

# Modeling of Energy Balance and Physical Activity Data

Danny Ries

Iowa State University

March 1, 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

# The “Fatal Flaw in Obesity Research”

It has been said the “Fatal Flaw in Obesity Research” is our inability to accurately measure how much someone eats (EI) in free living situations

- Current measures of EI; ie. self report, are clouded with (measurement) error
- Garbage in Garbage out
- Tough to understand dietary trends over the years
- Cannot measure adherence to clinically prescribed interventions

This error in measurement extends to EE and body composition, albeit not nearly as severe

# Modeling Energy Balance

## Energy Balance

The application of the first law of thermodynamics to nutrition/exercise science:

Change in Energy Stores ( $\Delta ES$ ) = Energy Intake (EI) - Energy Expenditure (EE)

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

We are now in a situation where we must measure both EE and  $\Delta ES$  in order to calculate EI

→ But gold standard measures (albeit very costly) and cheap measurements for both exist!

## Research Goal

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

- This will allow future research to calibrate cheaper measurements (when gold standard measures aren't used) and thus eliminate known biases
- Lots of research has been done for calibrating and evaluating measurement error for EI, some for EE, and little for  $\Delta$ ES
- 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 received DLW at end of 12 months, with additional DXA scan
- Demographic variables

Although these data don't have perfect replicates, it provides a baseline to start our modeling of measurement error in Energy Balance



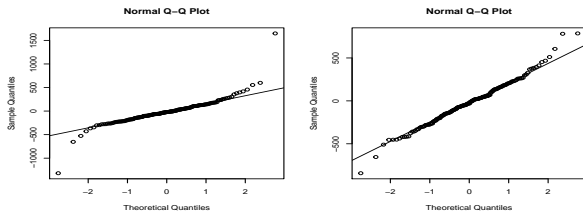


Figure: Differenced DXA  $\Delta$ ES, Females left, Males right

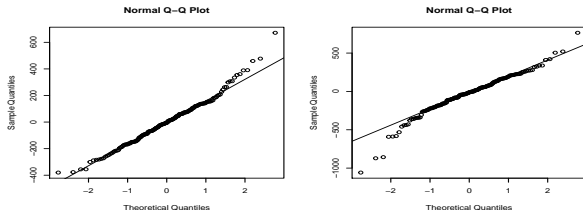


Figure: Differenced SWA EE , Females left, Males right

# Naïve Check for Biases

It seems reasonable that cheap measurement tools could be affected by factors other than the *truth*, ie. demographics

→ Fit the following multiple regression model with the subset data from EBS

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

Table: Regression of Sensewear Armband EE on DLW EE, Age, BMI, Gender using EBS data. Results show systematic biases could exist in cheap EE measurements.

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

Latent:

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

# Independence Assumptions

Given  $X_i^{EE}, Z_i$

- $Y_{ij}^{EE}$  are mutually independent for all  $i, j$
- $W_{ij}^{EE}$  are mutually independent for all  $i, j$
- $Y_{ij}^{EE}$  is independent of  $W_{ij}^{EE}$  for all  $i, j$
- $Y_{ij}^{EE}$  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  $i, j$

Same assumptions hold for reverse case (replace EE with  $\Delta ES$  and  $\Delta ES$  with EE)

# Model for Observed Variables

$$Y_{ij}^{EE} = m_{ee}(X_i^{EE}, Z_i) + \epsilon_{ij}^{EE}$$

$$Y_{ij}^{\Delta ES} = m_{es}(X_i^{\Delta ES}, Z_i) + \epsilon_{ij}^{\Delta ES}$$

$$W_{ij}^{EE} = X_i^{EE} + \nu_{ij}^{EE}$$

$$W_{ij}^{\Delta ES} = X_i^{\Delta ES} + \nu_{ij}^{\Delta ES}$$

$$E(\epsilon_{ij}^{EE}) = E(\epsilon_{ij}^{\Delta ES}) = E(\nu_{ij}^{EE}) = E(\nu_{ij}^{\Delta ES}) = 0$$

# Naïve Model

The Naïve Model assumes no measurement error in gold standard measurements (Note the part for EE is the same as what we used for our exploratory analysis)

$$\begin{aligned}(Y_{ij}^{EE} | W_{ij}^{EE}, Z_i, \theta_{\mathbf{yee}}) &\overset{\text{ind}}{\sim} N(\beta_{0,\text{ee}} + \beta_{1,\text{ee}} W_{ij}^{EE} + \gamma_{\text{ee}} Z_i, \sigma_{\epsilon}^2{}_{EE}) \\(Y_{ij}^{\Delta ES} | W_{ij}^{\Delta ES}, Z_i, \theta_{\mathbf{yes}}) &\overset{\text{ind}}{\sim} N(\beta_{0,\text{es}} + \beta_{1,\text{es}} W_{ij}^{\Delta ES} + \gamma_{\text{es}} Z_i, \sigma_{\epsilon}^2{}_{\Delta ES})\end{aligned}$$

# Linear Measurement Error Model

This is a basic modification to the Naïve model when there is measurement error in a covariate

$$\begin{aligned}
 (Y_{ij}^{EE} | X_i^{EE}, Z_i, \theta_{yee}) &\stackrel{ind}{\sim} N(\beta_{0,ee} + \beta_{1,ee} X_i^{EE} + \gamma_{ee} Z_i, \sigma_{\epsilon_{EE}}^2) \\
 (Y_{ij}^{\Delta ES} | X_i^{\Delta ES}, Z_i, \theta_{yes}) &\stackrel{ind}{\sim} N(\beta_{0,es} + \beta_{1,es} X_i^{\Delta ES} + \gamma_{es} Z_i, \sigma_{\epsilon_{\Delta ES}}^2) \\
 (W_{ij}^{EE} | X_i^{EE}, Z_i, \theta_{wee}) &\stackrel{ind}{\sim} N(X_i^{EE}, \sigma_{\nu_{EE}}^2) \\
 (W_{ij}^{\Delta ES} | X_i^{\Delta ES}, Z_i, \theta_{wes}) &\stackrel{ind}{\sim} N(X_i^{\Delta ES}, \sigma_{\nu_{\Delta ES}}^2) \\
 (X_i^{EE}, X_i^{\Delta ES} | \theta_X) &\stackrel{ind}{\sim} N\left(\begin{bmatrix} \mu_{EE} \\ \mu_{\Delta ES} \end{bmatrix}, \Sigma_X\right)
 \end{aligned}$$



# Extending the Linear Model

We would like to relax the assumption that the relationship between a cheap measurement and *usual* EE and  $\Delta$ ES is linear

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

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

# Free Knot Spline Model

$$f(Y_{ij}^{EE} | X_i^{EE}, Z_i, \theta_{yee}) \stackrel{ind}{\sim} N(s(X_i^{EE}; \beta_{ee}) + \gamma_{ee} Z_i, \sigma_{\epsilon}^2) \\ f(Y_{ij}^{\Delta ES} | X_i^{\Delta ES}, Z_i, \theta_{yes}) \stackrel{ind}{\sim} N(s(X_i^{\Delta ES}; \beta_{\Delta es}) + \gamma_{es} Z_i, \sigma_{\epsilon}^2)$$

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

In order to ensure the functions are monotone, constrain

$$\beta_{1,ee} \leq \beta_{2,ee} \leq \dots \leq \beta_{k_{ee}+3,ee}$$

# MCMC Algorithms

1. The Naïve model and Linear Measurement Error Model were fit using JAGS
2. The Spline Measurement Error Model required Reversible Jump MCMC (RJMCMC) and was written in C++ via Rcpp

# Gibbs Sampler

Priors were chosen to be conjugate to help simplify the sampler

For iteration  $k=1, \dots, K$ , sample from its full conditional:

1.  $\{\Sigma^{(k)}\} \mid \cdot \stackrel{ind}{\sim}$

$$Inv - Wish(d + n, \psi + (\mathbf{X}_i^{(k-1)} - \mu^{(k-1)})'(\mathbf{X}_i^{(k-1)} - \mu^{(k-1)}))$$

2.  $\{(\mu_{EE}^{(k)}, \mu_{\Delta ES}^{(k)})\} \mid \cdot \stackrel{ind}{\sim} N(M'_\mu, C'_\mu)$

$$C'_\mu = (C_\mu^{-1} + n\Sigma^{-1(k)})^{-1}$$

$$M'_\mu = C'_\mu (C_\mu^{-1}M + n\Sigma^{-1(k)}\bar{\mathbf{X}}^{(k-1)})$$

$$\bar{\mathbf{X}} = \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i$$

3. Update  $\mathbf{X}_i$  with a random walk

$$\{\mathbf{X}_i^{(k)} : i = 1, \dots, n\} \mid \cdot \text{ for } i = 1, \dots, n \text{ sample } \mathbf{X}_i^* \text{ from } N(\mathbf{X}_i^{(k)}, C_{X_i})$$

$$\text{Set } \mathbf{X}_i^{(k)} = \mathbf{X}_i^* \text{ with probability } \alpha_{X_i}, \text{ otherwise set } \mathbf{X}_i^{(k)} = \mathbf{X}_i^{(k-1)}$$

$$\alpha_{X_i} = \min \left( 1, \frac{f(\mathbf{X}_i^* \mid \cdot)}{f(\mathbf{X}_i^{(k-1)} \mid \cdot)} \right)$$

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

# RJMCMC

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

We extend method of Denison (1998) 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 approximation of BIC

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

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

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  $r^{(k-1)}$  to create the proposed knot locations  $r^*$

If death Step:

Sample one knot location from  $r^{(k-1)}$  at random and remove it.

If move step:

Sample one knot location from  $r^{(k-1)}$  at random, and change it to a new knot location at random from the set  $\mathcal{A}$

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

$$\alpha_{birth} = \min \left( 1, \text{Likelihood ratio} \times \frac{n - Z(k)}{n} \right)$$

$$\alpha_{death} = \min \left( 1, \text{Likelihood ratio} \times \frac{n}{n - Z(k)} \right)$$

$$\alpha_{move} = \min(1, \text{Likelihood ratio})$$

$$Z(k) = 2(\ell + 1) + k.(2\ell + 1)$$

$$k. = \text{length}(r^{(k-1)})$$

7. calculate mean function  $m(X^{(k)}, \beta^{(k)})$  using spline basis matrix



# Simulation Study

We performed a simulation study to assess the performance of the models we discussed in terms of parameter estimation and prediction (for calibration)

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:

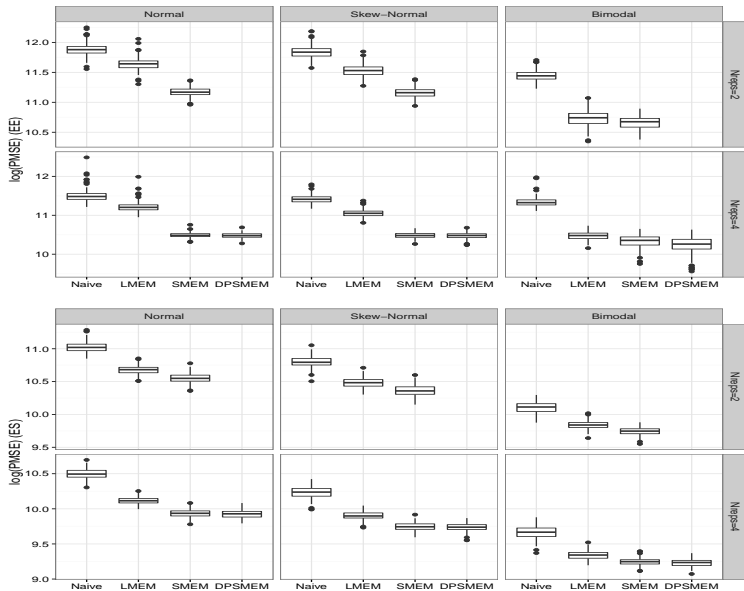
- Normal
- Skew-Normal
- Bimodal (50-50 mixture of two Normals)

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

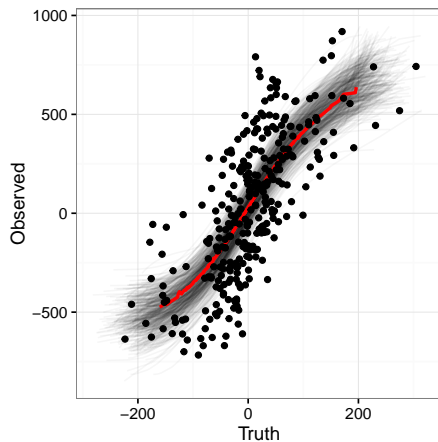
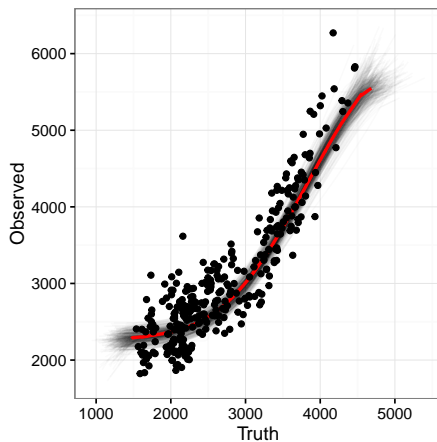
	$\sigma_{yee}$		$\sigma_{yes}$		$\gamma_{1,ee}$		$\gamma_{2,ee}$		$\gamma_{3,ee}$		$\gamma_{1,es}$		$\gamma_{2,es}$		$\gamma_{3,es}$
Replicates	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2
Mean Est	477.65	473.42	347.49	354.85	254.67	248.66	14.88	14.03	-4.14	-5.29	-200.53	-199.39	7.91	8.31	-4.9
Std Err	17.82	19.24	9.43	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

	$\sigma_{yee}$		$\sigma_{yes}$		$\sigma_{wee}$		$\sigma_{wes}$		$\gamma_{1,ee}$		$\gamma_{2,ee}$		$\gamma_{3,ee}$		$\gamma_{1,es}$
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	-199.5
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	-200.0

	$\sigma_{yee}$		$\sigma_{yes}$		$\sigma_{wee}$		$\sigma_{wes}$		$\gamma_{1,ee}$		$\gamma_{2,ee}$		$\gamma_{3,ee}$		$\gamma_{1,es}$
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	-200.0
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	-200.0



# Example of one data set



# Calibration

The major goal of this model is to be able to give calibration equations for cheap measurements

⇒ in the future if no gold standard measurements are taken, can correct cheap measurements using these calibration equations

Our goal is to get an estimate of the truth using cheap measurement of the truth  $y$  and covariates  $Z$ :

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

For  $r = 1, \dots, R$

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  $|m_i(x) - y_i^*|$  to choose the value of  $x$  that will minimize the aforementioned criteria, call this  $x_{i,calibrated}^{(r)}$ .  
 $m_i(x)$  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

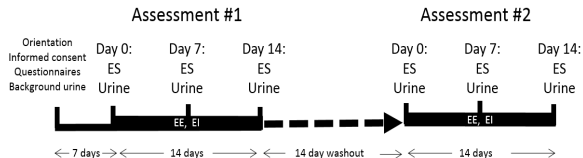
### 3 example calibrations

	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
$\Delta$ ES	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 of spline model
- 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 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
- 75 Minutes of Vigorous Physical Activity
- 75-150 Minutes during some combination of Moderate and Vigorous Physical Activity (MVPA)
- Full body resistance training two days per week
- Most benefits occur when activity is performed in at least 10 minute *bouts*

What proportion of adults adhere to these guidelines? 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

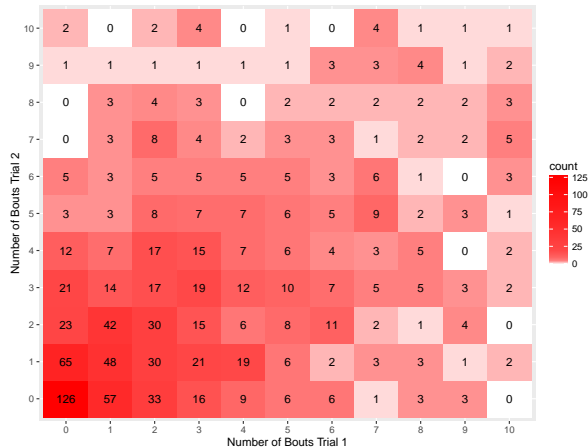
- Measurements of physical activity are over 24 hours
- Replicate measurements taken 2-3 weeks apart
- Minute by minute data from armband (in MET-minutes)
- Survery weights needed
- Measurement error involved

## Definition of a bout

- $Y_{1ij}$  : 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

$$Y_{2i1} - Y_{2i2} = \beta_0 + \beta_1(Y_{1i1} - Y_{1i2}) + \beta_2(Weekend_{i1} - Weekend_{i2}) + \epsilon_i$$

$$\epsilon_i \stackrel{iid}{\sim} N(0, \sigma^2)$$

Coefficient	Estimate	Std Error	P-value
$\beta_0$	23.305	11.10	0.036
$\beta_1$	86.827	3.489	<0.0001
$\beta_2$	22.145	18.425	0.230

Excluding individuals with  $Y_2 > 2300$  (5 out of 1085 individuals), the estimate for trial effect reduces to 13.62 with a p-value of 0.124, while other results remain similar

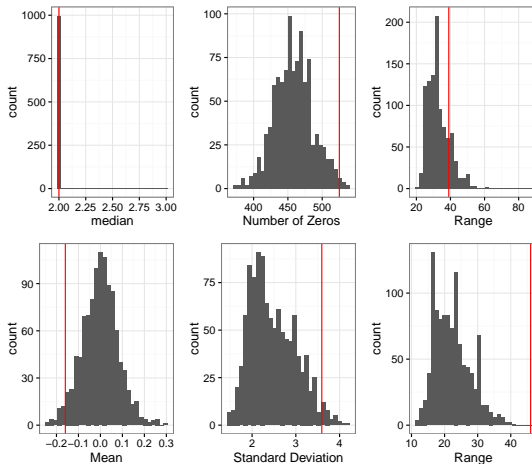
# Modeling Number of Bouts

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

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



# Assessing Overdispersion



## Two-Part Model for $Y_2$

After an appropriate adjustment for day effect, we model  $Y_2$  as

$$Y_{2ij}|X_{2i}, X_{1i} \stackrel{ind}{\sim} (1 - \pi_{ij}(\alpha))\delta_0(Y_{2ij}) + \pi_{ij}(\alpha)LN(\mu_{y,i}(\beta_y), \sigma_y^2)$$

$$X_{2i}|X_{1i}, Z_i \stackrel{ind}{\sim} LN(\mu_{x,i}(\beta_x), \sigma_x^2)$$

$$\pi_{ij}(\alpha) = \Phi(\mathbf{X}_i'\alpha)$$

$$\mu_{y,i}(\beta_y) = E(\log Y_{2ij} | Y_{2ij} > 0) = \mathbf{X}_i'\beta_y$$

$$\mu_{x,i}(\beta_x) = E(\log X_{2ij}) = \mathbf{Z}_i'\beta_x + X_{1i}\gamma$$

# Model Assessment

$$T_1(\mathbf{Y}_1) = \sum_{i=1}^n I(Y_{1i1} = Y_{1i2})$$

$$T_2(\mathbf{Y}_2) = \sum_{i=1}^n \sum_{j=1}^{j(i)} I(Y_{2ij} = 0)$$

$$T_3(\mathbf{Y}_2) = \sum_{i=1}^n I(Y_{2i1} = Y_{2i2} = 0)$$

$$T_4(\mathbf{Y}_2) = \frac{1}{N} \sum_{i=1}^n \sum_{j=1}^{j(i)} I\left(Y_{2ij} > \frac{450}{7}\right)$$

$$T_5(\mathbf{Y}_1, \mathbf{Y}_2) = \beta_1 \text{ coefficient from regression assessing exchangeability}$$

# Estimating Probability of Compliance on Average

For  $\ell$  from  $\ell = 1, 2, \dots, L$  do:

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