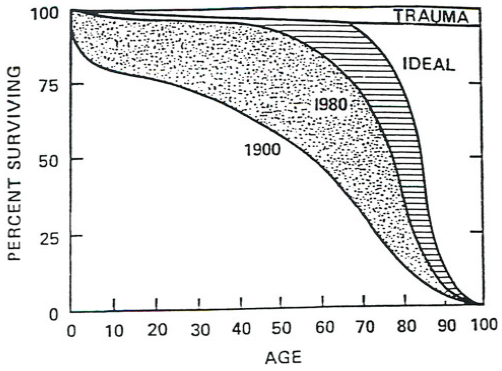
Evaluating Disease Predictors with Repeated Measurements: A Longitudinal Study of Alzheimer's Disease

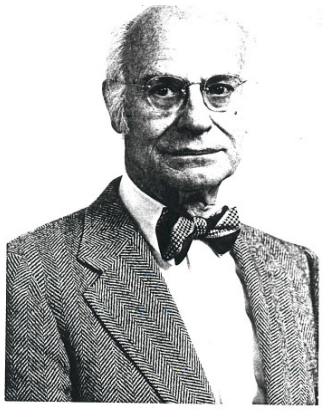
Larry J. Brant
National Institute on Aging (Retired)

SUCCESSFUL AGING

THE COMPRESSION OF MORBIDITY AND THE RECTANGULARIZATION OF SURVIVAL



Source: Fries et al. (1980), New England Journal of Medicine, 303: 130-135.

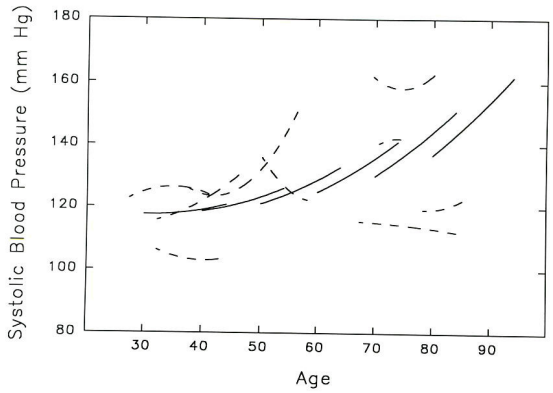


There is no general pattern of aging that applies to all performances, all organ systems, or all individuals.

Nathan Shock (1984)

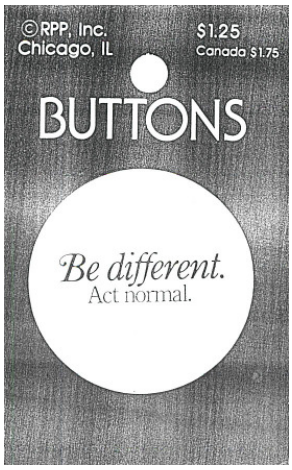
The challenge is to determine if this variability is informative.

To Vary is to be Normal



Age-specific population averages (solid lines) and estimated trends (dotted lines) for nine different individuals participating in the Baltimore Longitudinal Study of Aging

Source: Symposium on "Who is normal? Perspectives on Human Health and Variation" *Collegium Antropologicum* (v. 16, 1992), Brant, et al.



In biological aging, the question often arises: Who is normal or what is normal?

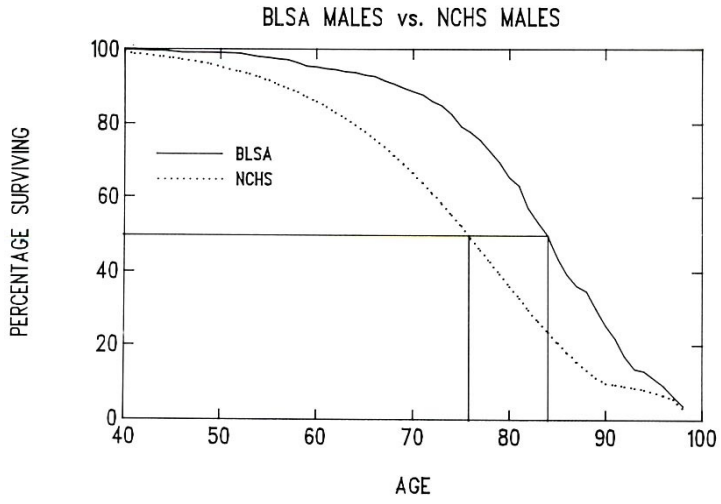
History of Disease Prediction Models

- Logistic Regression
  Cornfield, Gordon, Kannel (1967); Kannel, et al. (1976)
- Proportional Hazards Regression
  Cox (1972); Wilson et al. (1998)

Both of these approaches model a binary outcome for the presence or absence of disease using risk factor measurements from a single baseline examination.

- Mixed-Effects Regression
  Brant et al. (2003); Brant et al. (2005, 2010)

The risk factor or predictor variables are modeled as independent variables using a mixed-effects regression model with the binary outcome variable as a regressor variable as well as starting age and time. These results along with Bayes' theorem are used to compute posterior probabilities for disease prediction.



Survival probabilities for males from the Baltimore Longitudinal Study of Aging (BLSA) and the National Center for Health Statistics (NCHS) life table data for U. S. white males ages ≥ 40.

Disease Progression

Genetics ⟷ Risk Factors for Disease

↓

Preclinical Disease

↓

Detecting Preclinical Disease

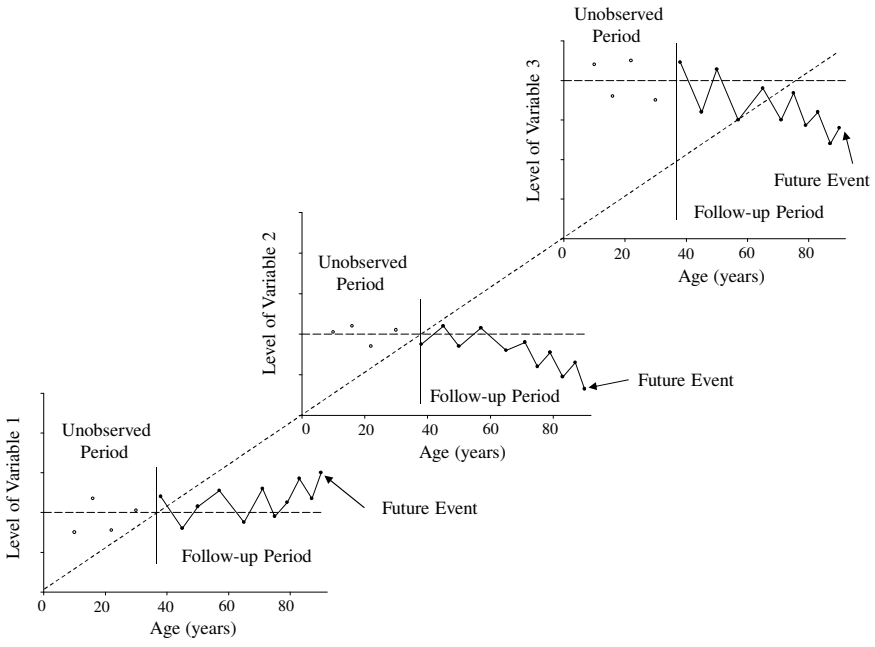
↓

Overt Clinical Disease

↓

Organ Damage or Failure

Longitudinal Data for an Individual



Mixed-Effects Regression Approach to Predicting Future Events

1) Model the longitudinal measurements for all individuals in the dataset and provide modeled trajectories for all the possible disease-related outcome categories.

2) Compute predicted marginal probability density values $f_{g(i)}(y_i)$ for each outcome category g one repeated examination at a time for each individual i.

3) Calculate posterior probabilities using Bayes' theorem

$$PP_{g(i)} = p_g f_{g(i)}(y_i) / \Sigma_g \, p_g f_{g(i)}(y_i)$$

where p_g are the prior probabilities for outcome category g.

Description of Study Population (1980 – 2010)

	Female (N = 786)		Male (N = 790)	
	APOE ε4 Carriers	APOE ε4 Non-Carriers	APOE ε4 Carriers	APOE ε4 Non-Carriers
Number of participants	227	559	217	573
Starting Age (yrs) Mean (Std Dev)	52.0 (14.2)	51.8 (16.4)	52.6 (16.2)	54.8 (16.7)
Length of follow-up (yrs) Mean (Std Dev)	14.1 (8.5)	13.5 (7.8)	14.2 (8.6)	15.6 (8.7)
Number of Alzheimer's Disease (AD) Cases (%)	18 (7.9)	32 (5.7)	17 (7.8)	39 (6.8)
Age of AD Diagnosis (yrs) Mean (Std Dev) Range	80.6 (7.1) 70.7 – 94.1	85.2 (7.5) 60.4 – 99.3	81.7 (4.9) 71.2 – 88.9	85.0 (6.0) 69.2 – 95.5

Possible Predictors for Alzheimer's Disease (AD)

Mini-Mental State Examination Score (MMSE)
Center for Epidemiologic Studies Depression Scale (CES-D)
Forced Expiratory Volume in One Second (FEV-1, L)
Forced Vital Capacity (FVC, L)
Body Mass Index (BMI, kg/m^2)
Systolic Blood Pressure (SBP, mm Hg)
Diastolic Blood Pressure (DBP, mm Hg)
Mean Arterial Pressure (MAP, mm Hg)
Pulse Pressure (PP, mm Hg)
Fasting Plasma Glucose (FPG, mg/dL)
Total Serum Cholesterol (TC, mg/dL)
High-Density Lipoprotein (HDL-C, mg/dL)
Low-Density Lipoprotein (LDL-C, mg/dL)
Triglycerides (TG, mg/dL)
Hemoglobin (HGB, g/dL)

Linear Mixed Effects (LME) Regression Model for LDL Cholesterol (LDL-C) in Male APOE ε4 Non-Carriers

For individual i at time j consider the LME model

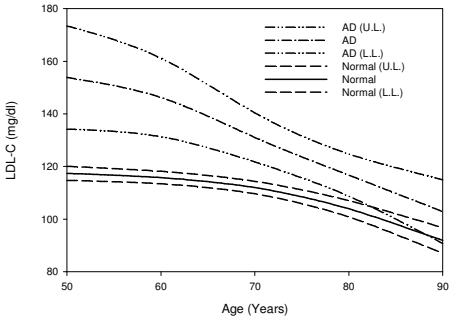
$$\begin{aligned}\text{LDL-C}_{ij} = (\beta_0 + b_{0i}) + \beta_1\, \text{fage}_i + (\beta_2 + b_{2i})\, \text{time}_{ij} + \beta_3 \text{AD}_i + \\ \beta_4 \text{fage}_i * \text{time}_i + \beta_5 \text{time}_i * \text{time}_i + \beta_6 \text{fage}_i * \text{AD}_i + \\ \beta_7\, \text{time}_{ij} * \text{AD}_i + \beta_8\, \text{time}_{ij} * \text{time}_{ij} * \text{AD}_i + \varepsilon_{ij}\end{aligned}$$

where the b and ε terms are independent with $\underline{b} \sim N(\underline{0}, D(\theta_D))$ and $\underline{\varepsilon} \sim N(0, \Sigma(\theta_\Sigma))$.

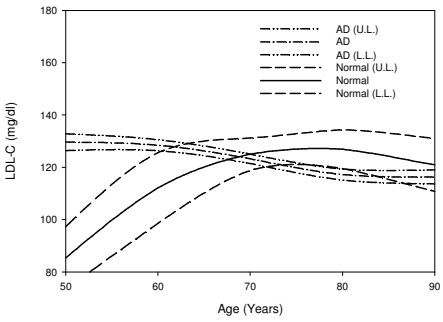
In general, the LME model is written $Y_i = X_i\beta + Z_i b_i + \varepsilon_i$ and so the marginal distribution (obtained by integrating out the b_i terms) has a normal distribution with mean $\underline{x}_i\beta$ and variance $V_i(\underline{\theta}) = Z_i D(\theta_D) Z_i' + \Sigma_i(\theta_\Sigma)$.

Male Predicted Population Average Trends for LDL Cholesterol (LDL-C) by APOE ε4 and AD Status*

Non-Carriers



Carriers



*Predictions made by mixed-effects model.

Calculation of Individual Alzheimer's Disease (AD) Prediction Probabilities

1) Compute multivariate normal probability density function values

$$f_g(y_i \mid \beta,\underline{\theta}) = (2\pi)^{-n_i/2} |V_i(\underline{\theta})|^{-1/2}\exp[-½(y_i - x_i\beta)'V_i^{-1}(\underline{\theta})\,(y_i - x_i\beta)]$$

from the marginal distribution and

2) calculate posterior probabilities using Bayes' theorem

$$PP_{g(i)} = p_g\, f_g(y_i \mid \beta,\underline{\theta}) \,/\, \Sigma_g\, f_g(y_i \mid \beta,\underline{\theta})$$

for $g = 0$ ($AD_i = 0$) and $g = 1$ ($AD_i = 1$) one examination at a time starting with the first examination until the individual is classified as an AD or non-AD case.

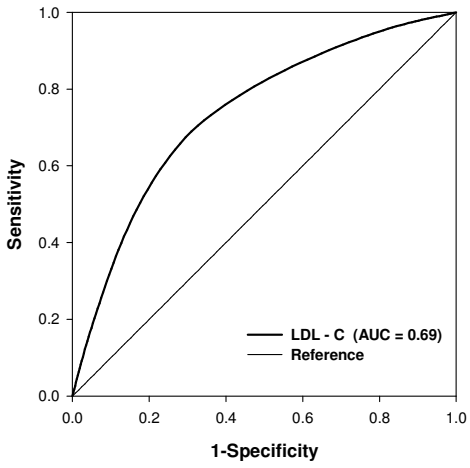
Classification Results

The posterior probabilities are used to create a classification table by considering a range of cutoff values for the posterior probabilities.

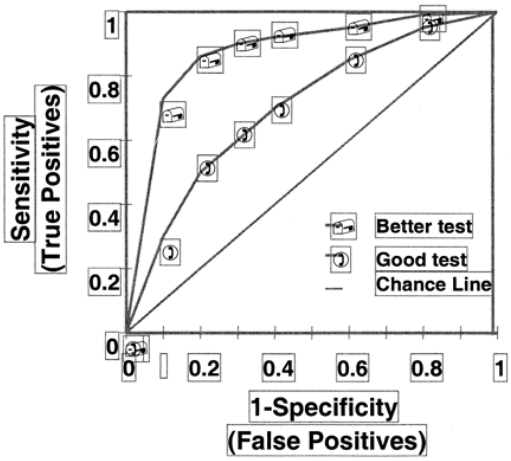
Note:
As the cutoff values decrease, the proportion of correctly classified outcomes (sensitivity) increases and the proportion incorrectly classified as positive outcomes (1 – specificity) increases.

These classification results for the different cutoff values are used to construct a receiver operator characteristic (ROC) curve.

Receiver Operator Characteristic (ROC) Curve for LDL-C (Male APOE ε4 Non-Carrier)

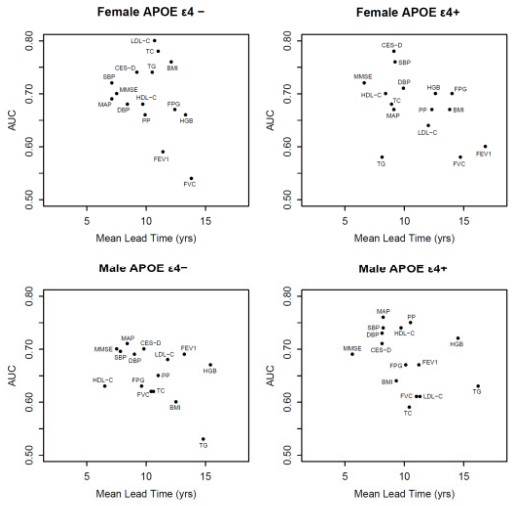


ROC Curve Guidelines



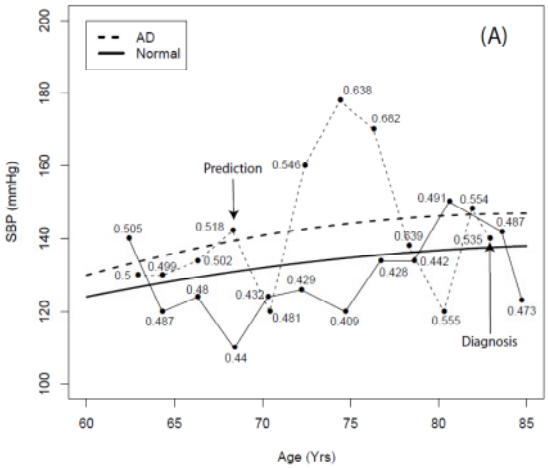
Source: American College of Cardiology, Identification of Coronary Heart Disease Risk, Task Force #1, JACC, 2003.

Classification Results by Sex and APOE ε4 Genotype



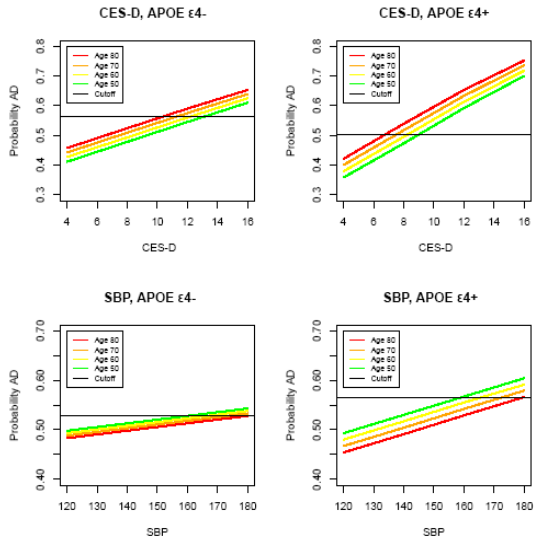
AUC = Area under receiver operator characteristic (ROC) curves.

Individual AD Predictions for Male AD Case* and Control (Systolic Blood Pressure (SBP) Measurements)



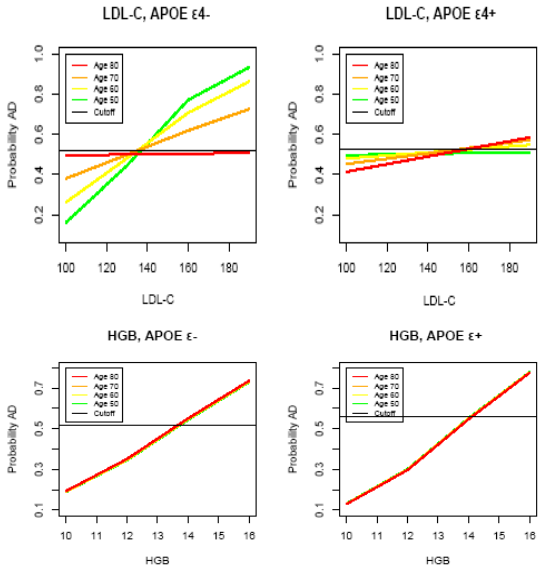
*No evidence of AD at first examination.
Remark: Both males are APOE ε4 non-carriers.

Female AD Prediction Probabilities* by APOE ε4 Genotype

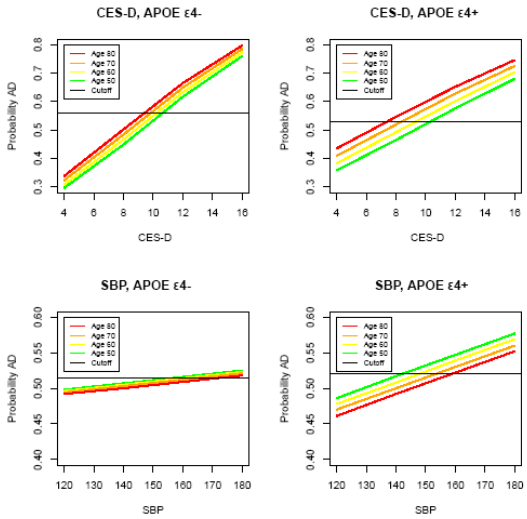


*Based on first examination values of age and predictor levels.

Female AD Prediction Probabilities by APOE ε4 Genotype

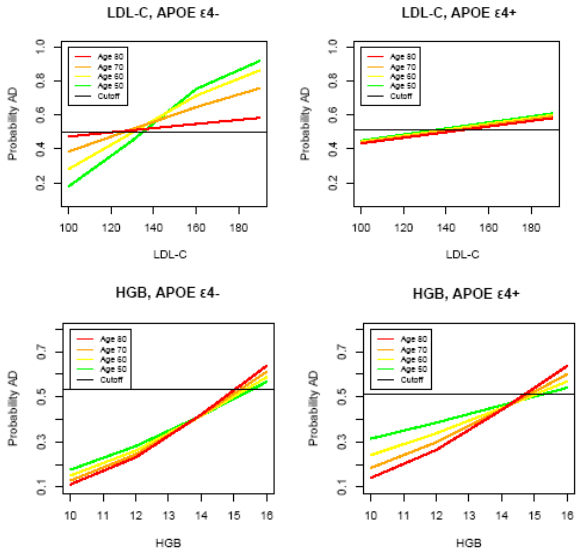


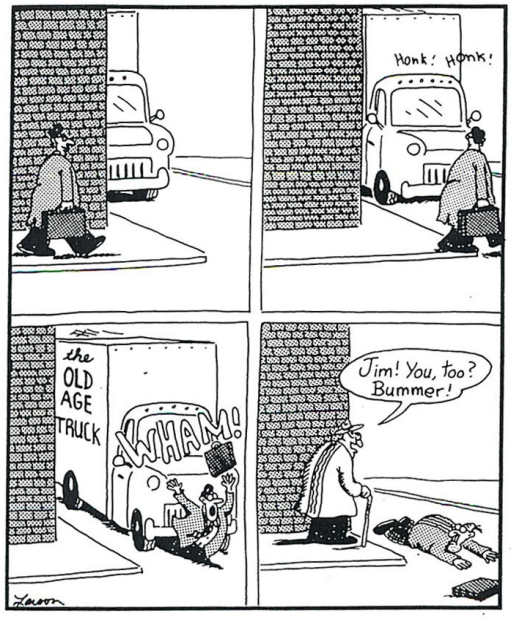
Male AD Prediction Probabilities* by APOE ε4 Genotype



*Based on first examination values of age and predictor levels.

Male AD Prediction Probabilities by APOE ε4 Genotype

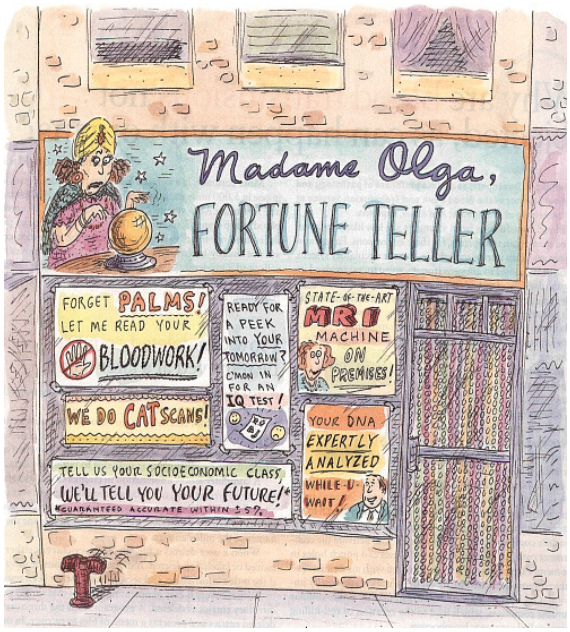




You never see it coming.

Conclusions

- Repeated measurements of common physiological and laboratory measures collected over the entire adult lifespan are useful for making individual predictions of AD.
- Methodology presented in this paper may be useful in identifying vulnerable populations and targeting them for midlife intervention studies, with the potential of dramatically reducing the projected prevalence of AD.
- Possibility exists to create a composite risk score for AD weighting predictor variable posterior probabilities using AUC and MLT values.
- If we look hard enough we may be able to see the Old Age Truck coming and perhaps slow it down a bit.



Scientific American, March 2004