Data Mining for Population Based Studies:

OPPORTUNITIES FOR PRE- AND POST-DOCTORAL STUDENTS TO MINE DATABASES AND MAKE CONTRIBUTIONS TO PUBLIC HEALTH AND FUTURE RESEARCH

Compstat Meeting
August 2010
Presenter: Dr. Stanley Azen

Pre-Docs, Post-Doc Collaborators: Kathy Sullivan, Julie K. Tilson, Steven Y. Cen, Jiaxiu He, and Cheryl Vigen

I. POPULATION BASED STUDY:
THE LOS ANGELES LATINO EYE STUDY (LALES)

- LALES was funded by the National Institutes of Health to determine the *prevalence , risk factors and impact on quality of life (QOL)* on ocular disease in 6082 non-institutionalized adult Latinos in urban Los Angeles County.

- In addition, LALES was designed to evaluate factors associated with the *utilization of eye care and medical services* , including *screening* for ocular disease.

LALES RESULTS REGARDING GLAUCOMA

- Open-angle glaucoma affects approximately 66.8 million people worldwide and is the second leading cause of blindness, afflicting 6.7 million people.

- LALES data indicate that 75% of the glaucoma cases were undiagnosed (based on diagnoses made by glaucoma experts using complete and complex evaluation methods = "gold standard".

- Hence, there is a need to better characterize glaucoma in the Latino population for the purpose of developing efficient screening strategies.

Standard Screening Tests
Used for Detecting Glaucoma

- **Standard Screening Tests using traditional standard cutpoints:**
    - **Humphrey visual field testing,**
    - **Frequency doubling technology,**
    - **Intraocular pressure (IOP),**
    - **Central corneal thickness (CCT)**
    - **Vertical cup to disk ratio (C/D)**
- **Traditional stepwise logistic regression analyses indicated that the best predictor was the C/D ratio, followed by expert visual field reading and glaucoma hemifield test (GHT)**
- **Specificity levels were good (0.7 to 0.9); but sensitivity levels were low (0.3 to 0.8)**

Data Mining:
Population Screening for Glaucoma

- **So, a postdoctoral biostatistician, Cheryl Vigen explored the benefit of using Classification and Regression Tree (CART) to develop an optimal screening algorithm.**

- **CART analyses chose the best cutpoint for each predictor, and determined the need for additional predictors in some situation.**

- **Sensitivity and specificity for the CART model were .92 and .92, respectively.**
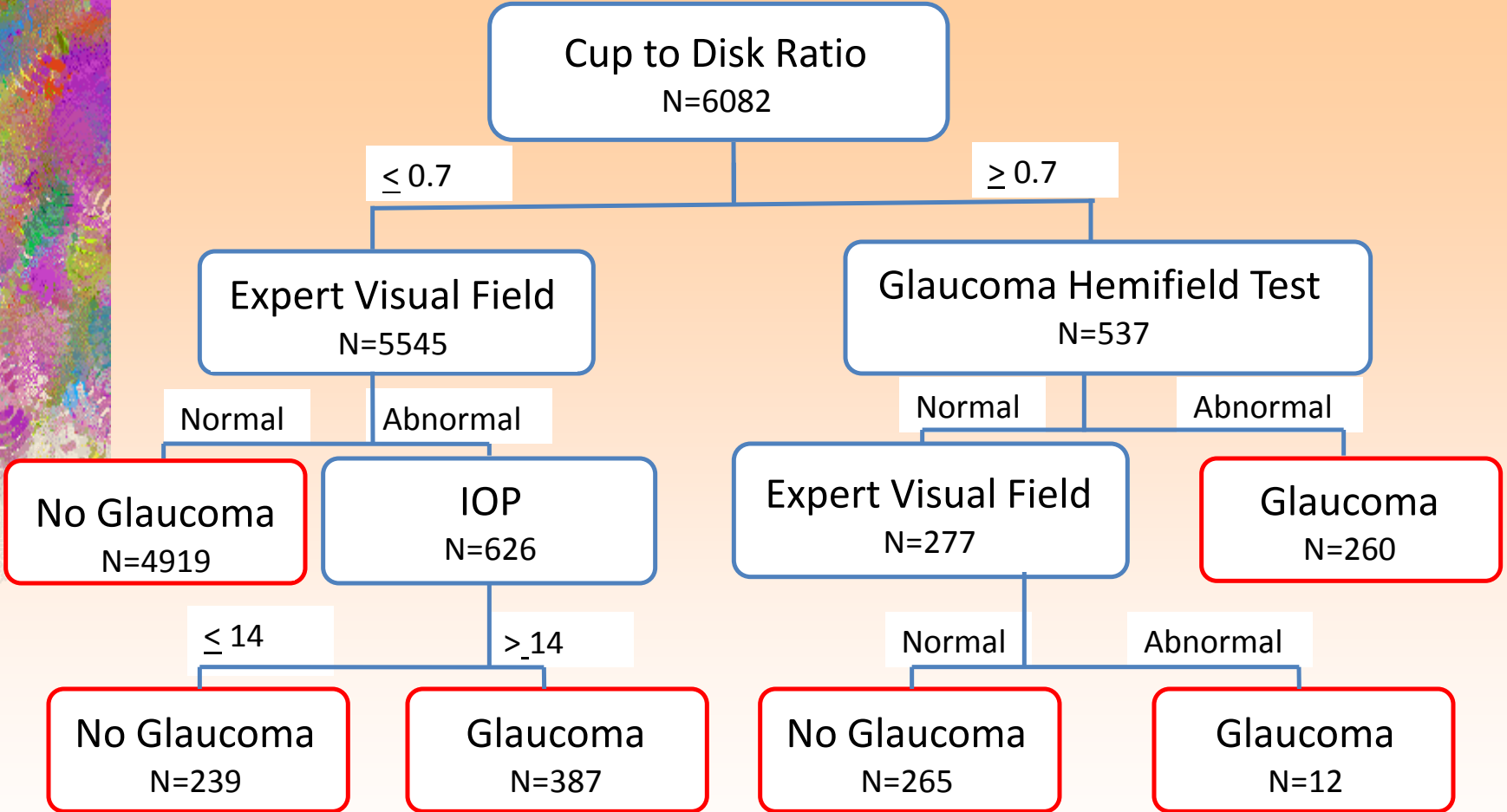
II. CLINICAL TRIAL:
## Hormonal Regulators of Muscle and Metabolism in Aging Males (HORMA)

- HORMA was a multi-centered clinical trial funded by the National Institutes of Health to determine whether endogenous testosterone (T) and growth hormone (rhGH) are important independent, but complementary regulators of skeletal muscle mass and function in the elderly.

- Results: Total lean body mass (LBM), appendicular skeletal muscle mass (ASMM), muscle performance, and stair combing power increased significantly with T and enhanced by rhGH.

- However, there was considerable variability in anabolic response as well as in changes of T and IGF-1 levels during treatment

DATA MINING: HORMA DATABASE

- The HORMA database provided the opportunity to examine relationships of a broad range of changes, including declines in anabolic hormone levels, and their effects on changes in lean tissue mass, muscle strength, performance and physical function.

- So, a predoctoral student in biostatistics (Jiaxiu Hawk He) explored the benefit of using Pathway Analysis to test the hypothesis that T and rhGH affected muscle mass directly and that a threshold change in lean tissue mass was needed to generate significant improvements in muscle performance and physical function.

Pathway Analysis: Results
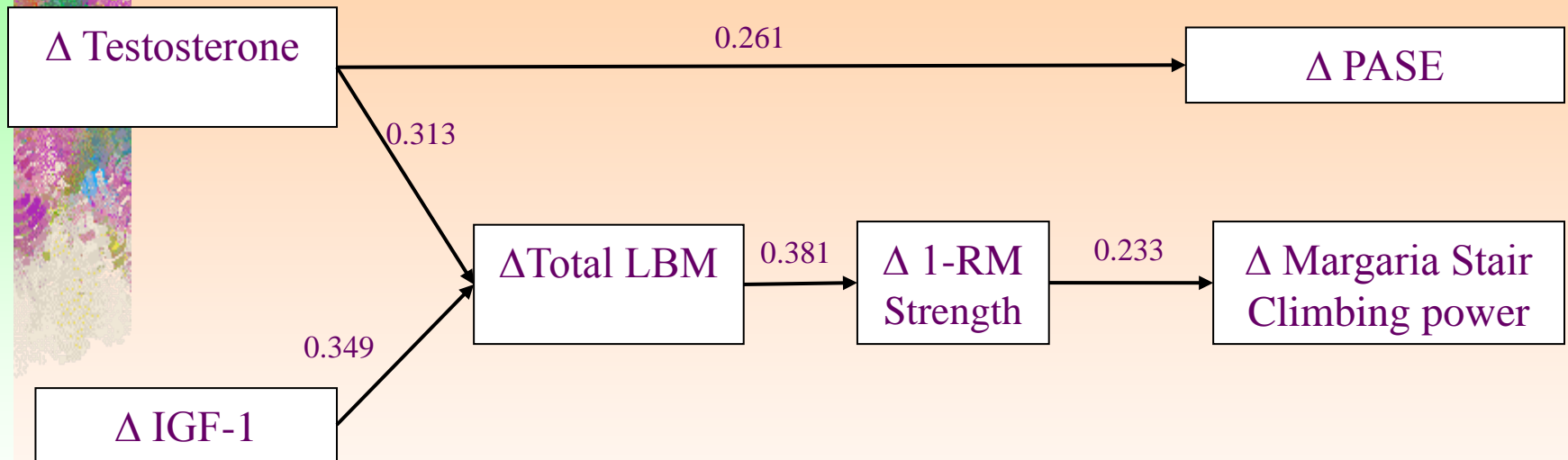
Results of Data Mining Analyses:
Strategies for Enhancing Muscle Strength and Physical Function in Aging Men

- Increase in total testosterone were necessary to achieve median increases in LBM and ASMM, which were required to significantly enhance 1-RM strength (by 30%).

- Pathway analyses supported the model that changes in testosterone and IGF-1 levels are related to changes in LBM needed to enhance muscle performance and physical function.

- Pathway analysis demonstrated that testosterone's effects on physical activity was mediated through a different pathway since testosterone directly affected Physical Activity Score of the Elderly (PASE).

III. CLINICAL TRIAL:
Locomotor Experience Applied Post Stroke (LEAPS)

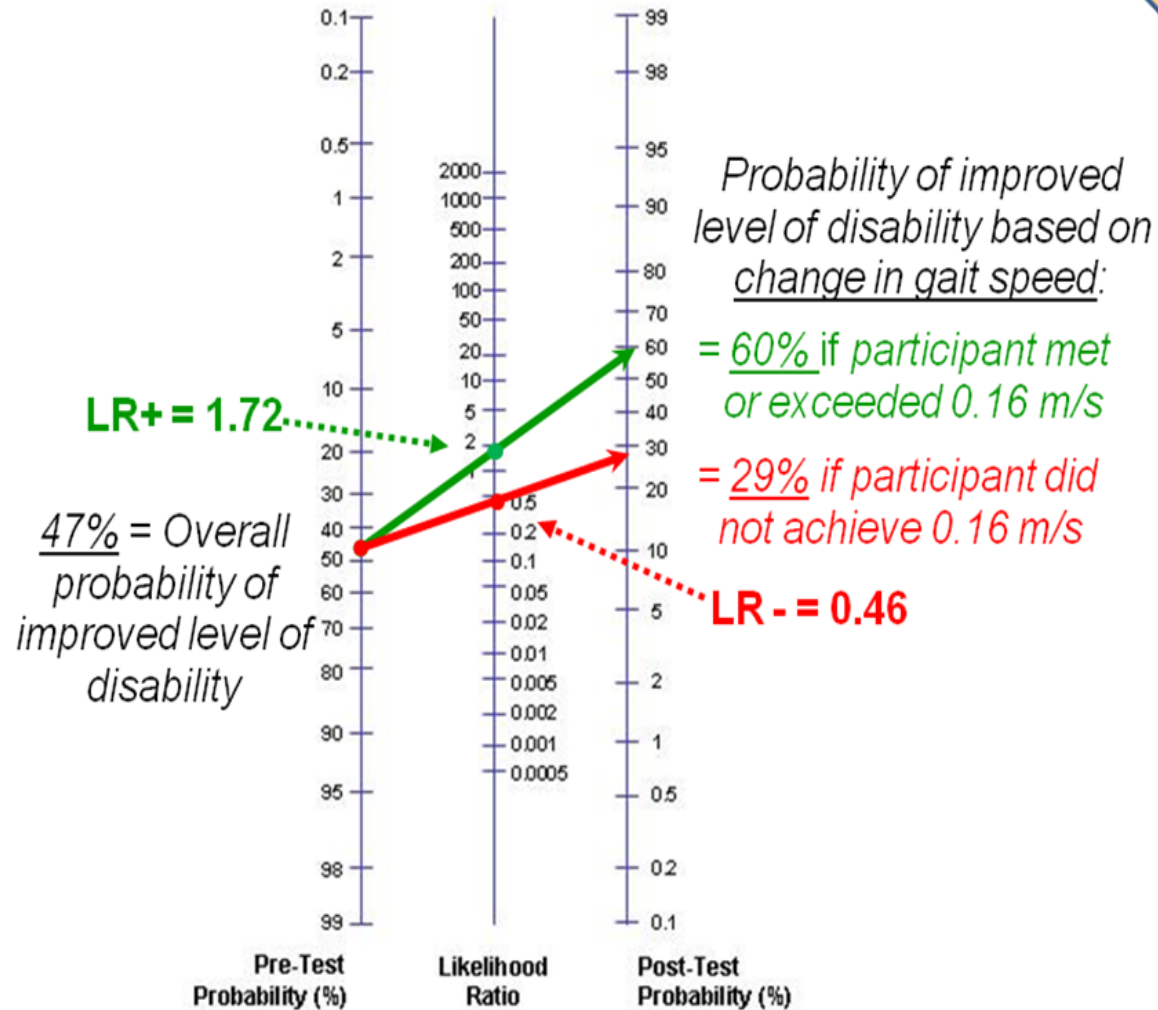- **LEAPS is a multi-centered clinical trial funded by the National Institutes of Health to compare two different therapist supervised treatments to improve walking speed after stroke. Treatments included:**
  - **1) a specialized locomotor training program (LTP) using body weight support and a treadmill.**
  - **2) progressive strength and balance exercises provided in the patient's home (HEP).**
- **Prior to completion of the study, one of the post-doc therapists, Dr. Julie Tilson, requested permission to mine the baseline information database to estimate the MCID = minimal amount of change in gait speed that is clinically meaningful and associated with an important difference in function. She was guided by Kathy Sullivan PhD (assistant professor) and Steven Cen PhD (statistician).**

III. CLINICAL TRIAL:

Locomotor Experience Applied Post Stroke (LEAPS)

- **Objective: To determine the minimal clinically important difference (MCID) for *comfortable gait speed (CGS)* associated with an *improvement in function* measured by the modified Rankin Scale (mRS) score.**

- **Cohort: 283 LEAPS participants between 20 to 69 days post stroke.**

- **Methods: CGS and mRS were measured at 20 and 60 days post stroke. Improvement of mRS$\geq$1 was used to detect MCID for CGS.**

Nomogram graphical representation of the probability that an individual with stroke will experience a meaningful change in disability level.

Probability of improved level of disability based on change in gait speed:

LR+ = 1.72

= 60% if participant met or exceeded 0.16 m/s

= 29% if participant did not achieve 0.16 m/s

LR − = 0.46

47% = Overall probability of improved level of disability

Pre-Test Probability (%)

Likelihood Ratio

Post-Test Probability (%)

LIKELIHOOD RATIO NOMOGRAM

III. CLINICAL TRIAL:
Locomotor Experience Applied Post Stroke (LEAPS)

- **Result 1. Among all participants, 47.3% experienced an improvement in disability level $\geq$1 on the mRS.**

- **Result 2: The MCID was estimated as an improvement in CGS of 0.16m/s anchored to the mRS.**

- **Conclusion: Patients with subacute stroke who increase gait speed >0.16 m/s are more likely to experience a meaningful improvement in disability level than those who do not.**

- **Impact on Stroke Rehabilitation: Clinicians can use this reference value to develop goals and interpret progress in patients with subacute stroke.**

Summary of Data Mining Examples

- **CART: An improved screening algorithm has been developed from LALES which could reduce the risk of undiagnosed glaucoma**

- **Path Analysis: Results from HORMA helped understand mechanisms needed to enhance muscle performance and physical function and quality of life.**

- **MCID: Results from LEAPS helped develop strategies for monitoring progress in patients with subacute stroke.**

- **Research resulted in three publications.**

QUESTIONS ?