Appendix B3

Dietary Intake Coding Procedures and Estimating Usual Intake

Appendix B3
Dietary Intake Coding Procedures and Estimating Usual Intake

B3.1 Dietary Intake Procedures for WIC ITFPS-2

The procedures for child dietary intake include a 24-hour dietary recall using the same system used in the National Health and Nutrition Examination Survey, What We Eat in America (NHANES, WWEIA) interview. This system consists of three components: the Automated Multiple Pass Method (AMPM) 24-hour recall interview system, the Post Interview Processing System (PIPS), and the SurveyNet coding application.[1] The system uses the U.S. Department of Agriculture (USDA) Food and Nutrient Database for Dietary Studies 5.0 (FNDDS5) as the source of the nutrient values.[2] The Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) Infant and Toddler Feeding Practices Study-2 (WIC ITFPS-2) collects the child's dietary intake from the child's caregiver at every interview age from 1 to 24 months, and then annually at 36, 48, 60 and 72 months. A 10 percent subsample of children at 13, 15, 18, 24, 36, 48, 60, and 72 months completes a second intake to allow estimation of "usual" intake.

B.3.1.1 AMPM Interview Data Entry

The AMPM interview asks the mother to recall all her child's dietary intake for the previous day in a systematic fashion. The interview guides the mother through the day and asks her to report all foods, beverages, and dietary supplements for each eating event during the 24-hour period; the interviewer records all responses. The interview produces a 24-hour snapshot of all foods, beverages, and dietary supplements consumed by the child. In preparation for the 13-month interview, participants received a package of measuring guides to help them report their child's portion sizes during the interview, and were asked to keep the measuring guides for the duration of

---

1 Raper, N., Perloff, B., Ingwersen, L., Steinfeldt, L., and Anand, J. (2004). An overview of USDA's dietary intake data system. *Journal of Food Composition and Analysis,* 17(3), 545-555. Available at: https://pubag.nal.usda.gov/catalog/20984.

2 Ahuja, J.K.A., Montville, J.B., Omolewa-Tomobi, G., Heendeniya, K.Y., Martin, C.L., Steinfeldt, L.C., Anand, J., Adler, M.E., LaComb, R.P., and Moshfegh, A.J. (2012). *USDA food and nutrient database for dietary studies, 5.0-Documentation and user guide.* Beltsville, MD: U.S. Department of Agriculture, Agricultural Research Service, Food Surveys Research Group.

the study. Additionally, if caregivers report that they do not know what the child ate while away from the caregiver, the dietary interviewer asked the caregiver to obtain the missing details about those foods from a knowledgeable source; afterward, the data retrieval interviewer contacted the caregiver within 2 working days to obtain the missing information.

B.3.1.2 Post Interview Processing System

Westat processes the recall data through the PIPS. During PIPS processing, approximately 70 percent of foods are auto-coded, meaning that the system assigns a food code and/or a portion quantity to the interview data. The PIPS also creates SurveyNet batches of no more than 20 intake days each, which the study team separated by recall month (3-, 5-, 7-, etc.). The online Coder Tracking System tracks each batch through the various coding and review steps. Dietary coders assign themselves batches and complete the coding for all intake days within a single assigned batch using SurveyNet.

B.3.1.3 Standard SurveyNet Processing

Assigning Food Codes

SurveyNet displays a shorthand version of each interview question and the selected response for all food description and portion data in a text box at the top of the food-coding screen. Dietary coders review this interview data, select the appropriate food code, and enter the quantity reported. In cases where the PIPS automatically assigns the food code or quantity, the dietary coder merely reviews the prefilled fields to ensure that there are no changes needed. Changes to these preassigned data may be required if the interviewer entered a comment or a text response in any field that would cause the coder to change the preassigned food code or quantity. For all foods not auto-coded during PIPS, the dietary coders review all question responses to determine the most appropriate food code to apply.

Recipe Modifications

Coding supervisors have the ability to create recipe modifications to more closely match the reported food. Coding supervisors follow the same modification guidelines used in NHANES,

which allow modification of a recipe for the type of fat used in cooking; the type of milk used in preparing selected foods (e.g., beverages, pudding, cooked cereal); the amount of liquid used to prepare condensed soup (when different from instructions); and the type of salad dressing used in salads such as coleslaw or chicken salad.

New Foods

The coders also flag new food items that they cannot link to an acceptable food code in SurveyNet. Coding supervisors do additional research to determine if the food could match an existing food code or if they need to flag the food for nutrient modification after analysis because the nutrient profile of the foods differs too much from existing food codes. The study team handled several food items in this way: agave syrup, almond milk, chia seed, edamame, Greek yogurt, hemp seed, and quinoa. Senior coding staff obtained nutrient information for these products from the USDA National Nutrient Database for Survey Research and corrected the information in the SurveyNet analysis files.

Coding Guidelines

The coders use NHANES coding guidelines to resolve common coding problems and to establish consistent coding methods. These guidelines contain rules for coding foods when not enough information is available (e.g., how much meat to code in a sandwich when the respondent did not report the amount and how to handle reports of nonstick spray). The study team developed a second set of coding guidelines for coding amounts of dietary supplements, since the default dose for non-children's supplements in the NHANES Dietary Supplement Database is generally appropriate for adults rather than infants and children. The study team develops additional guidelines throughout the study, as they encounter and resolve new issues. Coding staff document these guidelines in a decision log maintained throughout the study.

Entering Quantities

Once the food code is assigned or reviewed (in the case of auto-coding), coders review the auto-coded quantity or enter the amount of food reported. SurveyNet allows entry of portions using the same food models presented in the AMPM, and provides predetermined weights for foods in

commonly eaten portions (e.g., one-half grapefruit, one medium chicken leg). SurveyNet automatically converts food amounts entered as a shape, by dimensions (length, width, and height); volume; or weight in imperial units to a weight in grams. Coders also use SurveyNet to code imprecise measures, such as "handful," "medium bowl," or "swallow." When respondents report "Don't know" for the quantity consumed, coders are instructed to first consult the coding guidelines, which provide default amounts for items in a sandwich or salad, and other common combinations. If no coding guideline exists, coders select the "quantity not specified" portion option available in SurveyNet.[3]

Combinations

SurveyNet flags foods added to another food (e.g., milk added to cereal) or eaten in combination (e.g., the bread, meat, cheese, and spread on a sandwich) using combination codes. The system usually identifies combinations during data collection by AMPM, and PIPS assigns the combination code in SurveyNet. If coders need to add additional food codes to represent the reported food, the coder uses the combination code to link the foods.

Review

After the dietary coders assign food codes, coders and supervisors conduct quality control review by verifying, adjudicating, and editing the assigned food codes and portion amounts. Verifying involves a detailed review of coded intakes by a second coder. A coding supervisor reviews and adjudicates any notepad entries made by the second coder that highlight questions or disagreement between coders. The supervisor reviews and edits all adjudicated records and makes decisions on notepad questions and unfound foods. The adjudication process also allows evaluation of the accuracy of each coder's work. This quality control (QC) process selects two intakes from every batch for calculation of accuracy, assessing 10 percent of each coder's work. Coders must maintain 95 percent accuracy.

[3] For participants less than 2 years old, one-half of the "quantity not specified" amount was coded.

Analysis

Coding supervisors use SurveyNet to process the coded intakes and obtain the nutrient analysis. The system automatically generates error reports that document unresolved issues such as missing or invalid food codes, recipe modifications, or portion codes. Supervisors resolve all errors and rerun the analysis. The system produces two analysis data files: an "ANA" file, which contains one line of data for every food or supplement reported by the respondent on the intake day; and a "TOT" file, which contains one line of data for each respondent for a single intake day. The analysis files include 65 nutrients from the Food and Nutrient Database for Dietary Studies 5.0 (FNDDS5).

Quality Control Review

The study team performs standard QC checks on the analyzed data as a means of identifying errors. Outlier reports identify unusually high or low portions for key food items and high or low amounts of key nutrients. Coding supervisors review outliers and correct any deemed to be the result of coding errors. These outlier checks including the following:

Portion Outliers. Portion outlier reports identify errors in the reported amount of foods consumed. In addition, they serve as a check for intakes where coders applied an incorrect form of the food when specifying the amount.[4] The USDA SurveyNet software used to code AMPM intakes also identifies intakes where the portion of the reported food is either below or above established portion size range for that food item; these portion size ranges are specific for the age and gender of the respondent.

In addition to portion outliers, reports identify total calorie and macro- and micronutrient outliers. Coding supervisors examine all records flagged as outliers and correct any interviewer or coding errors. The records are reanalyzed prior to generating outlier reports for the remaining nutrients.

Minimum Criteria for Inclusion in Dataset. When conducting reviews of the intakes identified in any of the outlier reports, coding supervisors determine whether or not the intake met minimum

[4] For example, the coder entered 1 cup of rice as uncooked by mistake when the respondent reported cooked rice.

criteria. In general, an intake does not meet minimum criteria if any of the following situations are noted:

Interview is broken off prior to completing the time and occasion pass. For intakes other than those collected at the 1-, 3-, and 5-month recall, if the breakoff happens before the time and occasion is recorded for every food in the intake, the intake fails the minimum criteria and coding supervisors delete the intake from the dataset. Without time and occasion information for each food, it is not possible to determine that the reported foods span an entire day's intake. For intakes collected at 1, 3, and 5 months, the coders apply the coding guidelines developed for infant breast milk consumption; the guidelines do not require the time and occasion information.

Intake is judged as "unreliable." Although interviewers do not provide feedback on whether or not a respondent is reliable, coding supervisors implement guidelines developed in previous studies.

Meals with missing foods. Coding supervisors apply this flag when a respondent reports a meal but cannot recall foods eaten at the meal, and data retrieval for these foods was unsuccessful. For example, the respondent reports the child eating a meal at a friend's house but cannot recall the foods.

USDA Food Pattern Food Groups

The study team edits and finalizes all dietary recall data files before rerunning the SurveyNet analysis to obtain corrected nutrient values. Using the Food Patterns Equivalents Database (FPED) 2010-2011,[5] the study team appends food pattern equivalent (FPE) values to the dietary data. Coding supervisors identify food codes that do not have a match in the FPED and impute any needed food group values.

FITS Food Groups

In order to allow comparisons of the WIC ITFPS-2 dietary data to the Feeding Infants and Toddler Study (FITS), the study team assigns each FNDDS food code to one of the food groups developed for FITS 2002 and 2008.[6] The FITS adapted the food groups from the Continuing Survey of Food Intakes by Individuals (CSFII), a nationwide dietary intake study available at the time of the 2002

---

[5] Bowman, S.A., Clemens, J.C., Thoerig, R.C., Friday, J.E., Shimizu, M., and Moshfegh, A.J. (2013). *Food patterns equivalents database 2009-10: Methodology and user guide* [Online]. Beltsville, Maryland: Food Surveys Research Group, Beltsville Human Nutrition Research Center, Agricultural Research Service, U.S. Department of Agriculture. Available at: http://www.ars.usda.gov/nea/bhnrc/fsrg.

[6] Fox, M.K., Pac, S., Devaney, B., and Jankowski, L. (2004). Feeding infants and toddlers study: What foods are infants and toddlers eating? *Journal of the American Dietetic Association*, 104, 22-30.

FITS. The FITS adjusted some of the CSFII food groups to allow slightly different analysis of foods of interest to the diets of infants and toddlers. For example, because diets of young infants are largely milk-based, FITS moved yogurt, milk desserts, and cheese into other groups, leaving milk (breast milk, formula, cow's milk, and other fluid milks) in a group of its own.

B3.2 The National Cancer Institute Method for Analyzing Usual Intake Data

The study used methods recommended by the National Cancer Institute (NCI) for estimating "usual" intake. These methods rely on data from repeated administrations of a 24-hour dietary recall within a narrow time window. Both univariate and Markov Chain Monte Carlo (MCMC) models were used, with the latter primarily employed to estimate FPED food group values and Healthy Eating Index-2015 (HEI-2015) scores. For more information, see the NCI method for adjusting for dietary measurement error at https://prevention.cancer.gov/research-groups/biometry/measurement-error-impact/measurement-error-0. Tooze et al. (2010) offers an introduction to the model.[7]

Using 2 days of dietary recall information to estimate usual intake has several differences from an analysis based on a single recall. First, the repeated measures over time allow for the estimate of measurement variance (variability within person over time) separately from between-person variance. This results in adjustment of food and nutrient means and correlations and their associated standard errors for measurement error (i.e., the method estimates of what these values would be without measurement error). Second, the NCI method employs algorithms to transform the data to distribute outcomes more like a symmetric normal distribution.[8] This reduces the bias created by outliers (nutrient data is often highly skewed) and supports the validity of the assumption that errors are normally distributed, which is an assumption of the mixed model underlying the approach.[9] Third, the NCI method produces model-based estimates of distributions of food and nutrient

---

[7] Tooze, J.A., Kipnis, V., Buckman, D.W., Carroll, R.J., Freedman, L.S., Guenther, P.M., Krebs-Smith, S.M., Subar, A.F., and Dodd, K.W. (2010). A mixed-effects model approach for estimating the distribution of usual intake of nutrients: The NCI method. *Statistics in Medicine,* 29(27), 2857-2868.

[8] Box, G.E.P., and Cox, D. (1964). An analysis of transformations. *Journal of the Royal Statistical Society*, Series B, 26, 211-252.

[9] SAS Institute Inc. (2008). *SAS/STAT® 9.2 user's guide*. Cary, NC: SAS Institute Inc., Proc Genmod.

intakes that have decreased bias and error by using covariates to obtain outcome estimates. Fourth, the NCI method enables the valid estimation of "episodically" consumed food (i.e., foods not consumed on a daily basis) by employing a two-part model where one part of the model estimates the probability that the food will be consumed on a given day and the other part of the model estimates the amount of the food that is consumed if it is consumed at all.

All intakes were adjusted for the child's sex and the 12 key sociodemographic characteristics used throughout this report. In other words, these variables were included in the NCI usual intake models used. As a result, the usual intake estimates produced are tailored to this report. Alternative estimates of children's usual intake would result if different variables were included in the NCI models. Additionally, the NCI models generate a "pseudo-population." For this pseudo-population, the number of pseudo-individuals must be chosen; this report used 500 pseudo-individuals per observed respondent. Finally, the dietary data contain foods that are infrequently consumed by study children. Episodically consumed foods were those that had zero values for 10-90 percent of the sample. Some foods were so rarely consumed that they were considered "never" consumed. Foods considered never consumed had zero values for more than 90 percent of the sample.