Future Scenarios of Algorithm Building

With Illustration in the NIA/FNIH Sarcopenia Definitions and Outcomes Project

Thomas G Travison, PhD
Hebrew SeniorLife Institute for Aging Research
Harvard Medical School
State of Play in Non-alcoholic Steatohepatitis (NASH)

• Clinical gold standard (liver biopsy & histology) invasive and expensive

• Phenotype of NASH may be described by combination of pathological + clinical characteristics
  – Circulating analytes
  – Metabolic / prediabetic illnesses
  – Proteomics, metabolomics
  – Imaging
Objective

• Develop model-based / algorithmic equation or decision rule integrating some or many factors sensitive and specific for presence of NASH

• ‘Dual purpose’ of
  – Individual level diagnosis
  – Population-level research / monitoring (‘on average’)


The dual-purpose problem

• Classical methods (e.g. linear / nonlinear regression) adept at describing central tendency
  – ‘Typical’ ~ average

• Not everyone is typical; indeed the atypical patient may be the most important or informative
Tension between population trends and inter-individual variation

Frailty, Physical Activity, and Mobility in Patients With Cardiac Implantable Electrical Devices

Daniel B. Kramer, MD, MPH; Timothy Tsai, MPH; Poorna Natarajan, MBBS; Elise Tewksbury, RN; Susan L. Mitchell, MD, MPH; Thomas G. Travison, PhD

- J Am Heart Assoc. 2017;6:e004659. DOI: 10.1161/JAHA.116.004659
Strong and consistently significant association at the mean

**Multinomial Logistic Regression**

<table>
<thead>
<tr>
<th></th>
<th>Odds Ratio</th>
<th>95% Confidence Interval</th>
<th>p-val</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prefrail v. Robust</td>
<td>0.8</td>
<td>(0.68, 0.96)</td>
<td>0.015</td>
</tr>
<tr>
<td>Frail v. Robust</td>
<td>0.61</td>
<td>(0.44, 0.83)</td>
<td>0.002</td>
</tr>
</tbody>
</table>

*a Adjusted for Age and Body Mass Index*
But! Dramatic inter-individual variation might preclude use of activity as a sensitive patient-level classifier of frailty status.
Performance of a diagnostic cutpoint: composite predictor

- Gait Speed (m/s)

- F(Muscle Mass, Strength)

- Diagnostic Threshold

- True Negative
- False Negative
- False Positive
- True Positive

Mobility Limitation
Addressing the dual-purpose problem

• Potential strategy: use machine-learning / partitioning methods to divide population into subgroups at substantially differential risk
  – Deals less with ‘on average’ and more with rankings and individual inputs
  – Allows for high-dimension comparison of correlated and near-collinear candidate phenotypes
  – Deals intrinsically with nonlinearities and complex interactions

• Caveat: patient-level inference dramatically influenced by data quality, precision
  – Necessitates attention to specifics of data collection and aggregation
“Analyses of Datasets on Older Populations with High Prevalence of Mobility Disability to Develop Clinically Meaningful Diagnostic Cut-Points for Low Muscle Mass and/or Low Muscle Strength”
Sarcopenia Definition and Outcomes
Consortium Investigators

PIs: Peggy Cawthon and Shalender Bhasin

Steering Committee and Analysts: Sheena Patel, Rosaly Correa-De-Araujo, Roger Fielding, Jay Magaziner, Todd Manini, Thomas Travison, Karol Pencina, Hao Zhu

Project Investigators: Shehzad Basaria, Ellen Binder, Todd Brown, Ralph D’Agostino, Susan Greenspan, Tamara Harris, Douglas Kiel, Steven Kritchevsky, Joe Massaro, Robert McLean, Anne Newman, Denise Orwig, Marco Pahor, Adam Santanasto, Michelle Shardell, Qian-Li Xue

Other Contributing Investigators: Steven R. Cummings, Robert Cumming, Kristine Ensrud, Vasant Harini, Joanne Jordan, Magnus K. Karlsson, Timothy Kwok, Östen Ljunggren, Dan Mellström, Claes Ohlsson, Eric S. Orwoll, Laura Schaap, Jean Woo, Marjolein Visser

Funding from the National Institute on Aging and the Foundation for the NIH is gratefully acknowledged (U01)
Conceptual Problem

- Sarcopenia (age-related loss of muscle mass / strength / quality) difficult to measure

- Little consensus as to which aspects or in what combination constitute the phenotype

- Potential solution: establish predictive validity for patient-important endpoint (physical functioning / mobility)
Overall goals

• To develop diagnostic cut-points* for low muscle mass and strength for identifying those at risk of mobility limitation
  – *May be defined as a general function / combination of multiple inputs

• To establish utility of diagnostic cut-points for prediction of clinically important outcomes including mortality and falls
  – Once again ‘dual – purpose’
<table>
<thead>
<tr>
<th>Study</th>
<th>Participants</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Osteoporotic Fractures in Men (MrOS) Study</td>
<td>Ambulatory community dwelling men, 65+ y</td>
<td>5,835</td>
</tr>
<tr>
<td>Study of osteoporotic fractures (SOF)</td>
<td>Ambulatory community dwelling women, 65+ y</td>
<td>1,246</td>
</tr>
<tr>
<td>Health, Aging and Body Composition Study (Health ABC) Study</td>
<td>Non-disabled black and white men and women, 70-80 y</td>
<td>1,398</td>
</tr>
<tr>
<td>MrOS Sweden</td>
<td>Men in three Swedish communities, 70+ y</td>
<td>2,876</td>
</tr>
<tr>
<td>Mr&amp;MsOS Hong Kong</td>
<td>Men and women residing in Hong Kong, 65+ y</td>
<td>4,000</td>
</tr>
<tr>
<td>Concord Health and Aging in Men Project (CHAMP)</td>
<td>Men living near Concord, Australia, 65+ y</td>
<td>1,529</td>
</tr>
<tr>
<td>Cardiovascular Health Study (CHS)</td>
<td>Community dwelling men and women, 65+ y</td>
<td>1,509</td>
</tr>
<tr>
<td>Johnston County Arthritis Study</td>
<td>White and black residents of rural Johnston County, North Carolina, 45+ y</td>
<td>438</td>
</tr>
</tbody>
</table>
## Participant characteristics

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Men</th>
<th>Women</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>14157</td>
<td>5723</td>
</tr>
<tr>
<td>Age, yrs</td>
<td>75</td>
<td>75</td>
</tr>
<tr>
<td>White race</td>
<td>79%</td>
<td>44%</td>
</tr>
<tr>
<td>Weight, kg</td>
<td>78.7</td>
<td>64.2</td>
</tr>
<tr>
<td>Walk speed, m/s</td>
<td>1.16</td>
<td>0.92</td>
</tr>
<tr>
<td>Gait speed (4 or 6m) &lt; 0.8 m/s</td>
<td>8.7%</td>
<td>25.8%</td>
</tr>
<tr>
<td>Maximum grip strength, kg</td>
<td>39.6</td>
<td>22.1</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>26.6</td>
<td>26.3</td>
</tr>
<tr>
<td>Mobility complaints*</td>
<td>12%</td>
<td>26%</td>
</tr>
<tr>
<td>Percent fat (%)</td>
<td>27</td>
<td>38</td>
</tr>
<tr>
<td>ALM/height² (kg/m²)</td>
<td>7.88</td>
<td>6.47</td>
</tr>
<tr>
<td>ALM/BMI</td>
<td>0.89</td>
<td>0.61</td>
</tr>
</tbody>
</table>

*any difficulty walking or stair climbing*
Analytic Challenges and Approaches

- Variation of measurements across study populations.
  - Harmonization of body composition (DEXA) and other measurements across community-dwelling cohorts
  - Appropriate scaling of contributing variables (e.g. muscle / lean body mass)

- Uncertainty as to ‘best’ candidate measurements to be included in sarcopenia phenotype
  - Consensus development of candidate measurements of sarcopenia as functions of muscle mass and strength

- Multidimensionality of candidate predictive factors
  - Exploratory (ROC) and machine-learning approaches to derive candidate cutpoints

- Performance of candidate phenotypes against ‘hard’ endpoints
  - Sample-based estimation and regression modeling with eye to ‘dual purpose’ needs
Reduction in overall variation: Cross-calibration of DXA Measurements

Able to harmonize to ‘NHANES’ calibration on Hologic machines.

(Handles systematic, non-stochastic differences in measurement)

Classic to NHANES Pre-APEX 3.3

\[
\text{FATMass}_{\text{NHANES_PreAPEX 3.4}} = \text{FAT}_{\text{Classic}} + 0.054 \times (\text{Lean SoftTissueMass}_{\text{Classic}} + \text{BMC})
\]

\[
\text{Lean SoftTissueMass}_{\text{NHANES_PreAPEX 3.4}} = \text{TotalMass} - \text{BMC} - \text{FATMass}_{\text{NHANES_PreAPEX 3.4}}
\]

Classic to NHANES

\[
\text{FATMass}_{\text{NHANES}} = \text{FATMass}_{\text{Classic}} + 0.054 \times \text{Lean SoftTissueMass}_{\text{Classic}}
\]

\[
\text{Lean SoftTissueMass}_{\text{NHANES}} = \text{TotalMass} - \text{BMC} - \text{FATMass}_{\text{NHANES}}
\]

Shepherd et al JBMR 2012.
Data-driven approach for analytes

• Handles inter-cohort variation via statistical models of association, where systematic differences in assay methodology cannot be quantified by equations

Harmonized Reference Ranges for Circulating Testosterone Levels in Men of Four Cohort Studies in the United States and Europe

Thomas G. Travison,1 Hubert W. Vesper,3 Eric Orwoll,4 Frederick Wu,5 Jean Marc Kaufman,6 Ying Wang,4 Bruno Lapauw,6 Tom Fiers,7 Alvin M. Matsumoto,8 and Shalender Bhasin2

J Clin Endocrinol Metab, April 2017, 102(4):1161–1173
Defining muscle mass contribution – allometric scaling

- Absolute and proportionate scaled to:
  - Body mass
  - BMI
  - body surface area
  - height
  - height-squared or other
  - total & percent fat mass

- Consistency of relative lean mass or strength across populations and ranges of body size
Analytic Challenges and Approaches

• Variation of measurements across study populations.
  – Harmonization of body composition (DEXA) and other measurements across community-dwelling cohorts

• Uncertainty as to ‘best’ candidate measurements to be included in sarcopenia phenotype
  – Consensus development of candidate measurements of sarcopenia as functions of muscle mass and strength

• Multidimensionality of candidate predictive factors
  – Exploratory (ROC) and machine-learning approaches to derive candidate cutpoints

• Performance of candidate phenotypes against ‘hard’ endpoints
  – Sample-based estimation and regression modeling with eye to ‘dual purpose’ needs
## 36 Candidate sarcopenia variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 AGE</td>
<td>19 Maximum grip strength/BMI</td>
</tr>
<tr>
<td>2 ALM</td>
<td>20 Maximum grip strength/Height</td>
</tr>
<tr>
<td>3 Maximum grip strength</td>
<td>21 Maximum grip strength/Height**2</td>
</tr>
<tr>
<td>4 ARM lean mass (ARMLM), kg</td>
<td>22 Maximum grip strength/Total body fat</td>
</tr>
<tr>
<td>5 Appendicular fat mass (AFM), kg</td>
<td>23 ALM/Body Surface Area</td>
</tr>
<tr>
<td>6 Appendicular lean mass/Height</td>
<td>24 Maximum grip strength/Body Surface Area</td>
</tr>
<tr>
<td>7 Appendicular lean mass/BMI</td>
<td>25 Maximum grip strength/LLM</td>
</tr>
<tr>
<td>8 Appendicular lean mass/Total body fat</td>
<td>26 LLM/Body Surface Area</td>
</tr>
<tr>
<td>9 Appendicular lean mass/Weight</td>
<td>27 ALM/Percent Fat</td>
</tr>
<tr>
<td>10 Appendicular lean mass/ height**2, (kg/m2)</td>
<td>28 Maximum grip strength/Percent fat</td>
</tr>
<tr>
<td>11 Leg lean mass (LLM), kg</td>
<td>29 ALM/AFM</td>
</tr>
<tr>
<td>12 Leg lean mass/total body weight</td>
<td>30 Maximum grip strength/AFM</td>
</tr>
<tr>
<td>13 Leg lean mass/Height</td>
<td>31 Maximum grip strength/ARMLM</td>
</tr>
<tr>
<td>14 Leg lean mass/ Height**2</td>
<td>32 Maximum grip strength/ARMLM**2</td>
</tr>
<tr>
<td>15 Leg lean mass/BMI</td>
<td>33 Height</td>
</tr>
<tr>
<td>16 Leg lean mass/Total body fat</td>
<td>34 Weight</td>
</tr>
<tr>
<td>17 Maximum grip strength/ALM</td>
<td>35 Total body fat</td>
</tr>
<tr>
<td>18 Maximum grip strength/weight</td>
<td>36 BMI</td>
</tr>
</tbody>
</table>
Receiver operating curve (AUC)

- Area under (ROC) curve used to compare the accuracy of “diagnosing” low walk speed (<0.80 m/sec) Traditional accuracy ranking:
  - 1.0-0.90=excellent; 0.80-0.89=good; 0.79-0.70=fair; <0.70=poor
Methods: Machine Learning

• Classification and regression tree (CART) analysis
  – Designed to identify subgroups that maximize outcome homogeneity
  – Produces mutually exclusive & exhaustive population subgroups whose members are most similar with respect to the outcome.
Example: Detecting complex interactions

**CART: decline in physical performance**

Considered the following factors: α-tocopherol, vitamin B12, vitamin B6, folate, 25-hydroxyvitamin D, iron, age, sex, educational achievement, marital status, household composition, smoking, physical activity level, number of chronic conditions, body mass index, depression and cognitive function

Bartali et al., *JAMA* 2008.
Sarcopenia Results: MEN

13,877 males

GRBMI > 1.05

GRBMI < 1.05

Age < 80.9

Age ≥ 80.9

12,604 (90.8% of sample)
880 (7.0%) <0.80 m/sec

920 (6.6% of sample)
220 (23.9%) <0.80 m/sec

Additional node with
GRARMLM > 3.13
(see full CART report)

1273 (9.2% of sample)
382 (30%) <0.80 m/sec

Additional node with
ALMBMI > 0.66
(see full CART report)

353 (2.6% of sample)
162 (45.9%) <0.80 m/sec

AGE
74.46 +/- 4
BMM
30.09 +/- 4.89
GRIPMAX
27.26 +/- 5.51
GRBMI
0.91 +/- 0.12
ALMAFM
7.67 +/- 1.19
White
662 (71.96)
Black
47 (5.11)
Asian
201 (21.85)
Other
10 (1.09)

AGE
74.54 +/- 5.4
BMM
26.58 +/- 3.89
GRIPMAX
39.66 +/- 8.77
GRBMI
1.51 +/- 0.36
ALMAFM
3.28 +/- 1.39
White
10920 (78.69)
Black
576 (4.15)
Asian
2185 (15.75)
Other
196 (1.41)

AGE
77.23 +/- 5.89
BMM
29.47 +/- 4.77
GRIPMAX
26.6 +/- 5.44
GRBMI
0.9 +/- 0.13
ALMAFM
2.6 +/- 1.16
White
962 (75.57)
Black
67 (5.26)
Asian
233 (18.3)
Other
11 (0.86)

AGE
84.45 +/- 3.34
BMM
27.85 +/- 4
GRIPMAX
24.9 +/- 4.89
GRBMI
0.9 +/- 0.13
ALMAFM
7.67 +/- 1.19
White
300 (84.99)
Black
20 (5.67)
Asian
32 (9.07)
Other
1 (0.28)

White
9958 (79.01)
Black
509 (4.04)
Asian
1952 (15.49)
Other
185 (1.47)

White
962 (75.57)
Black
67 (5.26)
Asian
233 (18.3)
Other
11 (0.86)

White
10920 (78.69)
Black
576 (4.15)
Asian
2185 (15.75)
Other
196 (1.41)

White
300 (84.99)
Black
20 (5.67)
Asian
32 (9.07)
Other
1 (0.28)
Men

- Those with GRBMI values less than 1.05 were much older and had much higher BMI and much lower grip than those with GRBMI > 1.05.

- Age < 83.5 further discriminates slowness, although the group that the age cut-point applies to (about 10%) is a fairly small subset of the participants.
Results: Women

5,115 Females

GRTBF $\geq 0.65$

4261 (83.3% of sample)
1019 (23.9%) $< 0.80$ m/sec

AGE < 80.5

3460 (67.6% of sample)
712 (20.6%) $< 0.80$ m/sec

Terminal

Additional node with AGE

GRTBF $< 0.65$

854 (16.7% of sample)
426 (49.9%) $< 0.80$ m/sec

AGE > 80.5

801 (15.7% of sample)
307 (38.3%) $< 0.80$ m/sec

Additional node with AGE

GRARMLMSQ $\geq 0.62$

690 (13.5% of sample)
317 (45.9%) $< 0.80$ m/sec

Additional node with ALMAFM

GRARMLMSQ $< 0.62$

164 (3.2% of sample)
109 (66.5%) $< 0.80$ m/sec

Terminal
Women

• Grip strength/total body fat is the first discriminator
  – splits the sample such that 83% has higher strength to total body fat and 17% had lower strength to total body fat

• GRARMLMSQ < 0.62 further discriminated women with low strength to total body fat (3.2% of sample)

• AGE > 80.5 further discriminated women with high strength to total body fat (15.7% of sample)
Analytic Challenges and Approaches

• Variation of measurements across study populations.
  – Harmonization of body composition (DEXA) and other measurements across community-dwelling cohorts

• Uncertainty as to ‘best’ candidate measurements to be included in sarcopenia phenotype
  – Consensus development of candidate measurements of sarcopenia as functions of muscle mass and strength

• Multidimensionality of candidate predictive factors
  – Exploratory (ROC) and machine-learning approaches to derive candidate cutpoints

• Performance of candidate phenotypes against ‘hard’ endpoints
  – Sample-based estimation and regression modeling with eye to ‘dual purpose’ needs
10 y Mortality: Low Grip/BMI By Cohort

Age-adjusted Hazard Ratio (Cox PH Regression)

Men

Low Grip/BMI <1.05

Women

Low Grip/BMI <1.05

HABC, MrOS, CHS, Sweden, CHAMP, MrOS HK, SOF
Conclusions / Lessons Learned

• Combination of classical / and computing-based approaches may yield dividends
  – Attention to definitions, harmonization, and data quality is critical

• The ‘dual-purpose’ problem is highly context dependent
  – Individual-level inference relies critically on calibration of phenotypes

• Modern methods cannot solve – and may exacerbate – tension between complexity and clinical utility
  – Heuristic underpinning of results may not be clear
    • Partially addressed by profiling individuals in specific risk categories
  – Association with other endpoints may not ‘validate’ results to sufficient satisfaction
  – Sarcopenia: Controversy over exclusion of direct measure of muscle in main results (To be continued... )
Acknowledgments

• DEXA Harmonization
  – Karol Pencina, Brigham and Women’s Hospital
  – Kevin Wilson, Tom Kelly; Hologic, Inc.

• Allometric scaling
  – Todd Manini; University of Florida

• Machine Learning analyses:
  – Peggy Cawthon, Sheena Patel; UCSF, San Francisco Coordinating Center
We gratefully acknowledge funding for this project

National Institute on Aging
Foundation for the NIH

Funding for cohort studies included in this presentation was provided by NIA, NIAMS, NINDS, NCATS, NHLBI, NIH Roadmap, Research Grants Council (Hong Kong), The Chinese University of Hong Kong, Swedish Research Council, the Swedish Foundation for Strategic Research, the ALF/LUA research grant in Gothenburg, the Lundberg Foundation, the Torsten and Ragnar Söderberg's Foundation, Petrus and Augusta Hedlunds Foundation, the Västra Götaland Foundation, the Göteborg Medical Society, the Novo Nordisk Foundation, National Health and Medical Research Council (Australia), Ageing and Alzheimer’s Institute (Australia)

NIH grant and contract numbers: AG051421, AG027810, AG042124, AG042139, AG042140, AG042143, AG042145, AG042168, AR066160, TR000128, AR049439-01A1, AG005407, AR35582, AR35583, AR35584, AG005394, AG027574, AG027576, N01AG62101, N01AG62103, N01AG62106, N01-HC- 85079, N01-HC-85080, N01-HC-85081, N01-HC-85082, N01-HC-85083, N01-HC-85084, N01-HC-85085, N01-HC-85086; N01-HC-35129, N01 HC-15103, N01 HC-55222, N01-HC-75150, N01-HC-45133, N01-HC-85239, HHSN268201200036C, HL080295, HL087652, HL105756, HL103612, AG023629, AG15928, AG20098, AG027058