



## **Intake\_epis\_food(): An R Function for Fitting a Bivariate Nonlinear Measurement Error Model to Estimate Usual and Energy Intake for Episodically Consumed Foods**

**Adriana Pérez**  
The University of Texas

**Saijuan Zhang**  
Texas A&M University

**Victor Kipnis**  
National Cancer Institute

**Laurence S. Freedman**  
Gertner Institute for Epidemiology  
and Health Policy Research

**Raymond J. Carroll**  
Texas A&M University

---

### **Abstract**

We consider a Bayesian analysis using **WinBUGS** to estimate the distribution of usual intake for episodically consumed foods and energy (calories). The model uses measures of nutrition and energy intakes via a food frequency questionnaire along with repeated 24 hour recalls and adjusting covariates. In order to estimate the usual intake of the food, we phrase usual intake in terms of person-specific random effects, along with day-to-day variability in food and energy consumption. Three levels are incorporated in the model. The first level incorporates information about whether an individual reported consumption of a particular food item. The second level incorporates the amount of food consumption equalling to zero if not consumed, and the third level incorporates the amount of energy intake. Estimates of posterior means of parameters and distributions of usual intakes are obtained by using Markov chain Monte Carlo calculations which can be thought as mean estimates for frequentists. This R function reports to users point estimates and credible intervals for parameters in the model, samples from their posterior distribution, samples from the distribution of usual intake and usual energy intake, trace plots of parameters and summary statistics of usual intake, usual energy intake and energy adjusted usual intake.

*Keywords:* excess zero models, MCMC, nonlinear mixed models, R, **R2WinBUGS**, zero-inflation.

---

## 1. Introduction

There are many statistical challenges when modeling food intakes reported on two or more 24 hours (24h) recalls. Some of the challenges involve the presence of measurement error because estimating the distribution of usual intake of nutrients and foods in the population involves monitoring and measuring such intakes over time and their associated recall biases. In addition, many consumers report episodically consumed foods, foods that are not typically consumed every day. For example, fish may be reported in a particular day with a positive intake, but on a different day fish intake may equal zero. Consequently, it is difficult to estimate nutrition intake with recall surveys when one of these recalls incorporate excess zero measurements. Data of this nature is often modeled with measurement error models with zero inflated data.

Recently, in nutritional surveillance (Tooze *et al.* 2006) and nutritional epidemiology (Kipnis *et al.* 2009) two-part methods had been developed for analyzing episodically consumed foods. In the first part of the model, the probability of an episodically consumed food is estimated using logistic regression with a person-specific random effect. Then, in the second part of the model the amount of that episodically consumed food per day is modeled using linear regression on a transformed scale with also a person-specific effect. These two parts are linked allowing that the two person-specific effects are correlated as well as by allowing common covariates in both parts of the model. This method is known as the “NCI method” (available at <http://riskfactor.cancer.gov/diet/usualintakes/>) and described by Tooze *et al.* (2006) and Kipnis *et al.* (2009).

An extension into a three-part method that also incorporates the estimation of the amount of energy intake consumed per day is described in detail by Kipnis *et al.* (2011). These authors estimated this three-part method using nonlinear mixed effects models with likelihoods computed by adaptive Gaussian quadrature in SAS software. However, computationally it was found to have serious convergence issues within the context of nutritional epidemiology and nutritional surveillance as indicated by Kipnis *et al.* (2011) and Zhang *et al.* (2011).

The goal of this article is to present a function for the implementation of this nonlinear mixed effects model. The function `Intake_epis_food()` allows readers to input their data in R (R Development Core Team 2010) to generate and run the script of the three part model in **WinBUGS** (Spiegelhalter *et al.* 1999) and to run and obtain output simulations to R using the package **R2WinBUGS** (Sturtz *et al.* 2005). While the most important functions of the package **R2WinBUGS** are illustrated, we do not provide comprehensive documentation here; instead the reader is referred to the manual and online documentation with that package available from the Comprehensive R Archive Network at <http://CRAN.R-project.org/package=R2WinBUGS>. In the next sections, the function and corresponding algorithms are explained and an example is provided.

## 2. Method and parameter inference

The computational details of the Bayesian approach to fit the three-part model for an episodically consumed food and energy model (Kipnis *et al.* 2011) through Markov chain Monte Carlo (MCMC) techniques are provided by Zhang *et al.* (2011) with a brief summary given here. Our model takes into consideration measures of nutrition and energy intakes via a food frequency questionnaire (FFQ) along with two repeated 24h recalls and adjustment covari-

ates. These two repeated 24h recalls provide information on the amounts of food and energy consumed by each individual. Consequently, an indicator variable of whether the food was consumed can be generated from the reported amount of food consumed. In addition, because two or more 24h recalls require distinguishing between and within person random error (Eckert *et al.* 1997), a type of classical measurement error is needed. Generally nutritional data are skewed, which may require transformations to reach normality and standardization. For person  $i = 1, \dots, n$ , and for the  $k = 1, 2$  repeats of the 24h recall, the data are  $\tilde{Y}_{ik} = (Y_{i1k}, \dots, Y_{i3k})^\top$ , where

- $Y_{i1k}$  = Indicator of whether the food is consumed.
- $Y_{i2k}$  = Amount of the food consumed as reported by the 24h recalls, which equals zero if the food is not consumed.
- $Y_{i3k}$  = Amount of energy consumed as reported by the 24h recalls.

We will use a Box-Cox transformation to account for the skewness in the data. The Box-Cox transformation with transformation parameter  $\lambda$ , is  $h(y, \lambda) = (y^\lambda - 1)/\lambda$  if  $\lambda \neq 0$ , and  $h(y, \lambda) = \log(y)$  if  $\lambda = 0$ . We will allow for the user to specify different transformation parameters  $\lambda_F$  and  $\lambda_E$  for food and energy intake, respectively.

After Box-Cox transformation, we further standardize and center these transformed variables to have a mean of zero and variance of one. This is useful for making the prior distribution specifications given below to be sensible and allows rapid convergence of the posterior samples. Specifically, let  $\mu_{\lambda_F}$  and  $\sigma_{\lambda_F}$  be the mean and standard deviation of the transformed non-zero food data  $h(Y_{i2k}, \lambda_F)$ , and let  $\mu_{\lambda_E}$  and  $\sigma_{\lambda_E}$  be the mean and standard deviation of the transformed energy data  $h(Y_{i3k}, \lambda_E)$ . Then, our analysis is performed using the following transformations:

$$Q_{i1k} = Y_{i1k}; \tag{1}$$

$$Q_{i2k} = \sqrt{2} Y_{i1k} \{h(Y_{i2k}, \lambda_F) - \mu_{\lambda_F}\} / \sigma_{\lambda_F}; \tag{2}$$

$$Q_{i3k} = \sqrt{2} \{h(Y_{i3k}, \lambda_E) - \mu_{\lambda_E}\} / \sigma_{\lambda_E} = h_{\text{tr}}(Y_{i3k}, \lambda_E, \mu_{\lambda_E}, \sigma_{\lambda_E}); \tag{3}$$

There are also covariates such as age category, ethnic status and, in many cases the results of reported intakes from a food frequency questionnaire. In principle, the covariates can differ based on the three types of data, so we denote them as  $V_{i1}, V_{i2}, V_{i3}$ , which are vector-valued. To improve linearity and homoscedasticity of the model, we follow the recommendations of (i) implementing Box-Cox transformations on all or some covariates like intakes from a food frequency questionnaire (Kipnis *et al.* 2009), (ii) centering and (iii) scaling on all covariates. Without loss of generality, let  $\lambda_c$  represent a vector containing the Box-Cox transformation parameter used for covariates. Let  $\mu_{\lambda_c}$  and  $\sigma_{\lambda_c}$  be the vector means and standard deviations of the transformed covariates. Then, these transformed, center and scaled covariates are denoted by:

$$X_{i1} = \{h(V_{i1}, \lambda_c) - \mu_{\lambda_c}\} / \sigma_{\lambda_c} \tag{4}$$

$$X_{i2} = \{h(V_{i2}, \lambda_c) - \mu_{\lambda_c}\} / \sigma_{\lambda_c} \tag{5}$$

$$X_{i3} = \{h(V_{i3}, \lambda_c) - \mu_{\lambda_c}\} / \sigma_{\lambda_c} \tag{6}$$

Let  $(U_{i1}, U_{i2}, U_{i3}) = \text{Normal}(0, \Sigma_u)$  be random effects associated with consumption, amount (if the food is consumed), and energy. Similarly, for  $k = 1, 2$ ,  $(\epsilon_{i1k}, \epsilon_{i2k}, \epsilon_{i3k}) = \text{Normal}(0, \Sigma_\epsilon)$

accounts for day-to-day variation. The model for whether there is consumption can be stated as:

$$P(Q_{i1k} = 1 | X_{i1}, U_{i1}, U_{i2}, U_{i3}) = \Phi(\alpha_1 + X_{i1}^\top \beta_1 + U_{i1}), \quad (7)$$

where  $\Phi(\cdot)$  is the standard normal distribution function and  $(W_{i1k}, W_{i2k}, W_{i3k})$  represent their corresponding latent variables as follows. A food being consumed at visit  $k$  is equivalent to

$$Q_{i1k} = 1 \iff W_{i1k} = \alpha_1 + X_{i1}^\top \beta_1 + U_{i1} + \epsilon_{i1k} > 0. \quad (8)$$

The food when consumed and energy intake are modeled as

$$[Q_{i2k} | Q_{i1k} > 0] = W_{i2k} = \alpha_2 + X_{i2}^\top \beta_2 + U_{i2} + \epsilon_{i2k} \quad (9)$$

$$[Q_{i3k}] = W_{i3k} = \alpha_3 + X_{i3}^\top \beta_3 + U_{i3} + \epsilon_{i3k}. \quad (10)$$

The prior distribution of parameters  $\beta_1, \beta_2, \beta_3$  is assumed to be multivariate Normal with vector mean 0 and variance covariance  $\Sigma_u$ . An inverse Wishart denoted as  $IW(\Omega_u, m_u)$  prior was specified for  $\Sigma_u$ . We set  $\Omega_u$  to have an exchangeable correlation structure with starting values with diagonal entries all equal to 1 and correlations 0.5. We have two necessary restrictions on  $\Sigma_\epsilon$ : (i)  $\epsilon_{i1k}$  and  $\epsilon_{i2k}$  are independent; and (ii)  $\text{var}(\epsilon_{i1k}) = 1$ , so that  $\beta_1$  is identifiable, along with the distribution of  $U_{i1}$ . Furthermore, we constrained this covariance matrix using a *polar coordinate* representation with  $\gamma \in (-1, 1)$  and  $\theta \in (-\pi, \pi)$ . With these considerations  $\Sigma_\epsilon$  can be written as:

$$\Sigma_\epsilon = \begin{bmatrix} 1 & 0 & p_1 s_{33}^{1/2} \\ 0 & s_{22} & p_2 (s_{22} s_{33})^{1/2} \\ p_1 s_{33}^{1/2} & p_2 (s_{22} s_{33})^{1/2} & s_{33} \end{bmatrix}, \quad (11)$$

where  $p_1 = \gamma \cos(\theta)$  and  $p_2 = \gamma \sin(\theta)$ . The recommended prior distributions for  $s_{22}$  and  $s_{33}$  are  $\text{Uniform}(0, 3)$ , for  $\gamma$  is  $\text{Uniform}(-1, 1)$  and for  $\theta$  is  $\text{Uniform}(-\pi, \pi)$  (Zhang *et al.* 2011). The corresponding inverse of  $\Sigma_u$  is a Wishart distribution denoted as  $\Sigma_u^{-1} = W(\Omega_u^{-1}, m_u)$ , and the inverse of  $\Sigma_\epsilon$  can be written as:

$$\Sigma_\epsilon^{-1} = \begin{bmatrix} 1 + wp_1^2 s_{33} & wp_1 p_2 s_{33} s_{22}^{-1/2} & -wp_1 s_{33}^{1/2} \\ wp_1 p_2 s_{33} s_{22}^{-1/2} & s_{22}^{-1} + wp_2^2 s_{33} s_{22}^{-1} & -wp_2 s_{33}^{1/2} s_{22}^{-1/2} \\ -wp_1 s_{33}^{1/2} & -wp_2 s_{33}^{1/2} s_{22}^{-1/2} & w \end{bmatrix}, \quad (12)$$

where  $w = \{s_{33}(1 - p_1^2 - p_2^2)\}^{-1}$ .

### Usual intake analysis

Besides the parameters and the random effects  $(U_{i1}, U_{i2}, U_{i3})$  in this context it is vital to also have the posterior samples of usual intake  $T_{Fi}$ , usual energy intake  $T_{Ei}$ , and energy adjusted usual intake  $1000T_{Fi}/T_{Ei}$ , all on the original scale. We used the best-power method (Dodd *et al.* 2006) for estimating both usual energy intake and usual intake for person  $i$ . We first consider energy:

$$T_{Ei} = E \left\{ h_{\text{tr}}^{-1}(\alpha_3 + X_{i3}^\top \beta_3 + U_{i3} + \epsilon_{i3}, \lambda_E, \mu_{\lambda_E}, \sigma_{\lambda_E}) | X_{i3}, U_{i3} \right\} \quad (13)$$

$$\approx h_{\text{tr}}^* \left\{ \alpha_3 + X_{i3}^\top \beta_3 + U_{i3}, \lambda_E, \mu_{\lambda_E}, \sigma_{\lambda_E}, \Sigma_\epsilon(3, 3) \right\}. \quad (14)$$

where

$$h_{\text{tr}}^*\{v, \lambda, \mu, \sigma, \Sigma\} = h_{\text{tr}}^{-1}(v, \lambda, \mu, \sigma) + \frac{\Sigma \partial^2 h_{\text{tr}}^{-1}(v, \lambda, \mu, \sigma)}{2 \partial v^2}.$$

Similarly, a person's usual intake of the dietary component on the original scale is defined as

$$T_{Fi} = \Phi(\alpha_1 + X_{i1}^\top \beta_1 + U_{i1}) h_{\text{tr}}^* \left\{ \alpha_2 + X_{i2}^\top \beta_2 + U_{i2}, \lambda_F, \mu_{\lambda_F}, \sigma_{\lambda_F}, \Sigma_\epsilon(2, 2) \right\}. \quad (15)$$

When  $\lambda = 0$ , the back-transformation and second derivative are:

$$\begin{aligned} h_{\text{tr}}^{-1}(v, 0, \mu, \sigma) &= \exp(\mu + \sigma v / \sqrt{2}); \\ \partial^2 h_{\text{tr}}^{-1}(v, 0, \mu, \sigma) / \partial v^2 &= \sigma^2 h_{\text{tr}}^{-1}(v, 0, \mu, \sigma) / 2. \end{aligned}$$

Similarly, when  $\lambda \neq 0$ , the back-transformation and second derivative are:

$$\begin{aligned} h_{\text{tr}}^{-1}(v, \lambda, \mu, \sigma) &= (1 + \lambda \mu + \lambda \sigma v / \sqrt{2})^{1/\lambda}; \\ \partial^2 h_{\text{tr}}^{-1}(v, \lambda, \mu, \sigma) / \partial v^2 &= \frac{\sigma^2}{2} (1 - \lambda) (1 + \lambda \mu + \lambda \sigma v / \sqrt{2})^{-2+1/\lambda}. \end{aligned}$$

### 3. Using the `Intake_epis_food()` function in R

Users must install the **R2WinBUGS** (Sturtz *et al.* 2005) package in R and **WinBUGS** (Spiegelhalter *et al.* 1999) before running `Intake_epis_food()` function. This function assumes that there are no missing observations. `Intake_epis_food()` loads the packages **R2WinBUGS**, **stats**, and **timeSeries** (Würtz and Chalabi 2011) automatically.

The `Intake_epis_food()` function incorporates this three-part model, with truncated normal random variables for generating the latent  $W_{ik}$ , and Metropolis-Hastings computations for  $\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3$ , the elements of  $\Sigma_u$ , and  $\Sigma_\epsilon$ . This function generates and runs the script of this model in **WinBUGS** and returns a MCMC computation. We emphasize that the MCMC computation can either be thought of as a strictly Bayesian computation with ordinary Bayesian inference, or as a means of developing *frequentist* estimators of the crucial parameters. This MCMC computation produces estimates which are known to be asymptotically equivalent to maximum likelihood estimators, which are difficult to obtain (Zhang *et al.* 2011). In Bayesian terms we used the posterior means of the model. Users are not only interested in estimates of usual energy intake (Equation 14) and usual intake for person  $i$  (Equation 15) but also in the energy-adjusted usual intake  $1000T_{Fi}/T_{Ei}$ .

The `Intake_epis_food()` function produces seven output files. The first output file contains parameter estimates (posterior means) of:  $\alpha_1$  and  $\beta_1$  from Equation 8;  $\alpha_2$  and  $\beta_2$  from Equation 9; and  $\alpha_3$  and  $\beta_3$  from Equation 10. Also, the first output file contains parameter estimates for  $\Sigma_\mu$ ,  $\Sigma_\mu^{-1}$ ,  $\Sigma_\epsilon$ , and  $\Sigma_\epsilon^{-1}$ . The second output file contains summary statistics of Bayesian estimates of Equation 14, Equation 15, and the energy-adjusted usual intake  $1000T_{Fi}/T_{Ei}$ . The third output file contains the iterations from the MCMC computations. The fourth output file contains summary statistics of Bayesian estimates for  $U_{i1}, U_{i2}, U_{i3}$ . The fifth output file contains trace plots of the intercept and coefficients in Equation 8, Equation 9, and Equation 10. The sixth output file contains trace plots of the variance-covariances  $\Sigma_\mu$ ,  $\Sigma_\mu^{-1}$ ,  $\Sigma_\epsilon$  and  $\Sigma_\epsilon^{-1}$ . The seventh output file contains density plots of Equation 15 and the energy adjusted usual intake  $1000T_{Fi}/T_{Ei}$ .

The defaults for the `Intake_epis_food()` function in R are:

```
Intake_epis_food(data.file.name = "data.file.name",
  numcov = 4, df = 5, lambdaf = 0.32, lambdae = 0.23, lvar1 = 999,
  lvar2 = 999, lvar3 = 0.33, lvar4 = 0, n.iter = 11000, n.chains = 1,
  n.burnin = 1000, n.thin = 10, bugs.seed = 123456,
  working.directory = "working.directory", file.estimated = "estimates.csv",
  file.istats = "estimates_intake.csv", file.iterations = "iterations.csv",
  file.uis = "uis.csv", file.tracep = "trace_plots.pdf",
  file.tpvarcov = "trace_plot_var_cov_matrices.pdf",
  file.densityp = "density_plots.pdf",
  bugs.directory = "c:/Program Files/WinBUGS14/")
```

with the following arguments:

**Data.file.name:** Name of the file containing the data (default: "data.file.name") in the following order:

- The first variable identifies individuals, e.g., an ID number.
- The next two variables are the 24h recalls variables of food intake, in order.
- The next two variables are the 24h recalls variables of energy intake, in order.
- The next variables are the covariates, and cannot have a column of ones. We have set the number of covariates to four, given that each covariate may have a different Box-Cox transformation parameter.

**numcov:** Number of covariates in user's dataset (default: 4).

**df:** Number of the degrees of freedom of  $m_u$  (default: 5).

**lambdaf:** Value of  $\lambda_F$  to be used for transformation of food intake (default: 1 indicating no transformation done).

**lambdae:** Value of  $\lambda_E$  to be used for transformation of energy intake (default: 1 indicating no transformation done).

**lvar1:** Value of  $\lambda_c$  to be used for Box-Cox transformation of first covariable (default: 999 indicating no transformation done).

**lvar2:** Value of  $\lambda_c$  to be used for Box-Cox transformation of second covariable (default: 999 indicating no transformation done).

**lvar3:** Value of  $\lambda_c$  to be used for Box-Cox transformation of third covariable (default: 999 indicating no transformation done).

**lvar4:** Value of  $\lambda_c$  to be used for Box-Cox transformation of fourth covariable (default: 999 indicating no transformation done).

**n.iter:** Number of iterations of the MCMC. Please include the number of burning iterations also (default: 11000).

**n.chains:** Number of chains (default: 1).

**n.burnin:** Number of burn-in iterations of the MCMC (default: 1000).

- n.thin:** Thinning rate of the MCMC (default: 10).
- bugs.seed:** Random seed to be used for the MCMC (default: 123456).
- working.directory:** Drive and folder location for output files from **WinBUGS** (default: "C:").
- file.estimate:** Name of the file to save posterior summary statistics including mean, standard deviation as well as percentiles: 2.5th, 25th, 50th, 75th and 97.5th (default: "estimates.csv"), of  $\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3$  from Equation 8, Equation 9, and Equation 10 as well as  $\Sigma_\mu, \Sigma_\mu^{-1}, \Sigma_\epsilon,$  and  $\Sigma_\epsilon^{-1}$ .
- file.istats:** Name of the file to save summary statistics of usual food intake, Equation 15, usual energy intake, Equation 14, and usual food intake per 1000 calories ( $1000T_{Fi}/T_{Ei}$ ) (default: "estimates\_intake.csv").
- file.iterations:** Name of the file to save iterations (default: "iterations.csv"). The number of iterations saved corresponds to  $(n.iter - n.burnin)/n.thin$ .
- file.uis:** Name of the file to save summary statistics of Bayesian estimates (default: "uis.csv") for  $U_{i1}, U_{i2}, U_{i3}$  including mean, standard deviation as well as percentiles: 2.5th, 25th, 50th, 75th and 97.5th for person  $i = 1, \dots, n$ . The first variable given in the dataset that identifies individuals is included as the first variable in this file for linking purposes to users.
- file.tracep:** Name of the file to save trace plots of parameters (default: "trace\_plots.pdf").
- file.tpvarcov:** Name of the file to save trace plots of variance-covariances  $\Sigma_\mu, \Sigma_\mu^{-1}, \Sigma_\epsilon$  and  $\Sigma_\epsilon^{-1}$  (default: "trace\_plot\_var\_cov\_matrices.pdf").
- file.densityp:** Name of the file to save density plots of usual food intake and usual food intake per 1000 calories. Each one is generated individually and an overlaid plot is generated as well (default: "density\_plots.pdf").
- bugs.directory:** Drive and folder location of **WinBUGS** (default: "c:/Program Files/WinBUGS14/").

Once the data is input, `Intake_epis_food()` invokes **WinBUGS** from R. **WinBUGS** requires a file describing the model in Equation 8, Equation 9, Equation 10 in Section 2. This model is included as file `intake_program.txt` and is loaded in R using its `bugs()` argument. Although the default priors of the constrained variance-covariance matrix of random errors are geared to the pre-standardization of the data as described in Section 2, the user may change those prior entries within the file `intake_program.txt`, possibly changing any code described in Table 1.

Also, users may change any of the default values of `Intake_epis_food()` function when the function is called to run by replacing those arguments in the last sentence of the file accompanied by this paper. The MCMC posterior means of  $\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3$ , the elements of  $\Sigma_u, \Sigma_\epsilon, \Sigma_u^{-1}$  and  $\Sigma_\epsilon^{-1}$  are displayed on the screen of R by `Intake_epis_food()` and saved automatically in numeral matrices on file names previously listed. See Section 4 for an example.



Matrix notation	Default distribution and range	Code in intake_program.txt
$s_{22}$	Uniform(0, 3)	tau.e1 ~ dunif(0, 3)
$s_{33}$	Uniform(0, 3)	tau.e2 ~ dunif(0, 3)
$\gamma$	Uniform (-1, 1)	a.e1 ~ dunif(-0.99, 0.99)
$\theta$	Uniform (-1, 1)	a.e2 ~ dunif(-3.11, 3.11)

Table 1: Constrained variance-covariance matrix of random errors.

#### 4. Application of the Intake\_epis\_food() function

We simulated data using parameter values identified from the calibration sub-study of the National Institutes of Health (NIH) and the American Association of Retired Persons (AARP) Diet and Health Study (Schatzkin *et al.* 2001). The NIH-AARP study is a cohort composed of people who resided in one of six states: California, Florida, Pennsylvania, New Jersey, North Carolina, and Louisiana or in two metropolitan areas: Atlanta, Georgia and Detroit, Michigan. From 1995 through 1996, 3.5 million questionnaires were mailed to members of the AARP, aged 50–71 years. The questionnaire included a dietary section as well as some lifestyle questions. Over 500,000 people returned the questionnaire after three mailing waves. This, then, is the largest study of diet and health ever conducted in the USA. In 1996–1997, these participants received a risk-factor questionnaire which asked additional questions about lifestyle and behavior, and in 2004–2006 these participants received another follow-up questionnaire (available at <http://dietandhealth.cancer.gov/resource/>).

Participants for the calibration study were randomly selected from the 46,970 subjects who had responded to the first wave as of January 1996. Two thousand participants was the targeted sample size to attempt recruitment calls for a baseline and 24h follow up. The sample available to us included 920 men. Because the data are not publicly available, we simulated 920 food intakes that are similar in distribution to the men in the NIH-AARP calibration study. We used as  $X_{ik}^\top$  four covariates: Age, body mass index (BMI), consumption of servings of whole grains from the FFQ, and energy intake from the FFQ. Therefore,  $\Sigma_u^{-1}$  requires to estimate  $3 \times 3 = 9$  parameters, after subtracting four covariates estimated, then the degrees of freedom of the inverse Wishart were setup as  $m_u = 5$ . The latter two covariates were transformed by the cube root and the logarithm, respectively.

We present results from Intake\_epis\_food() for this simulated data example with a sample size of 920 men, modeling whole grains consumption. We used a burn-in of 1,000 steps followed by 10,000 MCMC iterations. These values were the same arbitrary values used empirically in our MATLAB code (Zhang *et al.* 2011). Our Intake\_epis\_food() function took approximately 6 hours and 16 minutes (Pentium computer D CPU 3.5GHz and 1.99GB of RAM). We used only every 10th value of the chain. The first output file contains the average over 10,000 MCMC of posterior mean, posterior standard deviation, posteriors: 2.5th percentile, 25th percentile, 50th percentile, 75th percentile and 97.5th percentile. The intercepts of the model are  $\alpha_1, \alpha_2, \alpha_3$  corresponding to each level of the three-part model. Also, estimates of regression parameters for each covariate appear for each level of the model. This is the notation used for the output:

**alpha\_1:** Represents the intercept in Equation 8.

**alpha\_2:** Represents the intercept in Equation 9.



`alpha_3`: Represents the intercept in Equation 10.

`beta_1_j`: Represents the estimated coefficient of the  $j$ th covariate in Equation 8,  
 $j = 1, \dots, \text{numcov}$ .

`beta_2_j`: Represents the estimated coefficient of the  $j$ th covariate in Equation 9,  
 $j = 1, \dots, \text{numcov}$ .

`beta_3_j`: Represents the estimated coefficient of the  $j$ th covariate in Equation 10,  
 $j = 1, \dots, \text{numcov}$ .

`tau_j_p`: Represents the  $(j, p)$  entry estimate of  $\Sigma_u^{-1}$ ,  $j = p = 1, 2, 3$ .

`sigma2_j_p`: Represents the  $(j, p)$  entry estimate of  $\Sigma_u$ ,  $j = p = 1, 2, 3$ .

`a.e1`: Represents the estimate of  $\gamma$ .

`a.e2`: Represents the estimate of  $\theta$ .

`sigmaemat_j_p`: Represents the  $(j, p)$  entry estimate of  $\Sigma_\epsilon$ ,  $j = p = 1, 2, 3$ , with  $\Sigma_{\epsilon 11} = 1$  and  
 $\Sigma_{\epsilon 12} = \Sigma_{\epsilon 21} = 0$ .

`tauemat_j_p`: Represents the the  $(j, p)$  entry estimate of  $\Sigma_\epsilon^{-1}$ ,  $j = p = 1, 2, 3$ .

We used  $\lambda_F = 0.32$  as the Box-Cox transformation parameter of food intake  $\lambda_E = 0.23$  as the Box-Cox transformation of energy intake. We setup our working directory in R with the command `setwd("C:")` and using `Intake_epis_food()` function:

```
R> Intake_epis_food(data.file.name = "Sim_AARP_wg_Men.csv",
+   numcov = 4, df = 5, lambdaf = 0.32, lambdae = 0.23, lvar1 = 999,
+   lvar2 = 999, lvar3 = 0.33, lvar4 = 0, n.iter = 11000, n.chains = 1,
+   n.burnin = 1000, n.thin = 10, bugs.seed = 123456,
+   working.directory = "C:/", file.estimate = "estimates.csv",
+   file.istats = "estimates_intake.csv",
+   file.iterations = "iterations.csv", file.uis = "uis.csv",
+   file.tracep = "trace_plots.pdf",
+   file.tpvarcov = "trace_plot_var_cov_matrices.pdf",
+   file.densityp = "density_plots.pdf",
+   bugs.directory = "c:/Program Files/WinBUGS14/")
```

After calculations, estimates of  $\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3, \Sigma_\mu, \Sigma_\mu^{-1}, \Sigma_\epsilon$ , and  $\Sigma_\epsilon^{-1}$ , are shown as well as summary statistics using a Bayesian approach for three variables are provided immediately in the console screen (a) food usual intake, (b) usual food intake per 1000 calories, and (c) energy usual intake. The following shows the results for whole grains.

```
Inference for Bugs model at ' Intake_program.txt ', fit using WinBUGS,
 1 chains, each with 11000 iterations (first 1000 discarded)
 n.thin= 10 n.sims= 1000 iterations saved
      mean sd 2.5% 25% 50% 75% 97.5%
alpha_1 0.89 0.06 0.79 0.85 0.89 0.92 1.00
```

alpha_2	-0.04	0.05	-0.13	-0.07	-0.04	-0.01	0.05
alpha_3	0.00	0.04	-0.07	-0.03	0.00	0.02	0.07
beta_1_1	0.16	0.05	0.07	0.13	0.16	0.19	0.25
beta_1_2	-0.08	0.05	-0.18	-0.12	-0.08	-0.05	0.00
beta_1_3	0.57	0.05	0.47	0.54	0.57	0.61	0.67
beta_1_4	-0.29	0.05	-0.38	-0.32	-0.29	-0.26	-0.19
beta_2_1	-0.07	0.04	-0.15	-0.09	-0.06	-0.04	0.01
beta_2_2	-0.08	0.04	-0.16	-0.11	-0.08	-0.06	-0.01
beta_2_3	0.42	0.04	0.34	0.40	0.42	0.45	0.50
beta_2_4	-0.11	0.04	-0.18	-0.13	-0.11	-0.08	-0.03
beta_3_1	-0.07	0.04	-0.15	-0.10	-0.07	-0.05	0.00
beta_3_2	-0.06	0.04	-0.14	-0.09	-0.06	-0.04	0.00
beta_3_3	0.00	0.04	-0.07	-0.03	0.00	0.03	0.08
beta_3_4	0.34	0.04	0.27	0.32	0.34	0.37	0.42
tau_1_1	2.05	0.57	1.26	1.66	1.94	2.31	3.50
tau_1_2	-0.33	0.60	-1.64	-0.66	-0.31	0.04	0.82
tau_1_3	-0.11	0.19	-0.47	-0.23	-0.11	0.01	0.29
tau_2_1	-0.33	0.60	-1.64	-0.66	-0.31	0.04	0.82
tau_2_2	3.29	0.79	2.21	2.77	3.13	3.62	5.31
tau_2_3	-0.57	0.28	-1.21	-0.73	-0.54	-0.38	-0.09
tau_3_1	-0.11	0.19	-0.47	-0.23	-0.11	0.01	0.29
tau_3_2	-0.57	0.28	-1.21	-0.73	-0.54	-0.38	-0.09
tau_3_3	1.53	0.16	1.26	1.42	1.51	1.63	1.89
sigma2_1_1	0.56	0.13	0.33	0.47	0.55	0.64	0.83
sigma2_1_2	0.06	0.09	-0.12	0.00	0.06	0.12	0.22
sigma2_1_3	0.06	0.06	-0.06	0.02	0.06	0.10	0.18
sigma2_2_1	0.06	0.09	-0.12	0.00	0.06	0.12	0.22
sigma2_2_2	0.37	0.07	0.23	0.32	0.36	0.41	0.51
sigma2_2_3	0.13	0.05	0.03	0.10	0.13	0.16	0.23
sigma2_3_1	0.06	0.06	-0.06	0.02	0.06	0.10	0.18
sigma2_3_2	0.13	0.05	0.03	0.10	0.13	0.16	0.23
sigma2_3_3	0.72	0.07	0.60	0.68	0.72	0.77	0.86
a.e1	0.17	0.05	0.09	0.14	0.17	0.20	0.26
a.e2	0.81	0.29	0.28	0.62	0.80	0.99	1.37
sigmaemat_1_3	0.12	0.05	0.03	0.09	0.12	0.16	0.23
sigmaemat_2_2	1.26	0.08	1.11	1.21	1.26	1.31	1.42
sigmaemat_2_3	0.14	0.05	0.04	0.11	0.14	0.18	0.24
sigmaemat_3_1	0.12	0.05	0.03	0.09	0.12	0.16	0.23
sigmaemat_3_2	0.14	0.05	0.04	0.11	0.14	0.18	0.24
sigmaemat_3_3	1.16	0.06	1.06	1.12	1.16	1.19	1.27
tauemat_1_1	1.02	0.01	1.00	1.01	1.01	1.02	1.05
tauemat_1_2	0.01	0.01	0.00	0.01	0.01	0.02	0.03
tauemat_1_3	-0.11	0.05	-0.21	-0.14	-0.11	-0.08	-0.03
tauemat_2_1	0.01	0.01	0.00	0.01	0.01	0.02	0.03
tauemat_2_2	0.81	0.05	0.72	0.78	0.81	0.84	0.91
tauemat_2_3	-0.10	0.04	-0.17	-0.13	-0.10	-0.08	-0.03
tauemat_3_1	-0.11	0.05	-0.21	-0.14	-0.11	-0.08	-0.03

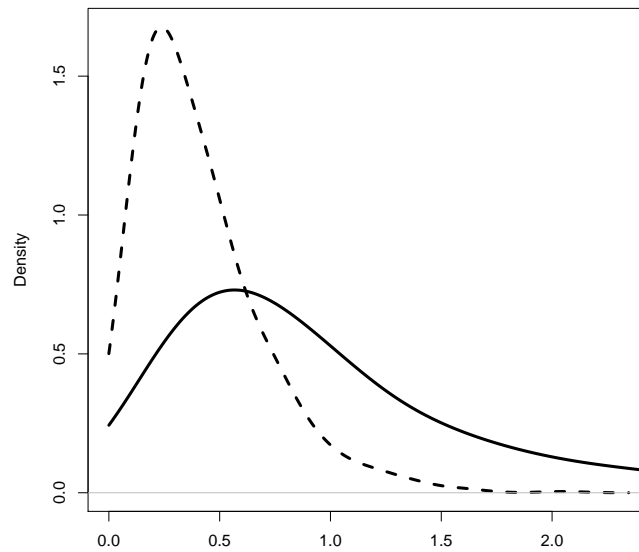


Figure 1: The solid line is the posterior density estimate of the mean for whole grains usual intake from 1000 MCMC. The dashed line is the posterior density estimate of the mean for whole grains usual intake per 1000 calories.

```
tauemat_3_2  -0.10 0.04 -0.17 -0.13 -0.10 -0.08 -0.03
tauemat_3_3   0.89 0.04  0.81  0.87  0.89  0.92  0.98
```

	Usual food intake	Usual food intake per 1000 calories	Energy usual intake
Mean	0.9359701	0.4072328	2298.3852
S.d.	0.6947905	0.2849965	446.3253
5th	0.1887330	0.0823138	1636.7102
10th	0.2632069	0.1156591	1755.9317
25th	0.4359173	0.1992969	1982.9681
50th	0.7421419	0.3408097	2267.2364
75th	1.2117088	0.5350990	2568.3280
90th	1.9042402	0.7920862	2891.3676
95th	2.3828280	0.9484864	3088.9744

This is the only output presented in the console screen. However, these results are saved in `file.estimate`s and `file.istats` respectively. Five additional output files are stored as mentioned before, but examples of these files are not presented here. Instead, we present the posterior density of the mean of usual whole grains intake plot and the posterior density of the mean of usual whole grains intake per 1000 calories in Figure 1 (`file.densityp`). We present trace plots of the intercept and coefficients in Equation 8, Equation 9, and Equation 10 in Figure 2 (`file.tracep`). We present trace plots of entry estimates saved in `file.tpvarcov` of variance-covariance matrices (i) for  $\Sigma_\mu$  in Figure 3 (first three rows), (ii) for  $\Sigma_\mu^{-1}$  in Figure 3 (last three rows), (iii) for  $\Sigma_\epsilon$  in Figure 4 and (iv) for  $\Sigma_\epsilon^{-1}$  in Figure 5.

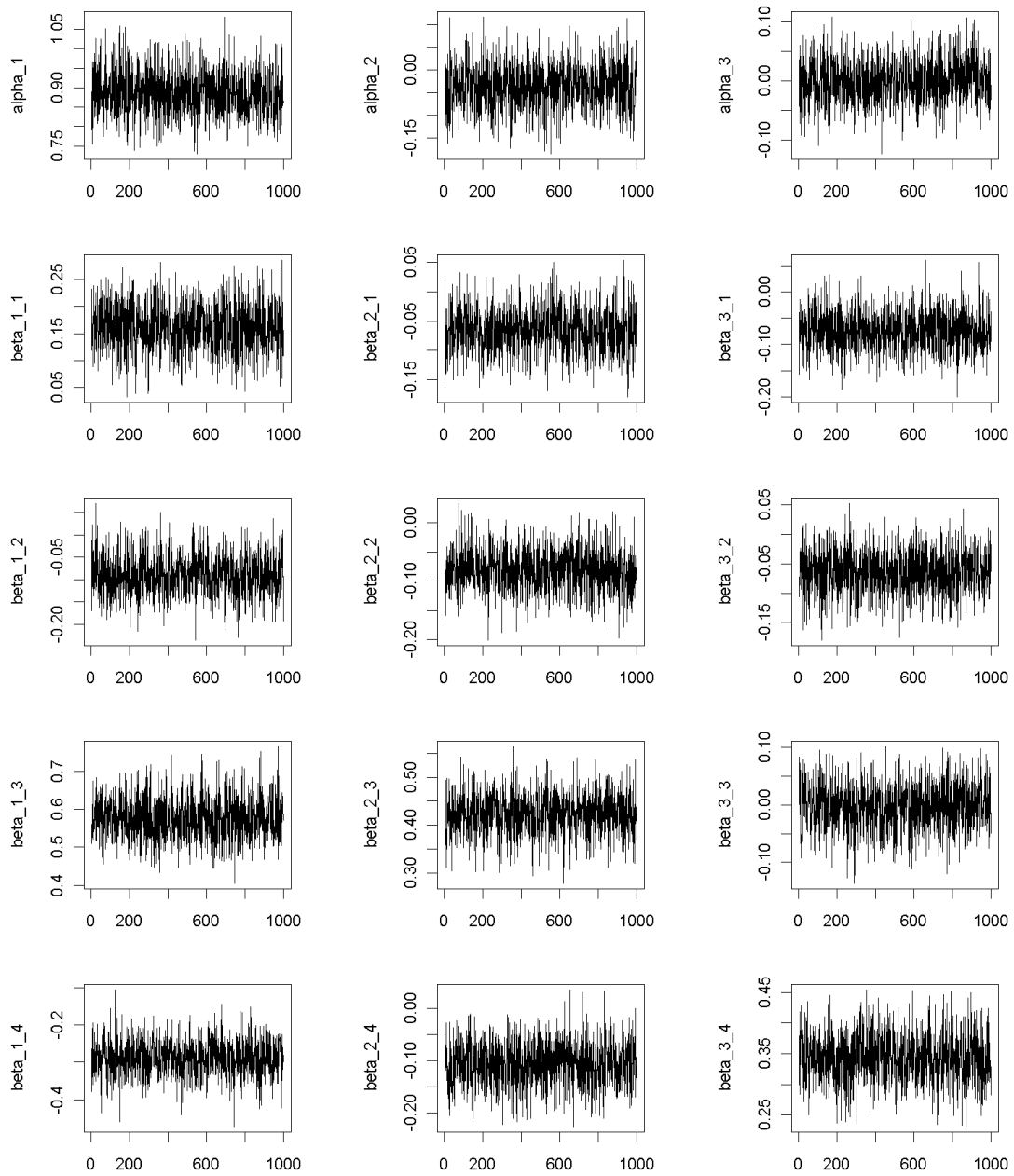


Figure 2: First column shows trace plots of parameters for the intercept, age, body mass index, consumption of whole grains and energy intake from whole grains in Equation 8. Second column shows trace plots of parameters for the intercept, age, body mass index, consumption of whole grains and energy intake from whole grains in Equation 9. Third column shows trace plots of parameters for the intercept, age, body mass index, consumption of whole grains and energy intake from whole grains in Equation 10. Trace plots were computed from 1000 MCMC of intercepts and estimated coefficients.

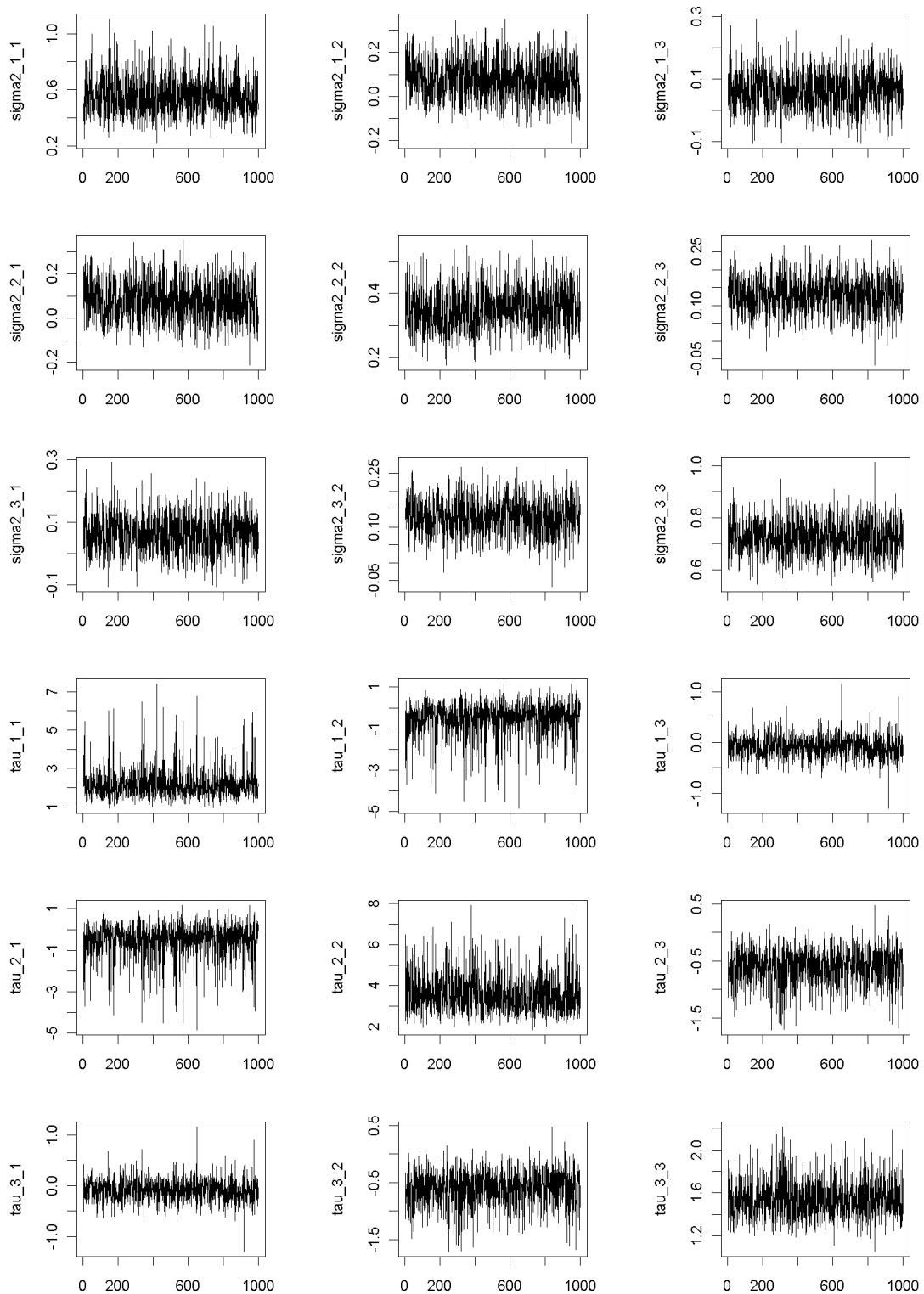


Figure 3: Trace plot of variance-covariance matrix estimates of  $\Sigma_\mu$  (first three rows) and  $\Sigma_\mu^{-1}$  (last three rows),  $j = p = 1, 2, 3$ , generated from 1000 MCMC for each entry.

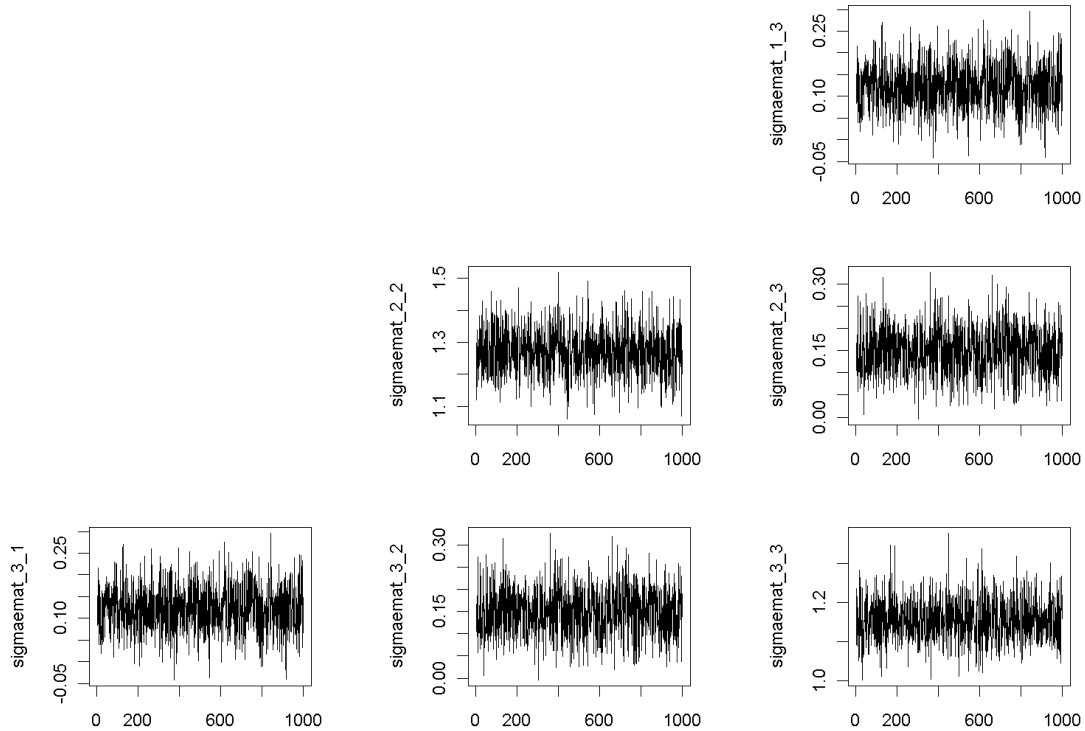


Figure 4: Trace plot of variance-covariance matrix estimates  $\Sigma_{\epsilon}$   $j=p=1,2,3$ , generated from 1000 MCMC for each entry. The following parameters are neither estimated nor plotted because they are fixed:  $\Sigma_{\epsilon 11}=1$  and  $\Sigma_{\epsilon 12}=\Sigma_{\epsilon 21}=0$ .

## 5. Discussion

This paper was motivated by the AARP calibration sub-study in nutritional epidemiology. Our main aim when implementing this function was to help users to estimate this recent nonlinear mixed three-part model of measurement error for an episodically food consumed in a Bayesian manner. Readers can evaluate if any of the parameters are different from zero by using the credible intervals reported. Similarly, not only posterior means are provided as well as medians and difference percentiles to facilitate interpretations of results for the users. This function has some limitations. This R function is implemented in its applicability to data with only two 24h recalls and four covariates. If users have more than 24h recalls and/or a different number of variates than four, users will need to modify not only the R code but also the **WinBUGS** model provided, accordingly. Although, our example was setup to run only one chain, due to the amount of time the program requires to run, users can use multiple chains. We caution users who work under Windows 7 operating system to make sure to read the instructions on how to make **WinBUGS** to run under it (available at <http://windows.microsoft.com/en-US/windows-vista/Make-older-programs-run-in-this-version-of-Windows>).

## Acknowledgments

This research was supported by a grant from the National Cancer Institute (R37-CA-057030).

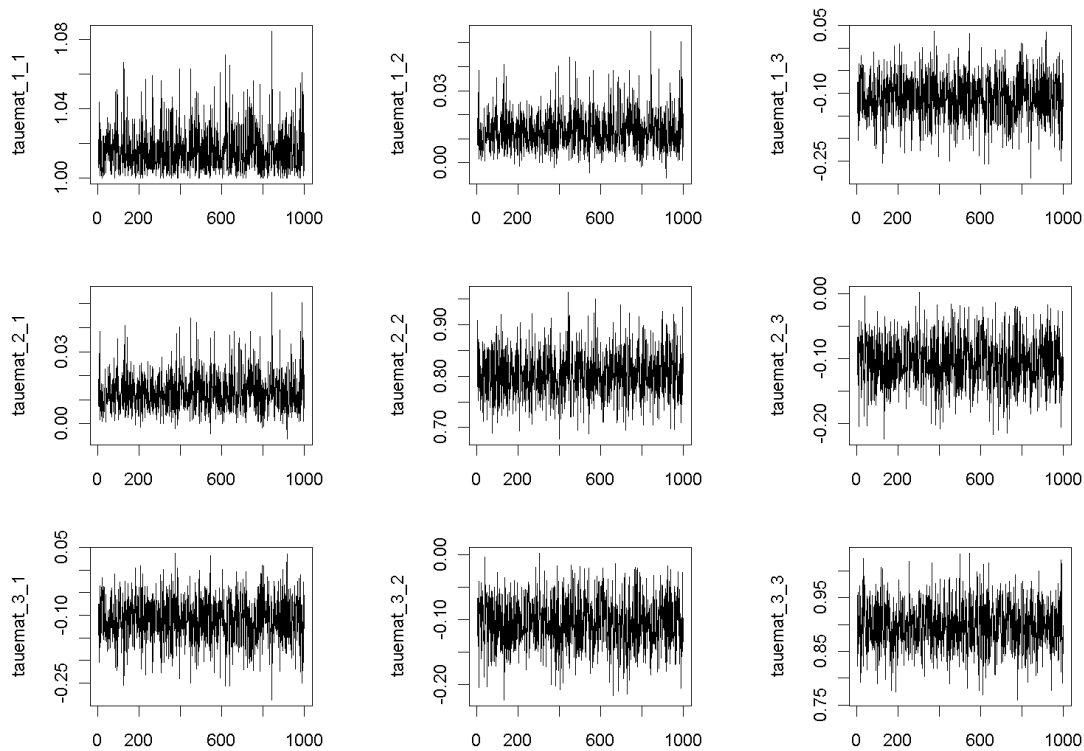


Figure 5: Trace plot of variance-covariance matrix estimates  $\Sigma_\epsilon^{-1}$   $j = p = 1, 2, 3$ , generated from 1000 MCMC each entry.

Dr. Pérez was also supported by a grant from the Michael and Susan Dell Foundation during the writing of this paper. We thank Laura Bellamy and Raja I. Malkani for proofreading this manuscript.

## References

- Dodd KW, Guenther PM, Freedman LS, Subar AF, Kipnis V, Midthune D, Tooze JA, Krebs-Smith SM (2006). “Statistical Methods for Estimating Usual Intake of Nutrients and Foods: A Review of the Theory.” *Journal of the American Dietetic Association*, **106**(10), 1640–1650.
- Eckert RS, Carroll RJ, Wang N (1997). “Transformations to Additivity in Measurement Error Models.” *Biometrics*, **53**, 262–272.
- Kipnis V, Freedman LS, Carroll RJ, Midthune D (2011). “A Measurement Error Model for Episodically Consumed Foods and Energy.” Preprint.
- Kipnis V, Midthune D, Buckman DW, Dodd KW, Guenther PM, Krebs-Smith SM, Subar AF, Tooze JA, Carroll RJ, Freedman, S L (2009). “Modeling Data with Excess Zeros and Measurement Error: Application to Evaluating Relationships between Episodically Consumed Foods and Health Outcomes.” *Biometrics*, **65**(4), 1003–1010.



R Development Core Team (2010). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.

Schatzkin A, Subar AF, Thompson FE, Harlan LC, Tangrea J, Hollenbeck AR, Hurwitz PE, Coyle L, Schussler N, Michaud DS, Freedman LS, Brown CC, Midthune D, Kipnis V (2001). “Design and Serendipity in Establishing a Large Cohort with Wide Dietary Intake Distributions: The National Institutes of Health-AARP Diet and Health Study.” *American Journal of Epidemiology*, **154**, 1159–1125.

Spiegelhalter DJ, Thomas A, Best NG (1999). *WinBUGS Version 1.2 User Manual*. MRC Biostatistics Unit, Cambridge, UK. URL <http://www.mrc-bsu.cam.ac.uk/bugs/>.

Sturtz S, Ligges U, Gelman A (2005). “R2WinBUGS: A Package for Running WinBUGS from R.” *Journal of Statistical Software*, **12**(3), 1–16. URL <http://www.jstatsoft.org/v12/i03/>.

Tooze JA, Midthune D, Dodd KW, Freedman LS, Krebs-Smith SM, Subar AF, Guenther PM, Carroll RJ, Kipnis V (2006). “A New Statistical Method for Estimating the Usual Intake of Episodically Consumed Foods with Application to Their Distribution.” *Journal of the American Dietetic Association*, **106**(10), 1575–1587.

Würtz D, Chalabi Y (2011). *timeSeries: Rmetrics – Financial Time Series Objects*. R package version 2130.92, URL <http://CRAN.R-project.org/package=timeSeries>.

Zhang S, Krebs-Smith S, Midthune D, Pérez A, Buckman DW, Kipnis V, Freedman LS, Dodd KW, Carroll RJ (2011). “Fitting a Bivariate Measurement Error Model for Episodically Consumed Dietary Components.” *The International Journal of Biostatistics*, **7**(1), Article 1.

### **Affiliation:**

Adriana Pérez  
Division of Biostatistics and Michael & Susan Dell Center for Healthy Living  
The University of Texas Health Science Center at Houston, School of Public Health  
1616 Guadalupe Street, Suite 6.340  
Austin, TX 78705, United States of America  
Telephone: +1-(512)-391-2524  
Fax: +1-(512)-482-6185  
E-mail: [adriana.perez@uth.tmc.edu](mailto:adriana.perez@uth.tmc.edu)

Saijuan Zhang, Raymond J. Carroll  
Department of Statistics  
Blocker Building, Room  
Texas A&M University  
3143 TAMU, College Station TX 77843-3143, United States of America  
Telephone: +1-979-845-3141  
Fax: +1-979-845-3144  
E-mail: [saijuanzhang@gmail.com](mailto:saijuanzhang@gmail.com), [carroll@stat.tamu.edu](mailto:carroll@stat.tamu.edu)

Victor Kipnis  
Division of Cancer Prevention  
National Cancer Institute  
6130 Executive Boulevard  
Bethesda, Maryland 20892-7354, United States of America  
Telephone: +1-301-496-7464  
Fax: +1-301-496-7463  
E-mail: [vk3b@nih.gov](mailto:vk3b@nih.gov)

Laurence S. Freedman  
Gertner Institute for Epidemiology and Health Policy Research  
Biostatistics Unit, Sheba Medical Center  
Tel Hashomer 52161, Israel  
Telephone: +972-3-530-5390  
Fax: +972-3-534-9607  
E-mail: [lsf@actcom.co.il](mailto:lsf@actcom.co.il)