A Measurement Error Model for Physical Activity Level Measured by a Questionnaire, with application to the NHANES 1999-2006 Questionnaire

Janet A. Tooze, Richard P. Troiano, Raymond J. Carroll, Alanna J. Moshfegh, Laurence S. Freedman

International Conference on Diet and Activity Methods
Rome, Italy, 16 May 2012
Objective

• To develop a model to evaluate the measurement error structure of the US National Health and Nutrition Examination Survey (NHANES) 1999-2006 Physical Activity Questionnaire (PAQ)
Physical activity: Definition

• Bodily movement that results in energy expenditure above resting
  • Time
  • Intensity
  • Measured by METs, the ratio of the activity level during activity to rest
Physical activity: Measures

- MET minutes
- Energy expenditure
  - Total – all energy expended (TEE)
  - Basal – energy expended at rest (BEE)
- Physical Activity Level (PAL)
  - PAL = TEE/BEE
The goal

Construct Physical Activity

Outcome Disease

\( \alpha_1 \)
The reality

Construct Physical Activity

= \neq \sim_1

Error-prone Measure Physical Activity

Outcome Disease
The questions

- What types of Error does the measure have?
- What are the implications of using this instrument?
Types of measurement error

\[ Q = \text{questionnaire} \]
\[ P = \text{true physical activity level (PAL)} \]
\[ \text{person } i \]
\[ (\text{on log scale}) \]

\[ Q_i = \beta_0 + \beta_1 P_i + \epsilon_i \]

Constant additive error

Random error

Proportional bias
Quantifying types of error

Error-prone Measure
Physical Activity

Unbiased Measure
Physical Activity

\[ Q_i = \beta_0 + \beta_1 P_i + \varepsilon_i \]

\[ M_{ij} = W_{ij} - E_i = P_i + \pi_{ij} + \gamma_i \]

From DLW
Provides unbiased measure of \( \ln(\text{TEE}), T_i \)

From Equation
Provides unbiased measure of \( \ln(\text{BEE}), B_i \)

Q=Questionnaire
P=True PAL
M=biomarker
W=Doubly labeled water
E=BEE from equation
T=true TEE
B=true BEE
\( \varepsilon, \pi, \gamma \) = Error terms

\[ P_i = T_i - B_i \]
Quantifying types of error

\[ W_{ij} = T_i + \pi_{ij} \]  
Classical measurement error

\[ B_i = E_i + \gamma_i \]  
Berkson error

\[ M_{ij} = W_{ij} - E_i = P_i + \pi_{ij} + \gamma_i \]

\[ \text{var}(M_{ij}) = \text{var}(P_i) + \text{var}(\pi_{ij}) + \text{var}(\gamma_i) + 2\text{cov}(P_i, \gamma_i) \]

Therefore, \( P_i \) may be correlated with \( \gamma_i \)

\[ \text{cov}(P_i, \gamma_i) \] obtained from external study that has an unbiased measure of \( B_i \)

\[ C_i = B_i + \eta_i \]

Assume \( \eta_i \) is independent of \( \pi_{ij} \)

\[ \text{cov}(P_i, \gamma_i) = \text{cov}(W_i, C_i) - \text{cov}(W_i, E_i) - \text{var} \gamma_i \]
Quantifying types of error

• Once we are able to estimate the variance components for our “biomarker” M, the random error for DLW W, Berkson error associated with BEE, and the covariance between physical activity and the error associated with BEE, we can calculate an estimate of the proportional bias $\beta_1$:

$$1 = \frac{\text{cov}(Q_i, M_{ij})}{\text{var}(M_{ij}) \cdot \text{var}(p_{ij}) \cdot \text{var}(\gamma) \cdot \text{cov}(P_i, \gamma)}$$

$Q$=questionnaire  
$P$=True PAL  
$M$=biomarker  
$\varepsilon, \pi, \gamma$ = Error terms  
$\text{cov} =$covariance
The questions

• What types of Error does the measure have?

• What are the implications of using this instrument?
Implications of measurement error

- **Attenuation factor**: the degree to which a regression coefficient is biased to the null (attenuated) due to measurement error
  - Closer to zero = more attenuation

\[ \beta_1 = \frac{\text{cov}(P_i, Q_i)}{\text{var}(Q_i)} = \frac{1}{\text{var}(Q_i)/\text{var}(P_i)} \]

- P=True PAL
- Q=Questionnaire
- \( \beta_1 \)=proportional bias
- cov=covariance
Implications of measurement error

- **Correlation between self-report and truth**: related to statistical power to detect activity-health relationships
  - Closer to zero = less powerful the study will be (i.e., need larger sample size)

\[
QP = \frac{\text{cov}(P_i, Q_{ij})}{\sqrt{\text{var}(P_i) \text{var}(Q_{ij})}} = \frac{1}{\sqrt{1 + \text{var}(Q_{ij})/\text{var}(P_i)}}
\]

- **P** = True PAL
- **Q** = Questionnaire
- **\( \beta_1 \)** = Proportional bias
- **cov** = covariance
Application
The OPEN Study

- Observing Protein and Energy Nutrition Study
- Participants:
  - 484 men and women
    - 433 with complete data
  - 40-69 years old
- Suburban Maryland
The OPEN Study: Measures

• Doubly labeled water (DLW) for total energy expenditure (TEE)

• A subsample (N=25) of participants were dosed with DLW a second time
Calculation of PAL from DLW

- Basal energy expenditure (BEE) was calculated from an equation from the USDA AMPM study that uses height, weight, age, gender

- \( \text{PAL}_{\text{DLW}} = \frac{\text{TEE}_{\text{DLW}}}{\text{BEE}} \)
The OPEN Study: NHANES PAQ

- PAQ queries activities over past 30 days
  - Frequency
  - Duration
- Multiple domains measured:
  - Transportation-related
  - Household
  - Moderate (specific)
  - Vigorous (specific)
Calculation of MET values from PAQ

• Each activity assigned a MET value:
  • Transportation: 4.0
  • House/Yard work: 4.5
  • Moderate and Vigorous: value from compendium (Ainsworth)

• Time spent in each activity multiplied by METs and summed to compute MET minutes per day

• Assumed sleep of 6.9 hrs per day = 1.0 MET

• “Usual” activity = 24 – 6.9 – time in activities

• Computed from usual activity question
Calculation of PAL from PAQ

• $\text{PAL}_Q = \frac{\text{TEE}_Q}{\text{BEE}}$

• Because METs are expressed as multiples of BEE, it cancels out of the equation.

• PAL from the questionnaire is a function of the minutes of reported activity and the METs for each activity, but not BEE
<table>
<thead>
<tr>
<th></th>
<th>Women (N=201)</th>
<th>Men (N=232)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40-49</td>
<td>39.8%</td>
<td>36.6%</td>
</tr>
<tr>
<td>50-59</td>
<td>36.8%</td>
<td>32.8%</td>
</tr>
<tr>
<td>60-69</td>
<td>23.4%</td>
<td>30.6%</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;=25</td>
<td>38.8%</td>
<td>24.6%</td>
</tr>
<tr>
<td>25-29.9</td>
<td>31.3%</td>
<td>46.6%</td>
</tr>
<tr>
<td>&gt;=30</td>
<td>29.9%</td>
<td>28.9%</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Hispanic White</td>
<td>76.6%</td>
<td>86.2%</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
</tr>
<tr>
<td>College Graduate</td>
<td>54.8%</td>
<td>71.9%</td>
</tr>
<tr>
<td></td>
<td>OPEN Study</td>
<td>AMPM Study</td>
</tr>
<tr>
<td>----------------------</td>
<td>-----------------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td></td>
<td>Women (N=201)</td>
<td>Men (N=232)</td>
</tr>
<tr>
<td>TEE (kcal/d)</td>
<td>2311 (394)</td>
<td>2899 (530)</td>
</tr>
<tr>
<td>BEE (IC)</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>BEE (equation)</td>
<td>1408 (160)</td>
<td>1749 (183)</td>
</tr>
<tr>
<td>PAL (IC)</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>PAL (equation)</td>
<td>1.64 (0.21)</td>
<td>1.65 (0.21)</td>
</tr>
<tr>
<td>PAL (Questionnaire)</td>
<td>1.61 (0.13)</td>
<td>1.62 (0.14)</td>
</tr>
<tr>
<td>Min activity/d</td>
<td>46 (44)</td>
<td>47 (47)</td>
</tr>
<tr>
<td>MET minutes/d</td>
<td>213 (210)</td>
<td>218 (210)</td>
</tr>
</tbody>
</table>
Error: “flattened slope”

Proportional Bias 0.23 (0.05)
Correlation Between PAQ and Truth

<table>
<thead>
<tr>
<th></th>
<th>Men</th>
<th>Women</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation</td>
<td>0.41 (0.12)</td>
<td>0.32 (0.09)</td>
</tr>
</tbody>
</table>

- Sensitivity analysis varying assumptions for sleep, METs for usual activity had little impact on the estimates.
- Correlations of this magnitude will have a large impact on power and sample size.
### Attenuation

<table>
<thead>
<tr>
<th></th>
<th>Men</th>
<th>Women</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attenuation</td>
<td>0.73 (0.12)</td>
<td>0.43 (0.12)</td>
</tr>
</tbody>
</table>

- Sensitivity analyses had little impact on the results.
- If physical activity decreases the risk of disease by 50% in fact, the PAQ estimates the decreased risk as 40% for men and 26% for women.
  
  $$\text{ObsRR} = (\text{TrueRR})^\lambda$$
  $$\text{ObsRR}_M = (0.5)^{0.73} = 0.60$$
  $$\text{ObsRR}_W = (0.5)^{0.43} = 0.74$$

- These estimated RRs may or may not be significant considering the noise due to unknown confounders.
Discussion

• Limitations of OPEN Study:
  • Did not measure BEE, which lead to Berkson error
    • Had to use external estimate
  • Not generalizable to US population

• Advantages
  • Large DLW sample of men and women
  • We used PAL to estimate physical activity, which is a global measure
  • A measurement error modeling framework may be applied to physical activity questionnaire data
Conclusion

• This physical activity questionnaire may not capture activity-disease relationships adequately due to the measurement error.

• Protective effects of physical activity may be greater than measured by the questionnaire based PAL.

• Regression calibration should be considered in studies of physical activity and health outcomes to adjust for measurement error.
Acknowledgements

• OPEN Study investigators
• David Bassett
• This work was funded by R21CA139291 from the National Cancer Institute