Introduction

Dietary surveillance is an integral part of efforts to improve nutritional health. Well-designed surveillance can suggest ways in which dietary guidance and other policy initiatives might be framed by identifying both beneficial and adverse trends in dietary behavior (1). A longstanding goal of dietary surveillance has been to estimate the proportion of the population with intakes above or below a target, such as a recommended level of intake (2,3). This seems simple enough but has actually been a difficult goal to achieve given the limitations of traditional dietary assessment methods (2–9).

Because dietary recommendations are intended to be met over time, it is the long-range or usual dietary intake that is often of interest in surveillance. Yet, national surveys have traditionally used 24-h recalls (24HR)4 to measure diet because of the details they capture. However, individuals do not eat the same thing day to day; thus, 24HR reflect excessive intra-individual variation that leads to an overestimation of the percentage of persons with very low or very high intakes, unless this variation is taken into account (4).

Over the past couple decades, statistical methods have been advanced to correct for intra-individual variation (5). The USDA Agricultural Research Service has applied methods that account for intra-individual variation to estimate distributions of usual nutrient intakes obtained from dietary sources (10). Estimating usual food intakes is somewhat more complex statistically, because foods are less likely than nutrients to be consumed on most days by most persons. The National Cancer Institute (NCI) has developed a method of estimating the distribution of usual intake of foods using information obtained from 2 24HR (6).

Apart from the statistical issues is the challenge of gauging how well intakes conform to dietary guidance. Foods reported in the survey are frequently reported as multi-ingredient foods or food combinations. Federal dietary guidance, on the other hand, is stated in terms of discrete food groups, with a relatively small discretionary calorie allowance that can be used for dietary extras such as added sugars and fats or additional servings of food groups (11,12). Therefore, reported intakes need to be translated into amounts of food groups used in federal guidance.

The purposes of this study are to demonstrate NCI’s method, using nationally representative data, and determine the percentages of the population and of various subgroups in the population who do not consume the recommended amounts of various food groups and who consume excess amounts of solid fats, added sugars, and alcohol. The foods and other dietary constituents examined in this study are highlighted in the 2005 Dietary Guidelines for Americans (13).
Dietary Guidelines for Americans (11) and the USDA food guidance system, known as MyPyramid (12).

Materials and Methods

Data. Data were obtained from the 2001–2004 NHANES for 17,311 persons aged ≥2 y for whom reliable dietary intake information was available. Family income, which is expressed in the dataset as the poverty income ratio (PIR), is an important predictor of food intakes and was included as a covariate in the models. Because PIR was missing for 973 individuals, they were excluded, leaving 16,338 individuals for analysis. The NHANES protocol was approved by the National Center for Health Statistics Research Ethics Review Board, Hyattsville, MD, and all participants provided informed consent.

Dietary data were collected via interviewer-administered 24HR conducted using the automated multiple pass method (13). In 2001, a single recall was requested from all individuals and a second recall was collected on a small subset of participants. In subsequent years, 2 recalls were requested of all participants. In each year of the survey, the first recall was obtained as part of a larger interview and examination in a mobile exam center and the second was obtained 3–10 d later via telephone. However, not all participants agreed to provide both recalls and confidentiality concerns precluded release of the second recalls from 2002. In this study, we used all publicly released 24HR except those deemed unreliable by survey staff. This resulted in a total of 23,633 24HR available for analysis: 2 each from 7295 individuals and 1 each from all other individuals. Values for all intake variables were assessed to detect any outliers that might exert undue influence as follows. The raw, non-zero values were Box-Cox transformed to approximate normality and extreme values were identified as those below the 25th percentile minus 2.5 multiples of the interquartile range of the transformed distribution and those above the 75th percentile plus 2.5 multiples of the interquartile range. Preliminary modeling indicated that the exclusion of these extreme values did not change the results of the analysis; thus, no exclusions or corrections were made in this regard for the final models. Any differences in data collection and processing procedures between survey cycles during the 2001–2004 time period were not accounted for. Further information about the methodology of the NHANES is available elsewhere (14).

MyPyramid food groups. Quantities of foods reported on the 24HR were translated into amounts of various food groups using USDA’s MyPyramid Equivalents Database (MPED), version 1.0 (for 2001–2002) and version 2.0 (for 2003–2004) (15,16). The MPED disaggregates all foods reported (represented by ~7000 unique codes) into their component ingredients, groups similar ingredients together, and provides estimates of the amounts of 32 food groups consumed by each respondent in terms that are consistent with the units of measure used in MyPyramid (12). For example, vegetables are measured in cup-equivalents, and they include the vegetables reported separately as well as those in mixed dishes, such as beef with broccoli.

MyPyramid generally recommends consuming foods in their lowest fat, most nutrient-dense form, without added fats or sugars. Recommendations are provided for fruits, vegetables, grains, milk, meat and beans, and oils and a small discretionary calorie allowance is given for extras, which can be obtained in the form of discretionary solid fats, added sugars, alcoholic beverages, or excess amounts of any of the food groups (12).

The MPED assigns all dietary components to appropriate food groups in accordance with MyPyramid recommendations. For instance, the skim milk portion of whole milk counts as milk and the fat portion counts as solid fat, a designation that MPED ascribes to discretionary fat derived from animal sources other than fish and from hydrogenated vegetable oils. Similarly, sugar added to flavored milk counts as added sugars, a designation that includes all sugars used as ingredients in processed and prepared foods. Further information about the MPED can be obtained from the documentation (15,16).

MyPyramid does not provide a single set of recommendations; rather, food group recommendations depend on each person’s appropriate level of energy intake, which in turn depends on a number of other factors, including his/her sex, age, and activity level. For this analysis, when examining the possibility of insufficient intakes, we used the sedentary level of activity, which meant we compared intakes to the minimum recommendation for the sex-age group (Tables 1 and 2). When examining the possibility of excessive intakes (Table 3), we used the active level, which meant we compared intakes to the maximum recommendation for the group. Therefore, in all cases, we erred on the side of providing conservative estimates of the prevalence of not meeting recommendations.

Statistical methods. Estimated usual intake distributions were modeled, using SAS macros developed at NCI (17), for the sex-age groups used for the Dietary Reference Intakes (7). Generating the usual intake distributions involves estimating the mean and the within- and between-person variance components for the amount of the food component consumed for each sex-age group. Stable estimation of the variance components requires a substantial number of individuals to have consumed the food component on multiple days. To meet this requirement, several sex-age groups were pooled for the purpose of variance estimation (the means were allowed to differ by sex-age group). This was achieved by stratifying the sample into 3 large groups (children aged 2–8 y, males 9 y and older, and females 9 y and older) and including categorical covariates for sex (necessary for children’s stratum only) and age group. Covariates were also included to account for sequence effect (in case the first recall is systematically different from the second recall) (6), weekend/weekday effect, PIR, and race/ethnicity. PIR is a measure of income in relation to family size; PIR variables were created using categories of <1.25, 1.25–3.49, and >3.49. Race/ethnicity groups were Mexican American, non-Hispanic Black, and all others. Because the NHANES uses a complex, multistage, probability design, variance estimation was carried out via the Balanced Repeated Replication technique using replicate weight sets developed by the USDA Agricultural Research Service specifically for the purposes of examining dietary intakes for NHANES 2001–2004.

A 1- or 2-part nonlinear mixed model was used, depending on whether or not the food in question was consumed daily by almost everyone (6). If the food was episodically consumed, the 2-part model was used. The first part of the model estimates the probability of consumption using logistic regression with a person-specific random effect and the second part specifies the consumption-day amount using linear regression on a transformed scale, also with a person-specific effect. The person-specific random effects are allowed to be correlated across the 2 parts, because the probability of consumption is often related to the amount consumed. If the food was ubiquitously consumed, the probability part of the model was not needed and the 1-part model was used. For each food group and stratum, the following criteria were employed to determine whether to use the 1- or 2-part model and, if the latter, whether to use the correlated or uncorrelated model. If <5% (unweighted) of the 24HR had zero intake of a food, zeroes were replaced with one-half the minimum observed nonzero value and an amount-only model was used. If >10% of the 24HR had zero intake of a food, the 2-part model was fit 2 times, once where correlation between the person-specific random effects was assumed to be zero (the reduced model) and once where the correlation was estimated from the data (the full model). In the case of milk, just over 6% of the 24HR had zero intake, so all 3 types of model were fit to the data. When multiple models were run, the best-fitting of these was selected as further described below.

One-part models were used in all 3 strata for total vegetables; total grains; meat and beans; oils; solid fats; added sugars; and energy from solid fats, alcoholic beverages, and added sugars. They were also used for milk in the children’s stratum, because <1.5% of the 24HR in that stratum had zero intake. The reduced 2-part model was used in all 3 strata for orange vegetables, because the full models did not converge. For the remaining food groups, if the Fisher’s Z transformation of the estimated correlation between random effects in the full 2-part model differed statistically from zero at the 5% significance level, that model was used; otherwise, the reduced model was used. Reduced models were used for legumes and starchy vegetables (all strata), dark green vegetables (children and males age 9 y and older), other vegetables

2 of 7  Krebs-Smith et al.
(children), and milk (males age 9 y and older). Full models were used for all other food group/stratum combinations, including alcoholic beverages, which were assessed only for individuals aged ≥19 y. The reduced models over the full models had negligible effects on estimated percentages of not meeting recommendations but provided a modest reduction to the SE.

Once models were fit for each stratum, usual intake distributions and the percentages meeting or exceeding recommendations were estimated for each sex-age group \((6,17)\). The methodology \((6)\) permits estimation of distributions not only for individual sex-age groups but also for collapsed groups such as all adult females.

### Results

A large majority of the population did not meet the minimum recommendations for every one of MyPyramid’s nutrient-rich food groups except total grains and meat and beans \((2)\). Even though the percentages of young children not meeting recommendations were unacceptably high, 2- to 3-y olds had the lowest rates of inadequacy among all sex-age groups for total fruit, whole fruit, orange vegetables, legumes, and milk. On the other hand, over 90% of women aged 19–30 y had usual intakes below recommendations for 8 food groups, including total fruits, whole fruits, several vegetable subgroups, whole grains, and milk.

Among the food groups, dark green vegetables, orange vegetables, legumes, and whole grains had the poorest showing, with nearly everyone in each sex-age group failing to meet recommendations. Total vegetable and milk recommendations were each unmet by over 90% of individuals in at least half of the 14 sex-age groups.

When usual intakes of energy from solid fats, added sugars, and alcoholic beverages were compared with maximum discrementary calorie allowances, the results revealed ubiquitous overconsumption of these sources of energy, with prevalences > 80% among persons age 71 y and older and > 90% among all other groups \((3)\). When these components were examined separately, it was clear that overconsumption of both solid fats and added sugars was widespread. Although the recommended limit on alcoholic beverages is twice as high for men as for women, the prevalence of intakes in excess of the limit was greater among men than among women for each age group. Among both sexes, the 31- to 50-y-old age group was most likely to exceed the alcohol limit.

### Discussion

This paper makes a contribution to the literature by reporting the prevalence of under- and overconsumption of food groups and empty calories relative to federal recommendations using data from a nationally representative dietary survey in a statistically appropriate manner. Employing 24 HR to estimate prevalence first requires estimating the distribution of usual food intakes, something that has eluded the field of dietary assessment for decades \((2,5,8,9)\). Because the mean of the population’s distribution of usual intake can be estimated from a sample of 24HR without sophisticated statistical adjustment, such measures of central tendency have been available and reported previously \((8,18)\); however, only the full distributions and their associated statistics can give a sense of how widespread the problems are. Percentiles of full distributions of usual intakes of food groups and energy from solid fat, alcoholic beverages, and added sugars among the U.S. population can be found elsewhere \((19,20)\).
<table>
<thead>
<tr>
<th>Sex-age group</th>
<th>Total</th>
<th>Whole grains</th>
<th>Milk</th>
<th>Meat and beans</th>
<th>Other</th>
<th>Legumes(^1)</th>
<th>Starchy vegetables</th>
<th>Dark green vegetables</th>
<th>Orange vegetables</th>
<th>Vegetable(^2)</th>
<th>Fruits</th>
<th>Other(^3)</th>
<th>Calories</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Children, (6-11) years</td>
<td>1041</td>
<td>62.3</td>
<td>3.67</td>
<td>983</td>
<td>3.31</td>
<td>873</td>
<td>1.43</td>
<td>784</td>
<td>2.59</td>
<td>70.1</td>
<td>5.43</td>
<td>31.6</td>
<td>4.61</td>
</tr>
<tr>
<td>Females, (6-11) years</td>
<td>210</td>
<td>73.0</td>
<td>3.13</td>
<td>168</td>
<td>3.88</td>
<td>156</td>
<td>3.27</td>
<td>147</td>
<td>2.98</td>
<td>177</td>
<td>2.06</td>
<td>177</td>
<td>2.08</td>
</tr>
</tbody>
</table>

1 Data are percentage of individuals whose usual intakes are below minimum recommended amounts of MyPyramid food groups, by sex and age, 2001–2004.\(^4\)

2 MyPyramid food group recommendations depend on appropriate level of energy, and energy needs vary by activity level, sex, and age. For this analysis, a sedentary activity level was assumed for each sex-age category.

3 Sex-age-specific recommendations were applied prior to estimating percentages for collapsed groups.

4 Estimated percentages of individuals whose usual intakes are below minimum recommended amounts of MyPyramid food groups, by sex and age, 2001–2004.\(^1\)
This analysis indicates that nearly the entire U.S. population consumes a diet with fewer vegetables and whole grains than recommended and that a large majority underconsume fruits, milk, and oils relative to recommendations. Young adults have the greatest tendency toward insufficient intakes, although no sex-age group is immune. The prevalences of not meeting the total grains and meat and beans group recommendations were low relative to the other food groups, yet the intakes of these groups fell short among a substantial proportion of persons in some sex-age groups. At the same time, the maximum recommended intakes of meat and beans were exceeded by over 10% of individuals in 10 of 14 sex-age groups, whereas over 10% of individuals in 12 of 14 sex-age groups exceeded the maximum recommendation for total grains (20). The results for solid fats, added sugars, and alcoholic beverages indicate pervasive overconsumption of these nutrient-poor energy sources. Although this study included no assessment of energy balance, this excess appears to be displacing nutrient-rich food groups and may be contributing to the obesity epidemic by providing many individuals with unnecessary empty calories. The poor quality of Americans' diets evident from this analysis is consistent not only with high rates of obesity and overweight (21) but also food insecurity, which is characterized by compromises in diet quality and variety (22).

This study was dependent in large part on the MPED, which provides both a strength and limitation to the analysis. The strength is the distinctive facility MPED provides to examine diets reported via 24HR in relation to federal food guidance. Such an endeavor requires that foods reported as consumed (including composite foods and mixtures) be disaggregated into their nutritionally relevant components and accounted for in units of measure consistent with guidance. The MPED and its predecessor, the Food Guide Pyramid Database, were uniquely designed to serve this purpose.

The limitation of using the MPED is that the most recent databases available at the time of this study represented the 2001–2004 timeframe. The NHANES is conducted in 2-y cycles and the 24HR data are not released until at least 12 mo following data collection. Beyond that, updates to the MPED database have fallen behind other data releases and are available only through the 2003–2004 survey. Given the MPED’s central role in dietary evaluation (23), dietary pattern analyses (24), and all other examinations of food group intakes for which it is desirable to have food mixtures disaggregated into their component parts, this is problematic. Future studies with relevance to dietary guidance and policy will benefit from more timely updates to this important resource.

The possible limitation of excluding individuals with missing family income data deserves mention. Although the prevalence of missing data did not differ significantly across gender or race/ethnicity categories, it did differ across age categories, with the result that older individuals were more likely to be excluded.
Weighted percentages of individuals with missing income data ranged from 3 to 7%.

These findings complement those of other analyses, adding another piece to the rather disturbing picture that is emerging of the nation’s diet. Assessments of diet quality in relation to the 2005 Dietary Guidelines using the Healthy Eating Index-2005 (23) suggest that the diets of individuals across a range of income groups and ages are far from optimal (25–27). When this same index of diet quality has been applied at a macro level, it is clear that the U.S. food supply has provided an overabundance of solid fats, added sugars, and sodium, and insufficient fruits, vegetables, whole grains, and milk over at least the last several decades (28). Previous analysis of food sources of energy, solid fats, and added sugars indicated that the population is deriving a large portion of these components from sugar-sweetened beverages, grain desserts such as cookies, pastries, and cakes, and other foods that are not nutrient-dense (29). Further information on how the population is obtaining MyPyramid food groups can be found elsewhere (30).

The stark contrasts observed between the diets of Americans as well as the U.S. food supply (28) and current dietary guidance underscore the need for individual- and environmental-level interventions to facilitate healthier dietary intake patterns. Without such interventions, the diets of most U.S. adults and children will continue to be markedly divergent from recommendations, a worrisome state in the context of the obesity epidemic and alarming rates of other diet-related chronic diseases.

Acknowledgments

The authors thank their fellow members of the NCI’s Surveillance Measurement Error Working Group—Dennis Buckman, Laurence S. Freedman, Victor Kipnis, Douglas Midthune, and Janet A. Tooze—for developing the statistical methods used for this analysis; Ruth Parsons and Stella Munuo for their invaluable programming assistance; Joseph Goldman for providing the BRR weights for NHANES; and Dave Castenson for producing the tables. S.M.K-S. conceived of the study, drafted the manuscript, and had primary responsibility for the final content. K.W.D. prepared the data and oversaw the statistical analysis. P.M.G., A.F.S., and S.I.K. contributed to the interpretation of results and provided substantive critical feedback on the manuscript. All authors read and approved the final paper.

Literature Cited


