Appendix C
Statistical and Reporting Guidelines

This report presents population means and proportions, standard errors of estimates, and percentiles of dietary intake distributions. Sample weights were used to account for sample design and nonresponse. Information about the NHANES-III survey design was used in estimating variances and testing for statistical significance.

Several software packages were used to produce the tabulations:

- ***C-SIDE: Software for Intake Distribution Estimation (Version 1.0)***—used to estimate means, percentiles, and standard errors for nutrient intake tables.

- ***SUDAAN (Version 7.5)***—used to calculate means, standard errors, and tests of statistical significance for non-nutrient tables, using the DESCRIPT procedure.

- ***SAS (Version 8.2)***—used to read the NHANES-III data files, call SUDAAN procedures, process SUDAAN output, and write SUDAAN results to ASCII files.

- ***TPL (Table Producing Language)***—this software produced all data tables in appendix D.

General Procedures

NHANES-III sample weights account for the fact that each sample person does not have an equal probability of selection into the sample. NHANES-III provides sample weights for three samples: the interviewed sample weight (WTPEQX6), the MEC-examined sample weight (WTPFEX6), and the MEC and home-examined sample weight (WTPFHX6). The sampling weight used for each table in this report was specific to the data item presented in the table, and is indicated by the source of data listed in the table footnote.

Variance is generally underestimated in a complex survey when information about the survey design is not used in variance estimation. For this report, two alternate methods were used to account for the sample design.

- Balance repeated replication (BRR)—this method was specified when using C-SIDE software to obtain estimates for nutrient tables. The BRR method used the 52 replicate weights provided in the NHANES-III data.

- Taylor series linearization—this method is used in SUDAAN procedures. The complex survey design is accounted for by specifying strata and PSU in the "nest" statement of SUDAAN procedures.

Coefficients of variation (CVs) and t-statistics were generated and examined, but are not provided in the tables. CVs were examined to determine the statistical reliability of estimates, as described below in the section on Reporting Guidelines. T-statistics were examined to determine the statistical significance of differences in means and proportions. When examining categorical data, t-statistics were used and the Bonferroni adjustment was applied to adjust for multiplicity of tests.

All tests for statistical significance are tests for differences between two independent samples defined by program participation and/or income-level. In volumes I and II, differences between

program participants and income-eligible nonparticipants are denoted by symbols on values for income-eligible nonparticipants; differences between program participants and higher-income nonparticipants are denoted by symbols on values for higher-income nonparticipants. In volumes III and IV, differences between the lowest-income group and the low-income group are denoted by symbols on values for the low-income group; differences between the lowest-income group and high-income group are denoted by symbols on values for the high-income group.

Differences in means and proportions were tested for statistical significance using α levels of 0.01, 0.05, and 0.001. For categorical data, differences involve multiple non-independent comparisons and were tested using α levels of 0.01, 0.05, and 0.001 adjusted using the Bonferroni method, by dividing α levels by the number of comparisons.

Age Standardization

Tables presented in appendix A include age-adjusted estimates for the total population (i.e., all age groups), calculated using the direct method (Klein, 2001). The age-adjusted estimates were obtained by weighting estimates for each age category by the year 2000 population distribution.

The population distribution used for age-adjustment is from *Monthly Estimates of the United States Population: April 2000*. Age-adjusted estimates were calculated by the SUDAAN software.

Nutrient Analyses

A primary goal for the analysis of dietary intake was to estimate the proportion of individuals whose intake is inadequate. Reference standards used to define adequate intake reflect expectations for usual intake. To apply these standards appropriately, it is necessary to have information about the distribution of intake in the population of interest. The variance of the distribution of observed intake is too large to produce reliable estimates of the prevalence of inadequate intake. This is because the variance of observed intake includes both within-person (day-to-day) and between-person variation. Methods have been established for adjusting observed intake distributions to estimate distributions of usual intake by removing within-person variation (NRC, 1986 and Nusser et al, 1996). These adjustments require two or more days of intake data for at least some subjects.

NHANES-III collected replicate 24-hour recalls on a convenience sample of approximately 5 percent of respondents. The nonrandom nature and small size of the replicate recall sample prohibited its use in estimating usual dietary intake. Instead, we used the Continuing Survey of Food Intake of Individuals (CSFII) 1994-96, to obtain estimates of within-person variation. CSFII is a nationally representative survey that includes two days of dietary intake data for all subjects.

CSFII data were used to estimate variance components for 96 demographic cells defined by age group (8), gender (male, female, both), and program participation or income (3 plus overall).[1] The variance components from CSFII were used to adjust observed intakes collected in the NHANES-III single-day dietary recalls. Estimation for all nutrients was done using *C-SIDE: Software for Intake Distribution Estimation* (Iowa State University, 1996). Because iron requirements for menstruating females are known to be asymmetrical, the adjustments performed by the C-SIDE software (using this "Iowa State Method") were not appropriate.

[1] Age groups correspond to the DRI age groups for volumes I, III, IV. CSFII used to estimate variance components for volume II (WIC participants and nonparticipants) were aggregated by year of age (4) and program participation or income (3 plus overall), but not by gender.

Therefore, distributions of iron intake were adjusted using the full probability approach as described in the IOM report *Dietary Reference Intakes: Applications in Dietary Assessment* (IOM, 2001). CSFII variance components are shown in table C1.

Reporting Guidelines

This report follows the recommendations in the NHANES-III Analytic Guidelines in the appendix titled "Joint Policy on Variance Estimation and Statistical Reporting Standards for NHANES-III and CSFII Reports: HNIS/NCHS Analytic Working Group Recommendations" (NCHS, 1996). The recommendations for presentation of statistical data call for estimates to be flagged if any of the following conditions are met:

1. **Inadequate sample size for normal approximation.** For means and for proportions based on commonly occurring events (where $0.25 < P < 0.75$), an estimate is flagged if it is based on a cell size of less than 30 times a "broadly calculated average design effect."

2. **Large coefficient of variation.** Estimates are flagged if the coefficient of variation (ratio of the standard error to the mean expressed as a percent) is greater than 30.

3. **Inadequate sample size for uncommon or very common events.** For proportions below 0.25 or above 0.75, the criteria for statistical reliability is that the cell size be sufficiently large that the minimum of nP and n(1-P) be greater than or equal to 8 times a broadly calculated average design effect, where n is the cell size and P is the estimated proportion. (I.e., an estimate is flagged when $n < 8 * (\text{avg design effect}) / \min(P,(1-P))$.) The coefficient of variation is not used in these cases.

For each data item, the design effect was calculated for each table cell as the ratio of the complex sampling design variance calculated by SUDAAN, to the simple random sample variance. The average design effect for a data item is the average of estimated design effects across age groups (pooled genders) within a demographic group, where demographic groups correspond to the columns of tables (groups defined by program participation and income).

Table C-1—CSFII variance components for 10 nutrients

	Total Children		Currently Receiving WIC Benefits		Income-eligible Nonparticipants		Higher-income Nonparticipants	
	Sample size	Within-individual variance	Sample size	Within-individual variance	Sample size	Within-individual variance	Sample size	Within-individual variance
Total energy								
1 year old	1,084	0.48900	306	0.45850	260	0.45459	512	0.54907
2 years old	1,107	0.54948	229	0.57212	328	0.51634	541	0.56212
3 years old	1,836	0.60511	348	0.50637	525	0.61904	944	0.64523
4 years old	1,859	0.60330	297	0.56819	644	0.60647	899	0.62004
Total, population adjusted	5,886	0.52208	1,180	0.46875	1,757	0.53582	2,896	0.54546
Vitamin C								
1 year old	1,084	0.59405	306	0.67637	260	0.58360	512	0.56983
2 years old	1,107	0.58606	229	0.62189	328	0.61011	541	0.56039
3 years old	1,836	0.62624	348	0.66059	525	0.67436	944	0.60359
4 years old	1,859	0.68635	297	0.79624	644	0.67490	899	0.66570
Total, population adjusted	5,886	0.62978	1,180	0.69246	1,757	0.64240	2,896	0.60805
Iron								
1 year old	1,084	0.51455	306	0.53153	260	0.50188	512	0.52476
2 years old	1,107	0.57238	229	0.69597	328	0.53098	541	0.55217
3 years old	1,836	0.66739	348	0.64471	525	0.74547	944	0.65861
4 years old	1,859	0.68272	297	0.73055	644	0.69444	899	0.68300
Total, population adjusted	5,886	0.59208	1,180	0.60402	1,757	0.61692	2,896	0.58602
Zinc								
1 year old	1,084	0.57894	306	0.54355	260	0.62037	512	0.59326
2 years old	1,107	0.64335	229	0.62776	328	0.70489	541	0.61708
3 years old	1,836	0.70109	348	0.68978	525	0.74988	944	0.68852
4 years old	1,859	0.71539	297	0.72768	644	0.76057	899	0.71794
Total, population adjusted	5,886	0.64428	1,180	0.60555	1,757	0.70509	2,896	0.63842
Calcium								
1 year old	1,084	0.42758	306	0.45101	260	0.39886	512	0.44847
2 years old	1,107	0.53634	229	0.59965	328	0.59377	541	0.49900
3 years old	1,836	0.57302	348	0.60608	525	0.67964	944	0.50818
4 years old	1,859	0.65281	297	0.70747	644	0.70275	899	0.61069
Total, population adjusted	5,886	0.55554	1,180	0.57100	1,757	0.61130	2,896	0.52236
Total fat								
2 years old	1,107	0.67828	229	0.67482	328	0.71323	541	0.67620
3 years old	1,836	0.73212	348	0.70363	525	0.77072	944	0.73979
4 years old	1,859	0.75730	297	0.82016	644	0.73778	899	0.78028
Total, population adjusted	4,802	0.72671	874	0.72620	1,497	0.73859	2,384	0.73788
Saturated fat								
2 years old	1,107	0.63957	229	0.71438	328	0.65009	541	0.60578
3 years old	1,836	0.70018	348	0.70013	525	0.72726	944	0.68959

Source: Variance components were estimated from two days of 24-hour recalls from the *Continuing Survey of Food Intakes by Individuals (CSFII)* using *C-SIDE: Software for Intake Distribution Estimation.*

Table C-1—CSFII variance components for 10 nutrients — Continued

	Total Children		Currently Receiving WIC Benefits		Income-eligible Nonparticipants		Higher-income Nonparticipants	
	Sample size	Within-individual variance	Sample size	Within-individual variance	Sample size	Within-individual variance	Sample size	Within-individual variance
Saturated fat								
4 years old	1,859	0.75756	297	0.79536	644	0.76249	899	0.75590
Total, population adjusted	4,802	0.70320	874	0.73021	1,497	0.72061	2,384	0.69121
Cholesterol								
2 years old	1,107	0.71330	229	0.74672	328	0.75802	541	0.71383
3 years old	1,836	0.70328	348	0.76400	525	0.66716	944	0.75804
4 years old	1,859	0.70384	297	0.77765	644	0.72813	899	0.70859
Total, population adjusted	4,802	0.70743	874	0.75450	1,497	0.71139	2,384	0.72676
Sodium								
2 years old	1,107	0.61045	229	0.67490	328	0.55385	541	0.61955
3 years old	1,836	0.66354	348	0.60198	525	0.66535	944	0.70296
4 years old	1,859	0.69900	297	0.71507	644	0.70761	899	0.70976
Total, population adjusted	4,802	0.64772	874	0.63806	1,497	0.64339	2,384	0.66511
Fiber								
2 years old	1,107	0.57529	229	0.57157	328	0.59539	541	0.57195
3 years old	1,836	0.68310	348	0.58998	525	0.76087	944	0.68554
4 years old	1,859	0.66612	297	0.66567	644	0.70181	899	0.64422
Total, population adjusted	4,802	0.64212	874	0.60212	1,497	0.68989	2,384	0.63225

Source: Variance components were estimated from two days of 24-hour recalls from the *Continuing Survey of Food Intakes by Individuals (CSFII)* using *C-SIDE: Software for Intake Distribution Estimation.*