The Fruit & Vegetable Screener in the 2000 California Health Interview Survey

(CHIS 2000)

November 2019

THE FRUIT & VEGETABLE SCREENER: CHIS 2000

This information was originally published online on the Epidemiology and Genomic Research Program, National Cancer Institute's website as a reference for the Fruit & Vegetable Screener in the 2000 California Health Interview Survey (CHIS 2000). The information contained in this document informs analyses of data from this screener. This information is archived and provided for reference purposes only.

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

Suggested citation for information contained in this report: The Fruit & Vegetable Screener in the 2000 California Health Interview Survey. Epidemiology and Genomics Research Program. National Cancer Institute. https://epi.grants.cancer.gov/diet/screeners/files. Updated November 20, 2019.

Acknowledgements

he 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 and Jennifer Lerman 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.

CONTENTS

TABLES

THE FRUIT & VEGETABLE SCREENER: CHIS 2000

1. Background

The Fruit and Vegetable Screener used in the 2000 California Health Interview Survey (CHIS) was derived from the Multifactor Screener in the 2000 National Health Interview Survey (NHIS) Cancer Control Supplement (CCS). The CHIS screener asks respondents for information about how frequently they consume foods in eight categories. No portion size questions are asked.

This screener does not attempt to assess total diet. The questions asked allow researchers to gather information about Pyramid servings of fruits and vegetables, using the 1992 definitions of servings from the Food Guide Pyramid.

You can view or print the CHIS 2000 Fruit and Vegetable Screener from the National Cancer Institute's (NCI) Register of Validated Short Dietary Assessment Instruments.

In CHIS 2000, we applied rules for excluding extreme data responses, described in Definition of Acceptable Dietary Data. The process of scoring the individual response data is described in Scoring Procedures. A description and guidelines for the appropriate uses of the screener-estimated dietary intakes is found in Uses of Screener Estimates. Validation data for the CHIS 2000 screener are presented in Validation Results. Finally, the various fruit and vegetable computed variables are found in Computed Variables.

NOTE: The dietary variables on the CHIS dataset are in their natural units. For analyses, however, they must be transformed, first, to approximate normal distributions. For all variable versions of servings of fruits and vegetables, use the square root transformation. After analyses, the result variables can be back-transformed for easier interpretation.

Different dietary screeners were used in the 2005 and 2009 CHIS, and these are also available in the Register of Validated Short Dietary Assessment Instruments.

2. Definition of Acceptable Dietary Data Values

Data collected on the California Health Interview Survey (CHIS) Fruit and Vegetable Screener are coded as frequency and time unit - times per day, week, or month. The data contain some values that are very unlikely. We used USDA's 1994-96 Continuing Survey of Food Intakes of Individuals (CSFII 94-96) data on reported intakes over two days of 24-hour recall to make judgments about reasonable frequencies of consumption that were reported on a per day basis.

Maximum daily average frequencies (averaged for each individual across his two days of report) in the CSFII ranged from 1.5 times to 12 times per day for the relevant food groups. We accepted frequency values reported in the CHIS that were reported on a per day basis up to the maximum average values (rounded to the next whole number).

Table 2- 1 Maximum daily average frequencies in CSFII

Food Group	Maximum Daily Acceptable Value
Fruit	12
Salad	5
Fried potatoes	3
Other white potatoes	3
Dried beans	3
Other vegetables	9
100% fruit juice	4
Tomato sauce/salsa	4

In addition, we applied judgment to determine the acceptability of frequency reports for the weekly and monthly time periods (see below). For example, a report of 25 times may be most logically associated with a month or year time period, but not so logically associated with a week time period. We applied this judgment to all foods.

Table 2- 2 Acceptability of frequency reports for the weekly and monthly time periods

Time Period	Acceptable Frequency
Week	14 / week: acceptable > 14: assign a missing value
Month	60 / month: acceptable > 60: assign a missing value

Depending on the intent of the analysis, the researcher could exclude a person with a missing value for any of the 8 foods, or only with missing values on foods needed to estimate a particular Fruit and Vegetable intake variable. In our analyses of CHIS data, we excluded individuals only for the Fruit and Vegetable variables in which they had missing food-level data.

3. Scoring Procedures

How Analytical Scoring Procedures Were Developed

Scoring procedures were developed to convert the individual respondent's screener responses to estimates of individual dietary intake for servings of fruits and vegetables using USDA's 1994-96 Continuing Survey of Food Intakes of Individuals (CSFII 94-96) dietary recall data.

For servings of fruits and vegetables:

E([Fruits and Veg]1/2) = b0+ b1([NFG1P1 + NFG2P2 + … + NFG8P8]1/2)

Servings of fruits and vegetables was square-root-transformed to approximate normality; NFGk is the usual number of times per day an individual consumed food group k; Pk is the median portion size of group k; and k indexes the 8 fruit and vegetable food groups. We calculated weighted least-squares estimates of the regression coefficients b0 and b1 on the adults (aged 18 and above) in the CSFII 94-96 sample, stratifying by gender and excluding extreme exposure values.

Scoring Procedures

After exclusion of extreme and missing values, we performed the following steps with the CHIS dietary data to estimate the individual's intake of servings of fruits and vegetables.

1. **Estimation of N_{FGk}**: All reported frequencies were standardized to a common unit of time by converting them to daily frequencies.

Table 3- 1 Daily frequencies

Time Period Reported	N_{FGk}: Daily Frequency
Day	As reported
Week	Reported frequency divided by 7
Month	Reported frequency divided by 30

2. **Estimation of P_k**: The median age- and gender-specific portion sizes for each food were estimated from CSFII 94-96. The units were in Pyramid servings (Table 3-2).

A Pyramid serving is defined by the U.S. Department of Agriculture in the Dietary Guidelines Food Guide Pyramid as:

- vegetables: 1 cup raw leafy, 1/2 cup of other vegetables, or 3/4 cup vegetable juice; and
- fruit: 1 whole fruit, 1/2 cup of cut-up fruit, or 3/4 cup fruit juice.

Table 3- 2 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
Tomato sauce/salsa (P_8)	0.626000	0.587000	0.579000	0.607000	0.533600	0.606667	0.405000

Food Group	Age Group						
	18-27	28-37	38-47	48-57	58-67	68-77	78-99
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
Tomato sauce/salsa (P_8)	0.410000	0.400000	0.402000	0.396000	0.477500	0.356000	0.252000

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

The model is: $E([\text{Dietary Factor}]^{1/2}) = b_0 + b_1 ([N_{FG1}P_1 + N_{FG2}P_2 + \ldots + N_{FG8}P_8]^{1/2})$

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

Table 3- 3 Estimated Regression Coefficients for Sum of Foods Predicting Servings of Different Versions of Daily Servings of Fruits and Vegetables, by Gender

Parameter	Men	Women
Summary Variable with French fries		
Intercept (b_0)	0.755105	0.711444
b_1	0.815466	0.775801
Summary Variable excluding French fries		
Intercept (b_0)	0.771919	0.694500
b_1	0.806806	0.783053
Summary Variable excluding French fries and beans		
Intercept (b_0)	0.773746	0.691176
b_1	0.796528	0.776859
Summary Variable excluding beans		
Intercept (b_0)	0.729300	0.703858
b_1	0.820729	0.772394

4. Uses of Screener Estimates in CHIS

Introduction

Dietary intake estimates from the California Health Interview Survey (CHIS) Fruit and Vegetable Screener are rough estimates of usual intake of fruits and vegetables. They are not as accurate as 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 CHIS could be used to augment national data using similar methods.

Variance-Adjustment Factor

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

Data from the CHIS Fruit and Vegetable 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 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 X is smaller than the variance of X 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 [1, 2]. 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 an external validation dataset available to us. The results indicate that the adjustment factors differ by gender: 1.2 for men and 1.1 for women. Under the assumption that the variance adjustment factors appropriate to the California Health Interview Survey are similar to those in the Eating in America's Table Study (EATS)[3], the variance-adjusted screener estimate of intake should have variance closer to the estimated variance of true intake than would have been obtained from repeat 24-hour recalls. For a slightly different fruit and vegetable screener (7 rather than 8 items) validated in the Observing Protein and Energy Nutrition (OPEN) Study, the variance adjustment factors are quite similar, which gives us some indication that these factors might be relatively stable from population to population. The OPEN Study screener is available in NCI's Register of Validated Short Dietary Assessment Instruments.

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:

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

This procedure is performed on the normally distributed version of the variable (i.e. Pyramid servings of fruits and vegetables is square-rooted). The results can then be squared, to obtain estimates in the original units.

A similar variance adjustment procedure is used to estimate prevalence of obtaining recommended intakes for the 2000 NHIS in [4].

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 4- 1 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.
Use the screener estimate as a continuous covariate in a multivariate regression model.	Use the unadjusted 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" [5], 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 EATS data (see below). If you use these factors to deattenuate estimated regression coefficients, note that the data come from a single study.

Table 4- 2 Attenuation factors for screener-predicted intake: EATS

Gender	Square-Root Fruit & Veg	Square-Root Fruit & Veg (excluding French Fries)	Square-Root Fruit & Veg (excluding French Fries and beans)
Men	0.80	0.88	0.76
Women	0.51	0.53	0.51

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 [6]. Although methods may be available to correct for this [7, 8] , 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 are 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.

5. Validation Results

Staff in NCI's Risk Factor Assessment Branch (RFAB) have assessed the validity of the CHIS Fruit and Vegetable Screener in the Eating at America's Table Study (EATS) [3]. In this study, multiple 24-hour recalls in conjunction with a measurement error model were used to assess validity. The screeners used in the EATS included additional foods and reported portion sizes. For comparison of the CHIS screener, we have used the similar questions in EATS and the scoring algorithms developed specifically for CHIS.

In EATS, estimates of median intake of Servings of Fruits and Vegetables were: Men: recalls - 5.8; screener - 5.5; Women: recalls - 4.2; screener - 4.5. These validation results suggest that dietary exposure estimates computed for the CHIS 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 were 0.68 for men and 0.49 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 for the Multifactor Screener, which includes a similar fruit and vegetable component, are previously reported in detail [9]. The Multifactor Screener is available in NCI's Register of Validated Short Dietary Assessment Instruments

National estimates based on the 2000 NHIS Multifactor screener are previously presented and compared with other national data [4].

6. Computed Variables

The computed diet variables for the 2001 CHIS Fruit and Vegetable Screener are available for download. There are two versions of the diet datasets available, each based on the 2001 CHIS Adult data: version A is based on the 02/15/2005 release; while version B is based on the 03/05/2004 release. Each version of the diet datasets is provided in two file formats -- SAS transport and comma separated values (CSV). Please see the CHIS 2001 Revised Sample Weights (PDF) for an overview of the differences between the two releases of the source data. The files include the following variables:

- Pyramid servings of fruits and vegetables
- Adjusted Pyramid servings of fruits and vegetables
- Pyramid servings of fruits and vegetables excluding French fries
- Adjusted Pyramid servings of fruits and vegetables excluding French fries
- Pyramid servings of fruits and vegetables excluding French fries and beans
- Adjusted Pyramid servings of fruits and vegetables excluding French fries and beans

These datasets, data dictionary, and SAS program for this screener are available on the NCI's Short Dietary Assessment Instruments website.

The datasets are sorted in ascending order by the ID variable PUF_ID. All numeric variables have been rounded to the nearest .000001.

Note: The datasets were updated on 08/29/2007. If you downloaded previous versions of the files, use the most recent updates provided below instead.

THE FRUIT & VEGETABLE SCREENER: CHIS 2000

A. Diet Datasets Based on the 02/15/2005 Release of 2001 CHIS Adult Data

These datasets were weighted using California Department of Finance population projections; includes Random Digit Dial (RDD) sample plus list oversample for Korean and Vietnamese.

- Comma Separated Values File (fv.2007_08_29a.csv.zip) - This zip file contains the comma separated values file which includes 9 variables, 56,270 records plus one record for variable names.

- SAS Transport File (fv.2007_08_29a.v8x.zip) - SAS transport file with 56,270 records and 9 variables. To access the SAS dataset, unzip the SAS transport file, then use proc cimport.

  For example:
  ```
  proc cimport file='fv.2007_08_29a.v8x' data=fv'
  ```

- **Other Documents**

  The following documents may also be of use:

  - Content tables of the SAS dataset (contents.fv.2007_08_29a.pdf)
  - SAS program that created the dataset (create.fv.2007_08_29a.sas)

B. Diet Datasets Based on the 03/05/2004 Release of 2001 CHIS Adult RDD Data

These datasets were weighted using US Census 2000 populations.

- Comma Separated Values File (fv.2007_08_29b.csv.zip) - This zip file contains the comma separated values file which includes 9 variables, 55,428 records plus one record for variable names.

- SAS Transport File (fv.2007_08_29b.v8x.zip) – The SAS transport file contains 55,428 records and 9 variables. To access the SAS dataset, unzip the SAS transport file, then use proc cimport.

  For example:
  ```
  proc cimport file='fv.2007_08_29b.v8x' data=fv'
  ```

- **Other Documents**

  The following documents may also be of use:

  - Content tables of the SAS dataset (contents.fv.2007_08_29b.pdf)
  - SAS program that created the dataset (create.fv.2007_08_29b.sas)

7. References

1. 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.
2. 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.
3. 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): 1089-99.
4. 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): 352-63.
5. 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): 1051-69; discussion 1071-3.
6. Flegal KM, Keyl PM, Nieto FJ. Differential misclassification arising from nondifferential errors in exposure measurement. *Am J Epidemiol*, 1991. 134(10): 1233-44.
7. Flegal KM, Brownie C, Haas JD. The effects of exposure misclassification on estimates of relative risk. *Am J Epidemiol*, 1986. 123(4): 736-51.
8. Morrissey MJ, Spiegelman D. Matrix methods for estimating odds ratios with misclassified exposure data: extensions and comparisons. *Biometrics*, 1999. 55(2): 338-44.
9. 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): 1097-105.