USDA NATIONAL NUTRIENT DATABANK FORMULATION ESTIMATIONS IMPROVE NUTRIENT DATA RELEASES

USDA, ARS, Beltsville Human Nutrition Research Center, Nutrient Data Laboratory, Beltsville, Maryland 20705

Linda Lemar, BS, MS (USDA/ARS), Rena Cutrufelli, BS (USDA/ARS), Gwen Holcomb, BS (USDA/ARS), Melissa Stup, BS (JHU), Marybeth Bingham, BS (JHU)

Objective: To compare estimated nutrient values with analytical data as part of an ongoing validation of the Nutrient Data Laboratory (NDL) formulation program. The formulation program is a valuable tool for estimating missing nutrient values for multi-ingredient commercial foods.

Methods and Materials: As part of the redesign of the National Nutrient Databank System, a software module for the development of a food formulation and subsequent estimation of missing nutrient values was developed. The formulation program gives a best-fit estimate of the proportion of ingredients in a food based on regression techniques minimizing differences between known nutrient values and those estimated by the program. In order for good predictions of nutrient values to be made, high quality ingredient nutrient data are needed as well as analytical information about cooking yields and nutrient retentions during food processing. Under a multi-year project, the National Food and Nutrient Analysis Program (NFNAP), analytical nutrient data are being obtained for many of the basic commodity foods and multi-ingredient foods which are major contributors to US nutrient intake. Many of the commodity foods being analyzed are also used as ingredients in the multi-ingredient foods. This affords a good opportunity to compare analytical values for multi-ingredient foods with the estimated values calculated through the formulation program using NFNAP ingredient data.

Results: Examples of foods from several different food matrices will be presented. In some cases, analytical yield and nutrient retention factors will have been applied to increase accuracy of the nutrient estimates. Model errors, showing the percent difference of actual and estimated values, will be given.

Significance: In the latest release of the USDA National Nutrient Database for Standard Reference (SR 16) the formulation program aided in the estimations of nutrients of emerging public health interest including carotenoids, total sugars, alpha-tocopherol, and Vitamin K. Limited funds are available for food analyses. The cost benefits from analyses of commodity foods are expanded by applying these high quality data in the calculation of nutrient values for a broader set of multi-ingredient foods.

Hypotheses:

- Use of NDL's National Food and Nutrient Analysis Program (NFNAP) analytical values (nationally representative, with rigorous quality control, and official methodologies) for matching purposes will increase the accuracy of estimations of the remaining nutrients.

- Use of constraints on the formulation program,
  - setting of upper and lower ingredient bounds,
  - moisture and fat losses and gains,
  - and use of retention factors to adjust for nutrient losses due to heating or other food preparations,
  will increase the accuracy of nutrient estimations.

Validation Process:

- Identify multi-ingredient food items, across food groups, that have been or will be analyzed through NFNAP national samplings.
- Confirm availability of analytical ingredient data and ingredient cooking yields and retentions prior to the final selection of food items for validation.
  - Items needing ingredient substitutions will be avoided. Ingredient substitutions, if necessary, will be made only for ingredients assumed to be less than 5% of the total formulation and shall be made per NDL standardized procedures.
- Obtain current label or company brochure information for each item. Availability of any manufacturer supplied analytical or calculated data will be determined.
- First, formulations will be estimated for each food item based on ingredient lists and label and/or manufacturer supplied analytical or calculated data.
- Next, formulations will be estimated for each food item using analytical NFNAP data as target for matching purposes. Only proximates and those nutrients supplied on food labels, will be used for matching.
- Food item moistures and/or component weight determinations made in the food lab may be applied to further refine formulation estimations.

Validation Process Cont.:

- Possible constraints added to the formulation include:
  - Setting upper and lower bound ingredient limitations which may be drawn from published food formularies and standards of identity
  - Moisture and fat loss and gains
  - Use of retention factors to adjust for nutrient losses due to heat processing, etc.
- An estimation is optimized by minimization of the sum of nutrient model errors. Nutrient model error is defined as the difference between the calculated values of the target nutrients and their known values, expressed as a percent of the known value.
- Nutrient value output from final formulation determinations of each item shall be compared to NFNAP analytical data using the Nutrient Databank System (NDBS) Food Item Nutrient Ratio Comparison Report.

Summary

As part of an on-going validation process, the NDL formulation program was used to estimate nutrient profiles for 3 multi-ingredient foods representing different food matrices: high fat (milk chocolate), low moisture/high carbohydrate (breakfast cereal), and a mixed matrix containing appreciable amounts of fat, protein, carbohydrate, and moisture (burritos). Ingredient foods in the National Nutrient Databank System were selected based on food label ingredient lists which were ranked in order of predominance. Many frequently used ingredient foods are being analyzed under NFNAP to improve estimations made by formulation. It is particularly important to have good analytical ingredient data to accurately estimate micronutrients. After ingredients are selected constraints may be used to guide the program in producing accurate nutrient estimations (Fig. 2, 4, and 6). Nutrient ratio reports were run to determine: 1) the ratio NFNAP analytical values to formulation values estimated using NFNAP target values and 2) the ratio of NFNAP analytical values to formulation values estimated using target values from labels. Ratios for target nutrients are shown in Tables 5, 6, and 7. For milk chocolate and burritos, NFNAP analytical data and formulation data estimated using NFNAP target values were virtually identical (ratios of 1.000 to 1.0197) for all nutrients except total sugars in milk chocolate (ratio 0.9347). Good, but somewhat lower, correlation was found between NFNAP analytical and NFNAP formulation estimations for breakfast cereal (ratios of 0.9589-0.9590). Sodium in breakfast cereal was better predicted from label formulation output, with a ratio of 1.0222. Contract or industry data are frequently lacking for some of the mandatory nutrients in the Survey Nutrient Database. In these cases, data must be imputed. For multi-ingredient commercial foods, imputing is often done through the formulation program. In SR16, 16% of the imputed values were derived by formulation; formulation estimates account for 6.3% of the total number of nutrient values in this database release. Two of the nutrients frequently estimated by formulation in breakfast cereals are magnesium and copper. In Table 8 it can be seen that analytical formulation output for magnesium and copper corresponded well with known analytical values. The ratio of known to formulation output for magnesium and copper was 1.1069 and 1.1826, respectively.

Fig. 1. Chocolate Formulation Ingredients

- Sugars, granulated
- Milk, dry, whole
- Oil, vegetable, cocoa butter
- Baking chocolate, unsweetened squares
- Vegetable oil, soybean lecithin
- Vanilla extract

Fig. 2. Constraints

- Lower bound on last ingredient
- Upper bound on first ingredient listed as equal to or less than 2% on label

Fig. 3. Burrito Ingredients

- Tortillas
- Refried beans, canned
- Cheese, cheddar
- Oil, soybean (hydrogenated) and cottonseed
- Salt, table
- Spices, pepper, red or cayenne

Fig. 4. Constraints

- Lower bound on last ingredient
- Retention factors
- Moisture loss

Fig. 5. Breakfast Cereal Ingredients:

- Rice, white, long-grain, regular, raw, unenriched
- Sugars, granulated
- Cocoa, dry powder, unsweetened, processed with alkali
- Candies, semisweet chocolate
- Salt, table
- Oil, soybean, salad or cooking, (hydrogenated)
- Syrups, corn, high-fructose
- Malt extract, dried

Fig. 6. Constraints

- Lower bound on last ingredient
- Retention factors
- Moisture loss
- Relaxed order for one ingredient

Table 1. Nutrient Values for Milk Chocolate (per 100 grams)

- High Fat
- Homogeneous
- Not Enriched

Target Nutrients	Target Data			Formulation Output Data		
	Analytical NFNAP Data	Aggregated Label Data	Brand A Label Data	Analytical NFNAP Data	Aggregated Label Data	Brand A Label Data
Total Fat, g	29.66	30.14	30.23	29.58	30.14	32.80
Carbohydrate, g	59.40	61.90	58.14	59.40	60.15	57.53
Total Sugars, g	51.50	55.57	51.16	55.10	55.57	52.20
Calcium, mg	189	158	186	189	160	186
Sat. Fatty Acids, g	14.252	18.227	20.930	13.977	17.904	20.553
Total Model Error				7.26	3.85	11.57

Table 4. Formulation Comparisons by Brand (run based on label nutrients)

Nutrients	Brand A	Brand B	Brand C
Total Fat, g	32.80	26.61	29.16
Carbohydrate, g	57.53	64.29	62.82
Total Sugars, g	52.20	60.29	58.39
Calcium, mg	186	143	154
Saturated Fatty Acids, g	20.553	14.031	19.120
Total Model Error	11.57	12.38	14.13

Table 5. Nutrient Ratio Comparisons

Target Nutrients	Analytical NFNAP Data to NFNAP Formulation Output	Analytical NFNAP Data to Aggregated Label Formulation Output	Analytical NFNAP Data to Brand A Formulation Output
Total Fat	1.0027	0.9839	0.9042
Carbohydrate	1.0000	0.9876	1.0326
Total Sugars	0.9347	0.9268	0.9866
Calcium	1.000	1.1811	1.0153
Saturated Fatty Acids	1.0197	0.7961	0.6935

Table 2. Nutrient Values for Bean and Cheese Burrito (per 100 grams)

- Mixed matrix
- Homogeneous

Target Nutrients	Target Data			Formulation Output Data		
	Analytical NFNAP Data	Aggregated Label Data	Tina's Burritos Label Data	Analytical NFNAP Data	Aggregated Label Data	Tina's Burrito Label Data
Protein, g	8.60	7.45	8.45	8.45	7.93	8.34
Total Fat, g	7.17	4.98	6.34	7.17	5.43	6.34
Water	41.77	41.77*	41.77*	41.77	42.27	41.76
Sodium, mg	637	498	422	637	530	434
Total Model Error				1.42	22.95	3.94

* Based on NFNAP data

Table 6. Nutrient Ratio Comparisons

Target Nutrients	Analytical NFNAP Data to NFNAP Formulation Output	Analytical NFNAP Data to Aggregated Label Formulation Output	Analytical NFNAP Data to Label Formulation Output
Protein	1.0141	1.0850	1.0312
Total Fat	1.000	1.3204	1.1312
Water	1.000	0.9882	1.002
Sodium	0.9997	1.2022	1.486

Table 3. Nutrient Values for Breakfast Cereal (per 100 grams)

- Low Moisture/High Carbohydrate
- Homogeneous
- Enriched

Target Nutrients	Target Data		Formulation Output Data	
	Analytical NFNAP Data	Label Data	Analytical NFNAP Data	Label Data
Total Fat, g	2.90	3.23	3.02	3.31
Total Sugars, g	34.00	45.16	35.41	46.32
Total Dietary Fiber, g	1.9	3.2	2.0	3.3
Sodium, mg	635	613	662	621
Total Model Error			17.10	9.05

Table 7. Nutrient Ratio Comparisons

Target Nutrients	Analytical NFNAP Data to NFNAP Formulation Output	Analytical NFNAP Data to Label Formulation Output
Total Fat	0.9590	0.8765
Total Sugars	0.9590	0.7332
Total Dietary Fiber	0.9589	0.5743
Sodium	0.9589	1.0222

Table 8. Nutrients Frequently Estimated by Formulation

Nutrients	Analytical NFNAP Value	Analytical Formulation Output	Label Formulation Output	Analytical NFNAP Data to Analytical Formulation Output	Analytical NFNAP Data to Label Formulation Output
Magnesium, mg	38	35	56	1.1069	0.6854
Copper, mg	0.332	0.281	0.450	1.1826	0.7378