United States
Environmental Protection
Agency

Office of
Radiation Programs
Washington, D.C. 20460

May 1984
EPA 520/1-84-015

Radiation

EPA

An Estimation of the Daily Food Intake Based on Data from the 1977–1978 USDA Nationwide Food Consumption Survey

AN ESTIMATION OF THE DAILY FOOD INTAKE BASED ON DATA FROM THE 1977-1978 USDA NATIONWIDE FOOD CONSUMPTION SURVEY

Christopher B. Nelson
You-yen Yang

May 1984

Office of Radiation Programs
U.S. Environmental Protection Agency
Washington, D.C. 20460

FOREWORD

The Office of Radiation Programs carries out a national program to evaluate the exposure of man to ionizing and nonionizing radiation and to develop guidance, standards, and criteria for the protection of public health and the quality of the environment.

This report describes a more refined methodology than heretofore used for estimating the daily food intake as food utilization factors in the assessment of environmental radionuclide intake by individuals through food consumption.

Readers are encouraged to bring to our attention any questions encountered in this report. Any comments and suggestions are also welcome. These comments should be directed to Christopher B. Nelson, Bioeffects Analysis Branch, Analysis and Support Division, Office of Radiation Programs (ANR-461C), U.S. Environmental Protection Agency, Washington, D.C. 20460.

Glen L. Sjoblom

Glen L. Sjoblom, Director
Office of Radiation Programs

ACKNOWLEDGEMENT

Numerous people helped in the preparation of this report, but particular acknowledgement is due to Drs. Eleanor M. Pao and Kathryn H. Fleming of USDA for their assistance and support relative to the USDA study and to Mary Anne Culliton of EPA for her editorial contributions.

Contents

APPENDICES

An Estimation of the Daily Food Intake Based on Data from the 1977-1978 USDA Nationwide Food Consumption Survey

Christopher B. Nelson
You-yen Yang

U.S. Environmental Protection Agency
Office of Radiation Programs

Summary

A Nationwide Food Consumption Survey was conducted by the U.S. Department of Agriculture in 1977-78 to investigate the food intake of various selected segments of the U.S. population and to identify changes in U.S. food intake patterns. EPA has used data from this survey to determine the average daily total intake of different food items. This daily average can be used to assess radionuclide intake by persons in various geographic categories within the United States.

Through analytical processes, we were able to estimate the daily intake of each food subclass, not only for the U.S. population as a whole, but also for subpopulations classified according to their geographical (census) characteristics. We also computed the standard error of each estimate for statistical inference. These estimates are weighted mean values of the daily intake and are adjusted for the age and sex distribution of the 1969-71 U.S. stationary population, region, division, primary sampling unit (PSU), and season.

All food items in the survey were grouped into 8 main classes which were further divided into 22 subclasses. We analyzed the data for these classes and subclasses to determine which factors in the statistical model (Appendix B) had significant effects on food intake.

Region was the most important factor. It significantly affected intake of practically all food subclasses. Liquid food, as opposed to solid food, accounted for most of the differences among regions. Urbanization (and the interaction of urbanization and geographic division) played a trivial role in food intake. While season had practically no significant effects on solid food intake, it affected beverage, tap water, and soup intake significantly.

Different studies on caloric intake for the U.S. population were compared and shown to be consistent with each other. The estimated mean caloric intake in this study is 1863 kcal, which is about 84% of the Recommended Dietary Allowances (RDA). This does not necessarily indicate either low bias in either the USDA survey or the EPA analysis.

I. Introduction

To assess the radionuclide intake by humans through food intake, we need estimates of the daily food intake of different food items. Historically, a series of ad hoc values for average and maximum food intakes has been used. These values are generally the result of a subjective interpretation of data reported in nutrition or marketing studies; Rupp's reports (Rupp, 1980A-B) are good examples. None of these reports, however, include significance tests for reported class effects or confidence intervals for estimated parameters.

EPA's study has used the data from the U.S. Department of Agriculture (USDA) 1977-78 Nationwide Food Consumption Survey (NFCS) (see Appendix A for more information about the Survey) to estimate the daily intake of different food classes which will be used for the assessment of environmental radionuclide intake through food consumption. We have also performed statistical analyses and computed the standard error of each estimate for statistical inference.

The USDA survey objectives included investigating the dietary intake of various segments of the U.S. population, determining the adequacy of diets as compared to the Recommended Dietary Allowances, and determining any changes in food intake patterns. Because EPA's objective was different, we had difficulties with the survey design, food classifications, and in applying the data to our statistical analysis. To overcome these problems, we reorganized the original data files so that we could classify food items according to the characteristics of radionuclide transport and thus be able to analyze the data for our purposes.

In this report, we use the term "daily intake" to mean "average daily total intake." We will present statistical findings, interpret their significance, and give the estimated mean daily intake of food classes per caput. This is the first of a series of reports concerning statistical analyses of the NFCS data. We discuss some geographical effects on food intake in this report.

II. Survey Design

In the Nationwide Food Consumption Survey (see Appendix A), the sampling unit was a household. For each household there were two parts to the survey: the household part investigated the quantities and sources of foodstuffs used in the preparation of meals served in the household, while the individual part surveyed the intakes of individual household members at each eating occasion over a three-day period. The data for the individual part of the survey were analyzed for this report. Since members of the same household would usually be eating substantially from the same food supply and would often have similar eating patterns, we could not expect data for members of the same household to be independent. Furthermore, the households in each segment of at least 100 households within a primary sampling unit (PSU) were allocated to four different quarterly samples--no samplings were performed for each season, and the households selected for the study were not statistically independent. For these reasons, we did not use the mean squares of "individual" or "household" as an estimate of variance of sampling error for hypothesis testing in the analysis of variance. We used, instead, the mean square of the PSU, or precisely, PSU nested in region, division, and urbanization, as an estimate of the variance of sampling error for the analysis. Thus, the analytical results we obtained are somewhat conservative.

Because of our different objectives, we encountered the following problems:

1) The protocol (USDA, 1981) for the spring quarter was different from that for the remaining quarters. In the spring quarter, every member in a sampled household was interviewed, but in the remaining quarters, all persons younger than 19 and one-half of all persons 19 and older were interviewed, i.e., a part of the household was subsampled. If there was only one person in the household, he or she was interviewed with certainty.

2) The protocol reads: "In order to provide seasonal information, the sample design called for identification of four independent interpenetrating samples to be implemented in consecutive quarters of data collection," i.e., households in each area segment were allocated to four quarterly samples. While this sampling procedure would be more economical than a random sample, it is very doubtful that these four quarterly samples would indeed be independent as stated in the protocol, because the samples were drawn systematically from the same area segment.

3) Because part of the data was collected through recollection, the data could be low-biased.

III. Food Classifications

Because concentrations of radionuclides in foods can vary widely, we classified foods by categories for which we can measure or calculate concentrations of radioactivity and which we know comprise significant dietary intake. We divided foods into eight classes; in turn, each class was divided into subclasses as shown in Table 1 below.

For example, we classified produce into three subclasses: leafy vegetables, produce exposed to the atmosphere, and protected produce. Leafy vegetables, such as lettuce, have a broad flat leaf surface for direct interception of atmospherically deposited material. The edible portion of the plants primarily consists of leaves and stems. Exposed produce, tomatoes for example, also intercepts atmospherically deposited material, but surface areas (for exposure) are typically smaller than those of leafy vegetables. Edible parts are usually reproductive organs which accumulate additional radionuclides through soil uptake to a smaller extent than vegetative parts. Protected produce, such as potatoes, are not directly exposed to atmospherically deposited material. Edible portions are principally reproductive or storage organs which either grow underground or are protected by pods, shells or nonedible peels. The accumulation of radionuclides takes place through uptake of radionuclides from soil or transfer from nonedible portions.

Similar considerations were extended to other classifications of foods so that each classification is pertinent to specific radionuclide pathways.

Every food item in the NFCS was assigned a 7-digit code (USDA, 1979A). The first three digits represent food groups; the last four are associated with ingredients added to, or the preparation of, the food items. The following illustrates this coding system for the first two digits.

First digit = major food groups.
1 = milk and milk products,
2 = meat, poultry, fish, and mixtures,
3 = eggs, mixtures, substitutes,
etc.

First and second digits = major food subgroups.
11 = milk and milk drinks,
12 = cream and cream substitutes,
13 = milk desserts,
14 = cheeses,
21 = beef,
22 = pork,
23 = lamb, veal, game, other carcass meat,
24 = poultry,
etc.

We classified all 3,735 food items in the survey into our eight major classes; each class again was divided into two to four subclasses. There were many occasions where the appropriate classification was not clear. To classify them, we made a judgment based on the objectives of the study and the characteristics of the food items under consideration. Reconstituted milk, for example, could be classified as other dairy products or as water-based drinks. We considered the item from a radiological assessment perspective: what were the ingredients of the food items and what was the most important ingredient; had it been processed or stored for some time before it was served? After making those considerations, we classified reconstituted milk into the other dairy products subclass.

Stew, for another example, consists of meat, vegetables, starch, water, and unspecified items. We considered the meat to be the most significant ingredient and identified the meat in the stew by using the USDA coding system. We separated bread and filling for sandwiches, including hamburgers, hot dogs, submarines, and tortillas. All 3,735 food items in the survey data file were classified as shown in Table 1.

Because there are so many varieties of vegetables and fruits, the following indicate important produce belonging to each subclass.

Leafy vegetables:	cabbage, cauliflower, broccoli, celery, lettuce, spinach.
Exposed produce:	apples, pears, berries, cucumber, squash, grapes, peaches, apricots, plums, prunes, string beans, pea pods, tomatoes.
Protected produce:	carrots, beets, turnips, parsnips, citrus fruits, sweet corn, legumes (peas, beans, etc.), melons, onions, potatoes.
Other produce:	Unspecified vegetables or fruits and mixtures of vegetables or fruits.

IV. Data Analyses

USDA recorded the information on food intake for every eating occasion during the three-day survey period. Since the objective of the EPA study was to estimate the daily intake of food classes and subclasses for each sampled individual, we added the quantities of a food subclass or class consumed on each occasion throughout the day, then took an average over the number of days the individual had participated in the survey.

If an individual had participated two days, for example, the average daily total would be obtained by dividing the total quantity by two for that food subclass or class. Most of the survey data had a complete three-day set of observations (93%), very few had a one-day (5%) or a two-day (2%) set. A total of 30,770 persons participated in the survey. Given the high proportion of individuals with three-day sets of observations, we chose not to weight the observations on the basis of the numbers of days of data (see Appendix B for more information).

V. Analytical Results

The general linear models procedure, GLM, in the Statistical Analysis System (1979), commonly known as SAS, was used to analyze the data. The results of analysis of variance are given in Table 2.

Table 1. Classification of the Food Items in the NFCS into Major Classes and Subclasses

Major Class	Subclass
Dairy products	Fresh cow's milk, Other dairy products (dry milk, butter, cheese, etc.).
Eggs	Fresh eggs, Other egg products (powdered eggs and other prepared egg products).
Meats	Beef, veal, Pork, Poultry (chicken, duck, other birds), Other meats (game, organ meats, meat mixtures, etc.).
Fish	Fin fish, Shellfish, Other seafood (mixtures and unspecified fish products).
Produce	Leafy vegetables, Exposed produce, Protected produce, Other produce (unspecified vegetables or fruits; mixtures of vegetables or fruits).
Grains	Breads, pasta, cakes, etc., Cereals, Other grains (wheat, rice, raw millet, raw rye, etc.).
Beverages	Tap water, Water-based drinks (coffee, tea, etc.), Soups, Other drinks (soft drinks, fruitades, alcoholic drinks).
Miscellaneous	No subclasses (chocolate, sugar, salad dressings, etc.).

The F statistics given in Table 2 can be expressed by:

$$F = MS(A)/MS(PSU),$$

where MS(A) and MS(PSU) are the mean squares of A and PSU, respectively. "A" represents any factor under hypothesis testing, such as region, division nested in region, or urbanization nested in region. The test statistic, under the null hypothesis, has an F-distribution with degrees of freedom f_A and f_{PSU}, i.e., $F(f_A, f_{PSU})$. For a given level of significance, say α, when

$$MS(A)/MS(PSU) \geq F_{1-\alpha}(f_A, f_{PSU}),$$

we reject the null hypothesis and declare that the test result is (statistically) significant at the α level.

For simplicity, when we mention division, urbanization, or season, we will mean division nested in region, urbanization nested in region, and so on. Note that since the quantities of other egg products and other fish products consumed were negligibly small, less than one gram/day, we have excluded them from the statistical analysis.

Census region and geographical division (see Appendix A) significantly affected food intake. Urbanization, interaction of division and urbanization are relatively unimportant on food intake. Season had practically no effect on solid food intake, but it significantly affected beverages, tap water, and soup intake.

The intake of two major food classes, dairy products and beverages, varied significantly by both region and geographical division (Table 2); for other food subclasses, it varied significantly only by region or division. Table 2 also shows that food intake patterns of food subclasses were overwhelmingly regional. In all but three subclasses, leafy vegetables, other grain products and other beverages, significant regional effects were detected. When hypothesis tests were performed for the major classes, significant effects could appear or disappear compared with the results obtained for subclasses. For example, all four subclasses of meats showed significant effects for region yet the class meats did not show a significant effect for region. Conversely, only one of the four subclasses of beverages showed a significant effect for geographical division yet beverages as a whole showed a significant effect for geographical division. This is because the aggregation can cause effects to be magnified or diminished. A few significant results were detected for season, urbanization, and the interaction of urbanization and geographical division.

The estimated mean daily intake, weighted by age, sex, region, division, PSU, and season, are given in Table 3. The values given in

Table 2. Analysis of Variance by Food Class and Subclasses
(F values for sources of variation)

Food class or Subclass	Source of variation: Region	Division	Urbanization	Division Urbanization	Season	Sampling Error (gram)
Degrees of Freedom	(3)	(5)	(8)	(10)	(12)	(84)
DAIRY PRODUCTS	17.40**	3.91**	1.94	.82	1.29	807.3
Fresh cow's milk	11.54**	4.38**	1.76	.83	1.67	753.7
Other	22.68**	1.00	1.69	.90	.42	179.4
EGGS	17.97**	1.79	2.07*	1.34	.45	75.7
MEATS	1.59	2.42*	1.15	1.07	.42	252.3
Beef & veal	7.12**	5.17**	1.18	.76	.66	176.0
Pork	8.75**	1.61	.71	1.21	.56	85.7
Poultry	8.57**	.97	1.87	1.40	.53	115.9
Other	14.63**	1.42	2.86**	1.86	.71	66.1
FISH	6.23**	2.14	.78	1.06	.57	81.5
Fin fish	4.12**	2.03	.65	1.00	.61	67.5
Shellfish	3.65*	1.32	.87	1.08	.93	35.3
PRODUCE	1.19	3.99**	.47	.25	1.35	538.0
Leafy	1.03	3.07*	2.64*	.86	1.08	116.8
Exposed	7.32**	4.22**	1.00	.87	7.09**	232.4
Protected	5.10**	3.28**	.96	.42	.42	348.6
Other	8.63**	1.69	1.12	2.08*	1.33	43.2
GRAIN	0.43	2.47*	.67	.67	1.20	451.5
Breads	7.90**	2.27	1.17	1.86	1.84	218.9
Cereals	2.86*	1.14	.38	.57	.81	194.3
Other	2.57	3.51**	1.67	1.44	.12	259.0
BEVERAGES	17.21**	3.85**	.30	.81	2.33*	2098.5
Tap water	30.91**	7.93**	1.57	2.55**	2.04*	1513.0
Water based	3.32*	.61	2.27*	1.66	.63	1034.0
Soups	9.80**	.94	1.07	.98	4.51**	186.0
Other	1.26	1.89	3.34**	.51	1.18	724.6
MISCELLANEOUS	9.14**	2.31	3.89**	1.51	.54	107.7

Notes: *: .01 < P ≤ .05 **: P ≤ .01.

Table 3. Mean and Standard Error for the Daily Intake of Food Class and Subclass Region (in grams)

Food class/ Subclass	U.S. Population	Region			
		Northeast	North Central	South	West
DAIRY PRODUCTS	308.6±5.3	318.6±10.4	336.1±10.0	253.6±8.4	348.1±12.3
Fresh cow's milk	253.5±4.9	256.1±9.7	279.7±9.4	211.0±7.8	283.5±11.5
Other	55.1±1.2	62.5±2.3	56.5±2.2	42.6±1.9	64.6±2.7
EGGS	26.9± .5	23.8±1.0	23.5± .9	31.0± .8	29.1±1.2
MEATS	172.2±1.6	169.9±3.3	176.9±3.1	171.9±2.6	168.6±3.9
Beef & veal	87.6±1.1	82.3±2.3	92.9±2.2	84.0±1.8	92.9±2.7
Pork	28.2± .6	28.8±1.1	29.6±1.1	30.1± .9	22.1±1.3
Poultry	31.3± .8	31.7±1.5	26.6±1.4	36.5±1.2	28.9±1.8
Other	25.1± .4	27.1± .9	27.8± .8	21.3± .7	24.7±1.0
FISH	17.5± .5	20.5± 1.1	14.7±1.0	17.0± .8	18.5±1.2
Fin fish	14.7± .4	16.7± .9	13.1± .8	14.2± .7	15.2±1.0
Shellfish	2.6± .2	3.6± .5	1.5± .4	2.6± .4	2.8± .5
PRODUCE	282.6±3.5	270.6±6.9	282.4±6.7	280.7±5.6	303.1±8.2
Leafy	39.2± .8	38.1±1.5	37.1±1.5	38.4±1.2	45.3±1.8
Exposed	86.0±1.5	88.5±3.0	87.8±2.9	76.9±2.4	95.5±3.6
Protected	150.4±2.3	137.2±4.5	150.1±4.3	160.1±3.6	152.5±5.3
Other	7.0± .3	6.9± .6	7.3± .5	5.4± .4	9.8± .7
GRAIN	200.0±3.0	203.5±5.8	192.8±5.6	202.2±4.7	202.6±6.9
Breads	147.3±1.4	153.1±2.8	150.9±2.7	143.9±2.3	139.5±3.3
Cereals	29.9±1.3	24.6±2.5	28.7±2.4	34.6±2.0	30.9±3.0
Other	22.9±1.7	25.9±3.3	13.3±3.2	23.7±2.7	32.1±4.0
BEVERAGES	1434.±13.7	1288.±27.0	1448.±26.0	1513.±21.8	1480.±32.1
Tap water	662.5±9.9	520.9±19.5	659.8±18.8	748.5±15.7	714.4±23.1
Water based	457.1±6.7	418.5±13.3	476.0±12.8	470.5±10.7	458.2±15.8
Soups	45.9±1.2	52.2± 2.4	48.9± 2.3	36.8± 1.9	48.1± 2.9
Other	269.9±4.7	297.2± 9.4	264.5± 9.0	258.1± 7.5	260.6±11.1
MISCELLANEOUS	34.6± .7	31.4± 1.4	37.0± 1.3	32.3± 1.1	39.2± 1.6

the second column, U.S. Population, of Table 3, are the expectation of the 1969-71 U.S. stationary population. It is interesting to note that the estimated standard error is small compared to the corresponding mean, although we have used a conservative sampling error in the statistical analyses. The age and sex distribution assumed is that of the 1969-71 U.S. stationary population (see Appendix B).

VI. Conclusions

Of the effects tested by analysis of variance, the regional factor is by far the most important. Except for leafy vegetables, other grain products, and other beverages, it significantly affects the intake of all food items under consideration. Food intake patterns are diverse as one can see in Table 3. The division factor is the second most important. It directly affects the intake of fresh cow's milk, beef and veal, leafy vegetables, exposed produce, protected produce, other grain products, and tap water. Urbanization and the interaction between division and urbanization do not significantly affect food intake. Almost no significant seasonal changes in solid food intake were detected.

All food subclasses can be classified as liquid or solid foods (Table 4). Of the total daily intake per caput (2478 grams), 68% (1689 grams) is liquid and 32% (789 grams) is solid, or the ratio of liquid foods consumed to solid foods is approximately 2 to 1. The total solid food intake is fairly constant across the regions whereas the liquid food intake has a wider range of variation. Overall intake is greatest in the West and least in the Northeast region.

When the regional intake is normalized by the national population intake for each major class and subclass, most regions are within ± 10% of the national levels. (Table 5 indicates the exceptions.)

The intake of shellfish in the Northeast region is 38% above the national average, while in North Central it is 42% below. The intake of other grains in the West is 40% above the national average; in North Central it was 42% below the national average. One may notice that the quantities of these two subclasses are small (see Table 3). It is interesting to note that only one subclass (pork) is more than 10% below the national average in the West, while there were five such subclasses in the South.

We have attempted to compare our findings to those of USDA (Pao, et al., 1982) and those of Rupp (Rupp, 1980). It is very difficult to make direct comparisons among these three sets of results. The USDA report presents average intakes per day and per eating occasion by sex and age. These averages are for individuals actually consuming the given food items; therefore, the values in the report are often quite different from the customary averages for the entire U.S. population.

Table 4. Average Total Daily Food Intake by Region (in grams)

Food type	Region				U.S. Population
	Northeast	North Central	South	West	
Solid food	782	784	777	825	789
Liquid food*	1545	1729	1725	1765	1689
TOTAL	2327	2513	2502	2590	2478

*Including fresh cow's milk.

Table 5. Exceptions to National Average Levels of Food Intake by Region

Northeast	North Central	South	West
	More Than 110% of National Average		
Other dairy products	Fresh cow's milk	Eggs*	Dairy products*
	Other meats	Poultry	Fresh cow's milk
Fish*		Cereals	Other dairy products
Fin fish		Tap Water	Leafy vegetables
Shellfish			Exposed produce
Other grains			Other produce
Soups			Other grains
Other beverages			Miscellaneous
	Less Than 90% of National Average		
Eggs*	Eggs*	Dairy products*	Pork
Cereals	Poultry	Fresh cow's milk	
Beverages*	Fish*	Other dairy products	
Tap water	Fin fish	Exposed produce	
	Shellfish	Other produce	
	Other grains	Soups	

*Major class.

Table 6. A Comparison of Studies by Rupp and EPA
(Mean daily intake per caput, in grams)

Food class	Rupp (Over 18 yrs)	EPA (All ages)
Dairy products	306	309
Meats	258	172
Produce	438	283
Grains	97	200
Fluid	1351*	1388**

*Other than soups and fresh cow's milk.
**Other than fresh cow's milk, but including soups.

Rupp's results, which are averages over a population, were obtained through a review of the literature and personal communications. Because Rupp's results are similar to ours, we summarize them in Table 6 for important major classes. One may note that the averages given by Rupp are for those individuals over 18. These values would normally be expected to be greater than those for all ages. There are considerable differences in intakes of meats, produce, and grains.

A difficult question to resolve in studies such as this is the possibility of bias. We recognize that even though study participants may have conscientiously recorded their best estimates of their intakes, the potential for recall bias clearly exists. Individuals may forget items they have eaten or systematically underestimate intake quantities. To our knowledge, no study has been performed to assess this effect.

One way to evaluate dietary intakes is to compare the resultant caloric intake with the food energy needs of the corresponding class of individuals. Zach, et al., have taken this approach and have concluded that ". . . the USNRC77 diets for maximum and average adults and teens fall well short of the expected caloric intake rates" (Zach, 1983). Their standard for comparison is primarily the dietary standard for Canada. We have estimated the daily caloric intake for an average U.S. individual using the caloric factors in Table 7.

These factors (col. 2) are based on caloric data reported in (USDA, 1979B). The resultant daily food intake for an average individual is 1853 kcal. For comparison, we have calculated the average food energy intake for the Health and Nutrition Examination Survey, 1971-74 (HANES)(USDHEW, 1979), the Nationwide Food Intake Survey-1977 (NFCS) (USDA, 1984), and the 1980 National Academy of Sciences Recommended Daily Allowances (USDA, 1980).

Table 7. Caloric Factors and Food Energy

Food Class	Caloric Factor (kcal/gram)	Daily Intake (gram)	Food Energy (kcal)
Milk	0.61	253.5	155
Other dairy products	1.76	55.1	97
Eggs	1.60	26.9	43
Beef	2.85	87.6	250
Pork	2.87	28.2	81
Poultry	1.85	31.3	58
Other meat products	2.43	25.1	61
Fin fish	1.90	14.7	28
Shellfish	0.95	2.6	2
Produce	0.81	282.6	229
Grains	3.42	200.0	684
Water based beverages	0.0074	457.1	3
Soups	0.42	45.9	19
Other beverages	0.40	269.9	108
Miscellaneous	1.00	34.6	35
Total			1853

Since our estimate is for an average individual in the 1970 stationary population, we have weighted the values according to age and sex in the same way for each set of data. HANES considers only ages 1 through 74. We have adjusted the value to all ages by presuming that the ratio of these two values would be the same for HANES as for NFCS (1897:1849). The results of this comparison are summarized in Table 8.

Caloric intakes in 1970s declined somewhat compared with those in the 1960s (USDA, 1980). The HANES result is consistent with this observation. In documenting HANES, USDHEW notes: "The recommended dietary allowances are designed to guide dietitians in formulating diets for the maintenance of good nutrition in healthy persons. They allow for some margin above what is really needed by most individuals, with the objective of maintaining good health" (USDHEW, 1979). It is interesting to note that the dietary standard for Canada (Zach, 1983), 2150-5350 kcal for adult males and 1750-4400 kcal for adult females, appears to be greater than for the United States, perhaps

Table 8. Comparisons of Caloric Intakes of the U.S. Population

	RDA*	HANES	NFCS	EPA
Year	1980	1971-74	1977-78	1977-78
kcal/day	2194	1925	1849	1853
Relative to RDA	--	0.88	0.84	0.84

*Recommended Dietary Allowances (see text).

refecting a somewhat colder climate than the U.S. average. We conclude that our result, 84% of the RDA, does not necessarily indicate either low bias or inadequate nutrient intakes.

Based upon the comparisons of different dietary parameters with the studies discussed above, we conclude that it is appropriate to use these estimated means of daily food intake as food utilization factors in the assessment of environmental radionuclide intake by individuals.

REFERENCES

USDA, (1979A), U.S. Department of Agriculture. Coding Manual to Handle Data from Nationwide Survey of Individuals. Spring 1977-78. Vols. I & II. National Technical Information Service, Springfield, Va.

USDA, (1979B), U.S. Department of Agriculture. Food Composition Data Values. CFE(Adm.)-347. Science and Education Administration, Consumer and Food Economics Institute.

USDA, (1980), U.S. Department of Agriculture. Food and Nutrient Intakes of Individuals in 1 Day in the United States, Spring 1977. Nationwide Food Consumption Survey 1977-1978, Preliminary Report No. 2. Science and Education Administration, Consumer and Food Economics Institute.

USDA, (1981), U.S. Department of Agriculture. Basic Sample Design. Personal communication from Brucy Gray.

USDA, (1984), U.S. Department of Agriculture. Food Intakes: Individuals in 48 States, Year 1977-78. Nationwide Food Consumption Survey 1977-78, Report I-1. Human Nutrition Information Service.

USDHEW, (1979). Caloric and Selected Nutrient Values for Persons 1-74 Years of Age: First Health and Nutrition Examination Survey, United States, 1971-1974. Office of Health Research, Statistics, and Technology, National Center for Health Statistics, U.S. Department of Health, Education, and Welfare.

Pao, Eleanor M., Kathryn H. Fleming, Patricia M. Guenther, and Sharon J. Mickle, (1982). "Foods Commonly Eaten by Individuals: Amount Per Day and Per Eating Occasion." U.S. Department f Agriculture, Home Economics Research Report, No. 44.

Rupp, E. M., (1980A). "Age Dependent Values of Dieta Intake for Assessing Human Exposure to Environmental Pollut .its," Health Physics 9, 151-63.

Rupp, E. M., Forest L. Miller, C. F. Baes, III, (1980B). "Some Results of Recent Survey of Fish and Shellfish Consumption by Age and Region of U. S. Residents," Ibid, 165-75.

SAS, (1979). SAS User's Guide, 1979 Edition, pp 244-263. SAS Institute Inc., Raleigh, N. C.

Zach R. and Mayoh, K.R., (1983). "Caloric Contents of Diets for Environmental Assessment Models: A Necessary Condition." Health Physics Vol. 45, No. 3, 786-789.

APPENDIX A

THE USDA 1977-78 NATIONWIDE FOOD CONSUMPTION SURVEY

APPENDIX A

THE USDA 1977-78 NATIONWIDE FOOD CONSUMPTION SURVEY

A.1 Design of the USDA Natiowide Food Consumption Survey

The basic sample design of the USDA Nationwide Food Consumption Survey (NFCS) is a multistage, stratified design covering a universe of all private households in the conterminous United States. This basic sampling design consists of four stages (USDA, 1981): stratification, formation and selection of primary sampling units (PSU), selection of area segments within PSUs, and sampling of housing units.

A.1.1. Stratification

The basic design consists of three levels of stratification:

First Level: Census Region

The conterminous United States, including the District of Columbia, was divided into four census regions: Northeast, North Central, South, and West. The following is the geographic grouping of states used by the Bureau of Census (1970):

Census Region	Geographic Division	State
Northeast	New England	Maine, New Hampshire, Vermont, Connecticut, Rhode Island.
	Middle Atlantic	New York, New Jersey, Pennsylvania.
North Central	East North Central	Ohio, Illinois, Indiana, Wisconsin, Michigan.
	West North Central	Minnesota, Iowa, Missouri, North Dakota, South Dakota, Nebraska, Kansas.

South	South Atlantic	Maryland, Delaware, District of Columbia, Virginia, West Virginia, North Carolina, South Carolina, Georgia, Florida.
	East South Central	Kentucky, Tennessee, Alabama, Mississippi.
	West South Central	Arkansas, Louisiana, Texas, Oklahoma.
West	Mountain	Montana, Idaho, Wyoming, Utah, Colorado, New Mexico, Arizona, Nevada.
	Pacific	Washington, Oregon, California.

The states within each of these divisions are, for the most part, fairly homogeneous in physical characteristics as well as in the characteristics of their population, economic, and social conditions; on the other hand, each division differs more or less sharply from most others in these respects.

Second Level: Urbanization

Each housing unit was classified by urbanization as: central city--in the central city of a standard metropolitan statistical area (SMSA); suburban--within an SMSA, but not in the central city; or nonmetropolitan--not within an SMSA.

A standard metropolitan statistical area consists of one or more entire counties, economically and socially integrated, that have a large population nucleus. The definition of an SMSA involves two considerations; first, a city or cities of specified population to constitute the central city and to identify the county in which it is located as the central county; second, economic and social relationships with contiguous counties which are metropolitan in character, so that the periphery of the specific metropolitan area may be determined. Standard metropolitan statistical areas may cross state lines to include qualified contiguous counties (Shryock, et al., 1976).

Third Level: Sampling Stratum

The third level of stratification groups census units with like geographic divisions and urbanization to produce sampling strata. One hundred fourteen (114) (Table A-1) strata were formed incorporating all housing units in the entire conterminous United States. The average stratum size is approximately 600,000 housing units.

Table A-1. Number of Strata in the Nationwide Food Consumption Survey by Region, Division, and Urbanization

Region	Division	Central City	Sub-urban	Nonmetro-politan	Total
Northeast	New England	2	3	2	7
	Middle Atlantic	8	9	4	21
North Central	East North Central	8	8	6	22
	West North Central	2	2	5	9
South	South Atlantic	4	6	7	17
	East South Central	2	1	4	7
	West South Central	4	2	5	11
West	Mountain	2	1	2	5
	Pacific	6	7	2	15
TOTAL		38	39	37	114

A.1.2. Formation and Selection of PSUs

Every stratum was divided into one or more primary sampling units, or PSUs, formed from cities, parts of cities, or counties, containing at least 10,000 housing units each. One PSU per stratum was randomly selected according to the following probablistic scheme:

$$p_{ij} = n_{ij}/N_i, \qquad i = 1,2,\ldots, 114; \quad j = 1,2,\ldots, k_i; \qquad \text{(A.1)}$$

where

p_{ij} = the probability of the jth PSU in the ith stratum being chosen,
n_{ij} = number of housing units in the jth PSU in the ith stratum,
k_i = number of PSU in the ith stratum,
N_i = total number of housing units in the ith stratum.

A.1.3. Selection of Area Segments within PSUs

Each selected PSU was [illegible]d into small clusters of housing units called area segments. The [illegible]ts, designed to contain 100 or more housing units, were based o[illegible] 1970 Census and usually consisted of one or more city blocks in urba[illegible]eas and a part of a Census Enumeration District elsewhere. From all PSUs, 2550 segments were randomly selected with the number of segments in a PSU being proportional to the size of the stratum in which the PSU was located. The probability that individual segments were drawn in a PSU was proportional to the ratio of the number of units in the segments to the total number of units in the PSU. There was one exception. Three strata, New York City, Chicago, and Los Angeles County, each with size well in excess of 600,000, were selected with certainty. Each of these was defined as two PSUs. Sampling from these three strata was done directly without further stratification by area segment.

A.1.4. Sampling of Housing Units

All 2550 area segments were prelisted to determine the number of occupied housing units. Then, the national increase or decrease in the number of housing units from 1970 to 1977 was estimated. This information, along with estimates of occupancy and completion rates, permitted sampling ratios to be calculated for the area segments that would yield a total of 3750 households per quarter.

For the first two quarters, an average sampling ratio of 2.3 households per segment was used. The housing units were ordered within their respective segments. For each quarter, a sample was systematically selected from each segment without replacement after a random start. By the end of the second quarter, the estimated completion rate had been adjusted, and an average sampling ratio of 2.86 households per segment was used for the last two quarters. Note that once the segments were chosen, different households would be sampled from the same segments for each season.

A.2. Household and Individual Survey

The Nationwide Food Consumption Survey consisted of two parts: household and individual.

Household Survey

For every sampled household, the eligible respondent was the person who usually planned and prepared the meals in the seven days before the interview. He or she was asked to keep dietary records, such as grocery shopping lists and receipts, menus, and labels from canned and packaged foods used for the meals. The information, in part, included:

-number of meals eaten during the past 7 days by each household member--from home supplies, not from home supplies,

-number of meals and refreshments served at home to nonhousehold members during the past 7 days--number, type, sex, age of consumers,

-household food consumption during the past 7 days--description of food, form, quantity, and money values, etc.

Individual Survey

Individuals from the sampled household were asked, as in the 7-day household survey, to keep records of their food intake. However, this survey was different from the household part in many respects:

-this was a 3-day instead of a 7-day survey. It included a one-day recall (yesterday), and a 2-day record (today and tomorrow) of food intake.

-in the spring quarter, every member of a sampled household was interviewed, but in the remaining quarters, all persons younger than 19, and one-half of all persons 19 and older were interviewed. If there was only one person in the household, he or she was included with certainty.

-food intake records were kept for all foods and beverages consumed.

The food intake information included:

-for each eating/drinking occasion, whether food was consumed at or away from home, information was collected for each item as to when, with whom, in what quantity and from what source. For away-from-home food, data on type of service and cost were also collected.

-intake of water.

REFERENCES

USDA (1981). U.S. Department of Agriculture. Basic Sample Design. Personal communication with Brucy Gray.

Department of Commerce, (1970). "Pocket data book, U.S.A. 1970."

Shryock, H. S. and J. S. Siegel and Associates (1976). "The Methods and Materials of Demography," Academic Press. New York.

APPENDIX B

STATISTICAL ANALYSIS

APPENDIX B

STATISTICAL ANALYSIS

B.1. Statistical Model

The data collected by USDA in this survey were very extensive. Many individuals were interviewed regarding their food intake; each was asked many questions. The data consisted of numerous variables and factors for a large number of people. For example, in addition to questions related to food consumption, there were 21 variables. Of these, seven, for the purpose of the EPA's study, might be contributing factors of food consumption. We attempted to discern whether there were any relationships between these factors and food intake by fitting a linear model to food intake incorporating these factors.

Data of this type must be subjected to careful preliminary examination before trying to fit a model involving many factors with several levels. A model that would normally have main effects could also include many high order interactions. As a result, the model could be very lengthy and complex. Before including a possible interaction we considered its practical significance and also the practicability of incorporating large numbers of (main) factors and interactions into the analysis.

After considering the design of the survey and the objective of our study, we adopted the following statistical model:

$$y_{ijklmpqr} = \mu + R_i + D_{j(i)} + U_{k(i)} + DU_{jk(i)} + S_{l(i)} + P_{m(ijk)} + Sex_p + A_q + (Sex * A)_{pq} + e_{ijklmpqr} \quad \text{(B.1)}$$

where:

$y_{ijklmpqr}$ = the food intake of the rth individual in the qth age group of pth sex (male: p = 1; female: p = 2) in the m(ijk)th PSU and in the lth season;

R_i = the ith census region, i = 1,2,3,4;

$D_{j(i)}$ = the jth geographic division in the ith census region, $j = 1,2,\ldots,n_i$ (since there are a total of 9 divisions, $\Sigma\, n_i = 9$);

$U_{k(i)}$ = the kth urbanization in the ith region, k = 1,2,3;

$DU_{jk(i)}$ = the interaction effect of the jth division with the kth urbanization in the ith region;

$S_{l(i)}$ = the lth season in the region i, l = 1,2,3,4;

$P_{m(ijk)}$ = the mth PSU in the ith region, the jth division and in the kth urbanization,

$m = 1,2,\ldots,T_{ijk}$ (since there are 111 PSUs, $\sum_{i,j,k} T_{ijk} = 111$);

Sex_p = male (p = 1) or female (p = 2)

A_q = the qth age group, q = 1,2,...,10.

$e_{ijklmpqr}$ = the random error associated with $y_{ijklmnpqr}$.

The geographic division in the nested classification was obvious; however, the urbanization and season in the nested classification were not that clear. Since each census region covers vast areas, winter, for example, in one region may not be same as in the others. The same is true for degree of urbanization. P (PSU) was treated as a random factor. Respondents were divided into 10 age groups by year: under 1, 1 to 4, 5 to 9, 10 to 14, 15 to 19, 20 to 24, 25 to 29, 30 to 39, 40 to 59, 60 and over. The random error, e, is assumed to be normally distributed with mean 0 and variance, σ^2.

One of the assumptions for the analysis of variance is that y and e in the model are independent and normally distributed. With this and other conditions, tests of hypothesis are possible, and confidence intervals can be derived. In an analysis of variance we chose the mean square of "PSU," instead of "segment," "household," or "individual," as an estimate of the variance of sampling error. The reasons were as follows:

1) Individuals were subsampling units in the summer, autumn, and winter quarters, but not in the spring quarter.

2) The intake data of individuals from the same household were correlated.

3) For each quarter, the sample of housing units was systematically selected, with a random start, from each area segment without replacement, i.e., the housing units were sampled repeatedly from the same segment for the different quarters. The housing units from the same segment in the different season were not totally independent.

4) The segments from which the housing units were selected were highly clustered. The segments selected were not randomly scattered in the PSU; therefore, neighboring segments were correlated.

Since, as described in A.1.2, PSUs were randomly selected we have used the mean square of PSU as an estimate of the sampling error; the analytical results we obtained would be somewhat conservative.

Because the original design was unbalanced, traditional analysis of variance methods were not applicable. Instead, we have used the general linear model procedures (Searle, 1971). The model in equation (B.1) can be rewritten in matrix form:

$$\underline{Y} = \underline{Xb} + \underline{e} \tag{B.2}$$

where:

$\underline{Y}$ = an N x 1 vector of observations, y, N is the total number of observations in the survey. N = 30770 in the present case. $\underline{Y}$ is normally distributed with mean $\underline{Xb}$, and variance, $\sigma^2 \underline{I}$, $\underline{I}$ is an N x N identity matrix.

$\underline{b}$ = a p x 1 vector of parameters; it is the vector of all parameters of the model.

$\underline{X}$ = an N x p design matrix, 0s and 1s throughout the matrix represent the incidence of terms of the model among the observations, i.e., the incidence of the classifications in which the observations lie.

$\underline{e}$ = an N x 1 vector of random error terms; $\underline{e}$ is normally distributed with mean $\underline{0}$ and variance $\sigma^2 \underline{I}$.

Note that $\underline{X}$ does not have full column rank, neither does $\underline{X}'\underline{X}$. The normal equations corresponding to the model can be derived by least squares, we obtain

$$\underline{X}'\underline{Xb}^\circ = \underline{X}'\underline{Y}. \tag{B.3}$$

Since $\underline{X}'\underline{X}$ is not of full rank, $\underline{X}'\underline{X}$ has no unique inverse, and the normal equations (B.3) have no unique solution. There are many possible solutions. To get any of them, we find a generalized inverse $\underline{G}$ of $\underline{X}'\underline{X}$, and write

$$\underline{b}^\circ = \underline{GX}'\underline{Y}. \tag{B.4}$$

Define

$$\underline{H} = \underline{GX}'\underline{X}. \tag{B.5}$$

The $\underline{b}^\circ$ in equation (B.4) for a solution to normal equation (B.3) emphasizes that what is derived by solving equation (B.3) is only a solution to the equations and not an estimator of $\underline{b}$. This point cannot be overemphasized. In a general discussion of linear models that are not of full rank, it is essential to realize that what is obtained as a solution of the normal equations is just a solution and nothing more. It is misleading and in most cases quite wrong for $\underline{b}^\circ$ to be termed an estimator of $\underline{b}$. The particular solution obtained for $\underline{b}^\circ$ depends entirely on which $\underline{G}$ is used.

Although $\underline{b}^\circ$ is not an estimator of $\underline{b}$, it is, in a practical sense, closely related to $\underline{b}$. In order to establish this relationship, the concept of estimable functions is essential. A linear function of the parameters, $\underline{q}'\underline{b}$, is defined as estimable if it is identically equal to some linear function of the expected value of the vector of observations $\underline{Y}$, i.e., for any $\underline{q}$, if there exists a vector $\underline{t}$ such that $\underline{t}'E(\underline{Y}) = \underline{q}'\underline{b}$, then $\underline{q}'\underline{b}$ is said to be estimable. (E is the expectation operator.) Thus, we are estimating $\underline{q}'\underline{b}$, not $\underline{b}$ alone. In order to estimate $\underline{q}'\underline{b}$, $\underline{q}'$ must satisfy the equation $\underline{q}'\underline{H} = \underline{q}'$, where $\underline{H}$ was defined in equation (B.5). The vector $\underline{q}'$ is, therefore, equal to $\underline{t}'\underline{X}$. The best, linear, unbiased estimator of the estimable function $\underline{q}'\underline{b}$ is $\underline{q}'\underline{b}^\circ$. From the above relationships, the functions of $\underline{Y}$ can be estimated even though $\underline{b}$ cannot. Since only estimable functions have best, linear, unbiased estimators that are invariant to the solution of the normal equations, they are the only functions for which confidence intervals are valid.

One concept closely related to estimable functions is the testable hypothesis. A testable hypothesis is one that can be expressed in terms of estimable functions. If a hypothesis cannot be expressed in terms of estimable functions, it cannot be tested. It cannot be overemphasized that one must first ascertain the estimability (Searle, 1979).

B.2. Main Effects in Model

The estimated means of "effects" in equation (B.1) were weighted means. The weights incorporated in the estimations were intimately related to the distributions of these effects in the population. The distributions of these effects can be divided into three major categories: 1) geographical effects, 2) seasonal effects, and 3) age and sex effects. The distributions of these three categories are:

1) Geographical effects

The geographic effects consist of region, division within region, urbanization within region, and the interaction effects of division and urbanization within region. The sample distributions of these effects were based on the number of strata in the region, division within region, and so on. The frequency distributions are given in Table A-1.

2) Seasons

Each season had the same length of duration, i.e., three months. Their distribution is, therefore, uniform.

3) Age and sex

We generated distributions of age and sex based on the 1969-71 U.S. stationary population (National Center for Health Statistics, 1975) with age- and sex-specific mortality rates and a constant male-to-female birth ratio of 1.051 (Shryock, et al., 1976). The distributions of the resulting stationary population by age and sex are given in Table B-1.

The stationary population was based on the life table technique. A stationary population is a population whose total number and distribution by age do not change with time. The birth rate remains constant for a long period of time, and each cohort of births experiences the current observed mortality rate throughout life. This hypothetical population provides age weighted values which are substantially corrected for those effects such as fluctuating birth or migration rates which affect the age distribution of a population but not its age-specific mortality.

The weighting factors described above were applied to Tables A-1 and B-1 to estimate the individual means shown in Table 3.

Table B-1. Percent Distribution of the Stationary Population by Age and Sex

Age (years)	Total (%)	Male (%)	Female (%)
TOTAL	100.00	48.56	51.44
Under 1	1.40	.72	.68
1 to 4	5.53	2.83	2.70
5 to 9	6.89	3.52	3.37
10 to 14	6.88	3.51	3.37
15 to 19	6.86	3.50	3.36
20 to 24	6.81	3.46	3.35
25 to 29	6.75	3.42	3.33
30 to 39	13.34	6.73	6.61
40 to 59	24.89	12.31	12.58
60 and over	20.65	8.56	12.09

REFERENCES

National Center for Health Statistics, (1975). "United States Life Tables, 1969-71", Vol. 1, No. 1. U.S. Department of Health, Education and Welfare.

Searle, S. R., (1971). "Linear Models," John Wiley & Sons, Inc. New York.

Shryock, H. S. and J. S. Siegel and Associates, (1976). "The Methods and Materials of Demography," Academic Press. New York.