Measurement Error Modeling of Energy Balance and Physical Activity Data

Danny Ries

Iowa State University

March 5, 2017

The Obesity Epidemic

Over 35% of Americans are obese, and over 75% of men are either overweight or obese. Obesity is linked to many different medical, psychological, emotional, and economic effects such as:

- Type 2 Diabetes
- Coronary Heart Disease
- High Blood Pressure
- Clinical Depression
- Anxiety
- Increased Health Care Costs
- Lost Wages
- Discrimination

2 Remedies to aid in Obesity Research

1. Accurately and efficiently measure Energy Intake (EI)
2. Assess compliance to *2008 Physical Activity Guidelines*

To address the difficulty in measuring EI:

Bayesian Semi-Parametric Energy Balance Measurement Error Model

Modeling Energy Balance

Energy Balance

Change in Energy Stores (ΔES) = Energy Intake (EI) - Energy Expenditure (EE)

where $\Delta \text{ES} = c_1 \frac{\Delta FM}{\Delta T} + c_2 \frac{\Delta FFM}{\Delta T}$

Current measures of EI; ie. self report, are clouded with (measurement) error, no true gold standard
$\rightarrow$ But gold standard measures (albeit very costly) and cheap measurements for EE and ΔES exist!

Research Goal

Create a joint statistical measurement model for gold standard and cheap measurements of both EE and ΔES in order to develop calibration equations for cheap measurements.

- Allows future researchers to calibrate cheaper measurements
- Lots of research has been done for calibrating and evaluating measurement error for EI, some for EE, and little for ΔES
- To the best of our knowledge, no research has been done in evaluating the measurement error and calibrating measurements jointly via the Energy Balance principle

Energy Balance Study

The Energy Balance Study (EBS) was conducted 2011-2012 at the University of South Carolina

- 430 male and females aged 20-35
- 5 DXA scans, one every 3 months
- Sensewear Armband measuring EE every 3 months (averaged across 10 days)
- Subset of 119 participants receieved DLW at end of 12 months, with additional DXA scan

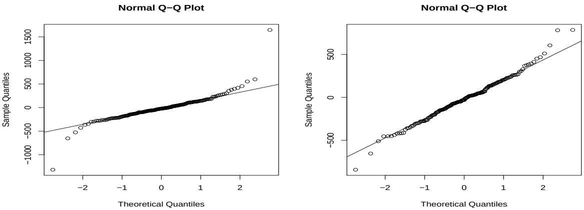


Figure: Differenced DXA ΔES, Females left, Males right

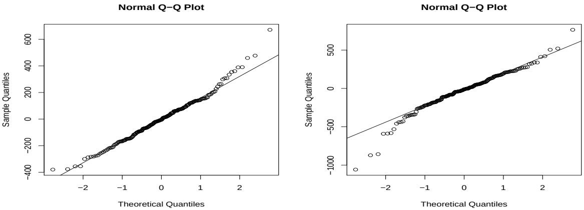


Figure: Differenced SWA EE , Females left, Males right

Naïve Check for Biases

$$\text{SWA EE} = \beta_0 + \beta_1 \text{DLW EE} + \beta_z\text{' Demographic Variables} + \epsilon$$

Coefficient	Estimate	Std Error	P-value
Intercept	878.422	154.113	<0.0001
DLW EE	0.558	0.040	<0.0001
Age	-7.351	3.999	0.0676
Male	305.582	43.258	<0.0001
BMI	14.146	3.988	0.0004

Notation

let i represent individual and j represent replicate number

Observable:

- W_{ij}^{EE} and $W_{ij}^{\Delta ES}$ represent gold standard measures of EE and ΔES
- Y_{ij}^{EE} and $Y_{ij}^{\Delta ES}$ represent cheap measures of EE and ΔES
- Z_i represent a $k \times 1$ vector of error free covariates

Latent:

- X_i^{EE} and $X_i^{\Delta ES}$ represent *usual* EE and ΔES

$$E(W_{ij}^{EE}|i) = X_i^{EE}$$
$$E(W_{ij}^{\Delta ES}|i) = X_i^{\Delta ES}$$

Independence Assumptions

We assume individuals are independent

Additionally, given X_i^{EE}, Z_i

- Y_{ij}^{EE} are independent for all j
- W_{ij}^{EE} are independent for all j
- Y_{ij}^{EE} is independent of W_{ij}^{EE} for j
- Y_{ij}^{EE} is independent of $W_{ij}^{\Delta ES}$ and $Y_{ij}^{\Delta ES}$ for all i, j
- W_{ij}^{EE} is independent of $W_{ij}^{\Delta ES}$ and $Y_{ij}^{\Delta ES}$ for all j

Same assumptions hold for reverse case (replace EE with ΔES and ΔES with EE)

Naïve Model

The Naïve Model assumes no measurement error in gold standard measurements

$$(Y_{ij}^{EE}|W_{ij}^{EE}, Z_i, \boldsymbol{\theta_{yee}}) \overset{ind}{\sim} N(\beta_{0,ee} + \beta_{1,ee} W_{ij}^{EE} + \boldsymbol{\gamma_{ee}}' Z_i, \sigma^2_{\epsilon^{EE}})$$
$$(Y_{ij}^{\Delta ES}|W_{ij}^{\Delta ES}, Z_i, \boldsymbol{\theta_{yes}}) \overset{ind}{\sim} N(\beta_{0,es} + \beta_{1,es} W_{ij}^{\Delta ES} + \boldsymbol{\gamma_{es}}' Z_i, \sigma^2_{\epsilon^{\Delta ES}})$$

Linear Measurement Error Model (LMEM)

Basic modification to the Naïve model when we account for measurement error in a covariate

$$
\begin{aligned}
(Y_{ij}^{EE}|X_i^{EE}, Z_i, \boldsymbol{\theta}_{\boldsymbol{yee}}) &\overset{ind}{\sim} N(\beta_{0,ee} + \beta_{1,ee}X_i^{EE} + \boldsymbol{\gamma}_{\boldsymbol{ee}}{}' Z_i, \sigma^2_{\epsilon_{EE}}) \\
(Y_{ij}^{\Delta ES}|X_i^{\Delta ES}, Z_i, \boldsymbol{\theta}_{\boldsymbol{yes}}) &\overset{ind}{\sim} N(\beta_{0,es} + \beta_{1,es}X_i^{\Delta ES} + \boldsymbol{\gamma}_{\boldsymbol{es}}{}' Z_i, \sigma^2_{\epsilon_{\Delta ES}}) \\
(W_{ij}^{EE}|X_i^{EE}, Z_i, \boldsymbol{\theta}_{\boldsymbol{wee}}) &\overset{ind}{\sim} N(X_i^{EE}, \sigma^2_{\nu_{EE}}) \\
(W_{ij}^{\Delta ES}|X_i^{\Delta ES}, Z_i, \boldsymbol{\theta}_{\boldsymbol{wes}}) &\overset{ind}{\sim} N(X_i^{\Delta ES}, \sigma^2_{\nu_{\Delta ES}}) \\
(X_i^{EE}, X_i^{\Delta ES}|\boldsymbol{\theta}_{\boldsymbol{X}}) &\overset{ind}{\sim} N\left(\begin{bmatrix} \mu_{EE} \\ \mu_{\Delta ES} \end{bmatrix}, \Sigma_X\right)
\end{aligned}
$$

Extending the Linear Model

Relax the assumption that the relationship between a cheap measurement and *usual* EE and ΔES is linear

We propose using cubic free knot splines to model the relationship between cheap measurement and *usual*

- Allows for a flexible nonlinear relationship
- No need to specify number or location of knots
- Using Reversible Jump MCMC, incorporates uncertainty in spline selection

Free Knot Spline Model (SMEMN)

$$(Y_{ij}^{EE}|X_i^{EE}, Z_i, \boldsymbol{\theta_{yee}}) \overset{ind}{\sim} N(s_{ee}(X_i^{EE}; \boldsymbol{\beta_{ee}}) + \boldsymbol{\gamma_{ee}} Z_i, \sigma^2_{\epsilon^{EE}})$$
$$(Y_{ij}^{\Delta ES}|X_i^{\Delta ES}, Z_i, \boldsymbol{\theta_{yes}}) \overset{ind}{\sim} N(s_{es}(X_i^{\Delta ES}; \boldsymbol{\beta_{\Delta es}}) + \boldsymbol{\gamma_{es}} Z_i, \sigma^2_{\epsilon^{\Delta ES}})$$

$$s_{ee}(X_i^{EE}; \boldsymbol{\beta_{ee}}) = \sum_{i=1}^{k_{ee}+3} b_{i,ee}(\mathbf{X^{EE}})\beta_{i,ee} = B_{ee}(\mathbf{X^{EE}})\boldsymbol{\beta_{ee}}$$
$$s_{es}(X_i^{\Delta ES}; \boldsymbol{\beta_{\Delta es}}) = \sum_{i=1}^{k_{es}+3} b_{i,es}(\mathbf{X^{\Delta ES}})\beta_{i,es} = B_{es}(\mathbf{X^{\Delta ES}})\boldsymbol{\beta_{es}}$$

In order to ensure the functions are monotone, constrain
$\beta_{1,ee} \leq \beta_{2,ee} \leq \dots \leq \beta_{k_{ee}+3,ee}$

Extending the Normality of Latent Variables

Allow more flexibility in modeling of latent variables, use Dirichlet process mixture prior (SMEMDP):

$$
\begin{aligned}
X_i^{EE}, X_i^{\Delta ES} | \zeta_i = h &\overset{ind}{\sim} N\left(\begin{bmatrix} \mu_{ee,h} \\ \mu_{es,h} \end{bmatrix}, \Sigma_h \right) \\
\zeta_i &\overset{iid}{\sim} Cat(H, \boldsymbol{\pi}) \\
V_h &\sim \text{Beta}(1, \alpha) \\
V_H &= 1 \\
\pi_h &= V_h \prod_{\ell < h} (1 - V_\ell)
\end{aligned}
$$

Gibbs Sampler

Priors were chosen to be conjugate to help simplify the sampler

For iteration k=1,...,K, sample from its full conditional:

1. $\{\Sigma^{(k)}\} \mid \cdot \overset{ind}{\sim}$
   $Inv - Wish(d + n, \psi + (\mathbf{X_i^{(k-1)}} - \boldsymbol{\mu}^{(\mathbf{k-1})})'(\mathbf{X_i^{(k-1)}} - \boldsymbol{\mu}^{(\mathbf{k-1})}))$
2. $\{(\mu_{EE}^{(k)}, \mu_{\Delta ES}^{(k)})\ \} \mid \cdot \overset{ind}{\sim} N(M'_\mu, C'_\mu)$
   $C'_\mu = (C_\mu^{-1} + n\Sigma^{-1(k)})^{-1}$
   $M'_\mu = C'_\mu(C_\mu^{-1}M + n\Sigma^{-1(k)}\bar{\mathbf{X}}^{(\mathbf{k-1})})$
   $\bar{\mathbf{X}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{X_i}$
3. Update $\mathbf{X_i}$ with a random walk
   $\{\mathbf{X_i^{(k)}} : i = 1, ..., n\} \mid \cdot$ for $i = 1, ...n$ sample $\mathbf{X_i^*}$ from $N(\mathbf{X_i^{(k)}}, C_{X_i})$
   Set $\mathbf{X_i^{(k)}} = \mathbf{X_i^*}$ with probability α_{X_i}, otherwise set $\mathbf{X_i^{(k)}} = \mathbf{X_i^{(k-1)}}$
   $\alpha_{X_i} = min\left(1, \frac{f(\mathbf{X_i^*}|\cdot)}{f(\mathbf{X_i^{(k-1)}}|\cdot)}\right)$

4. Update $\boldsymbol{\beta_{ee}}, \boldsymbol{\beta_{es}}, k_{ee}, k_{es}, r_{ee}, r_{es}, \boldsymbol{\gamma_{ee}}, \boldsymbol{\gamma_{es}}$ using RJMCMC described next. Calculate $s_{ee}(\mathbf{X^{EE(k)}}; \boldsymbol{\beta_{ee}^{(k)}})$ and $s_{es}(\mathbf{X^{\Delta ES(k)}}; \boldsymbol{\beta_{es}^{(k)}})$
5. $\sigma^{2(k)}_{\epsilon^{EE}}|\cdot \sim IG(a_{yee} + J \times \frac{n}{2}, b_{yee} + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{J}(Y_{ij}^{EE} - m_{ee}(X_i^{EE(k)}; \boldsymbol{\beta_{ee}^{(k)}}) - \boldsymbol{\gamma_{ee}}^{(k)} Z_i)^2)$
6. $\sigma^{2(k)}_{\epsilon^{\Delta ES}}|\cdot \sim IG(a_{yes} + J \times \frac{n}{2}, b_{yes} + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{J}(Y_{ij}^{\Delta ES} - m_{es}(X_i^{\Delta ES(k)}; \boldsymbol{\beta_{es}^{(k)}}) - \boldsymbol{\gamma_{es}}^{(k)} Z_i)^2)$
7. $\sigma^{2(k)}_{\nu^{EE}}|\cdot \sim IG(a_{wee} + J \times \frac{n}{2}, b_{wee} + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{J}(W_{ij}^{EE} - X_i^{EE(k)})^2)$
8. $\sigma^{2(k)}_{\nu^{\Delta ES}}|\cdot \sim IG(a_{wes} + J \times \frac{n}{2}, b_{wes} + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{J}(W_{ij}^{\Delta ES} - X_i^{\Delta ES(k)})^2)$

RJMCMC

RJCMCMC was introduced by Green (1995) as a way to “jump” between different models and dimensions as a means for model determination. Denison (1998) introduced a simple and effective implementation of a Bayesian free-knot spline

We extend method of Denison in three ways:

1. The spline covariate is a latent variable that is also sampled from
2. There are additional linear components to the mean function
3. We make the modification from DiMatteo (2001) so the likelihood ratio is an approixmation of BIC

Our algorithm is run independently for EE and ΔES regression functions due to conditional independence, so we will let $(\cdot)$ be a placeholder

$\mathcal{A}$=$\{x_i, i = 1, ...n : x_i$ is not currently a knot or within $\ell + 1$ locations of a current knot$\}$

1. Calculate $b_k = c \times min\left(1, \frac{p(k+1)}{p(k)}\right)$
   $d_k = c \times min\left(1, \frac{p(k)}{p(k+1)}\right)$
2. Select birth, death, or move step with probabilities $b_k, d_k, 1 - b_k - d_k$ respectively
3. **Knot Changes**
   If birth step:
   Select a new knot location at random from the set $\mathcal{A}$ and join with current knots $r_{\cdot}^{(k-1)}$ to create the proposed knot locations $r_{\cdot}^{*}$
   If death Step:
   Sample one knot location from $r_{\cdot}^{(k-1)}$ at random and remove it.
   If move step:
   Sample one knot location from $r_{\cdot}^{(k-1)}$ at random, and change it to a new knot location at random from the set $\mathcal{A}$

4. Calculate the spline basis matrix $B^*_\cdot(X^{\cdot(k)})$ using $X^{\cdot(k)}$ and proposed knot locations $r^*_\cdot$
5. Calculate proposed spline and linear regression coefficients $\boldsymbol{\beta}^*_\cdot, \gamma^*_\cdot$ by using OLS by regressing $Y^\cdot$ on $B^*_\cdot(X^{\cdot(k)}) + Z$
6. Accept proposed knots and coefficients with probability α. Otherwise set $r_\cdot^{(k)} = r_\cdot^{(k-1)}$, $\boldsymbol{\beta}_\cdot^{(\boldsymbol{k})} = \boldsymbol{\beta}_\cdot^{(\boldsymbol{k-1})}$, $\boldsymbol{\gamma}_\cdot^{(\boldsymbol{k})} = \boldsymbol{\gamma}_\cdot^{(\boldsymbol{k-1})}$

$$\alpha_{birth} = min\left(1, \text{Likelihood ratio} \times \frac{n - Z(k)}{n}\right)$$
$$\alpha_{death} = min\left(1, \text{Likelihood ratio} \times \frac{n}{n - Z(k)}\right)$$
$$\alpha_{move} = min\left(1, \text{Likelihood ratio}\right)$$
$$Z(k) = 2(\ell + 1) + k_\cdot(2\ell + 1)$$
$$k_\cdot = length(r_\cdot^{(k-1)})$$

7. calculate mean function $m_\cdot(X^{\cdot(k)}; \boldsymbol{\beta}_\cdot^{(\boldsymbol{k})})$ using spline basis matrix $B'_\cdot(X^{\cdot(k)})$ and $r_\cdot^{(k)}$

Simulation Study

Simulated 200 data sets for both 2 and 4 replicates per individual. Number of individuals was set to be 300. Three different measurement errors were used for each set of replicates:

- Normal
- Skew-Normal
- Bimodal (50-50 mixture of two Normals)

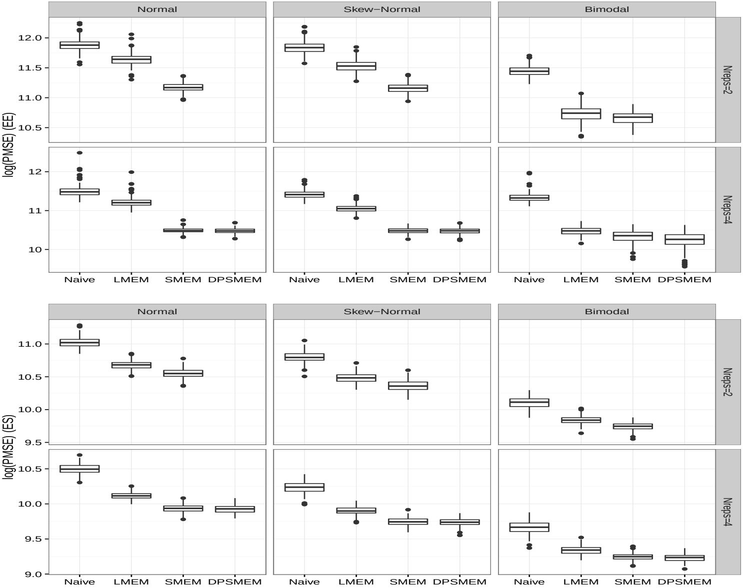
For each simulated data set, we ran the MCMC for 12,000 iterations, using the first 2000 as burn in

Issues with SMEMDP with 2 replicates

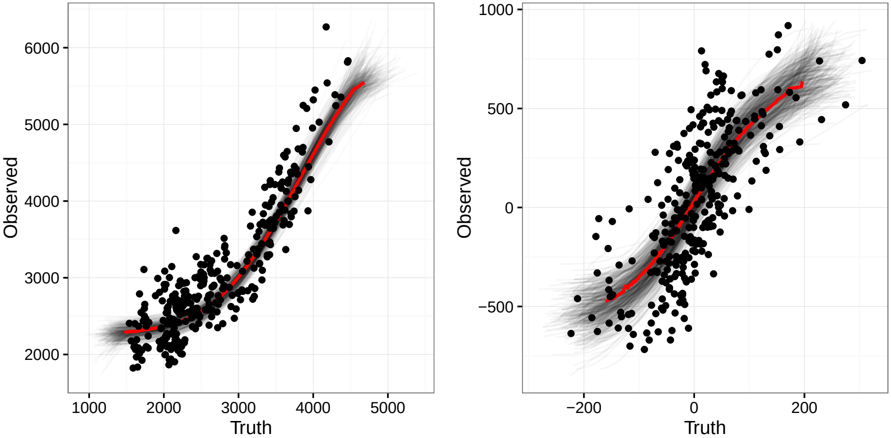
	σ_{yee}		σ_{yes}		$\gamma_{1,ee}$		$\gamma_{2,ee}$		$\gamma_{3,ee}$		$\gamma_{1,es}$		$\gamma_{2,es}$		$\gamma_{3,es}$
Replicates	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2
Mean Est	477.65	473.42	347.49	354.85	254.67	248.66	14.88	14.03	-4.14	-5.29	-200.53	-199.39	7.91	8.31	-4.9
Std Err	17.82	19.24	9.43	6.80	43.65	36.10	4.33	3.62	3.37	3.06	28.19	22.83	2.90	2.13	2.35
Bias	72.15	67.92	13.49	20.85	-45.33	-51.34	0.88	0.03	2.86	1.71	-0.53	0.61	-0.09	0.31	0.06
Truth	405.50	405.50	334.00	334.00	300.00	300.00	14.00	14.00	-7.00	-7.00	-200.00	-200.00	8.00	8.00	-5.0

	σ_{yee}		σ_{yes}		σ_{wee}		σ_{wes}		$\gamma_{1,ee}$		$\gamma_{2,ee}$		$\gamma_{3,ee}$		$\gamma_{1,es}$
Replicates	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2
Mean Est	444.34	446.63	320.41	338.26	255.70	255.85	69.18	71.81	249.50	240.63	14.30	13.67	-4.50	-5.28	-199
Std Err	16.84	14.22	10.76	7.53	10.74	6.33	2.27	1.56	43.44	37.02	4.25	3.60	3.38	3.04	28.3
Bias	38.84	41.13	-13.59	4.26	5.70	5.85	-3.68	-1.05	-50.50	-59.37	0.30	-0.33	2.50	1.72	0.73
Truth	405.50	405.50	334.00	334.00	250.00	250.00	72.86	72.86	300.00	300.00	14.00	14.00	-7.00	-7.00	-200

	σ_{yee}		σ_{yes}		σ_{wee}		σ_{wes}		$\gamma_{1,ee}$		$\gamma_{2,ee}$		$\gamma_{3,ee}$		$\gamma_{1,es}$
Replicates	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2
Mean Est	393.69	400.55	313.47	331.79	246.81	248.93	67.61	71.04	293.16	294.61	14.17	14.11	-6.86	-6.78	-200
Std Err	11.52	8.29	12.00	8.27	8.78	5.79	2.32	1.66	36.16	26.19	3.50	2.42	2.86	2.16	26.8
Bias	-11.81	-4.95	-20.53	-2.21	-3.19	-1.07	-5.25	-1.82	-6.84	-5.39	0.17	0.11	0.14	0.22	-0.0
Truth	405.50	405.50	334.00	334.00	250.00	250.00	72.86	72.86	300.00	300.00	14.00	14.00	-7.00	-7.00	-200


Normal
Skew−Normal
Bimodal
log(PMSE) (EE)
12.0
11.5
11.0
10.5
12
11
10
Nreps=2
Nreps=4
Naive
LMEM
SMEM
DPSMEM
Normal
Skew−Normal
Bimodal
log(PMSE) (ES)
11.0
10.5
10.0
9.5
9.0
Nreps=2
Nreps=4
Naive
LMEM
SMEM
DPSMEM

Example of one data set



Calibration

$$X_{calibrated} = s^{-1}(y - \gamma' Z)$$

For r = 1,...R

1. Calculate $y_i^* = y_i - \gamma^{(r)'} Z_i$, where Z_i are the covariate values for individual i
2. Use `optimize` for the function $|s_i(x) - y_i^*|$ to choose the value of x that will minimize the aforementioned criteria, call this $x_{i,calibrated}^{(r)}$. $s_i(x)$ is the predicted value of y_i for the given value x using the MCMC draw for the spline coefficients, latent variables, and knot locations from the r^{th} draw of the chain

3 example calibrations

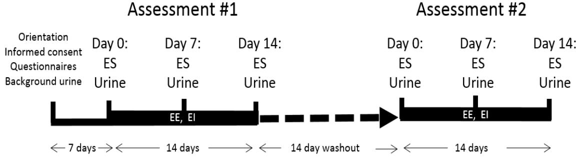
	Gender	BMI	Age
1	M	28.6	20.5
2	F	21.5	30.1
3	M	38.6	22.8

	Lower	Median	Upper	Observed	Truth
EE	2574.18	2666.00	2736.39	3028.89	2199.25
	3452.51	3525.18	3619.08	4119.26	3588.12
	2571.99	2665.46	2744.65	2555.86	2643.14
ΔES	25.15	42.35	60.57	142.30	64.17
	-104.21	-82.93	-63.90	-405.74	-21.08
	-8.41	3.91	17.83	96.06	-0.48

Table: 95% credible interval for calibration estimate for cheap measurements for Skewed Errors

Data Collection

- 30 participants
- Free living, no treatment
- M/F
- Ages 25-35
- BMI 18-30
- Convenience sample



EE: DLW (subcontract with Mass Spectrometry Core at Pennington Biomedical Research Center), Actigraph wGT3X-BT, Actigraph Link, Metalogic Lume
EI: 24-hr dietician administered dietary recalls (subcontract with Cancer Prevention and Control Program in the Arnold School of Public Health at the University of South Carolina
ES: DXA (administered by Nutrition and Wellness Research Staff), body weight, waist circumference, In-Body bioelectrical impedance, Hologic bioelectrical impedance

Discussion

We developed a Bayesian semi-parametric measurement error model for energy balance measurements

- Constructed a RJMCMC algorithm for free knot splines as a function of latent variables along with a linear component
- Developed calibration algorithm to correct for biases in cheap measurements
- Simulation study shows predictive power and mild robustness of SMEMN
- Data analysis (coming soon!)

To Assess Compliance with *2008 Physical Activity Guidelines*:

A Bayesian two-part model with measurement error: Assessing adult moderate to vigorous physical activity and compliance to 2008 Physical Activity Guidelines

2008 Physical Activity Guidelines

Because of the many benefits linked to prevention and treatment of numerous diseases The Department of Health and Human Services Recommends Adults 18-65 participate each week in at least:

- 150 Minutes of Moderate Physical Activity, or
- 75 Minutes of Vigorous Physical Activity, or
- 75-150 Minutes during some combindation of Moderate and Vigorous Physical Activity (MVPA)
- Resistance training targeting all major muscle groups twice
- Most benefits occur when activity is performed in at least 10 minute *bouts*

What proportion of adults adhere to these guidlines? Are there differences among different demographics?

Physical Activity Measurement Survery (PAMS)

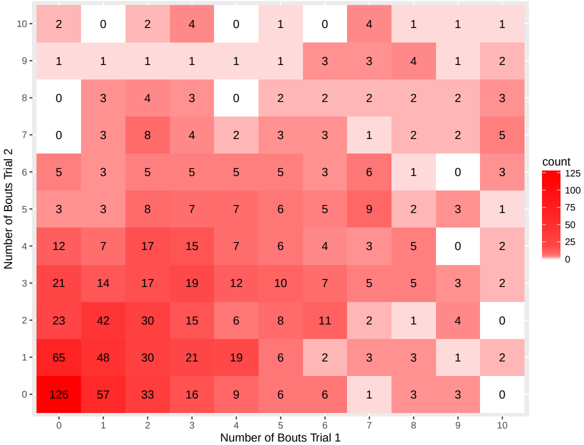
2 year studying starting in 2009 collecting 2 days of physical activity data via Sensewear Armband and 24 Hour Recall on adults in four Iowa counties

- Measurements of physical activity are over 24 hours
- Replicate measurements taken 2-3 weeks apart
- Minute by minute data from armband (in MET-minutes)
- Survery weights needed
- Measurement error involved

- Y_{1ij} : Number of bouts individual i during trial j participated in per day
- Y_{2ij} : Total MET-minutes individual i during trial j accrued in at least 10 minute bouts minues 30*Y_{1ij} per day (total excess MET-minutes)
- Z_i : Observed demographic variables for individual i
- X_{1i} Individual i's expected number of bouts per day
- X_{2i} Individual i's expced total excess minutes per day
- X_{3i} Individual i's expected total minutes in MVPA per day $(= 30 * X_{1i} + X_{2i})$

Exchangeability within person for Y_1

Bowker's test of symmetry: p-value = 0.1474



Exchangeability within person for Y_2

Total excess MET-minutes depends on number of bouts, need to account for it as well as eliminate individual effects

$$Y_{2i1} - Y_{2i2} = \beta_0 + \beta_1(Y_{1i1} - Y_{1i2}) + \beta_2(Weekend_{i1} - Weekend_{i2}) + \epsilon_i$$
$$\epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$$

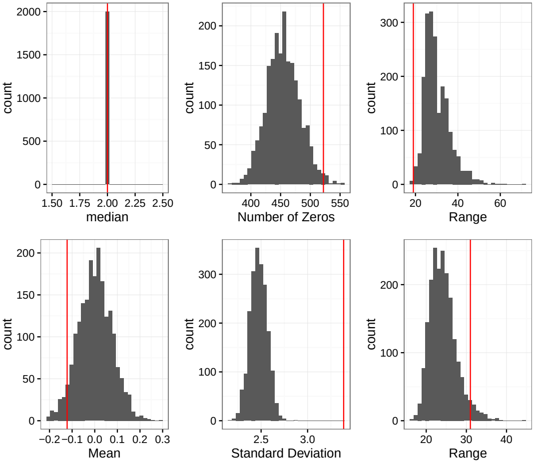
Coefficient	Estimate	Std Error	P-value
β_0	10.129	7.765	0.192
β_1	75.275	2.315	$<$0.0001
β_2	3.234	12.868	0.802

Modeling Number of Bouts

First thought is a Poisson-Gamma mixture with mean of Gamma as a regression on covariates **Z** (use BMI,age,gender,smoker)

$$
\begin{aligned}
Y_{1ij}|X_{1i} &\overset{ind}{\sim} \text{Poisson}(X_{1i}) \\
X_{1i}|\eta,\mu_i &\overset{ind}{\sim} \text{Gamma}\left(\eta, \frac{\eta}{\mu_i}\right) \\
\mu_i &= e^{Z_i'\boldsymbol{\beta}} \\
p(\eta) &\sim \text{Gamma}(a_1, a_2) \\
p(\boldsymbol{\beta}) &\sim N\left(\mathbf{0_k}, \frac{1}{v} I_{k\times k}\right)
\end{aligned}
$$

Assessing Overdispersion



Two-Part Model for Y_2

$$Y_{2ij}|X_{2i}, X_{1i} \overset{ind}{\sim} (1-\pi_{ij}(\boldsymbol{\alpha}))\delta_0(Y_{2ij}) + \pi_{ij}(\boldsymbol{\alpha})LN(\mu_{y,i}(\boldsymbol{\beta_y}), \sigma_y^2)$$
$$X_{2i}|X_{1i}, Z_i \overset{ind}{\sim} LN(\mu_{x,i}(\boldsymbol{\beta_x}), \sigma_x^2)$$

$$\pi_{ij}(\boldsymbol{\alpha}) = \Phi(\alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{2i})$$

$$\mu_{y,i}(\boldsymbol{\beta}) = E(\log\ Y_{2ij}|Y_{2ij} > 0) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i}$$

$$\mu_{x,i}(\boldsymbol{\gamma}) = E(\log\ X_{2ij}) = \gamma_0 + \gamma_1 X_{1i} + \boldsymbol{\gamma_z}'\mathbf{Z_i}$$

Model Assessment

$$T_1(\mathbf{Y_2}) = \sum_{i=1}^{1057} \sum_{j=1}^{2} I(Y_{2ij} = 0)$$

$$T_2(\mathbf{Y_2}) = \sum_{i=1}^{1057} I(Y_{2i1} = Y_{2i2} = 0)$$

$$T_3(\mathbf{Y_2}) = \frac{1}{2114} \sum_{i=1}^{1057} \sum_{j=1}^{2} I\left(Y_{2ij} > \frac{450}{7}\right)$$

$T_4(\mathbf{Y_1}, \mathbf{Y_2}) = \beta_1$ coefficient from regression assessing day effect of for Y_2

Estimating Probability of Compliance on Average

For ℓ from $\ell = 1, 2, ..., L$ do:

1. Sample $\boldsymbol{\theta}^{(\ell)}$ from the posterior distribution $p(\boldsymbol{\theta}|\mathbf{Y}, \mathbf{Z})$
2. Simulate $X_{1i}^{*(\ell)}$ from $p(X_{1i}|\boldsymbol{\theta}^{(\ell)}, Z_i)$ for $i = 1, ..., n$
3. Simulate $X_{2i}^{*(\ell)}$ from $p(X_{2i}|\boldsymbol{\theta}^{(\ell)}, X_{1i}^{*(\ell)}, Z_i)$ for $i = 1, ..., n$
4. Calculate $X_{3i}^{*(\ell)} = 30X_{1i}^{*(\ell)} + X_{2i}^{*(\ell)}$ for $i = 1, ..., n$
5. Calculate $p^{(\ell)} = \frac{1}{n}\sum_{i=1}^{n} I\left(X_{3i}^{*(\ell)} \geq \frac{450}{7}\right)$

Further work

We plan on the third chapter being a problem from the PAMS data or the data we will be collecting

Like the first two chapters, it will likely be centered around Bayesian measurement error modeling