The Multifactor Screener in the Observing Protein & Energy Nutrition (OPEN) Study

November 2019

THE MULTIFACTOR SCREENER: OPEN

This information was originally published online on the Epidemiology and Genomics Research Program, National Cancer Institute's website, as a reference for the Multifactor screener in the Observing Protein & Energy Nutrition (OPEN) Study. The information contained in this document informs analyses of data from this screener. This information is achieved and provided for reference purposes only.

This publication may be viewed and downloaded from the Internet at https://epi.grants.cancer.gov/diet/screeners/files.

Suggested citation for information contained in this report: The Multifactor screener in the Observing Protein & Energy Nutrition (OPEN) Study. Epidemiology and Genomics Research Program. National Cancer Institute. https://epi.grants.cancer.gov/diet/screeners/files. Updated November 20, 2019.

Acknowledgements

The Epidemiology and Genomics Research Program, Risk Factors Assessment branch, National Cancer Institute (EGRP, RFAB, NCI) acknowledge the work of Dr. Frances E. Thompson, Dr. Amy F. Subar, Douglas Midthune, Dr. Kevin Dodd, Dr. Victor Kipnis, and Lisa Kahle, whose work formed the basis of this document.

The EGRP, NCI would also like to acknowledge the important role of Dr. Edwina Wambogo, Jennifer Lerman, and Christie Kaefer who turned the webpages associated with this screener into this document.

If you have a comment or question about this document, please contact the EGRP, RFAB, NCI, at RFAB@mail.nih.gov.

THE MULTIFACTOR SCREENER: OPEN

CONTENTS

TABLES

THE MULTIFACTOR SCREENER: OPEN

1. Overview

What is the OPEN Study?

The Observing Protein and Energy Nutrition (OPEN) Study was an NCI-sponsored study designed to assess dietary measurement error by comparing results from self-reported dietary intake data with four dietary biomarkers: doubly labeled water and urinary nitrogen, sodium, and potassium. The study was conducted from July 1999 to March 2000 and included 484 men and women, aged 40-69 years old, living in Montgomery County, Maryland.

View a list of selected publications that have used data from the OPEN Study [1-14].

How did the OPEN Study assess the FFQ and 24-Hour Recall?

Over the course of the study, OPEN participants completed two Food Frequency Questionnaires (FFQ), two 24-recall interviews, and filled out several other health-related questionnaires. They were dosed with doubly labeled water (a biomarker used for measuring energy expenditure), provided several spot urine samples to complete the doubly labeled water assessment, completed two 24-hour urine collections and had their height and weight measured. Investigators analyzed the 24-hour urines for several nutrients: nitrogen, sodium and potassium, biomarkers that measure protein, sodium, and potassium intakes, respectively. The questionnaires and samples allowed study investigators to compare what participants said they ate and drank against the objective evidence provided by the biomarkers and thus, to get a better sense of the extent and nature of error in the FFQs and 24-hour recalls.

Multifactor Screener in the OPEN Study

The Multifactor Screener may be useful to assess approximate intakes of fruits and vegetables, percentage energy from fat, and fiber. The screener asks respondents to report how frequently they consume foods in 16 categories. The screener also asks one question about the type of milk consumed. No portion size questions are asked. This screener does not attempt to assess total diet.

The foods selected to compose the screener were identified through an analysis of USDA's 1994-96 Continuing Survey of Food Intakes of Individuals (CSFII), a nationally representative survey of the food intakes of the US population available from the USDA's Food Surveys Research Group. The NCI's Risk Factor Assessment Branch (RFAB) used stepwise regression to identify the food groups that would best predict the three dietary exposures. Some of the foods in the screener predict all three exposures; some predict only one or two of the exposures.

THE MULTIFACTOR SCREENER: OPEN

The Multifactor screener was self-administered in the OPEN study and was interviewer-administered in the 2000 NHIS Cancer Control Supplement.

You can view or print the Multifactor Screener in OPEN from the National Cancer Institute's (NCI) Register of Validated Short Dietary Assessment Instruments.

The process of scoring the individual response data is described in Scoring Procedures & Results. A SAS program is included on NCI's Short Dietary Assessment Instruments site. Guidelines for the appropriate uses of the screener-estimated dietary intakes are the same as those described in Uses of Screener Estimates. Validation data for the Multifactor Screener in OPEN is presented in Validation Results. The data dictionary for the screener is available in RTF and PDF formats on NCI's Short Dietary Assessment Instruments site.

2. Scoring Procedures & Results

How Analytical Scoring Procedures Were Developed

Scoring procedures were developed to convert a respondent's screener responses to estimates of individual dietary intake for percentage energy from fat, grams of fiber, and servings of fruits and vegetables, using USDA's 1994-96 Continuing Survey of Food Intakes of Individuals (CSFII 94-96) dietary recall data. The following equations were estimated in the CSFII 94-96:

For **percentage energy from fat** and **fiber**:

$$E(\text{Dietary Factor}) = b_0 + b_1 N_{FG1} P_1 + b_2 N_{FG2} P_2 + \ldots + b_{19} N_{FG19} P_{19}$$

E(Dietary Factor) indicates the expected values for percentage energy from fat and for fiber, and assumes a normal distribution. In the CSFII 94-96 dataset percentage energy from fat was normally distributed. However, fiber was positively skewed and required a cube-root transformation to approximate normality. N_{FGk} is the usual number of times per day an individual consumed food group k; P_k is the median portion size of group k; and k indexes the 19 food groups. These 19 food groups were formed to reflect the same food groups on the screener. We calculated weighted least-squares estimates of the regression coefficients b_k, k = 0, …, 19 on CSFII 94-96 adults aged 18 and above, stratifying by gender and excluding extreme exposure values. We first included all 19 food groups in the regression model. After examining the results, we dropped food groups that failed to attain statistical significance at the $\alpha = 0.25$ level to form more parsimonious final models. In both the percentage energy from fat model and the fiber model, the lettuce food group was dropped. Because of the complex survey design, the analysis was performed using SUDAAN (RTI Inc., Research Triangle Park, NC).

For Pyramid servings of **fruits and vegetables** (defined by USDA in the 1992 Dietary Guidelines Food Guide Pyramid):

$$E([\text{Fruits and Veg}]^{1/2}) = b_0 + b_1 ([N_{FG1}P_1 + N_{FG2}P_2 + \ldots + N_{FG7}P_7]^{1/2})$$

Pyramid servings of fruits and vegetables was square-root-transformed to approximate normality; N_{FGk} is the usual number of times per day an individual consumed food group k; P_k is the median portion size of group k; and k indexes the 7 fruit and vegetable food groups. We calculated weighted least-squares estimates of the regression coefficients b_0 and b_1 on the adults in the CSFII 94-96 sample, stratifying by gender and excluding extreme exposure values.

Scoring Procedures

1. **Estimation of N_{FGk}:** Express each reported frequency as a daily average. To do this, standardize the midpoint of each frequency category to the number of times per day.

Table 2- 1 Expressing reported frequencies as daily averages

Frequency Response	Times Per Day
Never	0.0
1-3 times per month	0.067
1-2 times per week	0.214
3-4 times per week	0.5
5-6 times per week	0.786
1 time per day	1.0
2 times per day	2.0
3 times per day	3.0
4 or more times per day	4.5

2. **Estimation of P_k:** The median age- and gender-specific portion sizes for each food were estimated from CSFII 94-96. For percentage energy from fat and fiber variables, the units were in grams (Tables 2-2); for fruit and vegetable servings variables, the units were in Pyramid servings (Table 2-3).

   A Pyramid serving is defined by the U.S. Department of Agriculture in the Dietary Guidelines Food Guide Pyramid (used from 1992-2004) as:

- vegetables: 1 cup raw leafy, ½ cup of other vegetables, or ¾ cup vegetable juice; and
- fruit: 1 whole fruit, ½ cup of cut-up fruit, or ¾ cup fruit juice.

Note: current dietary guidance uses cups rather than servings.

Table 2- 2 Median Portion Size (Pk) in Grams per Mention by Age for Percentage Energy from Fat and Fiber Analyses: Men

Food Group	Age Group						
	18-27	28-37	38-47	48-57	58-67	68-77	78-99
				Men			
Cold cereals (P_1)	74.666667	61.500000	57.500000	56.000000	46.000000	39.000000	33.000000
Whole milk (P_2)	305.000000	259.250000	306.710000	244.000000	244.000000	244.000000	203.333333
2% milk (P_3)	259.250000	305.000000	244.000000	244.000000	244.000000	183.000000	183.000000
1% milk (P_4)	341.600000	245.000000	245.000000	244.000000	213.500000	223.666667	183.000000
Skim milk (P_5)	366.666667	250.000000	250.000000	245.000000	214.375000	198.937500	160.725000
Bacon or sausage (P_6)	25.000000	40.250000	32.000000	32.000000	27.000000	26.000000	24.000000
Hotdogs (P_7)	114.000000	85.500000	88.000000	114.000000	57.000000	57.000000	57.000000
Whole grain bread (P_8)	56.000000	54.000000	52.000000	52.000000	51.000000	48.250000	48.000000
100% fruit juice (P_9)	372.000000	311.250000	249.000000	249.000000	248.000000	186.750000	186.750000
Fruit (P_{10})	131.750000	128.000000	123.200000	127.500000	122.000000	118.000000	114.250000
Salad dressing (P_{11})	23.543333	23.626667	22.030000	27.500000	24.585000	19.285000	15.600000
Fried potatoes (P_{12})	112.500000	114.000000	100.000000	100.000000	85.500000	85.500000	97.000000

THE MULTIFACTOR SCREENER: OPEN

Food Group	Age Group						
	18-27	**28-37**	**38-47**	**48-57**	**58-67**	**68-77**	**78-99**
Other white potatoes (P_{13})	210.000000	196.000000	184.000000	161.000000	145.000000	127.000000	107.000000
Dried beans (P_{14})	180.000000	130.000000	172.000000	172.000000	158.125000	175.000000	170.100000
Other vegetables (P_{15})	60.013333	73.000000	74.063333	79.833333	76.500000	73.000000	67.520909
Pasta (P_{16})	330.000000	280.000000	280.000000	247.500000	280.000000	210.000000	210.00000
Nuts (P_{17})	31.625000	58.000000	35.500000	54.665000	39.250000	17.130000	35.916667
Chips (P_{18})	40.000000	40.000000	31.895000	30.000000	26.000000	21.000000	17.500000

Table 2- 3 Median Portion Size (Pk) in Grams per Mention by Age for Percentage Energy from Fat and Fiber Analyses: Women

Food Group	Age Group						
	18-27	28-37	38-47	48-57	58-67	68-77	78-99
Women							
Cold cereals (P_1)	50.000000	49.500000	44.000000	43.500000	33.000000	33.000000	33.500000
Whole milk (P_2)	244.000000	244.000000	244.000000	244.000000	198.250000	198.250000	196.400000
2% milk (P_3)	244.000000	244.000000	244.000000	244.000000	183.000000	183.000000	152.500000
1% milk (P_4)	244.000000	244.000000	183.000000	152.500000	183.000000	183.000000	218.583333
Skim milk (P_5)	245.000000	245.000000	244.800000	229.690000	196.000000	183.750000	183.750000
Bacon or sausage (P_6)	26.000000	25.000000	24.000000	24.000000	18.000000	19.500000	16.000000
Hotdogs (P_7)	57.000000	57.000000	57.000000	114.000000	57.000000	57.000000	57.000000
Whole grain bread (P_8)	50.000000	48.000000	47.500000	45.000000	45.000000	42.400000	34.000000
100% fruit juice (P_9)	280.125000	249.000000	248.800000	233.250000	189.755000	186.600000	186.700000
Fruit (P_{10})	118.000000	118.000000	118.000000	118.000000	118.000000	112.427143	109.000000
Salad dressing (P_{11})	17.140000	20.626667	23.020000	21.873333	22.035000	18.335000	10.210000
Fried potatoes (P_{12})	79.500000	70.000000	70.000000	70.000000	66.000000	70.000000	64.000000

THE MULTIFACTOR SCREENER: OPEN

Food Group	Age Group						
	18-27	**28-37**	**38-47**	**48-57**	**58-67**	**68-77**	**78-99**
Other white potatoes (P_{13})	122.000000	127.000000	116.000000	122.000000	105.000000	105.000000	105.000000
Dried beans (P_{14})	126.500000	89.000000	126.500000	126.500000	126.500000	126.500000	173.000000
Other vegetables (P_{15})	53.750000	61.625000	61.500000	61.532500	63.165000	67.142857	71.333333
Pasta (P_{16})	217.500000	217.500000	182.525000	185.000000	165.000000	160.000000	175.000000
Nuts (P_{17})	18.000000	32.000000	20.655000	21.265000	18.250000	11.250000	25.500000
Chips (P_{18})	28.000000	24.333333	27.000000	26.000000	20.000000	18.000000	14.000000

Table 2- 4 Median Portion Size (Pk) in Pyramid Servings * per Mention by Gender and Age for Fruits and Vegetables Analyses

Food Group	Age Group						
	18-27	28-37	38-47	48-57	58-67	68-77	78-99
Men							
100% fruit juice (P_1)	2.000000	1.667500	1.335000	1.335000	1.334000	1.001000	1.001000
Fruit (P_2)	1.301000	1.301000	1.229571	1.227333	1.168000	1.168000	1.052333
Salad (P_3)	0.545000	0.708000	0.754500	0.750000	0.833500	0.750000	0.822500
Fried potatoes (P_4)	2.000000	2.000000	1.773000	1.710000	1.400000	1.250000	1.250000
Other white potatoes (P_5)	2.000000	2.000000	1.999000	1.999000	1.914000	1.544000	1.508000
Dried beans (P_6)	1.374000	1.047000	1.065000	1.227000	1.000000	1.000000	1.114000
Other vegetables (P_7)	0.750000	0.906000	0.974500	1.000000	1.000000	0.880000	0.833333
Women							
100% fruit juice (P_1)	1.500500	1.334000	1.334000	1.251250	1.019500	1.000500	1.000500
Fruit (P_2)	1.168000	1.168000	1.168000	1.168000	1.150500	1.083833	1.000000
Salad (P_3)	0.613500	0.572500	0.833333	1.000000	0.795500	0.625000	0.750000
Fried potatoes (P_4)	1.481000	1.365500	1.272000	1.400000	1.000000	1.026000	1.000000
Other white potatoes (P_5)	1.544000	1.544000	1.528000	1.544000	1.499000	1.516000	1.272000
Dried beans (P_6)	0.964000	0.684000	0.800000	0.687000	0.822000	0.807000	1.000000
Other vegetables (P_7)	0.702200	0.779333	0.792500	0.788500	0.774000	0.833000	0.856750

* Defined by [Dietary Guidelines Food Guide Pyramid (1992-2004).]

Table 2- 5 Median Portion Size (Pk) in Cup Servings ** per Mention by Gender and Age for Fruits and Vegetables Analyses

Food Group	Age Group						
	18-27	28-37	38-47	48-57	58-67	68-77	78-99
Men							
100% fruit juice (P_1)	1.499160	1.250580	1.000980	1.000980	1.000176	0.750735	0.750735
Fruit (P_2)	0.999580	0.933450	0.867300	0.867300	0.867300	0.774916	0.657060
Salad (P_3)	0.272700	0.353970	0.377235	0.374963	0.416640	0.375000	0.411323
Fried potatoes (P_4)	0.721125	0.727700	0.641000	0.641000	0.548055	0.480750	0.499980
Other potatoes (P_5)	1.000400	1.140030	0.999600	0.999600	0.999490	0.833175	0.754400
Dried beans (P_6)	0.717550	0.551540	0.566720	0.612360	0.500250	0.502285	0.575360
Other Vegetables (P_7)	0.387675	0.473920	0.499840	0.500240	0.499905	0.460585	0.416899
Women							
100% fruit juice (P_1)	1.124370	1.000960	1.000176	0.938130	0.764776	0.750728	0.750434
Fruit (P_2)	0.749235	0.867300	0.844838	0.789970	0.742350	0.712640	0.620475
Salad (P_3)	0.306788	0.286335	0.416625	0.499950	0.397688	0.312469	0.374963
Fried potatoes (P_4)	0.509595	0.455110	0.448700	0.448700	0.394856	0.444260	0.444260
Other white potatoes (P_5)	0.782020	0.876945	0.771260	0.771260	0.749700	0.771260	0.644235
Dried beans (P_6)	0.492150	0.341550	0.430530	0.345763	0.430685	0.430530	0.500400
Other Vegetables (P_7)	0.364468	0.395882	0.404303	0.408330	0.416913	0.436560	0.452214

** Defined by Dietary Guidelines for Americans, 2005

3. For **percentage energy from fat** and **fiber (grams)**, *estimation of b_k, $k = 0, \ldots, 19$*: the values for each parameter, for each gender, are in the following table:

Table 2- 6 Estimated Regression Coefficients for Foods as Predictors of Percentage Energy from Fat and Grams of Fiber, by Gender

Parameter	Men		Women	
	Percentage Energy from Fat	Fiber (cube root)	Percentage Energy from Fat	Fiber (cube root)
Intercept (b_0)	31.93268	2.08423	31.36357	1.89847
Cold cereals (b_1)	-0.02672	0.00209	-0.05797	0.00389
Whole milk (b_2)	0.00653	0.00013	0.00842	0.00009
2% milk (b_3)	0.00215	0.00013	0.00272	0.00011
1% milk (b_4)	-0.00149	0.00022	-0.00196	0.00024
Skim milk (b_5)	-0.00841	0.00028	-0.00867	0.00034
Bacon or sausage (b_6)	0.13831	-0.00139	0.23128	-0.00201
Hotdogs (b_7)	0.04078	0	0.10160	-0.00141
Whole grain bread (b_8)	0	0.00283	0	0.00337
100% fruit juice (b_9)	-0.00533	0.00019	-0.01011	0.00025
Fruit (b_{10})	-0.00932	0.00103	-0.01201	0.00105
Salad dressing (b_{11})	0.15036	0	0.23974	0
Fried potatoes (b_{12})	0.02734	0.00160	0.04272	0.00156
Other white potatoes (b_{13})	0.00580	0.00071	0.00618	0.00066
Dried beans (b_{14})	-0.00526	0.00275	-0.00608	0.00380
Other vegetables (b_{15})	0	0.00084	0	0.00093
Pasta (b_{16})	-0.00504	0.00075	-0.00540	0.00082

Parameter	Men		Women	
	Percentage Energy from Fat	Fiber (cube root)	Percentage Energy from Fat	Fiber (cube root)
Nuts (b_{17})	0.12454	0.00546	0.26018	0.00603
Chips (b_{18})	0.05376	0.00528	0.13144	0.00456

4. For Pyramid servings of **total fruits and vegetables**, *estimation of b_0 and b_1*:

The model is: $E([\text{Fruits and Veg}]^{1/2}) = b_0 + b_1 ([N_{FG1}P_1 + N_{FG2}P_2 + \ldots + N_{FG7}P_7]^{1/2})$

For Pyramid servings of **fruits and vegetables without French fries**, *estimation of b_0 and b_1*:

The model is: $E([\text{F\&V not FF}]^{1/2}) = b_0 + b_1 ([N_{FG1}P_1 + N_{FG2}P_2 + \ldots + N_{FG7}P_7]^{1/2})$

For Pyramid servings of fruits and vegetables, including and excluding French fries, for each gender, the estimates of the parameters are:

Table 2- 7 Estimated Regression Coefficients for Sum of Foods Predicting Pyramid Servings of Total Fruits and Vegetables and Fruits and Vegetables Excluding French Fries, by Gender

Parameter	Men	Women
Summary variable with French fries		
Intercept (b_0)	0.906793	0.819559
b_1	0.758560	0.730865
Summary variable excluding French fries		
Intercept (b_0)	0.940772	0.816265
b_1	0.739056	0.730219

For cups of fruits and vegetables (2005 MyPyramid definition), including and excluding French fries, for each gender, the estimates of the parameters are:

Table 2- 8 Estimated Regression Coefficients for Sum of Foods Predicting Cups of Total Fruits and Vegetables and Fruits and Vegetables Excluding French Fries, by Gender

Parameter	Men	Women
Summary Variable with French fries		
Intercept (b_0)	0.666228	0.611844
b_1	0.770652	0.733890
Summary Variable excluding French fries		
Intercept (b_0)	0.706696	0.616033
b_1	0.742255	0.727761

3.Validation Results

The NCI's RFAB have assessed the validity of the Multifactor Screener in several studies: NCI's OPEN Study, the Eating at America's Table Study (EATS), and the joint NIH-AARP Diet and Health Study. In all studies, multiple 24-hour recalls in conjunction with a measurement error model were used to assess validity. In general, the validation results reflect the Multifactor Screener's hierarchical design -- fruit and vegetable intake was estimated best by the screener, followed by percentage energy from fat, and lastly grams of fiber.

In the OPEN Study, estimates of median intake were:

- **Pyramid Servings of Fruits and Vegetables:**
  - Men: recalls - 6.3; screener - 5.3;
  - Women: recalls - 5.4; screener - 4.7.
- **Percentage Energy from Fat:**
  - Men: recalls - 31.8; screener - 32.0;
  - Women: recalls - 32.0; screener - 30.5.
- **Fiber:**
  - Men: recalls - 21.2; screener - 18.3;
  - Women: recalls - 16.5; screener - 14.1.

These validation results suggest that dietary exposure estimates computed from the Multifactor Screener may be useful to compare subgroup means, especially for populations consuming mainstream diets. The estimates may be less useful for populations with more ethnic diets, including Asian and possibly Latino populations.

At the individual level, correlations between the screener and estimated true intake ranged from 0.54 (fiber for men) to 0.76 (Pyramid servings of fruits and vegetables for women); about 25 to 50 percent of the variability in the true intake will be captured by the screener questions. Thus, although significant error may be associated with these estimates of diet, we believe the exposure estimates still substantially reflect what individuals are actually consuming.

Validation results are reported in detail in Thompson et al., [15].

APPENDIX

This content was originally posted online with information about the use of the Multifactor Screener in the 2000 National Health Interview Survey Cancer Control Supplement (NHIS 2000). The content on that website has been archived, including the section reproduced in this appendix, and can now be viewed in PDF format in its original context on the NCI's Short Dietary Assessment Instruments website.

A.Uses of Screener Estimates

Introduction

Dietary intake estimates derived from the Multifactor Screener are rough estimates of usual intake of fruits and vegetables, fiber, calcium, servings of dairy, and added sugar. These estimates are not as accurate as those from more detailed methods (e.g., 24-hour recalls). However, Validation Results suggests that the estimates may be useful to characterize a population's median intakes, to discriminate among individuals or populations with regard to higher vs. lower intakes, to track dietary changes in individuals or populations over time, and to allow examination of interrelationships between diet and other variables. In addition, diet estimates from the Cancer Control Supplement (CCS) could be used as benchmark national data for smaller surveys, for example, in a particular state.

Variance-Adjustment Factor

What is the variance adjustment estimate and why do we need it?

Data from the Multifactor Screener are individuals' reports about their intake and, like all self-reports, contain some error. The algorithms we use to estimate servings of fruits and vegetables, grams of fiber, mg of calcium, servings of dairy, and teaspoons of added sugar calibrate the data to 24-hour recalls. The screener estimate of intake represents what we expect the person would have reported on his 24-hour recall, given what he reported on the individual items in the screener. As a result, the mean of the screener estimate of intake should equal the mean of the 24-hour recall estimate of intake in the population. (It would also equal the mean of true intake in the population if the 24-hour recalls were unbiased. However, there are many studies suggesting that recalls underestimate individuals' true intakes).

When describing a population's distribution of dietary intakes, the parameters needed are an estimate of central tendency (i.e. mean or median) and an estimate of spread (variance). The variance of the screener, however, is expected to be smaller than the variance of true intake, since the screener prediction formula estimates the conditional expectation of true intake given the screener responses, and in general the variance of a conditional expectation of a variable is smaller than the variance of itself. As a result, the screener estimates of intake cannot be used to estimate quantiles (other than median) or prevalence estimates of true intake without an adjustment. Procedures have been developed to estimate the variance of true intake using data from 24-hour recalls, by taking into consideration within person variability [16, 17]. We extended these procedures to allow estimation of the variance of true intake using data from the screener. The resulting variance adjustment factor adjusts the screener variance to approximate the variance of true intake in the population.

How did we estimate the variance adjustment factors?

We have estimated the adjustment factors in the various external validation datasets available to us. The results indicate that the adjustment factors differ by gender and dietary variable. Under the assumption that the variance adjustment factors appropriate to National Health Interview Study (NHIS) are similar to those in Observing Protein and Energy Nutrition Study (OPEN) [2], the variance-adjusted screener estimate of intake should have variance closer to the estimated variance of true intake that would have been obtained from repeat 24-hour recalls. For Pyramid servings of fruits and vegetables, the variance adjustment factors in OPEN and Eating at America's Table Study (EATS) [18] are quite similar, which gives us some indication that these factors might be relatively stable from population to population.

Table A_ 1 Variance Adjustment Factors for the NHIS Multifactor Screener, from the OPEN Study

Nutrient	Gender	Variance Adjustment Factor
Total Fruit & Vegetable Intake (Pyramid Servings)	Male	1.3
	Female	1.1
Fruit & Vegetable Intake (excluding fried potatoes) (Pyramid Servings)	Male	1.3
	Female	1.2
Percentage Calories from Fat	Male	1.5
	Female	1.3
Fiber Intake (grams)	Male	1.2
	Female	1.2

How do you use the variance adjustment estimates?

To estimate quantile values or prevalence estimates for an exposure, you should first adjust the screener so that it has approximately the same variance as true intake.

Adjust the screener estimate of intake by:

- multiplying intake by an adjustment factor (an estimate of the ratio of the standard deviation of true intake to the standard deviation of screener intake); and
- adding a constant so that the overall mean is unchanged.

The formula for the variance-adjusted screener is:

> variance-adjusted screener = (variance adjustment factor)*(unadjusted screener - $\text{mean}_{\text{unadj scr.}}$) + $\text{mean}_{\text{unadj scr.}}$

This procedure is performed on the normally distributed version of the variable (i.e., Pyramid servings of fruits and vegetables is square-rooted, percentage energy from fat is untransformed, and fiber is cube rooted). For fruits and vegetables and fiber, the results can then be squared or cubed, respectively, to obtain estimates in the original units.

The variance adjustment procedure is used to estimate prevalence of obtaining recommended intakes for the 2000 NHIS in Thompson et al., [19]

When do you use variance adjustment estimates?

The appropriate use of the screener information depends on the analytical objective. Following is a characterization of suggested procedures for various analytical objectives.

Table A_ 2 Suggested procedures for various analytical objectives

Analytical Objective	Procedure
Estimate mean or median intake in the population or within subpopulations.	Use the unadjusted screener estimate of intake.
Estimate quantiles (other than median) of the distribution of intake in the population; estimate prevalence of attaining certain levels of dietary intake.	Use the variance-adjusted screener estimate.
Classify individuals into exposure categories (e.g., meeting recommended intake vs. not meeting recommended intake) for later use in a regression model.	Use the variance-adjusted screener estimates to determine appropriate classification into categories.

Analytical Objective	Procedure
Use the screener estimate as a continuous covariate in a multivariate regression model.	Use the unadjusted screener estimate.
Use the screener estimate as a response (dependent) variable.	Use the variance-adjusted screener estimate.

Attenuation of Regression Parameters Using Screener Estimates

When the screener estimate of dietary intake is used as a continuous covariate in a multivariate regression, the estimated regression coefficient will typically be attenuated (biased toward zero) due to measurement error in the screener. The "attenuation factor" [20] can be estimated in a calibration study and used to deattenuate the estimated regression coefficient (by dividing the estimated regression coefficient by the attenuation factor).

We estimated attenuation factors in the OPEN study (see below). If you use these factors to deattenuate estimated regression coefficients, note that the data come from a relatively small study that consists of a fairly homogeneous population (primarily white, well-educated individuals).

Table A_ 3 Attenuation factors for screener predicted intake: OPEN

Gender	Square-Root Fruit & Veg (Pyramid Servings)	Square-Root Fruit & Veg (excluding French Fries) (Pyramid Servings)	Percentage Energy From Fat	Cube-Root Fiber (grams)
Men	0.75	0.79	0.96	0.70
Women	0.81	0.87	0.88	0.69

If you categorize the screener values into quantiles and use the resulting categorical variable in a linear or logistic regression, the bias (due to misclassification) is more complicated because the categorization can lead to differential misclassification in the screener [21]. Although methods may be available to correct for this [22, 23], it is not simple, nor are we comfortable suggesting how to do it at this time.

Even though the estimated regression coefficients are biased (due to measurement error in the screener or misclassification in the categorized screener), tests of whether the regression coefficient is different from zero are still valid. For example, if one used the

SUDAAN REGRESS procedure with fruit and vegetable intake (estimated by the screener) as a covariate in the model, one could use the Wald F statistic provided by SUDAAN to test whether the regression coefficients were statistically significantly different from zero. This assumes that there is only one covariate in the model measured with error; when there are multiple covariates measured with error, the Wald F test that a single regression coefficient is zero may not be valid, although the test that the regression coefficients for all covariates measured with error are zero is still valid.

4. References

1. Subar, AF, et al. Checking for completeness of 24-h urine collection using para-amino benzoic acid not necessary in the Observing Protein and Energy Nutrition study. *Eur J Clin Nutr,* 2013. **67**(8): p. 863-7.
2. Subar, AF, et al. Using intake biomarkers to evaluate the extent of dietary misreporting in a large sample of adults: the OPEN study. *Am J Epidemiol*, 2003. **158**(1): p. 1-13.
3. Cook, CM, et al. Relation between holiday weight gain and total energy expenditure among 40- to 69-y-old men and women (OPEN study). *Am J Clin Nutr*, 2012. **95**(3): p. 726-31.
4. Kirkpatrick, SI, et al. Reactivity and its association with body mass index across days on food checklists. *J Acad Nutr Diet*, 2012. **112**(1): p. 110-8.
5. Bradley, D.P., et al. Effect of smoking status on total energy expenditure. *Nutr Metab (Lond)*, 2010. **7**: p. 81.
6. Millen, AE, et al. Differences between food group reports of low-energy reporters and non-low-energy reporters on a food frequency questionnaire. *J Am Diet Assoc*, 2009. **109**(7): p. 1194-203.
7. Yanetz, R, et al. Using biomarker data to adjust estimates of the distribution of usual intakes for misreporting: application to energy intake in the US population. *J Am Diet Assoc*, 2008. **108**(3): p. 455-64; discussion 464.
8. Tooze, JA, et al. Total daily energy expenditure among middle-aged men and women: the OPEN Study. *Am J Clin Nutr*, 2007. **86**(2): p. 382-7.
9. Tooze, JA, et al. Psychosocial predictors of energy underreporting in a large doubly labeled water study. *Am J Clin Nutr*, 2004. **79**(5): p. 795-804.
10. Lissner, L, et al. OPEN about obesity: recovery biomarkers, dietary reporting errors and BMI. *Int J Obes (Lond)*, 2007. **31**(6): p. 956-61.
11. Freedman, L.S., et al. Adjustments to improve the estimation of usual dietary intake distributions in the population. *J Nutr*, 2004. **134**(7): p. 1836-43.
12. Schatzkin, A, et al. A comparison of a food frequency questionnaire with a 24-hour recall for use in an epidemiological cohort study: results from the biomarker-based Observing Protein and Energy Nutrition (OPEN) study. *Int J Epidemiol*, 2003. **32**(6): p. 1054-62.
13. Kipnis, V, et al. Structure of dietary measurement error: results of the OPEN biomarker study. *Am J Epidemiol*, 2003. **158**(1): p. 14-21; discussion 22-6.

14. Kipnis, V, et al. Kipnis et al. Respond to "OPEN Questions. *Am J Epidemiol*, 2003. **158**(1): p. 25-26.
15. Thompson, FE, et al. Performance of a short tool to assess dietary intakes of fruits and vegetables, percentage energy from fat and fibre. *Public Health Nutr*, 2004. **7**(8): p. 1097-105.
16. National Research Council. *Nutrient Adequacy: Assessment Using Food Consumption Surveys*. Washington, DC: National Academy Press, 1986. Available from: https://www.nap.edu/catalog/618/nutrient-adequacy-assessment-using-food-consumption-surveys.
17. Institute of Medicine. *Dietary Reference Intakes: Applications in Dietary Assessment*. Washington, DC: National Academy Press, 2000. Available from: https://www.nap.edu/catalog/9956/dietary-reference-intakes-applications-in-dietary-assessment.
18. Subar, AF, et al. Comparative validation of the Block, Willett, and National Cancer Institute food frequency questionnaires : the Eating at America's Table Study. *Am J Epidemiol*, 2001. **154**(12): p. 1089-99.
19. Thompson, FE, et al. Dietary intake estimates in the National Health Interview Survey, 2000: methodology, results, and interpretation. *J Am Diet Assoc*, 2005. **105**(3): p. 352-63; quiz 487.
20. Rosner B, Willett WC, Spiegelman D. Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. *Stat Med*, 1989. **8**(9): p. 1051-69; discussion 1071-3.
21. Flegal KM, Keyl PM, Nieto FJ. Differential misclassification arising from nondifferential errors in exposure measurement. *Am J Epidemiol*, 1991. **134**(10): p. 1233-44.
22. Flegal KM, Brownie C, Haas JD. The effects of exposure misclassification on estimates of relative risk. *Am J Epidemiol*, 1986. **123**(4): p. 736-51.
23. Morrissey MJ, Spiegelman D. Matrix methods for estimating odds ratios with misclassified exposure data: extensions and comparisons. *Biometrics*, 1999. **55**(2): p. 338-44.