Title

Advanced Dietary Analysis and Modeling: A Deep Dive into the National Cancer Institute Method

Journal

Journal of Nutrition, 152(11)

ISSN

0022-3166

Authors

Luo, Hanqi
Dodd, Kevin W
Arnold, Charles D
et al.

Publication Date

2022-11-01

DOI

10.1093/jn/nxac144

Peer reviewed

The Journal of Nutrition
Methodology and Mathematical Modeling

Advanced Dietary Analysis and Modeling: A Deep Dive into the National Cancer Institute Method

Hanqi Luo,[1,2,3] Kevin W Dodd,[4] Charles D Arnold,[1,2] and Reina Engle-Stone[1,2]

[1]Institute for Global Nutrition, University of California, Davis, CA, USA; [2]Department of Nutrition, University of California, Davis, CA, USA; [3]Hubert Department of Global Health, Rollings School of Public Health, Emory University, Atlanta, GA, USA; and [4]National Cancer Institute, NIH, Bethesda, MD, USA

ABSTRACT

Background: The National Cancer Institute (NCI) method has been used widely by researchers to make inferences about usual dietary intake distributions of foods and nutrients based on a limited number of 24-h dietary recalls (24-HRs). Although the NCI method does not provide individual estimates of usual intake, it can be used to address many research questions, including modeling effects of nutrition interventions on population distributions of usual intake. Software for implementing the NCI method, and corresponding code examples, is publicly available in the form of SAS macros but little formal guidance exists for conducting advanced analyses.

Objectives: We aim to present advanced techniques for working with NCI macros to conduct both basic and advanced dietary analyses and modeling.

Method: We first present the 3 basic building blocks of analyses using the NCI method: *1*) data set preparation, *2*) application of the MIXTRAN macro to estimate parameters of the usual intake distribution, including effects of covariates, after transformation of 24-HRs to approximate normality, and *3*) application of the DISTRIB macro to estimate the distribution of usual nutrient intake. Then, we illustrate how researchers can employ these building blocks to answer questions beyond typical descriptive analyses.

Results: Researchers can adapt the building blocks to: *1*) account for factors such as demographic changes or nutrition interventions such as food fortification, *2*) estimate the prevalence of dietary inadequacy via the full probability method, *3*) incorporate nutrient intake from sources not always captured by 24-HRs, such as dietary supplements and human milk, and *4*) carry out multiple subgroup analyses. This article describes the theoretical basis and operational guidance for these techniques.

Conclusion: With this article as a detailed resource, researchers can leverage the basic NCI building blocks to investigate a wide range of questions about usual dietary intake distribution. *J Nutr* 2022;152:2615–2625.

Keywords: modeling, fortification, supplements, dietary intake, 24-h recalls

Introduction

Assessment of dietary intake is fundamental to evaluating a population's risk of nutrition-linked health outcomes and developing effective intervention strategies. Because dietary recommendations are intended to be met over time and diet-health hypotheses are typically based on dietary intakes over the long term (1), researchers, nutrition program planners and analysts, and policymakers often consider inference about the distribution of usual, or long-term average, daily intake of nutrients or foods. For example, the proportion of the population with usual intake above or below certain thresholds provides insight into the prevalence of inadequate or excessive dietary intake.

Dietary assessment tools such as FFQs directly query long-term intake behavior, whereas other tools query intake over shorter time periods. For example, the 24-h dietary recall (24-HR) is a popular assessment instrument that queries intake over the past day. Other short-term assessment methods include multi-day food records or diaries. All self-report dietary assessment tools measure their nominal target (long- or

This analysis was supported by award #OPP1170661 from the Bill & Melinda Gates Foundation, USDA National Institute of Food and Agriculture Hatch project 1013897, and by the National Cancer Institute.
Author disclosures: The authors report no conflicts of interest.
Address correspondence to HL (e-mail: Hanqi.luo@emory.edu).
Abbreviations used: BRR, balanced repeated replicates; EAR, estimated average requirement; d.f., degrees of freedom; IOM, Institute of Medicine; MINIMOD, Micronutrient Intervention Modeling project; NCI, National Cancer Institute; SIMPLE, Simulating Intake of Micronutrients for Policy Learning and Engagement; 24-HR, 24-h dietary recall.

Manuscript received March 22, 2022. Initial review completed April 18, 2022. Revision accepted June 20, 2022.
First published online June 29, 2022; doi: https://doi.org/10.1093/jn/nxac144.

short-term intake) with some degree of random and systematic measurement error, but the temporal mismatch between short-term assessments and usual intake distribution introduces additional within-person random variation (from day-to-day variation in dietary intake of an individual around that person's usual intake); intake on a particular day, even if exactly measured, cannot represent usual intake, which is by definition the average of many such exact measurements.

Dietary surveys routinely employ short-term assessments (most often 24-HRs) to assess diet, but the considerable effort and resources involved in conducting frequent visits to households generally limits data collection to only a few assessments per person. In such surveys, the within-person variation in 24-HRs is typically considerable, which implies that averaging only a few measurements per person will not yield an acceptable approximation to a longer-term average. Within-person variation in 24-HRs causes the population distribution of a single measurement per person to be wider than the distribution of the average of 2 measurements per person, which in turn is wider than the average of many measurements per person. Consequently, researchers (2–6) have developed statistical methods to estimate the distribution of usual intake in cases where only a few replicates of short-term assessments are available; these methods analytically account for within-person variation, rather than relying on simple averaging. These methods are widely employed, though they only account for within-person variation, and do not address systematic measurement error in the assessments.

The approach (6) used by researchers at the National Cancer Institute (NCI) is implemented via macros written in the SAS programming language, and can be applied to 24-HR assessments of foods and nutrients. In this article, we focus on the NCI method instead of other methods because the NCI method: *1*) can assess and account for the effects of multiple covariates on usual nutrient intake, *2*) can handle complex survey designs, and *3*) can be modified to accommodate many types of analyses. The NCI method applies equally well to other short-term data collection modalities such as diet records. For simplicity, we use "24-HRs" as a synecdoche throughout the article for short-term dietary assessments (24-HRs, food records, and food diaries).

The NCI macros allow for basic descriptive analyses of usual intake distributions and prevalence of inadequate or excessive intake, which are often considered sufficient for surveillance of food and nutrient intake at the population or subpopulation level. Although users of the NCI method are strongly encouraged to review the Measurement Error Webinar Series (7) to understand the underlying theory, and to exhaustively study more in-depth material and annotated examples on the NCI macro download webpages (8), users proficient in SAS can successfully apply the macros to generate results without a detailed understanding of what is happening within the macro or the default mechanisms in place for linking the macros.

Some researchers, analysts, and policymakers are interested in questions that require more complex analyses than the rudimentary examples provided on the NCI macro download webpages. For example, for nutrients that violate the assumptions for use of the estimated average requirement (EAR) cut-point method (e.g. iron), how can the full probability method be applied to estimate the prevalence of inadequate nutrient intake (9)? How can estimated nutrient intake from human milk (typically not assessed via 24-HR) be included in estimates of usual intake distributions? Researchers may also be interested in modeling the potential effect of hypothetical scenarios, such as initiating national micronutrient fortification or supplement distribution programs. These questions can be answered by modifying NCI-provided example code. Research groups such as Hamner and colleagues (10–13), Bailey and colleagues (14, 15), and the Micronutrient Intervention Modelling Project (MINIMOD) (16–20), have performed advanced analyses using the NCI method and national survey data from the USA and other countries. However, these advanced analyses require substantial time and technical expertise to develop, and no standard, detailed guidance is available for researchers who may wish to use these techniques.

The objectives of this article are to: *1*) thoroughly explain the basic ideas of the NCI dietary analysis method and familiarize users with the NCI SAS macros (i.e. the basic building blocks of the NCI method) and *2*) equip readers with the skills to modify these building blocks to conduct advanced analyses. We begin this article by providing a detailed description of the NCI macros. Then, we illustrate how users can work with these building blocks to answer questions beyond the standard analyses performed by the NCI examples. This article is intended for users who: *1*) have a basic knowledge of measurement error theory as it pertains to dietary assessment, 2) are proficient with the SAS programming language, and *3*) are interested in learning the nuts and bolts of the NCI method and modifying the code themselves to conduct new analyses. We expect this guidance to be helpful for beginners who are learning the NCI method for estimating usual intake distributions as well as for advanced users who aim to conduct complex analysis and modeling. We focus on the basic "amount-only" NCI method, which estimates the distribution of usual intake for a dietary component consumed by nearly everyone nearly every day ("nearly daily"), although these techniques can be adapted in analyses using the NCI method for episodically consumed foods, including bivariate/multivariate modeling. Some of the advanced functionality described in this article, such as modeling the potential effect of hypothetical nutrition programs, has been built into an open-source SAS macro, the "Simulating Intake of Micronutrients for Policy Learning and Engagement (SIMPLE) macro" (21). Therefore, this article also serves as a theoretical guidebook for SIMPLE macro users.

Typical use of the NCI method to estimate usual intake distributions

The NCI method for analyzing short-term dietary assessments to describe characteristics of usual intake distributions for nearly daily consumed nutrients has been illustrated in previous publications and through a webinar series on measurement error in dietary assessment (6, 7). The foundation of the method is a mixed effects model that represents transformed 24-HR measurements for a person as the sum of: *1*) a group mean in the normal scale, *2*) a normally distributed between-person deviate (reflecting variation of usual dietary intake among individuals) that explains how normal-scale usual intake for the person differs from the group mean, and *3*) normally distributed within-person deviates that explain how particular transformed assessments for a person differ from their normal-scale usual intake. Additional terms can be included to adjust for sampling characteristics such as sequence of recall, day of week, or season of assessment.

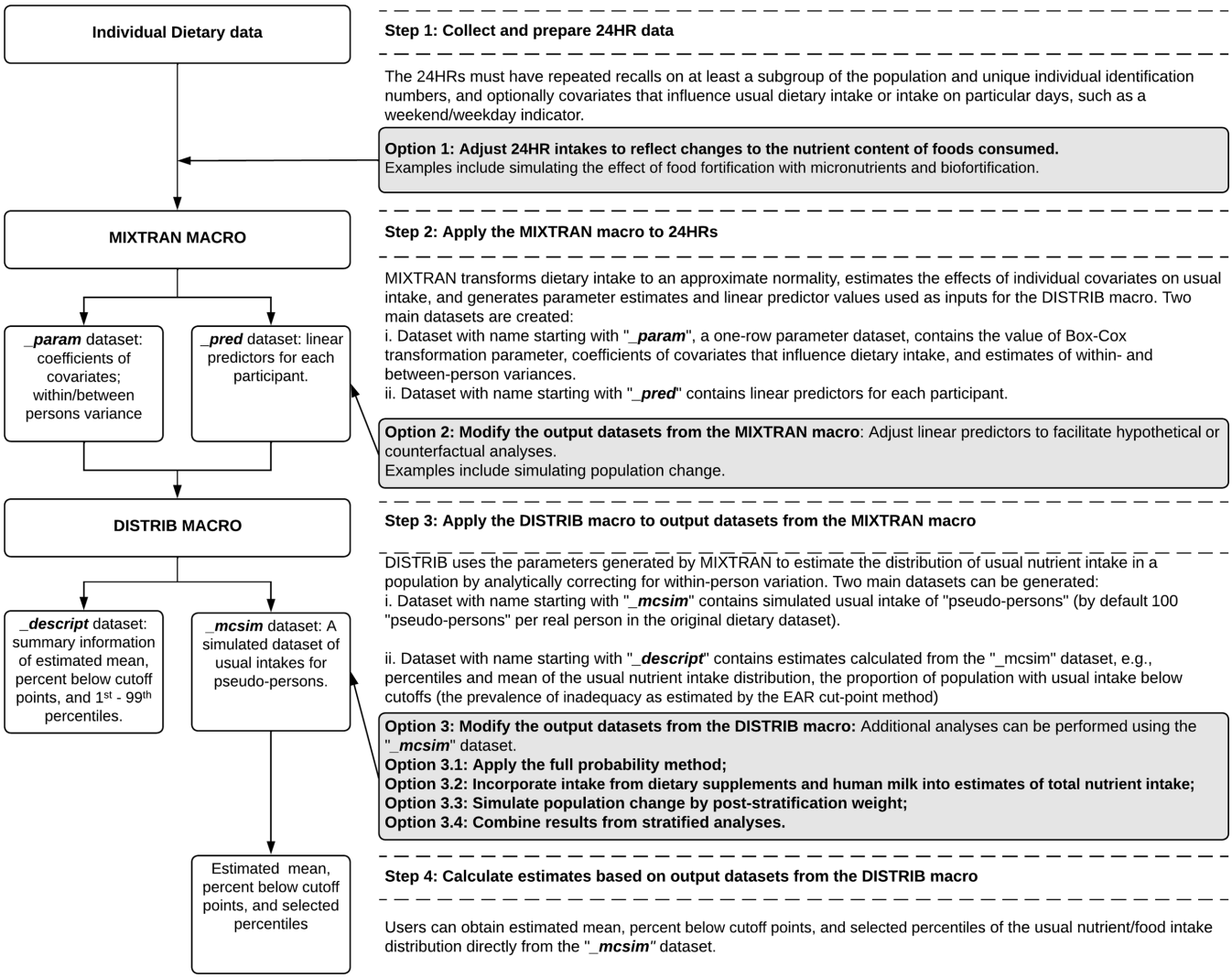

FIGURE 1 Steps of the NCI "amount-only" method.

Under the model assumptions (22), the distribution of usual intake can be derived analytically from estimates of the group mean, the between- and within-person variance components, and the transformation parameter. However, the NCI method uses a Monte Carlo approach to characterize the estimated usual intake distribution, using a random number generator to simulate usual intakes for a sample of "pseudo-persons" that are representative of the population, then taking the sample percentiles from the simulated data set as estimates of the theoretical percentiles. The advantage of the Monte Carlo approach is that it allows a general method to estimate percentiles of usual intake distribution when the group mean is a function of covariates, rather than a single parameter. In this case, because the population distribution of usual intake additionally depends on the distribution of the covariates, the Monte Carlo approach simulates samples of pseudo-persons for all covariate patterns in the population according to an appropriate covariate distribution, then combines them all to represent the total population. The estimated usual intake distribution reflects only between-person variation in usual intake, rather than a combination of between- and within-person variation in 24-HRs, and thus is narrower than the distribution of observed 24-HRs. Thus, the NCI method can be colloquially said to "shrink" the distribution of observed intake to obtain the distribution of usual intake.

The NCI amount-only method for estimating usual intake distributions is implemented via 2 SAS macros: MIXTRAN (version 2.21) and DISTRIB (version 2.2), which are available to download from the NCI website (8). A third macro, called INDIVINT, is also available for download from the website, but it is of use only when usual dietary intake (modeled from 24-HRs, and therefore subject to measurement error) is to be used as a predictor of a health outcome (assumed to be ascertained without error) using the regression calibration method to correct regression parameter estimates for the effects of measurement error in 24-HR-based dietary predictors (23). Despite its name, the INDIVINT macro does not produce output that represents true usual intake for specific persons, and therefore, users should not use the INDIVINT macro to estimate usual intake distributions.

Because the advanced use of the NCI method is based on modifying the inputs and outputs for the MIXTRAN and DISTRIB macros, we first describe each of the steps involved in applying the NCI amount-only method, from data set preparation to results output. We summarize these steps and the function, inputs, and outputs for each SAS macro in **Figure 1**. We also describe the input and output data sets

and their key variables in Table 1. This section is intended to help readers thoroughly understand the foundation of the NCI method and prepare them for more advanced analysis by modifying the core NCI method.

Step 1: Collect and prepare 24-HR data

To use the NCI amount-only method, the 24-HR data must be collected and organized following certain requirements. First, multiple days (2 at minimum) of reported intake must be available in at least a representative subset of the population, sampled persons must be identified via unique identification numbers, and repeated recalls must be identified by a day variable. The data must be organized in the "long" data set format, which means that each row represents 1 dietary record instead of 1 study participant.

Second, the number of individuals with repeat measurements and the number of repeats available influences the SEs of estimates derived from the method, and in general, more persons with 2 measurements is preferable to a smaller number of persons with >2 measurements. In the typical case, SEs are influenced primarily by the number of degrees of freedom (d.f.) associated with the within-person variance component, which is equal to the number of repeated (i.e. not the first per person) observations. A reasonable number (say 50) of d.f. can be achieved by, e.g. having 50 participants with 2 24-HRs or 10 participants with 6 24-HRs. The higher likelihood of achieving a representative subsample with more participants partly justifies the use of 2 measurements per participant. Furthermore, if the dietary variable is episodically consumed, the d.f. are based on the nonzero measurements only, so to contribute to the within-person variance component, a participant must have multiple positive 24-HR observations. Alternatively, because the d.f. for the between-person variance component is equal to the number of participants, data sets with very few participants may be problematic—25 participants with 3 24-HRs each would satisfy a 50 d.f. requirement for the within-person variance component but not for the between-person component.

Optionally, the data can contain a survey weight variable, covariates that may influence usual dietary intake, or a weekend variable indicating whether each 24-HR was collected on a weekend or a weekday. In the NCI method code, covariates must be either binary or continuous; categorical variables that contain >2 levels need to be recoded into multiple binary variables, leaving out 1 dummy variable as a reference. Although not the focus of this article, an adaptation of the NCI MIXTRAN macro for analysis of single-day data is available (24) and can be used to implement the methods described here.

Step 2: Apply the NCI MIXTRAN macro to 24-HRs

MIXTRAN transforms the target food or nutrient values from 24-HRs to approximate normality, estimates regression parameters for the effects of individual covariates on usual intake, and outputs parameter estimates and linear predictor values to be used as inputs for the DISTRIB macro. The linear predictor values are computed by simply multiplying the regression parameters by the associated covariate values for each observation, then summing the products and adding the result to the intercept term. The theory of the MIXTRAN macro has been described in previous publications (6, 23, 25).

After a successful run of the MIXTRAN macro, 4 data sets with names starting with "*_pred*," "*_param*," "*_parmsf2*," and "*_etas*" are created (in this article, the names of data sets and variables output from the NCI macros are italicized and in quotation marks, for which we use lower cases for data set names and upper cases for variable names) (Table 1). The "*_pred*" and "*_param*" data sets are the most important and are later used as the input data sets for the DISTRIB macro. The "*_pred*" data set contains linear predictors for each sampled person. By default, the "*_pred*" data set contains the linear predictor for the first observation for each person in the input data set. Researchers should keep this default behavior in mind when model covariates can vary at the observation level (i.e. each 24-HR), rather than only at the individual level. For example, if the first recall is considered the most accurate (e.g. because it is interviewer-assisted rather than by phone as in the NHANES), and therefore recall sequence is included as a model covariate, this behavior results in an estimated distribution with the first recall day as the reference level. If the default behavior is not appropriate, additional steps may be required to produce a suitable "*_pred*" data set, such as sorting the input data set in reverse sequence order, or by altering predictions to reflect a specific covariate pattern for the observation-level covariates. For example, if participants were asked whether each day is a "typical" consumption day, and the associated response variable was included as a covariate, usual intakes based on typical days could be generated by making a "*_pred*" data set that evaluates the linear predictor for each participant as if the typical day question was answered in the affirmative, even if the participant reported that all observed days were atypical. If a covariate for weekend is included, the "*_pred*" data set contains 2 linear predictors per person, 1 with and 1 without adding the regression parameter corresponding to the weekend covariate. The "*_param*" data set contains the Box-Cox transformation parameter (with zero representing the natural log transformation as a limiting case), estimated regression coefficients for covariates that may influence dietary intake, and the estimates of within- and between-person variance components. Note that the user has flexibility in the choice of which covariates to include, and to some extent the functional form of the relation between a covariate and mean normal-scale usual intake. For example, the user may want to model the covariate as a continuous variable or to use dummy variables if a categorized version of the covariate is more meaningful. An in-depth discussion of pros and cons of various methods that can be used to determine the covariates and their functional form in the final model is outside the scope of this article, but the remaining steps of the procedure assume that: *1*) the outputs from MIXTRAN reflect an appropriate model and *2*) the metrics of primary interest (e.g. fraction of the population with inadequate or excessive usual intake) are associated with the "downstream" distribution of usual intake, rather than the "upstream" MIXTRAN model.

Step 3: Apply the NCI DISTRIB macro to output data sets from the MIXTRAN macro

The DISTRIB macro uses the parameters stored in the "*_param*" and "*_pred*" data sets to simulate usual intakes according to the fitted MIXTRAN model for a sample of pseudo-persons that is representative of the target population. Characteristics (e.g. mean and percentiles) of the usual intake distribution are estimated from the sample of pseudo-persons. A successful run of the DISTRIB macro generates 2 data sets: *1*) the "*_descript*" data set containing estimates of the mean and percentiles of usual nutrient intake, and of the proportion of the population with intake below cutoffs used for applying the EAR cut-point method (26) (which estimates the prevalence of inadequate intake as the proportion of individuals with

TABLE 1 Input and output data sets and the associated key variables for the MIXTRAN and DISTRIB macros[1]

Data set	Description	Key variables
Input data set for the MIXTRAN macro		
24-HR	A data set containing variables representing nutrient intakes from a sample of 24-HRs where at least a subgroup of persons in the sample provide multiple 24-HRs. This data set must be organized in the "long" format, in which each row is a record of a dietary recall instead of a study participant. It is also an additional input dataset for the DISTRIB macro.	Required variables: • Unique identification number for each study participant. • Observed nutrient intake(s). • Repeat variable: values indicate the sequence of dietary recalls for a participant: 1 means it is the 1st, 2 means the 2nd, etc. Optional variables: • Covariates. • Weekend variable: a binary variable indicating that a dietary recall was collected either on a weekend (1) or weekday (0). • Survey weight.
Output data sets from the MIXTRAN macro applied to 1 of the nutrient intake variables in the 24-HR data set		
1)_pred	A prediction data set containing linear predictors for each participant. It is also an input dataset for the DISTRIB macro.	• Unique identification number for each study participant. • Survey weight variable (if applicable). • If the users do not specify the weekend variable to account for a weekend effect: ○ X2B2: predicted nutrient intake for each participant. • If the users specify the weekend variable to account for a weekend effect: ○ X2B2_0: predicted nutrient intake on a weekday for each participant. ○ X2B2_1: predicted nutrient intake on a weekend for each participant.
2)_param	A 1-row parameter data set containing the value of Box-Cox transformation parameter, coefficients of covariates that influence dietary intake, and estimates of within- and between-person variances. It is also an input dataset for the DISTRIB macro.	• A01_INTERCEPT: estimated intercept from the regression of transformed 24-HR intake on covariates. • Variables with names beginning with A02, A03, A04, …, etc.: estimated coefficients from the regression of transformed 24-HR intake on covariates. • MIN_AMT: the minimum positive intake value observed in the 24-HR intakes. • A_LAMBDA: Box-Cox transformation parameter. • A_VAR_U2: between-person variance. • A_VAR_E: within-person variance.
3)_parmsf2	A "long-format" version of the "*_param*" data set containing the key parameters.	• PARAMETERS: the name of the parameter, such as a coefficient from the regression of transformed 24-HR intake on covariates or the Box-Cox transformation parameter. • ESTIMATE: the estimated values of the parameter.
4) etas	A 1-row data set containing the regression equation of transformed 24-HR intake on covariates.	• ETA_2: textual specification of the left-hand side of the regression equation, including repeat variable and between-person error terms. • SHORTETA2: a shortened textual specification of the left-hand side of the regression equation, excluding repeat variable and between-person error terms.
Output data sets from the DISTRIB macro		
1)_mcsim	A simulated data set containing "true" dietary intake of "pseudo-persons."	• Unique identification number for "pseudo-persons" that represent the study population. • NUMSIMS: the number of "pseudo-persons" per real person (by default, the value of "numsims" is 100). • MCSIM_WT: survey weight of the _mcsim data set. ○ If the users specify a survey weight variable in 24-HR recalls for the MIXTRAN macro, mcsim_wt is equal to the value of the survey weight variable divided by the "NUMSIMS" value. ○ If the users do not specify a survey weight variable in 24-HR recalls for the MIXTRAN macro, mcsim_wt is equal to the reciprocal of the of "NUMSIMS" value. • MC_T: "true" dietary intake of the "pseudo-persons."

(*Continued*)

TABLE 1 (*Continued*)

Data set	Description	Key variables
2)_descript	A summary data set containing estimates calculated from the "*_mcsim*" data set, e.g. percentiles and mean of the usual nutrient intake distribution, the proportion of population with usual intake below cutoffs (e.g. prevalence of inadequacy estimated by the EAR cut-point method).	• MEAN_MC_T: mean of the usual intake distribution. • TPERCENTILE0-TPERCENTILE100: minimum, the 1st, 2nd, …, 99th, percentiles, and the maximum of the usual intake distribution. • CUTPROB1- CUTPROBN: proportion of the usual intake distribution below the corresponding cut-point. • NUMSUBJECTS: the number of study participants • Subgroup variable (if available).

[1]The same analysis also applies to other short-term dietary intake data, such as a food record. EAR, estimated average requirement; 24-HR, 24-h dietary recall.

intake below each individual's age- and sex-specific EAR for a given nutrient), and 2) the "*_mcsim*" data set containing the simulated usual intakes per pseudo-person. By default, 100 pseudo-persons are generated for each sampled person in the "*_pred*" data set, but users can override the default – increasing the number makes the estimated percentiles and fractions below cutoffs more stable, but at the expense of computing time and data storage requirements. The choice of 100 has historically been used for large national-level data sets, with several thousand participants, so that the "*_mcsim*" data set has several hundred thousand observations. Users may wish to alter the number of pseudo-persons to maintain a comparable "*_mcsim*" data set size – using 1000 instead of 100 if the data set contains only hundreds of participants. Users should note that in the DISTRIB macro, they should either leave the "*call_type*" macro parameter blank (the default), or specify either "*call_type* = *FULL*" or "*call_type* = *MC*" to ensure that the "*_mcsim*" data set is generated. If a weekend variable is in the MIXTRAN model, usual intake for a pseudo-person in the "*_mcsim*" data set is a weighted average of usual weekend intake and usual weekday intake using the default (3/4) or a user-specified weekend to weekday ratio.

Step 4: Calculate estimates based on output data sets from the DISTRIB macro

From the "*_descript*" data set, researchers can directly obtain parameters such as the estimated mean and selected percentiles of the usual nutrient intake distribution, and prevalence of inadequate or excessive intake. For implementation of advanced methods, it is important to note that researchers can calculate these (and other) estimates from the "*_mcsim*" data set. Researchers can estimate the population mean and percentiles of usual intake distribution by calculating the weighted (by variable "*MCSIM_WT*") mean and percentiles of the variable "*MC_T*". Using the EAR cut-point method (26), researchers can construct a binary variable (e.g. "*INADEQUACY*") that equals 1 if "*MC_T*" is less than the EAR cut-off point, and equals zero otherwise. The estimated prevalence of inadequate intake is the survey weighted mean of the binary variable "*INADEQUACY.*" Similarly, a binary variable (e.g. "*EXCESS*") that equals 1 if "*MC_T*" is greater than the tolerable upper intake level (UL) can be used to estimate the prevalence of excessive intake.

Advanced use of the NCI method

Researchers can use the existing NCI building blocks to answer questions not included in the standard "*_descript*" output. These advanced analyses include, but are not limited to, simulating the effects of nutrition programs such as fortification and supplementation, forecasting changes in dietary intake associated with changes in population characteristics, incorporating reported nutrient intake from human milk or dietary supplements to obtain estimates of intake from sources other than foods and drinks reported on 24-HRs, or estimating nutrient inadequacy using the full probability method (9). In this section, we discuss how to modify the NCI building blocks described above using various options in the analysis, as illustrated in Figure 1. The data set names and variable names referred to in this section are described in Table 1.

Option 1: Adjust 24-HR intakes to changes to the nutrient content of foods consumed

Some intervention programs operate by altering nutrient profiles of selected foods (such as staple foods or condiments) via fortification (adding extra nutrients from an external source) or biofortification (altering the nutrient profile of a staple crop via genetic engineering or selective breeding). Under the assumption that food consumption patterns do not change after implementing the program, researchers can model the effects of fortification by adjusting nutrient intakes from 24-HR reports that reflect hypothesized changes in nutrient profiles. The intake for the 24-HR for person i on day j must first be adjusted using the equation:

$$\begin{aligned}&\text{nutrient intake after simulated fortification}_{ij}\\&\quad=\text{nutrient intake without fortification}_{ij}\\&\qquad+\text{amount of fortified food consumed}_{ij}\\&\qquad\times\text{fortification level}\end{aligned}\tag{1}$$

The amount of fortified food consumed is measured in grams, and the fortification level is expressed on a per-gram basis.

Then, researchers need to apply the MIXTRAN and DISTRIB macros to the modified 24-HRs to estimate the distribution of usual intake with a fortification program in place. The result can be compared to the distribution obtained from the unmodified data (although formal statistical testing of differences between the 2 estimates should recognize that because the nutrient profiles for only a subset of foods are altered, the 2 distributions are not estimated independently). This example assumes that the unfortified food (e.g. wheat flour) is completely replaced by the fortified one. However, researchers can also model scenarios where only a proportion of the unfortified food is replaced by the fortified version. For example, researchers could model a scenario where only certain brands of foods are fortified, if information on the brand or other

characteristics of the food is available, or where food sold only in some geographic regions is fortified. Because fortification and biofortification both alter the nutrient profiles of selected foods (often with the assumption that consumption patterns are unaltered), the method for fortification simulations can be applied to biofortification as well. Using the same approach, researchers can also estimate the effect of hypothesized changes in food consumption, such as replacing whole fat milk with nonfat milk. This approach of modifying the 24-HRs and then applying the MIXTRAN and DISTRIB macros is sometimes referred to as the "add-then-shrink" approach because the additional nutrient intake is added to the 24-HRs before the NCI method is applied (27) to obtain an estimated distribution more narrow (or "shrunken") than the sample distribution of 24-HR measurements. Because fortification involves changes to the nutrient content of foods which are captured by a 24-HR, it is appropriate to modify 24-HR observations to reflect a hypothetical fortification scenario before applying the NCI method (27). More complicated hypothesized changes can be modeled by altering the nutrient profiles for some, rather than all, 24-HRs, if only some fraction of the population will be affected by changes in the nutrient profiles.

Option 2: Modify the output data sets from the MIXTRAN macro

The implicit assumption in **Step 2** of the NCI method is that the sample data set is representative of the target population. However, some analyses consider hypothetical, or counterfactual, scenarios where the target population is different from the sampled population. For example, in comparing nutrient intakes between the USA and Canada, Kirkpatrick and colleagues conducted sensitivity analyses to estimate nutrient adequacy in a hypothetical USA where food and nutrition assistance programs (e.g. USDA's Special Supplemental Nutrition Program for Women, Infants, and Children [WIC] or Supplemental Nutrition Assistance Program [SNAP]) do not exist (28). Estimates for an unobserved scenario can be obtained after modification of the linear predictor output ("*_pred*" data set) from MIXTRAN. In the analyses by Kirkpatrick et al. (28), an indicator for participation in assistance programs was included as a covariate in the MIXTRAN macro, and the corresponding parameter estimate was subtracted from the linear predictors of individuals who were enrolled on the program. After modifying the "*_pred*" data set in such a way, the rest of the analysis proceeded as usual and the desired counterfactual estimates were produced. The same could be done for continuous characteristics. For example, individual linear predictors for each person could be recalculated while holding income at a fixed value.

Option 3: Modify the output data sets from the DISTRIB macro

Researchers can also modify the "*_mcsim*" data set generated by the DISTRIB macro to conduct a diverse range of advanced analyses. We describe the following 4 scenarios that modify the "*_mcsim*" data set.

Option 3.1: Apply the full probability method.

Use of the EAR cut-point method to estimate prevalence of nutrient adequacy requires several assumptions, including: *1*) the distribution of requirements is approximately symmetrical, *2*) between-person variability in usual intake is greater than the variability in the requirements distribution, *3*) prevalence is neither very low nor very high, and *4*) usual intake is independent of requirements (26). Under these assumptions, the prevalence of nutrient inadequacy (the proportion of the population with usual intake less than requirements) is estimated as the proportion of the population with usual intake below the EAR (26), which can be estimated as previously described. Because empirical information on the distribution of requirements is unknown for many nutrients, researchers typically assume that the distribution of requirements is normal, and the EAR cut-off point method is applied to most nutrients.

When the distribution of nutrient requirements is known to be skewed (as in the case of iron), the "full probability method" can be used (9) in conjunction with the "*_mcsim*" data set. The full probability approach requires creating a variable (e.g. "*INADEQUACY*") that is the assigned probability of inadequate iron intake for **each** usual intake in the "*_mcsim*" data set according to a reference table appropriate to age, sex, and oral contraceptive use (only applicable to women of childbearing age) (29). For example, if analyzing women of childbearing age, the usual intake value of 7.0 mg/d for a woman aged between 19 and 50 y and not using oral contraceptives is assigned a probability of inadequacy of 0.65 based on the Institute of Medicine (IOM) reference (29). It is important to note that the default level of fractional iron absorption assumed in the IOM tables is 18%. Users are advised to adjust the iron requirement distribution based on the expected level of iron absorption in their study population. Then, after assigning the probability of inadequacy to each observation in the "*_mcsim*" data set, the estimated prevalence of inadequate iron intake can be calculated as the survey weighted (using variable "*MCSIM_WT*") mean of the "*INADEQUACY*" variable. Additional details (e.g. choice of reference table for probability of adequacy or modifications to assumed fractional iron absorption) for applying the full probability method to assess adequacy of iron intake using the NCI method have been described previously (Supplemental Methods 1 of the SIMPLE macro [21]).

Option 3.2: Incorporate intake from dietary supplements and human milk into estimates of total nutrient intake.

The NCI method is intended to adjust for within-person variation in intakes measured by short-term assessments. However, some contributors to intake may be assessed through other modalities. Examples include nutrients from dietary supplements that may be queried through questionnaires with a medium- or long-term recall period (e.g. 30 d), or human milk consumption, which may be assessed via the isotope dilution method that typically represents a period of 2 wk or more (30). Although medium- or long-term assessments are subject to within-person variation that could be accounted for by collecting replicates, in practice researchers tend to assume that such assessments capture a sufficiently long-term average that the incremental benefit of fully adjusting for the effects of within-person variation is not worth the required effort and expense. Thus, estimating the distribution of long-term intake from multiple sources (potentially with different sources of measurement error) may best be served by combining quantities representing usual intake distribution from 24-HR-based sources with quantities from other sources. This approach allows independent adjustments, if possible, for measurement error in multiple sources of intake. In contrast to the "add-then-shrink" approach of **Option 1** where 24-HR measurements are adjusted to reflect additions to intake, some researchers (15, 27) suggest estimating distributions of usual total intake from food and dietary supplements using a "shrink-then-add" approach

where 24-HRs are first used to produce an estimated usual food-based intake distribution that is then further augmented with the estimated distribution of usual dietary intakes from dietary supplements. Garriguet (27) demonstrated that the "shrink-then-add" approach was better, compared to the "add-then-shrink" approach, at reflecting evidence that supplement users may have different food consumption patterns than nonusers, e.g. because they take supplements to make up for shortfalls in their food-based intakes, or (if particularly health conscious) they take supplements in addition to having a diet that provides sufficient nutrient levels on its own.

Other characteristics of dietary supplement intake argue for the use of "shrink-then-add" even when supplement intake is queried via 24-HRs. First, each dietary supplement formulation provides a fixed dosage of a nutrient (generally some fraction or multiple of the RDA that may be much larger than would be provided from typically consumed amounts of food sources). Distributions of reported amounts of intake from supplements often have multiple "spikes" that reflect common dosages across formulations. Adding supplement intake (regardless of measurement modality) to 24-HR nutrient intakes from food sources yields spiky distributions that cannot be easily transformed to normality, because the supplement-provided amounts overwhelm the smaller contributions from food sources. Second, some dietary supplements are consumed episodically, either by design (e.g. a high-dose of vitamin D once a week) or by personal habit (e.g. sometimes forgetting to take a daily multivitamin, or taking vitamin C or calcium-containing antacids only when feeling ill). Total 24-HR intakes (obtained by adding together 24-HR-based intakes from supplements and food sources) for infrequent supplement users have larger within-person variance than either frequent or never-users of supplements because the first group can have a mix of large (from supplements) and small (from food) intakes, compared with consistently large or consistently small intakes for the other groups. This feature violates one of the main assumptions of the NCI method (22), namely that there is a common within-person variance component that applies to the population represented by the input data. For these reasons, the "shrink-then-add" approach is considered most appropriate for estimating total intake distributions including nutrients from both food sources and supplements.

The shrink-then-add approach can be implemented in a 4-step process. The first step is to run the MIXTRAN and DISTRIB macro on nutrient intake solely from food and obtain the "*_mcsim*" data set. The second step is to merge the "*_mcsim*" data set with the data set containing nutrient intake from dietary supplements by the unique identification numbers. The third step is to generate a new variable representing total nutrient intake by adding the simulated usual intake from food ("*MC_T*" variable) to the nutrient intake from other sources (e.g. dietary supplements). Finally, researchers can analyze this new variable following the guidance for **Step 4** of the typical use of the NCI method (described above) to obtain estimates of the mean and percentiles of usual total intake distribution, as well as estimates of the prevalence of inadequate or excessive usual total intake.

Option 3.3: Simulate population change by poststratification weight.

Researchers can also simulate changes in population characteristics, such as urbanization, by modifying the survey weight ("*MC_WT*" variable) in the "*_mcsim*" data set. Following the principle of poststratification (31), researchers can (preferably with the aid of a survey statistician) create a new profile of a hypothetical population and estimate characteristics of the usual nutrient intake distribution following the guidance for **Step 4**. For example, if the original sample population included 30% urban residents, the survey weights could be adjusted to model a hypothetical population with 50% urban residents. Variables used to define the poststratification cells (urbanization in the example) must be included in the MIXTRAN model as covariates.

Option 3.4: Combine results from stratified analyses.

Even when population-level estimates are of interest, there are situations where the NCI method is best applied separately to subgroups of a population. For example, older people can have completely different diet profiles compared with young children, and covariates that influence dietary intake can be different as well. Thus, for a study population with a wide age range, (e.g. from the age of 0 to 99 y), the full sample might be split into groups: young children (under the age of 5 y), preteens (aged 5–12 y), teenagers (aged 13–18 y), and adults (aged 19+ y); the MIXTRAN and DISTRIB macros can then be run separately on these different subgroups to generate different sets of result data sets (i.e. "*_mcsim*" and "*_descript*" data sets for young children, preteens, teenagers, and adults, respectively). Researchers can append multiple "*_mcsim*" data sets to obtain a combined data set that allows calculation estimates for the entire population, or for combinations of subgroups that were analyzed separately. This approach complements the ability to perform subgroup analysis via inclusion of covariates. Using the example splits given above, separate distribution estimates can be obtained for the populations of young children <2 y and 3–4 y via inclusion of an appropriate binary or categorical variable for subgroup and the use of the "subgroup" option of the DISTRIB macro applied only to the "young children" portion of the split sample. However, overall results for individuals aged 3–14 y (an age range that spans 2 portions of the split sample) would require a combined "*_mcsim*" results data set.

In summary, to carry out these advanced analyses, researchers need to modify either the input data set for the NCI macros or output data set(s) from the macros that comprise the NCI method (Figure 1). Researchers can also model the combined effects of different interventions by modifying multiple options on the NCI method pathway. For example, researchers can simulate the combined effect of micronutrient fortification and supplementation programs by using **Options 1** and **3.2**. Note that the preceding material focuses on obtaining the estimates from advanced analyses; obtaining SEs essential to inference for estimates from MIXTRAN and DISTRIB is not straightforward. For some complex survey designs, SEs can be calculated using the Balanced Repeated Replication (BRR) method (21); for other survey designs or for simple randomized samples, a bootstrap approach can be used. Both methods entail running the MIXTRAN/DISTRIB combination many times on modified input data, saving the desired output each time, and then estimating SEs from the replicated output using code such as that provided by the NCI BRR_PVALUE_CI macro (8). A substantial amount of SAS programming is required to automate this approach for application to a variety of nutrients and/or data sets, with careful attention paid to properly combining output from the MIXTRAN and DISTRIB macros.

Advanced application and modification of the NCI macros requires intermediate to advanced statistics and programming skills. The open-source SIMPLE macro and its variants provide

a user-friendly structure that streamlines basic and some advanced analyses. In particular, they allow: *1*) typical use of the NCI method, *2*) adjustment of 24-HR intakes to reflect changes to the nutrient content of foods consumed (Option 1), *3*) application of the full probability method (Option 3.1), and *4*) incorporation of intake from dietary supplements and human milk into estimates of total nutrient intake (Option 3.2), while simplifying the calculation of SEs using the BRR or bootstrap methods (21). However, some modifications need individual customization of the NCI macros based on specific research questions, such as modifying the data sets output from the MIXTRAN macro (Option 2), so they have not been incorporated into the SIMPLE macros.

Discussion

The collection and appropriate analysis of high-quality dietary intake data are necessary to characterize nutrition problems and inform solutions. Advanced methods of analysis and modeling applied to 24-HR data allow researchers to answer a wide range of questions about the distributions of usual intake; however, the complexity of the methods presents a barrier to full utilization. Among the various methods available to estimate distributions of usual intake from 24-HR data, a feature of the NCI method is its flexibility; however, users must be very familiar with the method to take advantage of this feature to conduct advanced analyses. In this article's "deep dive" into the NCI macros, we aimed to make the standard NCI macros as well as advanced analyses and modeling methods more accessible. Understanding and manipulating the basic building blocks of the NCI method can empower researchers to investigate a wider range of research questions. Although the techniques presented in this article have been applied by researchers to address specific research or policy questions (10–13, 15–20, 32), manuscripts rarely have space or scope to include the theory or practical details necessary for a new analyst to replicate the technique. This article serves as a reference for individuals who are familiar with dietary analysis theory and wish to apply the NCI macros to their own data as well as for experienced users to conduct advanced analyses. In addition, users of the SIMPLE macro (21), which relies on some of the techniques described here, may consult this article for more in-depth understanding of the tool and the results it generates.

Although the methods described here permit a wide range of analyses, some limitations must be noted. We focus on the nearly daily (amount-only) version of the NCI models, which is the most used. Although we do not specifically cover the analysis of foods or nutrients that are consumed episodically, a similar approach may be employed with the corresponding NCI macros for analyzing episodically consumed food (25). We also do not cover estimation of the joint distribution of usual intake for several nutrients or foods (33), for which other published NCI macros are used.

It is also important to note that the applications of the NCI method presented here only estimate the distribution of usual intake of 1 nutrient; it does not attempt to estimate the usual intake of each person in the sample. Direct estimation of usual intake for a person is not practical; it has long been established that, depending on the nutrient, large, very large, or tremendous numbers of 24-HRs are required for the average thereof to be reliably close to the true long-term average for a person (34). Statistical modeling cannot overcome this challenge. Instead, the NCI method uses the limited data available to estimate characteristics of a group, not of a single individual. Even the unfortunately named INDIVINT macro mentioned earlier does not estimate individual usual intakes (23); instead, it estimates the average usual intake for all possible individuals that would have provided the same 24-HR values that a particular sampled person provided (again, estimating a characteristic of a group, not of an individual). Most importantly, the NCI method relies on the assumption that 24-HRs are unbiased measures of intake. For the few nutrients for which an (at least approximately) unbiased biomarker exists, this assumption about 24-HRs has been shown not to hold (35, 36). Thus, any application of the NCI method will hopefully provide an improved analysis but should be interpreted with due care.

This article provides the theoretical foundation to model the potential impacts of nutrition intervention programs, such as food fortification or supplementation programs, for which it is important to assess the likely impacts of such programs on population dietary adequacy or excess. When applied and interpreted properly, these results can be used in many ways. For example, Smith and colleagues analyzed the USA National Health and Nutrition Examination Surveys to determine an optimal fortification level of ready-to-eat cereal to provide key nutrients that can minimize prevalence of nutrient inadequacy and excessive intake among all populations (37). The MINIMOD research group analyzed the Cameroon National Micronutrient Survey (38) to assess the cost-effectiveness of various combinations of vitamin A fortification and supplementation programs at national and subnational levels (18) and to examine the potential contribution to dietary adequacy of fortifying commercially available biscuits (19). Similarly, analysis of the Ethiopian National Food Consumption Survey (39) suggested that fortifying imported edible oils in Ethiopia is a cost-effective way to achieve dietary vitamin A adequacy and save lives of children aged under 3 y (20). In all, this information is useful to justify, plan, and manage successful micronutrient interventions.

In conclusion, substantial efforts and investments are required for collection of dietary intake data, so there is incentive to maximize the use of this information. Where research questions may be addressed by conducting advanced data analyses, this path may be more efficient compared with additional primary data collection. Although the NCI method is very powerful and flexible, the many capabilities of the method are not fully apparent without understanding the details of the method and its implementation. With these methods in hand, users can work with the building blocks of the NCI method to conduct a range of analyses, from accounting for human milk and supplement intake to predictive modeling of the impact of micronutrient fortification or supplementation programs. This article can be used together with the SIMPLE macro to better understand the modeling, or experienced users can do the coding themselves (21). Increasing the range of analyses that can be conducted will help facilitate the availability of dietary intake results and model predictions to inform nutrition research and policy.

Acknowledgments

The authors appreciate comments from Stephen Vosti, Caitlin French, Demewoz Woldegebreal, Justin Kagin, Katherine Adams, and Kenneth Brown from the MINIMOD project. The authors' responsibilities were as follows—all authors: contributed to developing the idea of the manuscript; HL and KWD: wrote the introduction and method sections of the

manuscript; CDA: wrote the section "Option 2" and provided editing and statistical support; RES: provided overall guidance on the project and wrote most of the discussion; and all authors: critically revised the manuscript and read and approved the final version

Data Availability

We did not describe any data or have any codes associated with this manuscript. This paper is a method/theory paper.

References

1. Carriquiry AL. Estimation of usual intake distributions of nutrients and foods. J Nutr 2003;133(2):601S–8S.
2. Nusser SM, Carriquiry AL, Dodd KW, Fuller WA. A semiparametric transformation approach to estimating usual daily intake distributions. J Am Statist Assoc 1996;91(436):1440–9.
3. Iowa State University. Software for Intake Distribution Estimation. 2003 [cited 2020 Oct 27]. Available from [Internet]: http://www.side.stat.iastate.edu/pc-side.php.
4. Dekkers AL, Verkaik-Kloosterman J, van Rossum CT, Ocké MC. SPADE, a new statistical program to estimate habitual dietary intake from multiple food sources and dietary supplements. J Nutr 2014;144(12):2083–91.
5. Harttig U, Haubrock J, Knüppel S, Boeing H. The MSM program: Web-based statistics package for estimating usual dietary intake using the multiple source method. Eur J Clin Nutr 2011;65(S1): S87–91.
6. Tooze JA, Kipnis V, Buckman DW, Carroll RJ, Freedman LS, Guenther PM, et al. A mixed-effects model approach for estimating the distribution of usual intake of nutrients: the NCI method. Stat Med 2010;29(27):2857–68.
7. EGRP/DCCPS/NCI/NIH. Measurement error webinar series. 2020 [cited 2020 Oct 27]. Available from [Internet]: https://epi.grants.cancer.gov/events/measurement-error/
8. National Cancer Institute. Single Regularly-Consumed or Episodically-Consumed Food or Nutrient. Division of Cancer Prevention. 2019 [cited 2020 Oct 27]. Available from [Internet]: https://prevention.cancer.gov/research-groups/biometry/measurement-error-impact/software-measurement-error/single-regularly-consumed-or-0.
9. National Research Council (US) Subcommittee on Criteria for Dietary Evaluation. The Probability Approach. Nutrient Adequacy: Assessment Using Food Consumption Surveys. Washington (DC): National Academies Press (US) 1986 [cited 2020 Oct 27]. Available from Internet: https://www.ncbi.nlm.nih.gov/books/NBK217531/
10. Hamner HC, Mulinare J, Cogswell ME, Flores AL, Boyle CA, Prue CE, et al. Predicted contribution of folic acid fortification of corn masa flour to the usual folic acid intake for the US population: National Health and Nutrition Examination Survey 2001–2004. Am J Clin Nutr 2009;89(1):305–15.
11. Hamner CH, Tinker SC, Berry RJ, Mulinare J. Modeling fortification of corn masa flour with folic acid: the potential impact on exceeding the tolerable upper intake level for folic acid, NHANES 2001–2008. Food Nutr Res 2013;57:19146. doi:10.3402/fnr.v57i0.19146.
12. Tinker SC, Devine O, Mai C, Hamner HC, Reefhuis J, Gilboa SM, et al. Estimate of the potential impact of folic acid fortification of corn masa flour on the prevention of neural tube defects: modeled impact of corn masa flour fortification. Birth Defects Research Part A: Clinical and Molecular Teratology 2013;97(10):649–57.
13. Hamner HC, Tinker SC. Fortification of corn masa flour with folic acid in the United States: an overview of the evidence. Ann NY Acad Sci 2014;1312(1): 8–14.
14. Bailey RL, Gahche JJ, Lentino CV, Dwyer JT, Engel JS, Thomas PR, et al. Dietary supplement use in the United States, 2011. J Nutr 2011;141(2):261–6.
15. Bailey RL, Dodd KW, Gahche JJ, Dwyer JT, Cowan AE, Jun S, et al. Best practices for dietary supplement assessment and estimation of total usual nutrient intakes in population-level research and monitoring. J Nutr 2019;149(2):181–97.
16. Engle-Stone R, Nankap M, Ndjebayi AO, Vosti SA, Brown KH. Estimating the effective coverage of programs to control vitamin A deficiency and its consequences among women and young children in Cameroon. Food Nutr Bull 2015;36(3_suppl): S149–71.
17. Engle-Stone R, Perkins A, Clermont A, Walker N, Haskell MJ, Vosti SA, et al. Estimating lives saved by achieving dietary micronutrient adequacy, with a focus on vitamin A intervention programs in Cameroon. J Nutr 2017;147(11):2194S–2203S.
18. Vosti SA, Kagin J, Engle-Stone R, Luo H, Tarini A, Clermont A, et al. Strategies to achieve adequate vitamin A intake for young children: options for Cameroon. Ann NY Acad Sci 2020;1465(1): 161–80.
19. Haile D, Luo H, Vosti SA, Dodd KW, Arnold CD, Engle-Stone R. Micronutrient fortification of commercially available biscuits is predicted to have minimal impact on prevalence of inadequate micronutrient intakes: modeling of national dietary data from Cameroon. Current Developments in Nutrition 2020;4(9): nzaa132.
20. Vosti S, Engle-Stone R, Woldegebreal D, Luo H, Asfaw E, Kagin J, et al. Estimated effectiveness and cost-effectiveness of a fortified edible oils program in Ethiopia. Current Developments in Nutrition 2020;4(Supplement_2):1735.
21. Luo H, Dodd KW, Arnold CD, Engle-Stone R. Introduction to the SIMPLE macro, a tool to increase the accessibility of 24-h dietary analysis and modelling. J Nutr 2021;151(5):1329–40.
22. Kirkpatrick SI, Guenther PM, Subar AF, Krebs-Smith SM, Herrick KA, Freedman LS, et al. Using short-term dietary intake data to address research questions related to usual dietary intake among populations and subpopulations: assumptions, statistical techniques, and considerations. J Acad Nutr Diet 2022;122(7): 1246–62. doi:10.1016/j.jand.2022.03.010.
23. Kipnis V, Midthune D, Buckman DW, Dodd KW, Guenther PM, Krebs-Smith SM, et al. Modeling data with excess zeros and measurement error: application to evaluating relationships between episodically consumed foods and health outcomes. Biometrics 2009;65(4): 1003–10.
24. Luo H, Dodd KW, Arnold CD, Engle-Stone R. A new statistical method for estimating usual intakes of nearly-daily consumed foods and nutrients through use of only one 24-hour dietary recall. J Nutr 2019;149(9):1667–73.
25. Tooze JA, Midthune D, Dodd KW, Freedman LS, Krebs-Smith SM, Subar AF, et al. A new statistical method for estimating the usual intake of episodically consumed foods with application to their distribution. J Am Diet Assoc 2006;106(10):1575–87.
26. Institute of Medicine (US) Subcommittee on interpretation and uses of Dietary Reference Intakes; Institute of Medicine (US) standing committee on the scientific evaluation of Dietary Reference Intakes. Dietary Reference Intakes (DRIs): estimated average requirements. Washington (DC): National Academies. Available from [Internet]: https://ods.od.nih.gov/HealthInformation/nutrientrecommendations.aspx#dri.
27. Garriguet D. Combining nutrient intake from food/beverages and vitamin/mineral supplements. Health Rep 2010;21(4):71–84.
28. Kirkpatrick SI, Dodd KW, Parsons R, Ng C, Garriguet D, Tarasuk V. Household food insecurity is a stronger marker of adequacy of nutrient intakes among Canadian compared to American youth and adults. J Nutr 2015;14(7):1596–603.
29. Institute of Medicine (US) Panel on Micronutrients. Iron. Dietary Reference Intakes for vitamin A, vitamin K, arsenic, boron, chromium, copper, iodine, iron, manganese, molybdenum, nickel, silicon, vanadium, and zinc. Washington (DC): National Academies Press(US); 2001 [cited 2020 Oct 27]. Available from [Internet]: https://www.ncbi.nlm.nih.gov/books/NBK222309/.
30. Lucas A, Ewing G, Roberts SB, Coward WA. Measurement of milk intake by deuterium dilution. Arch Dis Child 1987;62(8):796–800.
31. Little RJA. Post-stratification: a modeler's perspective. J Am Statist Assoc 1993;88(423):1001–12.
32. Du M, Luo H, Blumberg JB, Rogers G, Chen F, Ruan M, et al. Dietary supplement use among adult cancer survivors in the United States. J Nutr 2020;150(6):1499–508.
33. Zhang S, Krebs-Smith SM, Midthune D, Perez A, Buckman DW, Kipnis V, et al. Fitting a bivariate measurement error model for episodically consumed dietary components. The International Journal of Biostatistics 2011;7(1):1–32.

34. Basiotis PP, Welsh SO, Cronin FJ, Kelsay JL, Mertz W. Number of days of food intake records required to estimate individual and group nutrient intakes with defined confidence. J Nutr 1987;117(9):1638–41.
35. Freedman LS, Commins JM, Moler JE, Arab L, Baer DJ, Kipnis V, et al. Pooled results from 5 validation studies of dietary self-report instruments using recovery biomarkers for energy and protein intake. Am J Epidemiol 2014;180(2):172–88.
36. Freedman LS, Commins JM, Moler JE, Willett W, Tinker LF, Subar AF, et al. Pooled results from 5 validation studies of dietary self-report instruments using recovery biomarkers for potassium and sodium intake. Am J Epidemiol 2015;181(7):473–87.
37. Smith JD, Jain N, Bailey RL. Ready-to-eat cereal fortification: a modelling study on the impact of changing ready-to-eat cereal fortification levels on population intake of nutrients. Public Health Nutr 2020;23(12):2165–78.
38. Engle-Stone R, Ndjebayi AO, Nankap M, Brown KH. Consumption of potentially fortifiable foods by women and young children varies by ecological zone and socio-economic status in Cameroon. J Nutr 2012;142(3):555–65.
39. Ethiopian Public Health Institute. Ethiopia National Food Consumption Survey. Addis Ababa (Ethiopia): Ethiopian Public Health Institute; 2013 Aug.