

# Sample Size and Repeated Measures Required in Studies of Foods in the Homes of African-American Families<sup>1–3</sup>

June Stevens,<sup>4,5\*</sup> Maria Bryant,<sup>7</sup> Chin-Hua Wang,<sup>4</sup> Jianwen Cai,<sup>6</sup> and Margaret E. Bentley<sup>4</sup>

<sup>4</sup>Department of Nutrition, <sup>5</sup>Department of Epidemiology, and <sup>6</sup>Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC; and <sup>7</sup>Clinical Trials Research Unit, University of Leeds, Leeds, UK

## Abstract

Measurement of the home food environment is of interest to researchers because it affects food intake and is a feasible target for nutrition interventions. The objective of this study was to provide estimates to aid the calculation of sample size and number of repeated measures needed in studies of nutrients and foods in the home. We inventoried all foods in the homes of 80 African-American first-time mothers and determined 6 nutrient-related attributes. Sixty-three households were measured 3 times, 11 were measured twice, and 6 were measured once, producing 217 inventories collected at ~2-mo intervals. Following log transformations, number of foods, total energy, dietary fiber, and fat required only one measurement per household to achieve a correlation of 0.8 between the observed and true values. For percent energy from fat and energy density, 3 and 2 repeated measurements, respectively, were needed to achieve a correlation of 0.8. A sample size of 252 was needed to detect a difference of 25% of an SD in total energy with one measurement compared with 213 with 3 repeated measurements. Macronutrient characteristics of household foods appeared relatively stable over a 6-mo period and only 1 or 2 repeated measures of households may be sufficient for an efficient study design. *J. Nutr.* 142: 1123–1127, 2012.

## Introduction

Measurement of the amount of foods in the home and the nutrient content of those foods is of interest to investigators examining food availability as a predictor of dietary intake (1), obesity (2), or food security (3). Investigators developing interventions that target changes in foods in the home to promote healthier diets also need accurate measures of home food availability. Similar to dietary intake, the changing nature of home food availability from day to day must be considered when developing protocols so that valid and reliable estimates of usual availability within a home can be obtained. Variability in measurements is influenced by temporal instability (i.e., true change in availability) as a result of purchasing behaviors and food consumption and by measurement error. It is reasonable to hypothesize that day to day variability of foods in the home may be less than that of daily food intake due to repetition in the foods prepared at home, the periodic nature of shopping, and the storage of the foods purchased in one shopping episode over multiple days, weeks, or months.

Sisk et al. (4) recently presented a pilot study of 9 households in which foods in the home were assessed on up to 5 occasions over 30 d by study staff using a 251-item home observation guide. They found that the presence of foods in some categories tended to vary over time and that fresh fruits and vegetables, milk, canned vegetables, and processed meats were not usually observed during >3 of the 5 measurement episodes. They noted that only a limited number of investigators collected repeated measurements of home food availability (5–9) and that it is not known how many households need to be assessed or how many repeated measures of each household must be obtained to achieve an adequately precise measurement of home food availability.

The purpose of this research was to provide guidance to investigators on the number of households and repeated measurements needed to obtain specified levels of precision in the estimation of nutritional attributes of foods in the home. Results are shown with and without covariates included in models. To our knowledge, this is the first study to examine the study design and sample size implications of within- and between-household variation in the availability of foods.

## Materials and Methods

The participants and data collection methods used in the current study were previously described (10–12). Briefly, households were enrolled through the Infant Care Project, a longitudinal study of African American first-time mother/infant dyads who were observed in their

<sup>1</sup> Supported in part by NIH grants R21CA125735, Development of a tool to measure food availability in the home (to J.S.) and R01 HD0422119, Infant care, feeding, and risk of obesity (to M.E.B.).

<sup>2</sup> Author disclosures: J. Stevens, M. Bryant, C-H. Wang, J. Cai, and M. E. Bentley, no conflicts of interest.

<sup>3</sup> Supplemental Table 1 is available from the "Online Supporting Material" link in the online posting of the article and from the same link in the online table of contents at <http://jn.nutrition.org>.

\* To whom correspondence should be addressed. E-mail: June\_Stevens@unc.edu.

home environments (12). Mothers were recruited through clinics for the Supplemental Food Program for Women, Infant and Children in Wake and Durham counties in North Carolina. This study was approved by the University of North Carolina public health institutional review board on research involving human subjects.

The Exhaustive Home Food Inventory was used to measure all foods and drinks in participant homes (10). Staff systematically collected information by scanning Universal Product Codes on packages of food and drink items. Scanned Universal Product Code information was transferred to a laptop computer using a FoxPro data entry program (V6.0 The Sage Group) and linked to a reference database containing food identification and nutrient information. (USDA Food and Nutrient Database for Dietary Studies, 3.0. 2008; Agricultural Research Service, Food Surveys Research Group). Items without barcodes were entered by hand or using surrogate barcodes, permitting entry of all items in the home. Over a period of ~1 y in 2006 and 2007, we aimed to perform 3 assessments in each household, each separated by ~2 mo. Using this methodology, we assessed the number of food items and the household availability of total energy, dietary fiber, fat, percent energy from fat, and energy density (kcal/g).

Mother's age, presence of maternal grandmother, household size, shopping frequency, and the number of days since the last shopping were assessed by trained staff during the home visits. For analytic purposes, a weighted score was created to indicate household size adjusted for differences in energy needs. To do this, we used the sum of scores generated by each individual in the home using the age- and gender-appropriate energy intake from the Dietary Reference Intakes (13,14). The score for each household member was their estimated energy allowance relative to (divided by) that of an adult female (2200 kcal/d). Shopping behaviors were measured using 3 items that queried usual shopping frequency and the number of days since the last shopping. Maternal height and weight was measured and BMI [weight (kg)/height (m)<sup>2</sup>] was calculated.

**Statistical methods.** Distributions of the nutrient and food data were skewed for the number of food items, total energy, fiber, and fat. Log transformations created an approximately normal distribution based on the Anderson-Darling test (15). Percent energy from fat and energy density were normally distributed in the original scale.

Mixed models were used to calculate within- and between-household variability and the intra-class correlation (ICC). These quantities were then used to calculate the number of repeated measurements in one household needed for the mean of the observations in that household to lie within a specified percentage of the true value with 95% probability. Standard formulas (16,17) were used for these and other calculations (Supplemental Table 1). The specified percentages were calculated as 10, 20, and 30% of the true mean. For log-transformed variables, the required number of repeated measurements within a household for 10% deviation from the true mean in the original scale was calculated by designating the relevant interval as log (1.1) unit deviation from the true mean of the log-transformed measurements. Similarly, the calculation for 20 and 30% deviation in the original scale is based on log (1.2) and log (1.3) unit deviation in the long-transformed scale, respectively. We also calculated the number of repeated measures within a household needed to achieve a given correlation coefficient between the observed and true mean. In addition, we estimated the number of households required to detect a mean difference of a given size between 2 groups with  $\alpha$  set at 0.05 and  $\beta$  set at 0.20 (power at 80%). Estimates were made based on models with and without adjustment for covariates.

We calculated the number of single measurements of households needed in each of 2 groups to detect an effect size equivalent to 10, 25, 50, and 75% of 1 SD of the food or nutrient. The number of households needed to be measured to detect differences between groups given 2, 3, or 10 repeated measures in each household was also determined with and without adjustment for covariates. Again,  $\alpha$  was set at 0.05 and  $\beta$  at 0.20.

Analyses were performed using SAS (version 9.2, SAS Institute).

## Results

For the current study, we contacted 112 eligible mothers with 12- to 18-mo-old infants and 80 (71%) agreed to participate. Of

the 80 participating households, 63 were successfully measured 3 times, 11 were measured twice, and 6 were measured once, in sum producing 217 inventories. The main reason for not participating in the repeat assessments was change in residence. Descriptive variables were reassessed at each household visit, but because values changed little in repeated visits, only the levels at the first measurement visit are shown (Table 1).

The CV was smallest for log total energy and largest for percent energy from fat (Table 2). As expected, the number of repeated measures needed decreased as the size of the designated deviation around the mean increased. The numbers of observations needed for a specified level of deviation from the mean showed little change when covariates were added.

Percent energy from fat was the only one of 6 outcomes studied for which the variance within households was larger than the variance between households (Table 3). The within- or between-variance ratios and the ICC estimates changed little as a result of covariate adjustment and there was little or no impact on the number of repeated measures needed within households to achieve a correlation between 0.7 and 0.9 between the mean levels assessed and the true values. To obtain a correlation in this range, only one observation per household was required for most outcomes; however, as many as 7 were needed for one outcome (percent energy from fat) to obtain a correlation of 0.9.

The number of households that need to be studied to detect a difference between 2 groups of equal size was calculated assuming conventional levels of significance and power (Table 4). To detect a very small difference in the log number of food items between groups (10% of total SD) 1570 households would have to be studied if each household was measured once and covariates were not included. In contrast, only 28 households would be needed to detect a difference as large as 75% of 1 SD.

Because the assumed effect size was calculated as a percent of the SD of the outcome, the number of repeated measures needed per household was identical for each nutrient or index in calculations that included only one measurement of each household. Increasing the number of repeated measures decreased the number of households needed to detect effects; however, in no case was the total number of measurements required reduced by increasing the number of repeated measures. For instance, with each household studied only once, 63 households in each of 2 groups (a total of 126 measurement

**TABLE 1** Description of households at first measurement<sup>1</sup>

	%	Mean ( <i>n</i> = 80)	SD	Median	IQR
Maternal age, <i>y</i>		24.7	4.3	23.0	6.0
Grandmother living in home, %	16.3				
Adjusted household size, <i>n</i>		3.0	1.4	2.9	2.0
Shopping frequency, %					
Weekly	12.5				
Biweekly	40.0				
Monthly	32.5				
Other	15.0				
Time since last shopping trip, <i>d</i>		7.2	7.3	4.0	11.5
Food items, <i>n</i>		194	97	177	100
Total energy, <i>MJ</i>		990	502	872	595
Total fiber, <i>kg</i>		1.62	0.928	1.41	1.20
Total fat, <i>kg</i>		10.5	6.31	8.54	7.33
Energy from fat, %	38.3		7.55	37.9	11.4
Energy density, <i>kJ/g</i>		9.21	1.47	9.42	1.72

<sup>1</sup> Households, *n* = 80. Household inventories, *n* = 217.



**TABLE 4** ICC and number of households per group required to detect a mean difference of 10, 25, 50k and 75% of SD

Nutrient or index	No covariates						Adjusted for 5 covariates <sup>1</sup>					
	ICC	Effect size % of SD	Number of households for a fixed number of repeated measures				ICC	Effect size % of SD	Number of households for a fixed number of repeated measures			
			1	2	3	10			1 <sup>2</sup>	2	3	10
Food items, <sup>3</sup> <i>n</i>	0.79	10	1570	1408	1354	1279	0.78	10	-	1395	1337	1256
		25	252	226	217	205		25	-	224	214	201
		50	63	57	55	52		50	-	56	54	51
		75	28	26	25	23		75	-	25	24	23
Total energy, <sup>3</sup> <i>MJ</i>	0.77	10	1570	1391	1332	1248	0.74	10	-	1370	1303	1210
		25	252	223	213	200		25	-	220	209	194
		50	63	56	54	50		50	-	55	53	49
		75	28	25	24	23		75	-	25	24	22
Total fiber, <sup>3</sup> <i>g</i>	0.83	10	1570	1439	1395	1334	0.80	10	-	1413	1361	1287
		25	252	231	224	214		25	-	227	218	206
		50	63	58	56	54		50	-	57	55	52
		75	28	26	25	24		75	-	26	25	23
Total fat, <sup>3</sup> <i>g</i>	0.69	10	1570	1326	1245	1131	0.64	10	-	1288	1194	1062
		25	252	213	200	181		25	-	206	191	170
		50	63	54	50	46		50	-	52	48	43
		75	28	24	23	21		75	-	23	22	19
Energy from fat, %	0.41	10	1570	1108	955	739	0.39	10	-	1089	928	704
		25	252	178	153	119		25	-	175	149	113
		50	63	45	39	30		50	-	44	38	29
		75	28	20	17	14		75	-	20	17	13
Energy density, <i>kJ/g</i>	0.52	10	1570	1197	1072	898	0.54	10	-	1206	1085	915
		25	252	192	172	144		25	-	193	174	147
		50	63	48	43	36		50	-	49	44	37
		75	28	22	20	16		75	-	22	20	17

<sup>1</sup> Adjusted for mother's age, presence of maternal grandmother, household size, shopping frequency, and days since last shopping.

<sup>2</sup> Number of households is the same as shown for "1" under No covariates.

<sup>3</sup> Estimates are presented for the log-transformed data.

able to detect a significant difference between the 2 groups if they differed by 107 items. Thus, investigators estimating the number of households and repeated measures to be collected to obtain a desired level of precision and accuracy will need to estimate both the differences expected and the expected mean in 1 of the 2 groups in order to apply our estimates to a calculation of statistical power for untransformed nutrients. In addition, log transformation will need to be planned in the subsequent data analyses, assuming that the transformation makes the data more normal.

Researchers often use simplifying assumptions in the calculation of power for a planned study. One of the assumptions commonly made is to ignore the covariates that will be used to adjust estimates in the final analyses. Palaniappan et al. (16) compared the within-person:between-person ratio in daily dietary intakes with and without adjustment for 6 covariates. They found that covariate adjustment resulted in a larger ratio for all the nutrients examined and therefore a greater number of days of intake needed to be observed to obtain fixed levels of precision and accuracy. In women, the ratios for energy were 2.04 compared with 1.76 in the adjusted compared with the unadjusted models. The authors noted that the higher ratios were due to a reduction in between-subject variance and cautioned that not adjusting for covariates in a power calculation could result in the study having insufficient power to detect differences in intakes when examined in multivariate analyses. We found that covariate adjustment resulted in the same or a slightly lower within-household CV, ICC, and number of households that

needed to be examined for the same power. The list of covariates examined in our work differed from that used by Palaniappan et al. (16) and this difference, differences in daily nutrient intake compared to nutrients in the home, and numerous other differences in the 2 studies may account for the discrepancy. More work needs to be done to better understand the impact of statistical adjustment on the sample size requirement for an adequately powered study of nutrients in the home.

This work was limited to the study of macronutrients pertinent to our interest in obesity and listed on the food facts label. Willett et al. (19) and several others (18,20–21) have shown that the number of repeated measures needed for a given level of precision and accuracy for estimation of macronutrients is often smaller than that needed for estimation of micronutrients in the diet. This is because in a mixed group of different types of foods (meats, eggs, breads, vegetables, fruits, sweets, etc.), the distribution of micronutrients tends to be more varied than macronutrients. It is possible that a larger number of repeated visits would be required for the examination of micronutrients in the home food supply. Another limitation of this work was that although repeated assessments of households were separated by an average of 2 mo and households were studied over the course of a year, seasonal effects on foods in the home were incompletely evaluated. In addition, caution should be exercised in the application of our results to populations different from the one studied here.

A great strength of home food surveys compared with diet records or recalls is the validity of the data collected. It is well

known that self-reported dietary intake is prone to reporting bias, that obese individuals report a lower energy intake than normal-weight participants, and that correlations of reported intake with objective markers such as energy from doubly labeled water are moderate at best (22). It is a refreshing contrast that repeated measures of food in the home show very high validity when assessed using trained examiners. Recent reports on the repeatability and validity of self-reported data on foods in the home from checklists indicate promising levels of precision and accuracy using instruments that are quick and inexpensive to administer compared with the Exhaustive Home Food Inventory (23–26).

Exhaustive surveys of all foods in the home provide a method of validating home food availability checklists as well as providing detailed, precise, and accurate data for analysis of hypotheses concerning the causes and consequences of foods in the home. To our knowledge, studies determining intra- and/or inter-household level food variability have not been conducted previously. This work as well as the work of others (27–30) indicate that direct observation of foods in the home provides a feasible and valid opportunity to gain insights into the influence of the home food environment on diet, obesity, food security, and other diet related factors.

### Acknowledgments

J.S., M.B., J.C., and M.E.B. designed research; M.E.B. provided an essential database; C.W. analyzed data; J.S. and M.B. wrote the paper; and J.S. had primary responsibility for final content. All authors read and approved the final manuscript.

### Literature Cited

- van der Horst K, Oenema A, Ferreira I, Wendel-Vos W, Giskes K, van Lenthe F, Brug J. A systematic review of environmental correlates of obesity-related dietary behaviors in youth. *Health Educ Res.* 2007;22:203–26.
- Campbell KJ. Associations between the home food environment and obesity-promoting eating behaviors in adolescence. *Obesity (Silver Spring).* 2007;15:719–30.
- Rose D, Bodor JN. Household food insecurity and overweight status in young school children: results from the early childhood longitudinal study. *Pediatrics.* 2006;117:464–73.
- Sisk C, Sharkey J, McIntosh W, Anding J. Using multiple household food inventories to measure food availability in the home over 30 days: a pilot study. *Nutr J.* 2010;9:19.
- Weinstein JL, Phillips V, MacLeod E, Arsenaault J, Ferris AM. A universal product code scanner is a feasible method of measuring household food inventory and food use patterns in low-income families. *J Am Diet Assoc.* 2006;106:443–5.
- Terry K, Beck S. Eating style and food storage habits in the home: assessment of obese and nonobese families. *Behav Modif.* 1985;9:242–61.
- Turrini A, Saba A, Perrone D, Cialfa E, D'Amicis A. Food consumption patterns in Italy: the INN-CA Study 1994–1996. *Eur J Clin Nutr.* 2001;55:571–88.
- Kendall A, Olson CM, Frongillo EA. Validation of the Radimer/Cornell measures of hunger and food insecurity. *J Nutr.* 1995;125:2793–801.
- Baranowski T, Davis M, Resnicow K, Baranowski J, Doyle C, Lin L, Smith M, Wang D. Gimme 5 fruit, juice and vegetables for fun and health: outcome evaluation. *Health Educ Behav.* 2000;27:96–111.
- Stevens J, Bryant M, Wang L, Borja J, Bentley ME. Exhaustive measurement of food items in the home using a universal product code scanner. *Pub Health Nutr.* 2011;14:314–8.
- Bryant M, Stevens J, Wang L, Tabak R, Borja J, Bentley ME. Relationship between home fruit and vegetable availability and infant and maternal dietary intake in African American families: evidence from the Exhaustive Home Food Inventory. *J Am Diet Assoc.* 2011;111:1491–7.
- Sacco LM, Bentley ME, Carby-Shields K, Borja JB, Goldman BD. Assessment of infant feeding styles among low-income African-American mothers: comparing reported and observed behaviors. *Appetite.* 2007;49:131–40.
- Institute of Medicine. *Dietary Reference Intakes: the essential guide to nutrient requirements.* Washington, DC: The National Academies Press; 2006.
- Schulze MB, Hoffmann K, Kroke A, Boeing H. An approach to construct simplified measure of dietary patterns from exploratory factor analysis. *Br J Nutr.* 2003;89:409–19.
- Anderson TW, Darling DA. A test of goodness of fit. *J Am Stat Assoc.* 1954;49:765–9.
- Palaniappan U, Cue RI, Payette H, Gray-Donald K. Implications of day-to-day variability on measurements of usual food and nutrient intakes. *J Nutr.* 2003;133:232–5.
- Frison L, Pocock SJ. Repeated measures in clinical trials: analysis using mean summary statistics and its implications for decision. *Stat Med.* 1992;11:1685–704.
- Beaton GH, Milner J, Corey P, McGuire V, Cousins M, Stewart E, de Ramos M, Hewitt D, Grambsch PV, Kassim N, et al. Source of variance in 24-h dietary recall data: implications for nutrition study design and interpretation. *Am J Clin Nutr.* 1979;32:2546–59.
- Willett W. *Nutritional epidemiology.* 2nd ed. New York: Oxford University Press; 1998.
- Nelson M, Black AE, Morri JA, Cole TJ. Between and within-subject variation in nutrient intake from infancy to old age: estimating the number of days required to rank dietary intakes with desired precision. *Am J Clin Nutr.* 1989;50:155–67.
- Gibson RS. *Principles of nutritional assessment.* New York: Oxford University Press; 1990.
- Prentice AM, Black AE, Coward WA, Davies HL, Goldberg GR, Murgatroyd PR, Ashford J, Sawyer M, Whitehead RG. High levels of energy expenditure in obese women. *Br Med J (Clin Res Ed).* 1986;292:983–7.
- Bryant M, Ward D, Hales D, Vaughn A, Tabak R, Stevens J. Reliability and validity of the Healthy Home Survey: a tool to measure factors within homes hypothesized to relate to overweight in children. *Int J Behav Nutr Phys Act.* 2008;5:23.
- Crockett SJ, Potter JD, Wright MS, Bacheller A. Validation of a self-reported shelf inventory to measure food purchase behavior. *J Am Diet Assoc.* 1992;92:694–7.
- Hearn MD, Baranowski T, Baranowski J, Doyle C, Smith M, Lin L, Resnicow K. Environmental influences on dietary behavior among children: availability and accessibility of fruits and vegetables enable consumption. *J Health Educ.* 1998;29:26–32.
- Miller C, Edwards L. Development and validation of a shelf inventory to evaluate household food purchases among older adults with diabetes mellitus. *J Nutr Educ Behav.* 2002;34:261–7.
- Byrd-Bredbenner C, Abbot Jaclyn M, Cussler E. Nutrient profile of household food supplies of families with young children. *J Am Diet Assoc.* 2009;109:2057–62.
- Byrd-Bredbenner C, Bryant M, Stevens J. Household food and beverage purchasing. Letter to the Editor. *J Am Diet Assoc.* 2009;109:983–4, author reply 984–6.
- Byrd-Bredbenner C, Maurer Abbot J. Differences in food supplies of U. S. households with and without overweight individuals. *Appetite.* 2009;52:479–84.
- Scheffke SD, Bellows AC, Byrd-Bredbenner C, Cuitea CL, Rapporta H, Vivara T, Hallman WK. Nutrient analysis of varying socioeconomic status home food environments in New Jersey. *Appetite.* 2010;54:384–9.