PHYSICAL FITNESS FACTORS THAT IMPACT BIOMARKERS OF LONGEVITY

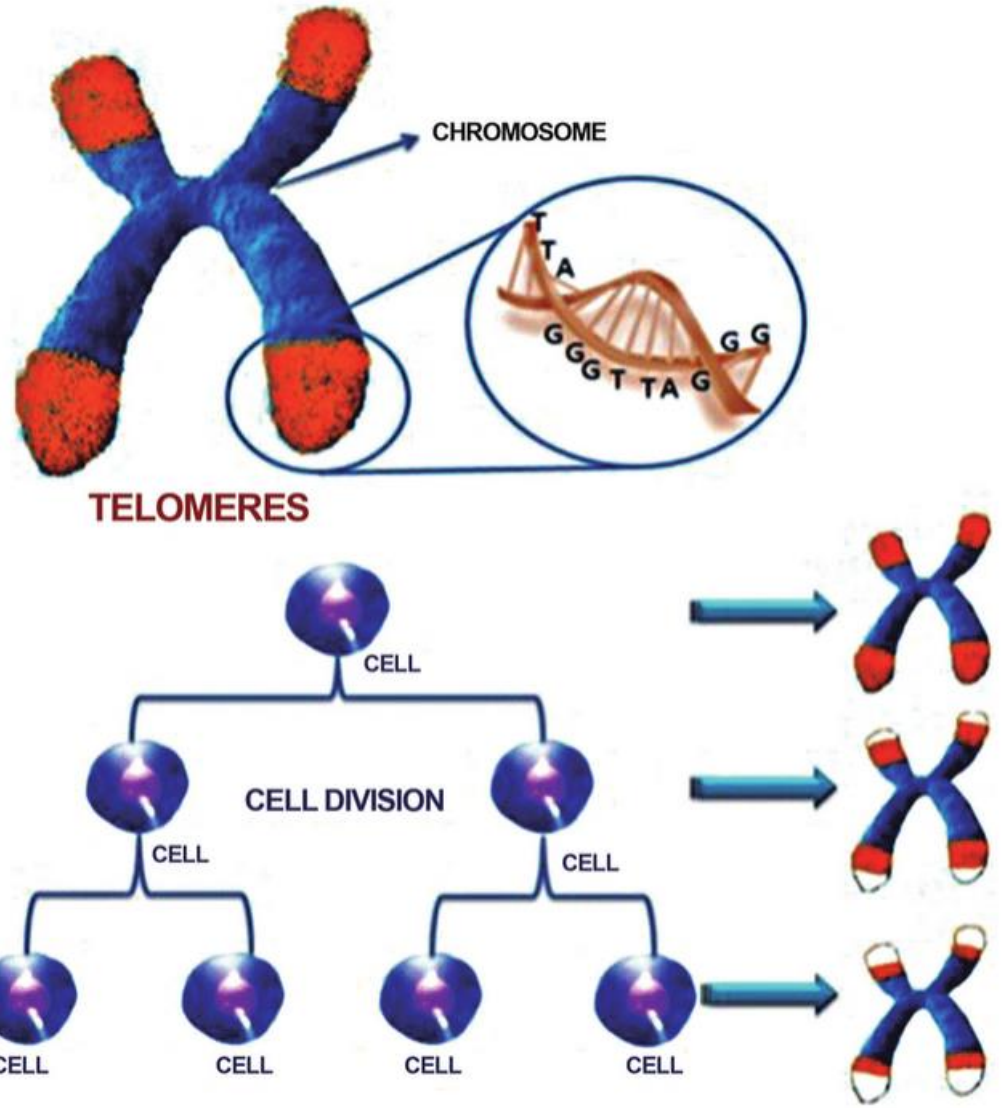
SUPERVISED LEARNING
CAPSTONE

OVERALL FINDINGS

- Age is the most important predictor of telomere length
- Fitness level does not predict longevity, but sedentary behavior might!

MEASURING LONGEVITY THROUGH DNA

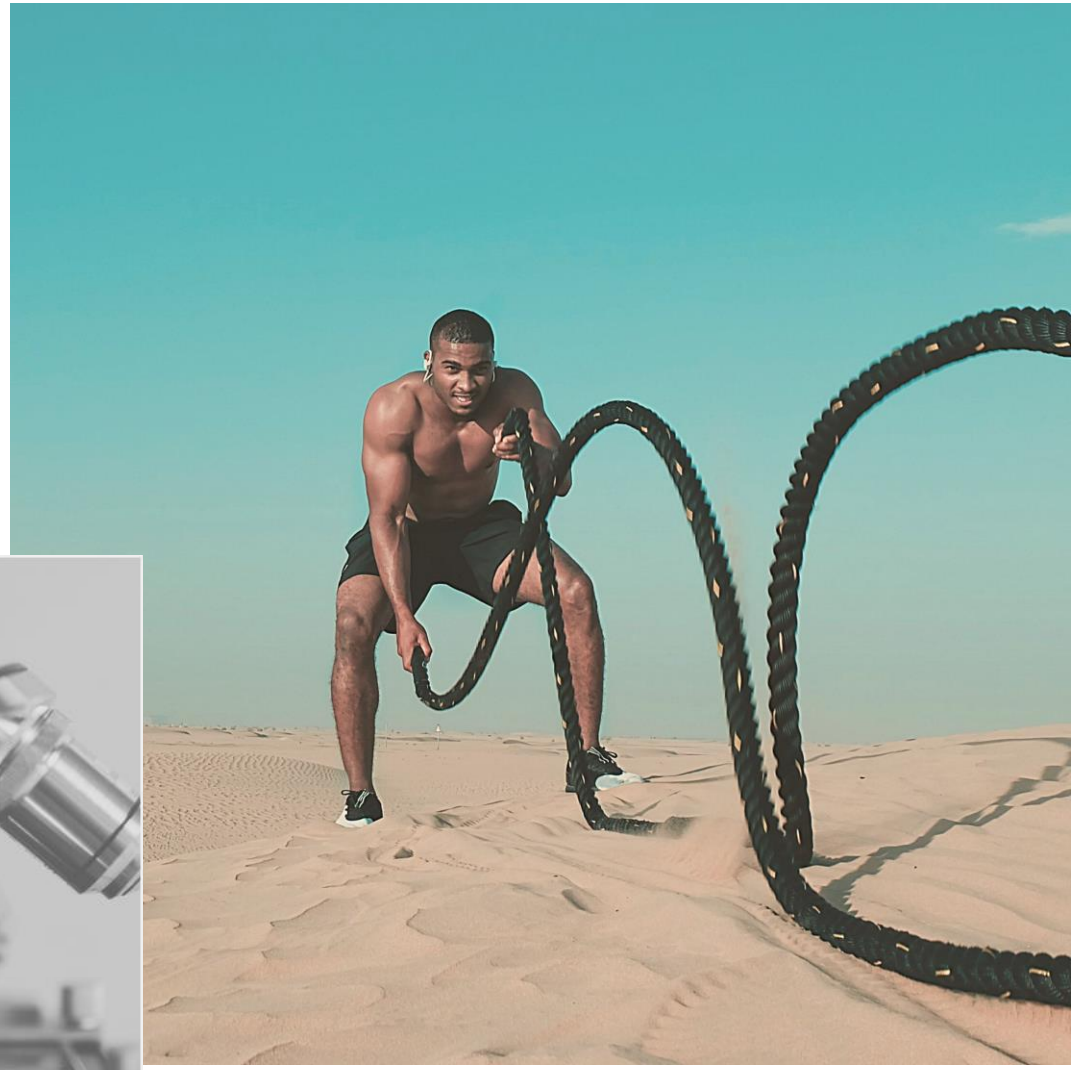
- Protective DNA at the ends of chromosomes
- Short telomere length in particular cells has been shown to predict shortened longevity
- Telomere length as a biomarker for aging



Musumeci et al., Histology and histopathology (2014)

CAN WE INFLUENCE LONGEVITY?

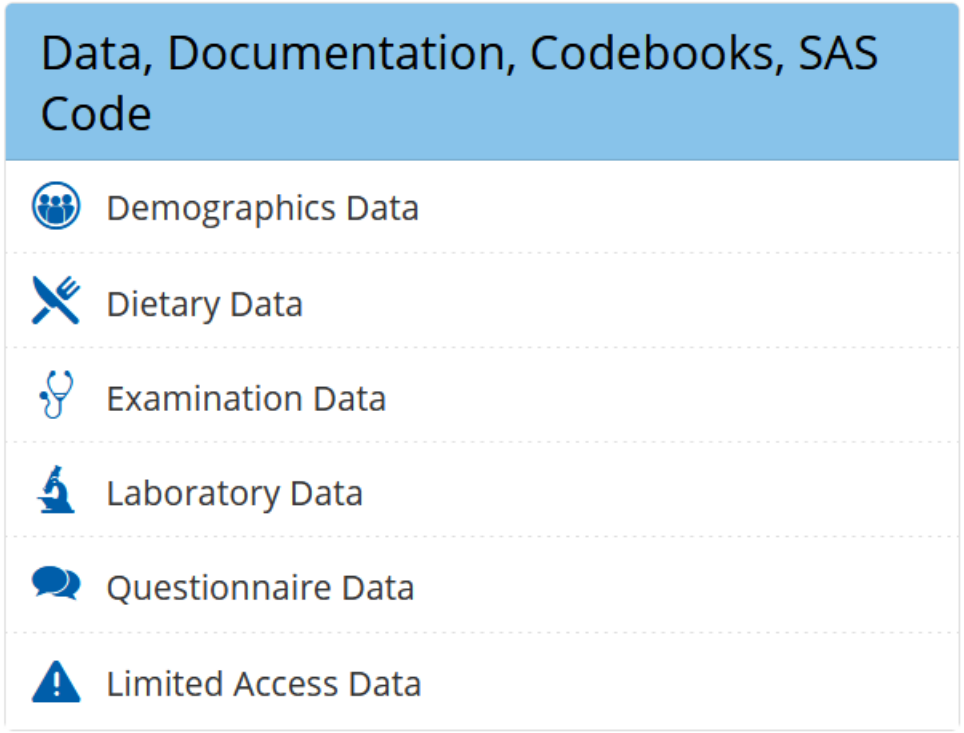
- Physical activity is associated with delayed aging
  - Longer telomeres in immune cells of people who engage in regular, moderate exercise
  - Exercise increases activity of telomerase in mouse heart



PREDICTING BEST METHODS TO INCREASE TELOMERE LENGTH

- Predict length of telomeres by measuring health factors, including physical activity measures, exercise tests indicative of health

CDC NATIONAL HEALTH AND NUTRITION EXAMINATION SURVEY DATASET



- Two-year period of 1999 and 2000
- Ten tables of most interest to the hypothesis
- 3570 records

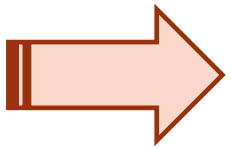
DATA CLEANUP AND PREPARATION FOR MODELING

- Demographics, Balance, Cardiovascular Fitness, Muscle Strength, Physical Activity, Physical Functioning
- Feature selection versus full dataset
- Null values systematically dropped, resulting in 1837 records and 154 features
- Telomere length target variable outliers capped

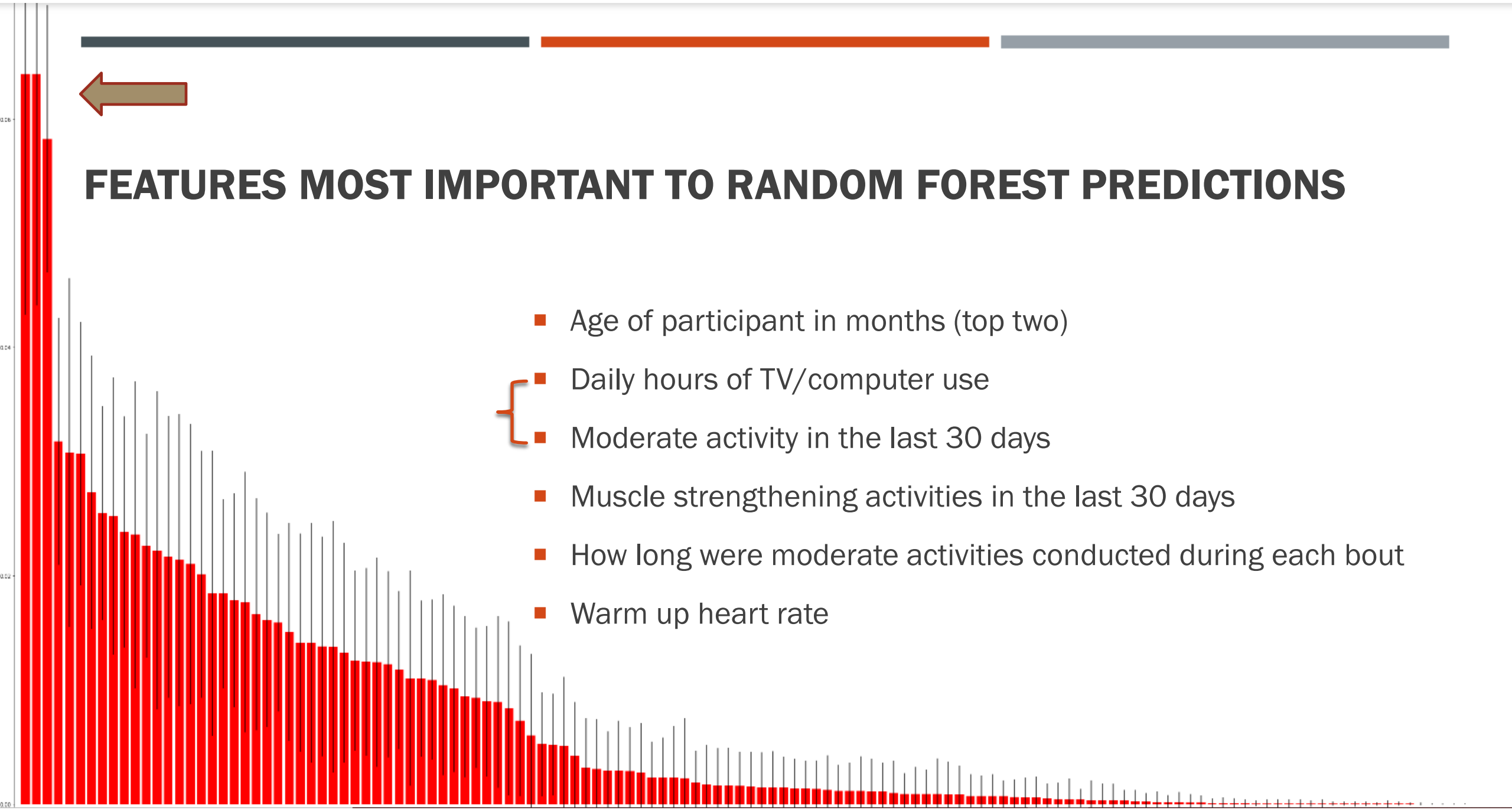
	TELOMEAN
count	1837.000000
mean	1.087386
std	0.329323
min	0.000000
25%	0.897950
50%	1.052693
75%	1.235469
max	9.420415

MODELING THE EFFECTS OF THE FEATURES ON TELOMERE LENGTH

Model	Training R-squared Score	MAE	MSE	RMSE	MAPE
OLS Regression	0.200420	0.195099	0.060282	0.245524	19.135288
Ridge Regression	0.167293	0.187128	0.054569	0.233599	18.363417
Lasso Regression	0.000000	0.197083	0.057665	0.240135	19.554895
ElasticNet Regression	0.000000	0.197083	0.057665	0.240135	19.554895
K Nearest Neighbors	1.000000	0.195011	0.057481	0.239752	19.215721
Decision Tree	1.000000	0.279368	0.125149	0.353764	26.888574
Random Forest	0.865980	0.186159	0.053077	0.230383	18.318748
Support Vector	0.078897	0.193155	0.057887	0.240597	18.608731
Gradient Boosting	0.393732	0.189445	0.055405	0.235383	18.588954

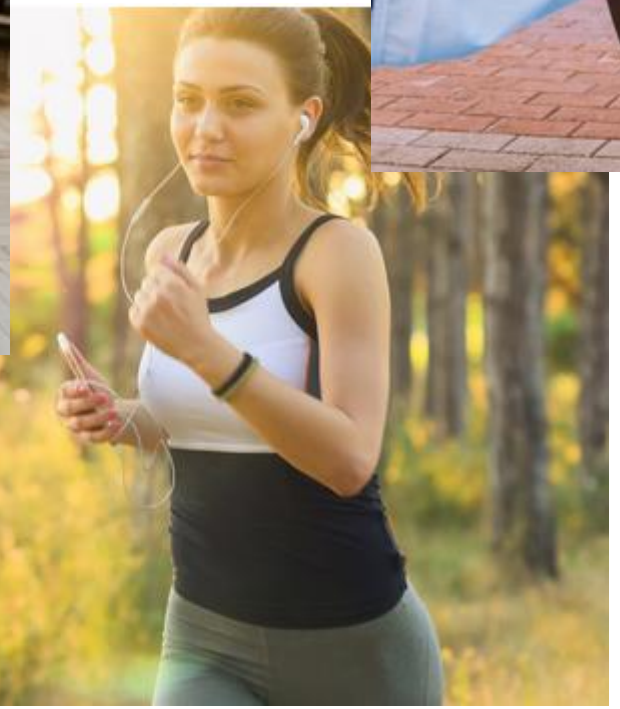


FEATURES MOST IMPORTANT TO RANDOM FOREST PREDICTIONS

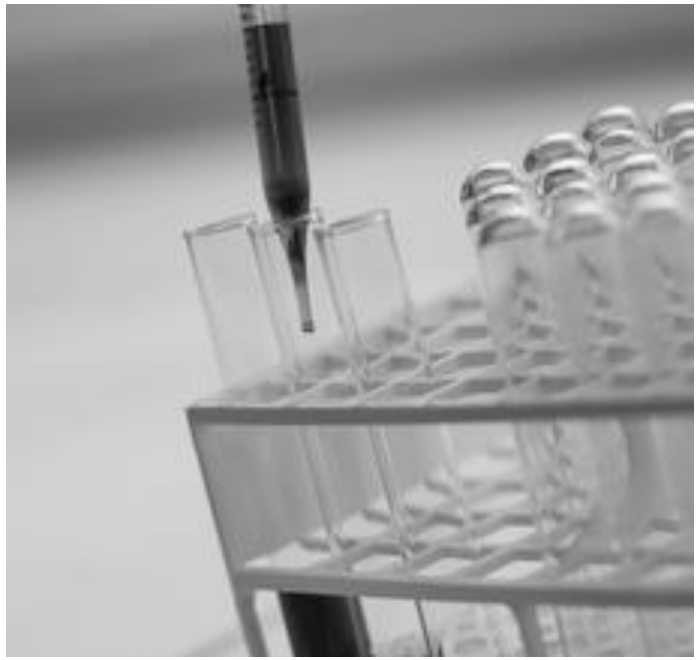


- Age of participant in months (top two)
- Daily hours of TV/computer use
- Moderate activity in the last 30 days
- Muscle strengthening activities in the last 30 days
- How long were moderate activities conducted during each bout
- Warm up heart rate

MOVE MORE, MOVE MODERATELY



FUTURE WORK



- Diet, other health behaviors, chronic illness
- Rerun models with strength features only
- Effects of overtraining and extreme exercise
- Population effects

QUESTIONS

- Thank you for your attention!