking, LVH, diabetes, BP, family history**

# Example

- male
- age 55
- sbp=150 mmHg
- smoker
- total cholesterol=210
- HDL=60
- no diabetes
- no LVH
- with family history
- light physical activity
- albumin=4.0
- fibrinogen=300
- lipoprotein(a)=10
- not use aspirin
- homocysteine=10
- CRP=0.8

● Framingham prediction 11.8%

● BioSignia prediction 14.8%

- male
- age 55
- sbp=150 mmHg
- smoker
- total cholesterol=210
- HDL=60
- no diabetes
- no LVH
- no family history
- high physical activity
- albumin=4.5
- fibrinogen=310
- lipoprotein(a)=5
- not use aspirin
- homocysteine=15
- CRP=0.5

● Framingham prediction 11.8%

● BioSignia prediction 8.5%

# Advantage of Synthesis Analysis

---

- Uses a novel technique to take full advantage of all the current scientific knowledge about disease risk
- Continuously updateable

# Validation

---

- Validity of each study that they are synthesized from
- Validity of the synthesis algorithms
- Validity of the statistical assumptions

# Validation of the Synthesis Methodology

---

## Validate the algorithm

- Patent # 6110109
- Publication
- Evaluation by experts

## Empirical validation

- NHANES III data
- Framingham data

# Comparison of an Empirical Model with the Synthesis Model

---

Build an empirical CHD model  
using Framingham data

$$P1 = 1 / (1 + \exp(-4.6225 + 0.0142 * \text{age} + 0.8595 * \text{sex} - 0.0623 * \text{smoke} + 0.00674 * \text{cholesterol} + 0.0107 * \text{sbp}))$$

Build a synthesis CHD model  
using the correlation from NHANES III data

$$P2 = 1 / (1 + \exp(-0.9765 + 0.7876 * \text{sex} + 0.0316 * (\text{age} - 44) + 0.005863 * (\text{cholesterol} - 224) + 0.002 * (\text{sbp} - 137) - 0.002 * (\text{smoke} - 0.577)))$$

Do a correlation test between p1 and p2

$$r = 0.95$$

# Comparison of Synthesis CHD Model with the Framingham Model

---

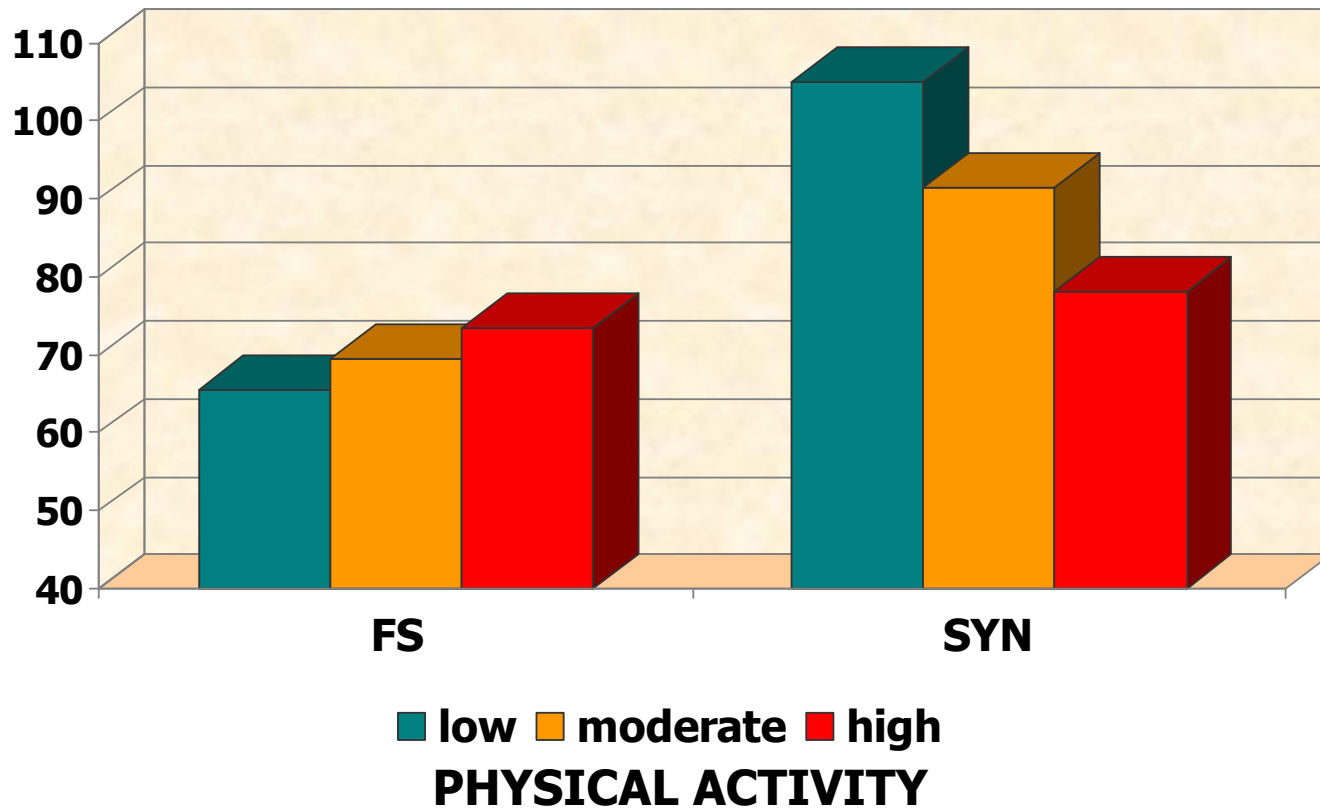
- Using NHANES III data to prove that a synthesis model cannot be worse than the Framingham model.
  - Synthesis model is as good as the Framingham model
  - Synthesis model is better than the Framingham model

# Synthesis Model Is As Good As the Framingham Model

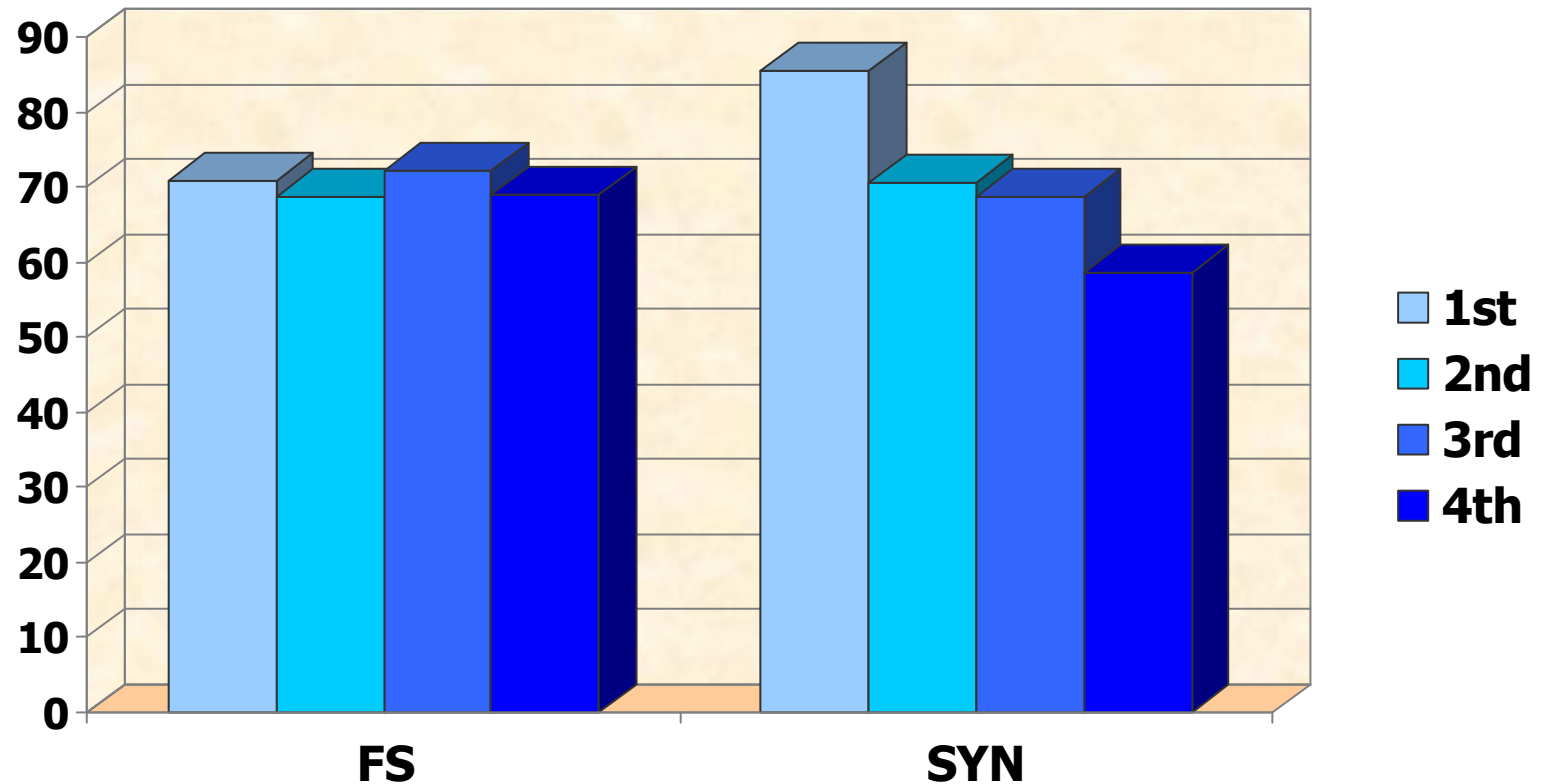
## Number of people will have CHD in 5 years among NHANSE III population

	Framingham model	Synthesis model
<b>Gender</b>		
Female	181 (160-200)	183 (161-202)
Male	372 (351-400)	372 (351-400)
<b>Age</b>		
35-44	23	23
45-54	73	72
55-64	139	138
65-74	182	183
75-84	135	138
<b>Smoking</b>		
YES	387	389
NO	166	166
<b>HDL/cholesterol quartile</b>		
1st	202	201
2nd	163	164
3rd	118	119
4th	69	70

# Synthesis Model Is Better Than the Framingham Model

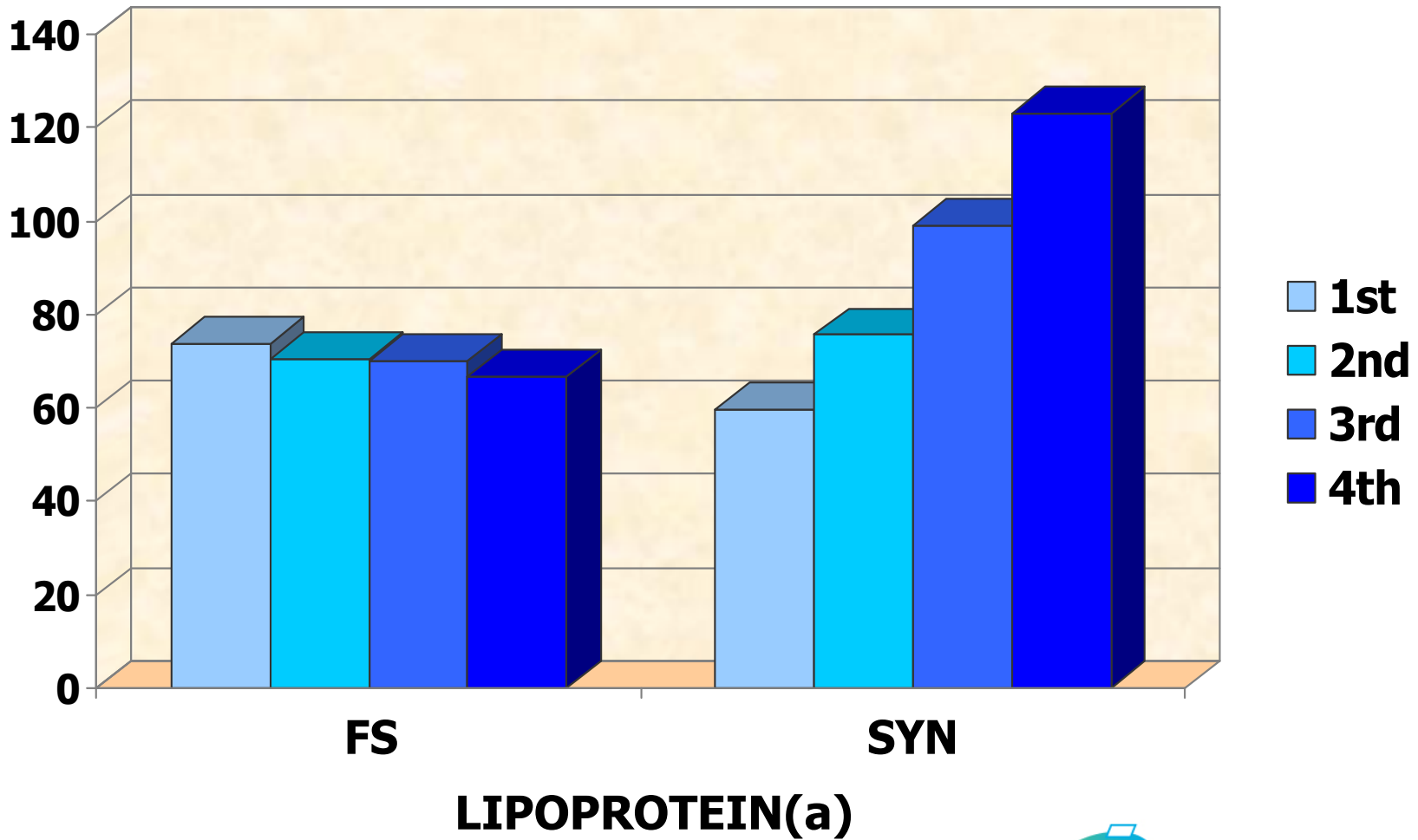


# Synthesis Model Is Better Than the Framingham Model

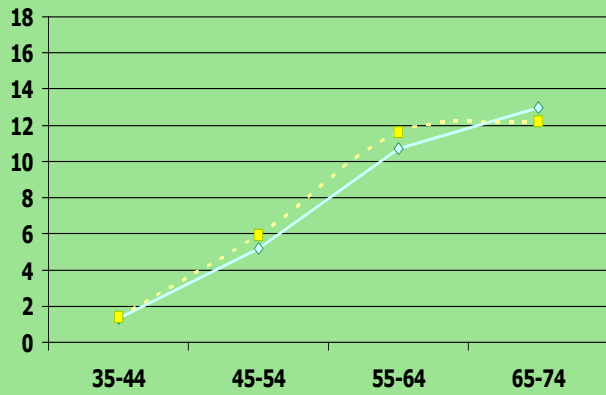


**ALBUMIN**

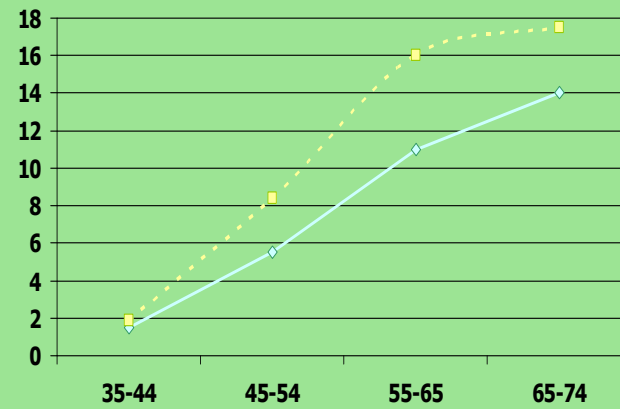
# Synthesis Model Is Better Than the Framingham Model



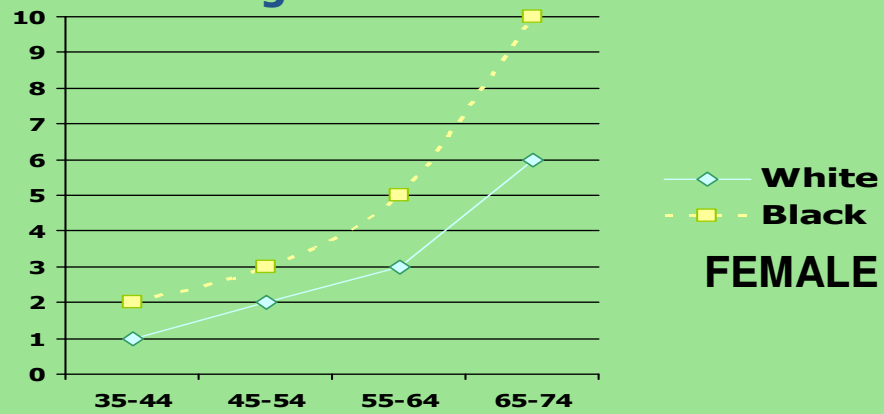
### Framingham model



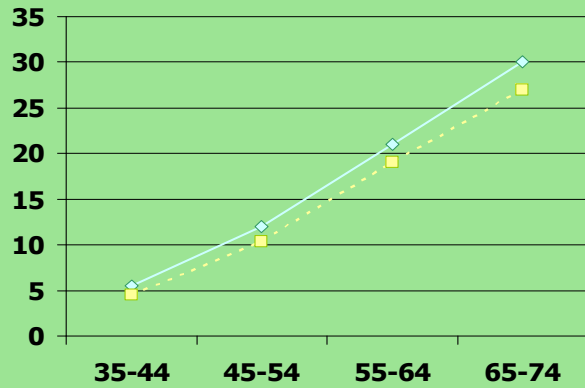
### Synthesis model



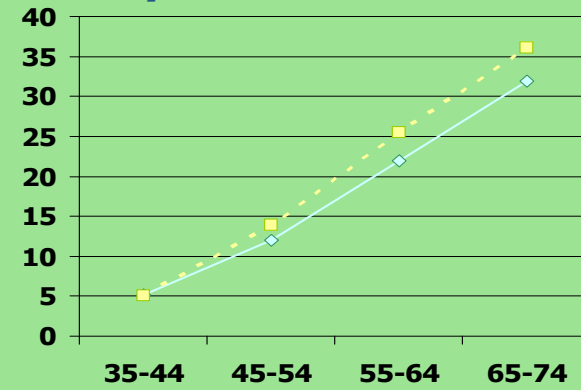
### Findings from ARIC



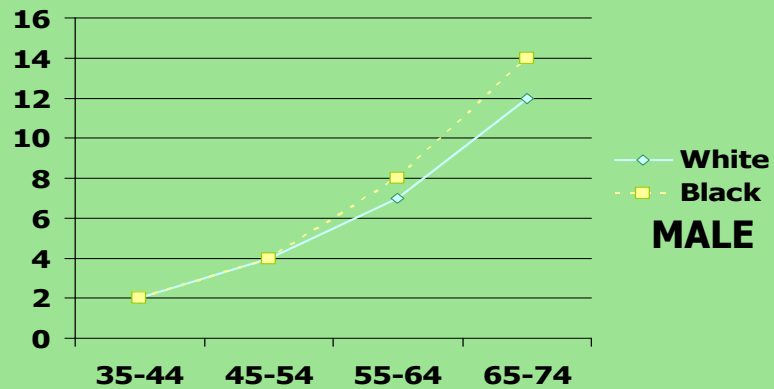
### Framingham model



### Synthesis model



### ARIC findings



# Increment of discriminate validity by adding more risk factors in the model

Start from mean incidence 7.5%

- gender  $\pm 3.5\%$  e.g. (M 11%, F 4%)
  - age  $\pm 8.5$
  - smoking  $\pm 4.9$
  - BP  $\pm 3.7$
  - HDL/chol  $\pm 11$
  - diabetes  $\pm 0.8$
- 
- family history  $\pm 0.1$
  - physical  $\pm 0.7$
  - HRT  $\pm 0.5$
  - albumin  $\pm 0.9$
  - fibrinogen  $\pm 1.2$
  - lipoprotein(a)  $\pm 2.3$

**FS**

**SYN**