# 21.0 REFERENCES

**Table of Contents**

**Section 21.0**


Page

Questionnaire Development and Methodologic and Statistical Issues

Methodologic Issues and Reviews . . . . . 21-1

Diet Questionnaire Development . . . . . 21-1

Validation and Reproducibility Studies . . . . . 21-2

Descriptive Data on Dietary Factors . . . . . 21-4

Epidemiology . . . . . 21-7

Other Useful References . . . . . 21-9

Article Reprint: A Data-Based Approach to
Diet Questionnaire Design and Testing . . . . . Appendix A

## QUESTIONNAIRE DEVELOPMENT AND METHODOLOGIC AND STATISTICAL ISSUES

### Methodologic Issues and Reviews


1. Block G. Dietary assessment issues related to cancer for NHANES III. In: Briefel RB, Sempos CT, eds. Dietary Methodology Workshop for the Third National Health and Nutrition Examination Survey. Hyattsville, MD: National Center for Health Statistics, Series 4, No. 27, 1992; 24-31.

2. Block G, Hartman AM. Issues in reproducibility and validity of dietary studies. Am J Clin Nutr 1989; 50:1133-1138.

3. Block G, Sinha R, Gridley G. Collection of dietary supplement data and implications for analyses. Am J Clin Nutr 1994: (In press).

4. Brown CC, Kipnis V, Freedman LS, Hartman AM, Schatzkin A, Wacholder S. Energy adjustment method for nutritional epidemiology: the effect of categorization. Am J Epidemiol 1994; 139: (In press).

5. Fox TA, Heimendinger J, Block G. Telephone surveys as a method for obtaining dietary information: a review. J Am Diet Assoc 1992; 92:729-732.

6. Harlan LC, Block G. Use of adjustment factors with a brief food frequency questionnaire to obtain nutrient values. Epidemiol 1990; 1:224-231.

7. Hartman AM, Block G. Dietary assessment methods for macronutrients. In: Micozzi MS, Moon TE, Eds. Macronutrients Investigating their Role in Cancer. New York: Marcel Dekker, Inc. 1992; 87-124.

8. Hartman AM, Brown CC, Palmgren J, Pietinen P, Verkasalo M, Myer D, Virtamo J. Variability in nutrient and food intakes among older middle-aged men: implications for design of epidemiologic and validation studies using food recording. Am J Epidemiol 1990; 132:999-1012.

9. Kipnis V, Freedman LS, Brown CC, Hartman A, Schatzkin A, Wacholder S. Interpretation of energy adjustment models for nutritional epidemiology. Am J Epidemiol 1993; 137:1376-1380.

10. Potosky A, Block G, Hartman AM. The apparent validity of diet questionnaires is influenced by number of diet-record days used for comparison. J Am Diet Assoc 1990; 90:810-813.


### Diet Questionnaire Development


1. Block G. A review of validations of dietary assessment methods. Am J Epidemiol 1982; 115:492-505.

2. Block G, Clifford C, Naughton MD, Henderson M, McAdams M. A brief dietary screen for high fat intake. J of Nutr Educ 1989; 21:199-207.

3. Block G, Hartman AM, Dresser CM, Carroll MD, Gannon J, Gardner L. A data-based approach to diet questionnaire design and testing. Am J Epidemiol 1986; 124:453-469.

4. Block G, Hartman AM, Naughton D. A reduced dietary questionnaire: development and validation. Epidemiol 1990; 1:58-64.

5. National Health Interview Survey (NHIS). Cancer Epidemiology Supplement, Sections C, D, and F (Food Frequency; Vitamin and Mineral Intake; and Food Knowledge). National Center for Health Statistics, Hyattsville, Maryland, 1992.

6. Public Use Data Tape Documentation, Model gram and Nutrient Composition, Tape Numbers 5702 and 5703, National Center for Health Statistics, Hyattsville, Maryland, 1982.

7. Smucker R, Block G, Coyle L, Harvin A, Kessler L. A dietary and risk factor questionnaire and analysis system for personal computers. Am J Epidemiol 1989; 129:445-449.

8. United States Department of Agriculture. Human Nutrition Information Service. Composition of Foods: Raw, Processed, Prepared (Revised Handbooks # 8-1, Dairy and Egg Products-1976 to # 8-21, Fast Foods - 1988). Nutrition Monitoring Division. Washington, DC: U.S. GPO, 1976-1988.

**Validation and Reproducibility Studies**

1. Block G, Hartman AM, Naughton D. A reduced dietary questionnaire: development and validation. Epidemiol 1990; 1:58-64.

2. Block G, Thompson FE, Hartman AM, Larkin FA, Guire KE. Comparison of two dietary questionnaires validated against multiple dietary records collected during a 1-year period. J Am Diet Assoc 1992; 92:686-693.

3. Block G, Woods M, Potosky A, Clifford C. Validation of a self-administered diet history questionnaire using multiple diet records. J Clin Epidemiol 1990; 43:1327-1335.

4. Coates RJ, Eley JW, Block G, Gunter EW, Sowell AL, Grossman C, Greenberg RS. An evaluation of a food frequency questionnaire for assessing dietary intake of specific carotenoids and vitamin E among low-income Black women. Am J Epidemiol 1991; 134:658-671.

5. Cummings SR, Block G, McHenry K, Baron RB. Evaluation of two food frequency methods of measuring dietary calcium intake. Am J Epidemiol 1987; 126:796-802.

6. DiSogra C, Block G. WIC Dietary Assessment Validation Study. 1993:(UnPub).

7. Hartman AM, Block G, Chan W, Williams J, McAdams M, Banks Jr. WL, Robbins A. Reproducibility of a self-administered diet history questionnaire over different seasons. (Submitted for publication 1993).

8. Leighton J, Neugut AI, Block G. A comparison of face-to-face food frequency interviews with telephone interviews and self-administered questionnaires. Am J Epidemiol 1988; 128:891. (Abstract).

9. Mares-Perlman JA, Klein BEK, Klein R, Ritter LL, Fisher MR, Freudenheim JL. A diet history questionnaire ranks nutrient intakes in middle-aged and older men and women similarly to multiple food records. J Nutr 1993; 123:489-501.

10. Pietinen P, Hartman AM, Haapa E, Rasanen L, Haapakoski J, Palmgren J, Albanes D, Virtamo J, Huttunen JK. Reproducibility and validity of dietary assessment instruments: I. A self-administered food use questionnaire with a portion size picture booklet. Am J Epidemiol 1988; 128:655-666.

11. Pietinen P, Hartman AM, Haapa E, Rasanen L, Haapakoski J, Palmgren J, Albanes D, Virtamo J, Huttunen JK. Reproducibility and validity of dietary assessment instruments: II. A qualitative food frequency questionnaire. Am J Epidemiol 1988; 128:667-676.

12. Sobell J, Block G, Koslowe P, Tobin J, Andres R. Validation of a retrospective questionnaire assessing diet 10-15 years ago. Am J Epidemiol 1989; 130:173-187.

## DESCRIPTIVE DATA ON DIETARY FACTORS


1. Block G. Antioxidant intake in the U.S. J Toxicology and Ind Health 1993; 9:295-301.

2. Block G. Nutrient sources of carotene in the American diet. Am J Epidemiol 1994; (In press).

3. Block G, Abrams B. Vitamin and mineral status of women of childbearing potential. Ann NY Acad Sci 1993; 678:244-254.

4. Block G, Cox C, Madans J, Schreiber GB, Licitra L, Melia N. Vitamin supplement use, by demographic characteristics. Am J Epidemiol 1988; 127:297-309.

5. Block G, Dresser CM, Hartman AM, Carroll MD. Nutrient sources in the American diet: quantitative data from the NHANES II survey. I. Vitamins and minerals. Am J Epidemiol 1985; 122:13-26.

6. Block G, Dresser CM, Hartman AM, Carroll MD. Nutrient sources in the American diet: quantitative data from the NHANES II survey. II. Macronutrients and fats. Am J Epidemiol 1985; 122:27-40.

7. Block G, Lanza E. Dietary fiber sources in the United States by demographic group. J Natl Cancer Inst 1987; 79:83-91.

8. Block G, Mangels AR, Levander OA, Patterson BH, Campbell W, Norkus E, Taylor PR. Plasma reduced and oxidized ascorbate attained on various levels of dietary intake. FASEB J 1991; 5:A1444 (Abstract).

9. Block G, Norris JC, Mandel RM, DiSogra C. Sources of 7 nutrients in the diets of low-income Hispanic-American women and children: Quantitative data from the HHANES survey, 1982-1984. J Am Dietet Assoc 1994; (In press).

10. Block G, Rosenberger WF, Patterson BH. Calories, fat, and cholesterol: intake patterns in the U.S. population by race, sex and age. Am J Public Health 1988; 78:1150-1155.

11. Block G, Sorenson A. Vitamin C intake and dietary sources by demographic characteristics. Nutr Cancer 1987; 10:53-65.

12. Block G, Subar AF. Estimates of nutrient intake from a food frequency questionnaire: The 1987 National Health Interview Survey. J Am Diet Assoc 1992; 92:969-977.

13. Dickinson VA, Block G, Russek-Cohen E. Supplement use, other dietary and demographic variables, and serum vitamin C in NHANES II. J Am Col Nutr 1994; (In press).

14. Flagg EW, Coates RJ, Eley W, Jones DP, Gunter EW, Byers TE, Block G, Greenberg RS. Dietary glutathione intake in humans and the relation between intake and plasma glutathione level. Nutr Cancer (In press).

15. Hartman AM, Brown CC, Palmgren J, Pietinen P, Verkasalo M, Myer D, Virtamo J. Variability in nutrient and food intakes among older middle-aged men: implications for design of epidemiologic and validation studies using food recording. Am J Epidemiol 1990; 132:999-1012.

16. Jones DP, Coates RJ, Flagg FW, Eley JW, Block G, Greenberg RS, Gunter EW, Jackson B. Glutathione in food listed in the National Cancer Institute's food frequency questionnaire. Nutr Cancer 1992:17:57-75.

17. Kant AK, Block G. Dietary vitamin B-6 intake and food sources in the U.S. population: NHANES II, 1976-1980. Am J Clin Nutr 1990; 52:707-716.

18. Kant AK, Block G, Schatzkin A, Ziegler R, Nestle M. Dietary diversity in the U.S. population: NHANES II, 1976-80. J Am Diet Assoc 1991; 91:1526-1531.

19. Kant AK, Schatzkin A, Block G, Ziegler R, Nestle M. Food group intake patterns and associated nutrient profiles of the U.S. population. J Am Diet Assoc 1991; 91:1532-1537.

20. Lanza E, Jones DY, Block G. Dietary fiber intake in the U.S. population. Am J Clin Nutr 1987; 46:790-797.

21. Mangels AR, Block G, Frey CM, Patterson BH, Taylor PR, Norkus EP, Levander OA. The bioavailability to humans of ascorbic acid from oranges, orange juice and cooked broccoli is similar to that of synthetic ascorbic acid. J Nutr 1993; 123:1054-1061.

22. Mangels AR, Block G, Levander OA, Taylor PR, Morris VC, Patterson BH. Differences in plasma beta-carotene (P-BC) maintained despite similar dietary beta-carotene. FASEB J 1991; 5:A1322 (Abstract).

23. Mares-Perlman JA, Klein BEK, Klein R, Ritter LL, Freudenheim JL, Luby MH. Nutrient supplements contribute to the dietary intake of middle- and older-aged adult residents of Beaver Dam, Wisconsin. J Nutr 1993; 123:176-188.

24. Mares-Perlman JA, Subar AF, Block G, Greger JL, Luby MH. Zinc intake and sources in the U.S. adult population: 1978-1980. 1994; (In preparation).

25. Murphy SP, Subar AF, Block G. Vitamin E intakes and sources in the United States. Am J Clin Nutr 1990; 52:361-367.

26. Patterson BH, Block G. Food choices and the cancer guidelines. Am J Public Health 1988; 78:282-286.

27. Patterson BH, Block G, Rosenberger WF, Pee D, Kahle LL. Fruit and vegetables in the American diet: data from the NHANES II survey. Am J Public Health 1990; 80:1443-1449.

28. Pickle LW, Hartman AM. Indicator foods for vitamin A assessment. Nutr Cancer 1985; 7:3-23.

29. Sinha R, Block G, Taylor PR. Determinants of plasma ascorbic acid in a healthy male population. Cancer Epidemiology, Biomarkers and Prevention 1992; 1:297-302.

30. Sinha R, Block G, Taylor PR. Problems with estimating vitamin C intakes. Am J Clin Nutr 1993; 57:547-550.

31. Sinha R, Patterson BH, Block G, Mangels AR, Levander OA, Taylor PR. Determinants of plasma Vitamin E in healthy males. Cancer Epidemiology, Biomarkers and Prevention 1993; 2:473-479.

32. Subar AF, Block G. Use of vitamin and mineral supplements: demographics and amounts of nutrients consumed. The 1987 Health Interview Survey. Am J Epidemiol 1990; 132:1091-1101.

33. Subar AF, Block G, James LD. Folate intake and food sources in the U.S. population. Am J Clin Nutr 1989; 50:508-516.

34. Vanderslice JT, Higgs D, Hayes JM, Block G. Ascorbic acid and dehydroascorbic acid content of food-as-eaten. J Food Comp and Anal 1990; 3:105-118.

## EPIDEMIOLOGY


1. Block G. The data support a role for antioxidants in reducing cancer risk. Nutr Rev 1992; 50:207-213.

2. Block G. Vitamin C and cancer prevention: The epidemiologic evidence. Am J Clin Nutr 1991; 53:270S-282S.

3. Block G. Vitamin C and reduced mortality. Epidemiol 1992; 3:189-191.

4. Block G, Hartman AM. Understanding the results of epidemiologic studies. Seminars in Oncology 1983; 10:257-263.

5. Block G, Menkes M. Ascorbic acid in cancer prevention. In: Moon TE, Micozzi MS., eds. Diet and Cancer Prevention. Investigating the Role of Micronutrients. New York: Marcel Dekker, Inc., 1989; 341-388.

6. Block G, Patterson B, Subar A. Fruit, vegetables, and cancer prevention: A review of the epidemiologic evidence. Nutr Cancer 1992; 18:1-29.

7. Gridley G, McLaughlin JK, Block G, Blot WJ, Gluch M, Fraumeni JF. Vitamin supplement use and reduced risk of oral and pharyngeal cancer. Am J Epidemiol 1992; 135:1083-1092.

8. Gridley G, McLaughlin JK, Block G, Blot WJ, Winn DM, Greenberg RS, Schoenberg JB, Preston-Martin S, Austin DF, Fraumeni JF, Jr. Diet and oral and pharyngeal cancer among blacks. Nutr Cancer 1990; 14:219-225.

9. Harlan LC, Coates RJ, Block G, Greenberg RS, Ershow A, Forman M, Austin DF, Chen V, Heymsfield SB. Estrogen receptor status and dietary intakes in breast cancer patients. Epidemiol 1993; 4:25-31.

10. Hyman L, He O, Grimson R, Oden N, Leske MC, Schachat AP, Block G and Risk Factors Study Group. Nutritional factors and age-related macular degeneration (AMD): results of a case-control study. International Journal of Eye Research 1991; (Abstract).

11. Jones DY, Schatzkin A, Green SB, Block G, Brinton LA, Ziegler RG, Hoover R, Taylor PR. Dietary fat and breast cancer in the National Health and Nutrition Examination Survey I Epidemiologic Follow-up Study. J Natl Cancer Inst 1987; 79:465-471.

12. Kant AK, Schatzkin A, Harris TB, Ziegler RG, Block G. Dietary diversity and subsequent mortality in the First National Health and Nutrition Examination Survey Epidemiologic Follow-Up Study. Am J Clin Nutr 1993; 57:434-440.

13. McLaughlin JK, Gridley G, Block G, Winn DM, Preston-Martin S, Schoenberg JB, Greenberg RS, Stemhagen A, Austin DF, Ershow AG, Blot WJ Fraumeni JF. Dietary factors in oral and pharyngeal cancer. J Natl Cancer Inst 1988; 80:1237-1243.

14. Schiffman MH, Andrews AW, Van Tassell RL, Smith L, Daniel J, Robinson A, Hoover RN, Rosenthal J, Weil R, Nair PP, Schwartz S, Pettigrew H, Block G, Batist G, Shaw R, Wilkins T. A case-control study of colorectal cancer and fecal mutagenicity. Cancer Res 1989; 49:3420-3424.

15. Taylor P, Schiffman M, Jones DY, Judd J, Schatzkin A, Nair PP, Van Tassell R, Block G. Relation of changes in amount and type of dietary fat to fecapentaenes in premenopausal women. Elsevier 1988; 206:3-9.

## OTHER USEFUL REFERENCES


1. American Cancer Society, Prevention Study II, 1983.

2. American Thoracic Society/NHLBI Division of Lung Diseases, Epidemiology Standardization Project, 7/78. (Also published in Am Rev Resp Dis 1978; 118(6):7-53.

3. Berkman LF and Syme SL. Social networks, host resistance, and mortality: A nine-year follow-up study of Alameda county residents. Am J Epidemiol 1979; 109:186-204.

4. Carriquiry AL, Jensen HH, Dodd KW, Nusser SM, Guenther PM, Fuller WA. Estimating usual nutrient distributions. Submitted to Am J Clin Nutr 1993.

5. Chug-Ahuja JK, Holden JM, Forman MR, Mangels AR, Beecher GR, Lanza E. The development and application of a carotenoid database for fruits, vegetables, and selected multicomponent foods. J Am Diet Assoc 1993; 93:318-323.

6. Decoulfe P. "Occupation", in Schottenfeld D and Fraumeni JF. Cancer Epidemiology and Prevention, WB Saunders Company, Philadelphia: WB Saunders Company, 1982.

7. Doll R and Peto R. The causes of cancer: Quantitative estimates of avoidable risks of cancer in the United States today. J Natl Cancer Inst 1981; 66:1191-1308.

8. Freedman LS, Carroll RJ, Wax Y. Estimating the relation between dietary intake obtained from a food frequency questionnaire and true average intake. Am J Epidemiol 1991; 134:310-320.

9. Gordis L. Assuring the quality of questionnaire data in epidemiologic research. Am J Epidemiol 1979; 109:21-24.

10. Guidelines for Statistics and Information on Effects of the Environment on Health, National Center for Health Statistics, Hyattsville, Maryland, 1980.

11. Haenszel W, Loveland DB, Sirken MG. Lung cancer mortality as related to residence and smoking histories. I. White males. J Natl Cancer Inst 1962; 28:947-1001.

12. House JS, Robbins C, and Metzner HL. The association of social relationships and activities with mortality: Prospective evidence from the Tecumseh Community Health Study. Am J Epidemiol 1982; 116:123-140.

13. Lilienfeld AM and Lilienfeld DE. Foundations of Epidemiology. New York: Oxford University Press, Inc., 1980; pp. 14 and 291.

14. Mangels AR, Holden JM, Beecher GR, Forman MR, Lanza E. Carotenoid content of fruits and vegetables: An evaluation of analytic data. J Am Diet Assoc 1993; 93:284-296.

15. McGuire C and White GD. The measurement of social status. Research Paper in Human Development #3, Department of Educational Psychology, University of Texas, 1955.

16. NHANES II, National Health and Nutrition Examination Survey, National Center for Health Statistics, Hyattsville, Maryland, 1976-1980.

17. National Health Interview Survey. Current Estimates from the National Health Interview Survey; United States 1981, Series 10, Number 141, National Center for Health Statistics, Hyattsville, Maryland.

18. National Health Interview Survey (NHIS) Health Practices Supplement. National Center for Health Statistics, Hyattsville, Maryland, 1983.

19. National Health Interview Survey (NHIS) Core questionnaire, National Center for Health Statistics, Hyattsville, Maryland, 1979.

20. National Health Interview Survey (NHIS) Smoking supplement, National Center for Health Statistics, Hyattsville, Maryland 1979.

21. National Health Interview Survey (NHIS) Cancer Epidemiology Supplement, Section H: Smoking Habits, National Center for Health Statistics, Hyattsville, Maryland, 1992.

22. Spiegelman D, Gray R. Cost efficient study designs for binary response data with Gaussian covariate measurement error. Biometrics 1991; 47:851-869.

23. Surgeon General's Report on Smoking and Health, 1979.

24. Walker AM, Blettner M. Comparing imperfect measures of exposure. Am J Epidemiol 1985; 121:783-790.

25. Walter Reed Army Institute of Research, A Comprehensive Health Hazards and Protectors Inventory, 12/81.

26. Warner WL, Meeker M, and Eells K. Social Class in America; Chicago Social Research Associates, 1949.

27. Willett W. Nutritional Epidemiology. New York: Oxford University Press, 1989.

28. Wingard DL, Berkman LF, Brand RJ. A multivariate analysis of Health related practices: A nine-year, mortality, follow-up of the Alameda County Study. Am J Epidemiol 1982; 116:765-775.

29. Wynder EL and Gori GB. Contribution of the environment to cancer incidence: An epidemiologic exercise. J Natl Cancer Inst 1977; 58:625-632.

AMERICAN JOURNAL OF EPIDEMIOLOGY
Vol. 124, No. 3
Printed in U.S.A.
Copyright © 1986 by The Johns Hopkins University School of Hygiene and Public Health
All rights reserved

# A DATA-BASED APPROACH TO DIET QUESTIONNAIRE DESIGN AND TESTING

GLADYS BLOCK,[1] ANNE M. HARTMAN,[1] CONNIE M. DRESSER,[2] MARGARET D. CARROLL,[2] JANET GANNON,[3] AND LILLY GARDNER[3]

**Block, G. (NCI, Bethesda, MD 20892-4200), A. M. Hartman, C. M. Dresser, M. D. Carroll, J. Gannon, and L. Gardner. A data-based approach to diet questionnaire design and testing. Am J Epiderniol 1986; 124:453-69.**

**A self-administered diet history questionnaire has been developed for epidemiologic and clinical use. Both the food list and the nutrient values to be associated with it were developed using dietary data from 11,658 adult respondents to the Second National Health and Nutrition Examination Survey (NHANES II). Food items were selected on the basis of their contribution to total population intake of energy and each of 17 nutrients in the NHANES II data, and represent over 90% of each of those nutrients. Associated nutrient composition values were determined from the NHANES II database using frequency of consumption data in that survey. Portion sizes to be associated with each food item were derived from observed portion size distributions in NHANES II, based on three-dimensional models. The resulting food list and its corresponding brief data base, when used to calculate nutrients from a diet record, yielded correlations of $r >$ 0.70 with the more detailed method. Field administration produced mean values comparable to national data.**

**diet; epidemiologic methods; food; nutrition; questionnaires**

Current dietary assessment approaches are each appropriate in different research situations. The single 24-hour recall is relatively easy to administer, but does not provide a valid reflection of an *individual's* usual diet. Multiple 24-hour recalls, a traditional diet history interview, or a multiday diet record may provide valid assessments of an individual's usual intake, but require a high level of commitment of nutritional expertise and respondent time (1, 2). In many epidemiologic and clinical research situations, however, what is needed is an instrument which is both valid at the individual level and is easier and less expensive to administer. In order to meet this need, many investigators have developed frequency questionnaires and checklists (3-8). The instrument that we describe in this paper, a self-administered diet history, is a

Received for publication August 12, 1985 and in final form January 8, 1986.

Abbreviation: NHANES II, Second National Health and Nutrition Examination Survey.

[1] Division of Cancer Prevention and Control, National Cancer Institute, Bethesda, MD.

[2] Nutrition Statistics Branch, Division of Health Examination Statistics, National Center for Health Statistics, Hyattsville, MD.

[3] Capital Systems Group, Washington, DC.

Send reprint requests to Dr. Gladys Block, NCI/DCPC, Blair Building, Room 515,9000 Rockville Pike, Bethesda, MD 20892-4200.

Acknowledgements: Dietary data based on actual self-administration of the questionnaire, reported in tables 6 and 7, were obtained from the following contributors: the Isotretinoin Basal Cell Cancer Prevention Study, a multi-center trial funded by the intramural program of the Division of Cancer Prevention and Control, National Cancer Institute; the Beta-carotene, Retinol and Lung Cancer Chemoprevention Study conducted at the University of Texas Health Center at Tyler; and the Massey Cancer Center Diet Study, Medical College of Virginia. The contribution of these investigators is gratefully acknowledged. The authors also wish to acknowledge the assistance of Michael Selmon in the software development, and of Lisa Carter for her help with the manuscript.

continuation of those efforts. We believe that the methodology used to develop the instrument may be as useful as the instrument itself, in that it made use of a data-based approach to select not only the food list, but also the quantitation which must accompany such a list, i.e., the nutrient content of each food, and the portion-size attributions. Such a data-based approach may permit investigators to take advantage of large-scale representative national surveys to design improved dietary questionnaires.

For epidemiologic and clinical research, the ability to rank individuals by relative levels of nutrient intake permits the evaluation of etiologic hypotheses and interactions. However, for public health and clinical purposes, as well as for the refinement of etiologic hypotheses, the estimation of absolute levels of nutrient intake is also important. Both of these functions, the ability to rank and the ability to estimate absolute nutrient level, were considered in the design of this instrument. A data-based approach may make this latter goal more achievable than it has been thought to be in the past.

A number of other purposes and constraints have guided the design of this questionnaire. First, and most importantly, the questionnaire must be representative of an *individual's* usual diet, so that associations of dietary factors with clinical signs or health outcomes can be fully and validly interpreted. Second, it should be relatively brief, 20-25 minutes, since, in many epidemiologic studies or clinical trials, investigators are reluctant to impose additional burdens on their subjects or staff.

Third, the instrument should be capable of assessing nutrients as well as foods or food groups, since the expansion and refinement of our understanding requires that we examine hypotheses at the nutrient level. In addition, it must permit the assessment of a broad range of nutrients of future as well as current interest. Only if an instrument can be analyzed for a wide range of nutrients can the role of those nutrients be examined, or the complexities of nutrient, biochemical, and physiologic interactions begin to be understood.

Finally, the questionnaire was also designed with the intention that it would be able to assess the diet of a variety of demographic groups, through inclusion of foods important to such groups. Such an instrument would allow for direct comparability between studies, and permit better analyses of apparent risk differences.

## METHODS

Both the list of foods and the nutrients to be associated with them have been developed by means of a data-based approach, using dietary intake data from the Second National Health and Nutrition Examination Survey (NHANES II) (9-11). The NHANES II survey was a large-scale stratified multi-stage probability sample representative of the United States population. The survey collected 24-hour dietary recalls, with data on generic and brand names and methods of food preparation, as well as portion size of each reported food using three-dimensional food models. After omission of imputed, unreliable or surrogate data, detailed dietary intake data were available on 11,658 adults ages 19–74 years. Sample weights calculated by the National Center for Health Statistics were used in these calculations, to eliminate potential biases resulting from nonresponse and intentional oversampling.

### *Development of the list of foods*

Foods reported by these NHANES II respondents (2,244 different food codes) were grouped into 147 conceptually similar food items (for example, 11 green bean codes became one food item, "green beans"). Criteria for grouping or keeping foods separate included conceptual similarity, respondents' ability to make the necessary distinctions, similarity in nutrient content per *usual serving (not* per 100 g), importance of a particular food to our ability to correctly classify an individual with respect to nutrient intake, and approximate

number of persons at risk of such misclassification.

For energy and each of the 17 nutrients on the NHANES II database (listed in table 2), the per cent of that nutrient which was contributed by each food item was determined, and foods were ranked in order of their contribution. This work has been described in detail elsewhere (10, 11). The amount contributed by any food item thus takes into account not only its nutrient composition, but also the frequency of its consumption by the population, and typical portion size. This approach permits the detection of some foods which are sometimes overlooked as important nutrient contributors. For example, of all the vitamin A or provitamin A consumed by these 11,658 respondents, 4.0 per cent of it was contributed by vegetable or tomato soup, and this item ranked fourth in importance among all contributors of vitamin A. "Vegetable or tomato soup" is thus an important candidate for inclusion, as a separate item, in the food list. This approach leads to inclusion of some infrequently eaten foods, such as liverwurst, if their frequency and nutrient content results in an important population contribution. Other infrequently eaten foods which are sometimes included on questionnaires, such as mangoes (eaten by 32 per 10,000 (11)), do not meet this test and are not included.

Foods were included on the questionnaire if they made an important contribution to the population's intake of energy and each of 17 nutrients in the NHANES II database. Attention to an adequate capture of total calories also would enhance the capture of a wide range of nutrients which could become of interest in the future. For example, we can be reasonably confident that vitamin "x" is well-represented, since the food list contains foods representing 93 per cent of the nation's total caloric intake, over 90 per cent of each of a wide range of nutrients, and 94 per cent of all food items reported in the NHANES II data. For vitamin "x" to be poorly represented in the food list, it would have to be found primarily in foods infrequently eaten and in foods which are not important to the contribution of calories, fat, protein, carbohydrates, or the 14 other important nutrients.

The food list was then augmented to ensure adequate assessment of dietary fiber intake, intake of major cruciferous vegetables, foods with suspected health implications (e.g., coffee/tea or artificial sweeteners), foods important in geographic or ethnic subgroups (e.g., chili peppers), and foods whose omission might lead to misclassification of the relatively few persons who consume them (e.g., winter squash).

Finally, the food list was modified following a pretest. Some fruits, for example, were separated into "fresh, in season" and "canned/frozen", because respondents had difficulty with a combined item.

In addition to the line-item food list, the questionnaire includes a prompted open-ended question where the respondent can indicate other frequently eaten foods. This permits the capture of additional foods important to a particular individual or demographic group. Many common responses are preceded for easy data handling.

Table 1 presents the list of foods included as line items or open-ended prompts on the questionnaire. Table 2 presents, for energy and each nutrient in the NHANES II database, the percentage of the total population intake which is represented by this food list. The basic list, excluding items in the open-ended section, contains foods which represent 93 per cent of the population's intake of calories, 96 per cent of vitamin A, 96 per cent of vitamin C, 95 per cent of total fat, and at least 90 per cent of each of the other nutrients.

Figure 1 presents an example of the format used with the self-administered diet history food list. The respondent may indicate his usual pattern of consumption as number of times per day, week, month, or year. (Accompanying instructions provide a clear illustration of the manner of response. ) Provision of a "year" category permits the capture of some foods which may make an important contribution to average

TABLE 1

*Foods included in main food list or mentioned in prompted open-ended section of the self-administered diet history*

*LINE-ITEM FOOD LIST*

*Fruits*

Apples, applesauce, pears
Bananas
Peaches, apricots (canned or frozen, whole year)
Peaches, fresh (in season)
Cantaloupe (in season)
Cantaloupe (out of season)
Watermelon (in season)
Strawberries (frozen or canned, whole year)
Strawberries (fresh, in season)
Oranges
Orange juice
Grapefruit, grapefruit juice
Tang, Start breakfast drinks
Other fruit juices, fortified fruit drinks
Any other fruit, including berries, fruit cocktail

*Vegetables*

String beans, green beans
Peas
Other beans such as baked beans, pintos, kidney beans, limas
Corn
Mixed vegetables
Winter squash, baked squash
Tomatoes, tomato juice
Red chilies, powdered/crushed, sauces—exclude picante, taco sauce
Broccoli
Cauliflower or brussels sprouts
Spinach (raw)
Spinach (cooked)
Mustard greens, turnip greens, collards
Cole slaw, cabbage, sauerkraut
Carrots
Green salad
French fries and fried potatoes
Sweet potatoes, yams
Other potatoes, including boiled, baked, mashed
Rice
Any other vegetable, including cooked onions, summer squash
Butter, margarine or other fat added to vegetables

*Meat*

Hamburgers, cheeseburgers, meat loaf
Beef—steaks, roasts
Beef stew or pot pie with carrots, other vegetables
Liver, including chicken livers
Pork, including chops, roasts
Fried chicken
Chicken or turkey, baked, stewed or broiled
Fried fish or fish sandwich
Tuna fish, tuna salad, tuna casserole
Shellfish (shrimp, lobster, crab, oysters, etc)
Other fish, broiled, baked

*Mixed dishes*

Spaghetti, lasagna, other pasta with tomato sauce
Pizza
Mixed dishes with cheese (such as macaroni and cheese)
Liverwurst
Hot dogs
Ham, lunch meats
Vegetable soup, vegetable beef, minestrone, tomato soup
Other soups

*Breads/salty snacks/spreads*

White bread, rolls, crackers (including sandwiches)
Dark bread, including whole wheat, rye, pumpernickel
Corn bread, corn muffins, corn tortillas
Salty snacks (such as chips, popcorn)
Peanuts, peanut butter
Butter on bread or rolls
Margarine on bread or rolls
Salad dressing, mayonnaise (including on sandwiches)
Gravies made with meat drippings, or white sauce

*Breakfast foods*

High fiber, bran or granola cereals, shredded wheat

*Breakfast foods—cont'd.*

Highly fortified cereals, such as Product 19, Total, or Most
Other cold cereals, such as Corn Flakes, Rice Krispies
Cooked cereals
Sugar added to cereal
Eggs
Bacon
Sausage

*Sweets*

Ice cream
Doughnuts, cookies, cake, pastry
Pumpkin pie, sweet potato pie
Other pies
Chocolate candy
Other candy, jelly, honey, brown sugar

*Dairy products*

Cottage cheese
Other cheeses and cheese spreads
Flavored yogurt
Whole milk and beverages with whole milk
2% milk and beverages with 2% milk
Skim milk, 1% milk or buttermilk

*Beverages*

Regular soft drinks
Diet soft drinks
Beer
Wine
Liquor
Decaffeinated coffee
Coffee, not decaffeinated
Tea (hot or iced)
Nondairy creamer in coffee or tea
Milk in coffee or tea
Cream (real) in coffee or tea
Sugar in coffee or tea
Artificial sweetener in coffee or tea

*PRE-CODED FOODS FOR OPEN-ENDED SECTION*

| | | |
|---|---|---|
| | Milkshake | Beets |
| Veal, lamb | Other dairy product | Pineapple or pineapple juice |
| Chili con carne | Other dessert, sweet | Prunes or prune juice |
| Tofu | Sour cream, dips | Hi-C |
| Mixed dish with meat | Diet salad dressing | Cranberry juice cocktail |
| Mixed dish with chicken | Catsup | Grapes |
| Chinese dishes | Green chilies, jalapenos | Mangoes |
| Mexican dishes | Salsa picante, taco sauce | Papayas |
| Seafood creole | Onions | Honeydew/cassaba melon |
| Refried beans or bean burritos | Summer squash | Lemone/lemon juice |
| Polish or Italian sausage | Asparagus | Nuts and seeds |
| Cream soups | Sweet green peppers | Bran |
| Noodles | Sweet red peppers | Other vegetable/fruit |
| Pancakes, waffles | Bean sprouts | Other not mentioned here |
| Instant breakfast, Metrecal | Avocado, guacamole | |
| Pudding | | |

intake even though infrequently eaten, such as liver or sweet potatoes. Furthermore, some respondents appear to prefer this flexibility, since it permits them to report sporadic or nonregular consumption of some foods, such as seasonal items. The format also requires the respondent to indicate whether his usual serving size is small, medium, or large with respect to a stated medium serving. Such a portion size assessment is important for at least some age-sex groups and nutrients (12).

The questionnaire includes questions on restaurant foods, frequency and type of fat added in cooking or at the table, and the consumption of the skin on chicken or the fat on meat. In the computer program developed to accompany the questionnaire, nutrient estimates are adjusted based on responses to these questions. Questions on special diets and vitamin supplements are also included.

### *Development of the nutrient quantities*

The quantitative values to be assigned for each food on the questionnaire are as

TABLE 2

*Per cent of total United States population intake represented by line-item food list of the self-administered diet history, for energy and 17 nutrients. Based on NHANES II dietary intake data on 11,658 persons, 1976-1980*

| Energy or nutrient | % of total population intake |
|---|---|
| Energy (Kcal) | 92.9 |
| Protein | 95.0 |
| Total fat | 94.7 |
| Carbohydrate | 92.1 |
| Calcium | 96.5 |
| Phosphorus | 95.3 |
| Iron | 92.4 |
| Sodium | 92.2 |
| Potassium | 94.3 |
| Vitamin A | 96.4 |
| Thiamin | 93.6 |
| Riboflavin | 95.2 |
| Niacin | 94.6 |
| Vitamin C | 95.7 |
| Saturated fat | 95.0 |
| Oleic acid | 91.9 |
| Linoleic acid | 90.0 |
| Cholesterol | 96.9 |

| FRUITS & JUICES | Medium Serving | Your Serving Size S | M | L | How Often? Day | Week | Month | Year | Rarely/Never | Office Use |
|---|---|---|---|---|---|---|---|---|---|---|
| EXAMPLE Apples, applesauce, pears | (1) or ½ cup | | ✓ | | | 4 | | | | |
| Apples, applesauce, pears | (1) or ½ cup | | | | | | | | | 11* |
| Bananas | (1) medium | | | | | | | | | 15 |
| Peaches, apricots (canned, frozen or dried, whole year) | (1) or ½ cup | | | | | | | | | 19 |
| Peaches, apricots, nectarines (fresh, in season) | (1) medium | | | | | | | | | 23 |
| Cantaloupe (in season) | ¼ medium | | | | | | | | | 27 |
| Cantaloupe (out of season) | ¼ medium | | | | | | | | | 31 |
| Watermelon (in season) | 1 slice | | | | | | | | | 35 |
| Strawberries (frozen or canned, whole year) | ½ cup | | | | | | | | | 39 |
| Strawberries (fresh, in season) | ½ cup | | | | | | | | | 43 |
| Oranges | 1 medium | | | | | | | | | 47 |
| Orange juice | 6 oz. glass | | | | | | | | | 51 |
| Grapefruit, grapefruit juice | (½) or 6 oz. | | | | | | | | | 55 |
| Tang, Start breakfast drinks | 6 oz. glass | | | | | | | | | 59 |

FIGURE 1. An example of part of the self-administered diet history questionnaire.

crucial to the success of the instrument as the food list itself. These values were also developed in a data-based manner, making use of the frequency and portion-size information from the NHANES II survey. A person reporting consumption of one medium serving of green beans on the self-administered diet history is assigned a nutrient value which is the product of the median serving size of that food, in grams, and the median nutrient per 100 g from the NHANES 11 data, both of which were developed as described below.

*Median serving size in grams.* For each food item on the questionnaire (e.g., among all of the servings of green beans reported by the 11,658 NHANES II respondents), the portion sizes of all respondents were ranked, and the median serving size in grams was determined. Small and large portions were also determined from the actual distribution of portion sizes in these data. (It should be noted that this is not simply a mechanical process; a considerable element of judgment remains important in the determination of portion sizes.) This data-based approach permits more accurate attribution of nutrient amounts in a quantitative frequency questionnaire, by taking advantage of the extensive NHANES II population-based data which used three-dimensional models to determine portion size. These calculations were also carried out to determine portion sizes for six age-sex groups (male/female, 19–29, 30–54, 55+ years) to reflect observed age and sex differences in intake. Thus, portion sizes used by the computer program can be age-sex-specific.

*Nutrients per 100 g.* The 11 green bean codes could each have had a different value of, for example, IUS of vitamin A per 100 g, and thus it is necessary to select one nutrient content value to be used in the calculations. The error inherent in assigning a single value—unavoidable in an instrument of this sort—was minimized by using the median value, that nutrient content for which half of the consumers of green beans ate a variety of green beans richer in vitamin A, and half ate a variety poorer in vitamin A. Again, this took advantage of the population-based data on frequency of consumption, so that codes representing infrequently-eaten foods were not permitted to exert an undue influence on the nutrient content which would be assigned. This approach minimizes the error considerably more than does a weighted mean.

Table 2 indicates that the list is comprehensive. However, the ability of a question-

naire to make accurate nutrient estimates depends not only on its comprehensiveness, but on the precision of the nutrient assignments which can be made from that list of foods. For example, a food item such as "pasta" may be comprehensive but affords little precision, since it cannot permit the differentiation between carotenoids (in pasta with tomato sauces), retinol (in pasta with cheese sauces), or neither (in plain pasta). The accuracy of the nutrient estimates also depends intimately on the exact nature of the quantitation method for each food. Consequently, the precision of the nutrient assignments derivable from the food list was evaluated, also in a data-based approach, as described below.

### *Data-based evaluation of the precision of nutrient assignments*

The performance of a diet history or frequency questionnaire depends on the answers to two broad questions: 1) How accurately can the individual report on his frequency of consumption and his portion sizes? And 2) how adequate is the food list itself, and its associated quantitation? Investigators who combine these two questions deny themselves the opportunity to perfect their instruments to the greatest extent possible prior to adding the imperfection of the human response to it. The following section describes the process by which the instrument itself—the food list and quantitation—were evaluated and improved. It does not address the question of how well an individual can report on his frequency (a question which arises for all self-reports of diet). Thus, the results should not be taken to imply that this instrument can *necessarily* capture an individual's usual diet, since that requires some (as yet unknown) degree of accuracy in the human response. This approach does, however, permit investigators to evaluate and improve the instrument itself. It addresses a limited but essential question: "If respondents were able to respond accurately about their diet, could this food list and associated quantitation adequately represent individual nutrient intake, in spite of the diversity of dietary behavior?" Having demonstrated the adequacy of the specific food list and the associated portion size and nutrient composition assignments for each food, one can then proceed to field validations in which the human component is evaluated. Validations which combine these two stages into a single step always leave unresolved the uncertainty as to whether fair or poor performance of the instrument resulted from the inability of people to report their diet reliably, or from an inadequate food list or quantitation. If the stages are kept separate, the sources of error can be better pinpointed, and improvements more appropriately targeted.

The methods and results presented in tables 3–5 represent such a first-stage or developmental evaluation, in a population different from that in which they were developed, of the adequacy of the food list and the nutrient values assigned to each food. Second-stage or field validations have also been carried out, and are reported elsewhere (13, 14).

In this first-stage evaluation, the food list and quantitation were examined with respect to their ability to produce nutrient estimates similar to those of a 24-hour diet record. Diet records had been obtained previously from 50 healthy individuals, 33 women and 17 men, ages 25–77 years. They were coded by a trained coder, using the respondents' reported portion sizes and a detailed nutrient data bank based on US Department of Agriculture data (15). These same diet records were then recoded using the food list and quantitations developed for the self-administered diet history food list, as *if* the respondents had used it to report on their diet for that 24-hour period. In this way, imperfect human response is eliminated as a source of error, permitting a clean evaluation and improvement of the instrument itself. (One-day records may be preferable to longer records for this stage of the evaluation, since what is being evaluated is the food list and portion size/

content assumptions. If longer records were used, the agreement of the two methods might be inflated simply because of the impact of frequency of consumption, and the precision of the instrument's list and quantitations could not be as accurately evaluated.)

A comparison of the nutrient estimates calculated by the two methods permits an evaluation of three major sources of error inherent in the use of a diet instrument of this sort-the adequacy of the food list itself, the nutrient content assumptions for each food, and the portion size attributions for each food. Thus, the nutrient estimates calculated by the diet record and the self-administered diet history food list quantitations could fail to agree for any of three reasons. First, if the food list did not contain an item into which a reported food could be coded, the individual would not be credited with any nutrient values for that food on the diet history. For example, an individual who had consumed veal would get the appropriate nutrient values on the diet record, but would get no nutrient values for that food in the calculation based on the self-administered diet history food list, since veal is not a line item in the questionnaire. (It is, however, prompted for in the open-ended section.) Second, if, for example, an individual had eaten an extremely lean or extremely fat cut of beef, he would get the appropriate value on the diet record calculations, while the self-administered diet history calculations would assume an intermediate value directed by the NHANES II data on the proportion consuming various cuts of beef. And, third, an individual who reported eating a 12-ounce cut of beef would get the appropriate nutrient values for 12 ounces on the diet record, while in the self-administered diet history calculation he would get nutrient values resulting from the median portion size for "beef".

Thus, a comparison of the resulting nutrient totals represents an assessment of whether the food list for the self-administered diet history is extensive and inclusive enough to adequately represent nutrients as consumed in the varied diets of individuals, and of whether the standardized nutrient assignments can adequately represent the nutrient intake of individuals who consume a wide range of portion sizes and specific food types.

Two types of quantitation of the self-administered diet history are evaluated here: 1) portion sizes which consist of a single median value for each food based on population medians observed in the NHANES II data, and 2) portion sizes which comprise small, medium, and large values, again derived from the NHANES II data.

The latter quantitation represents an attempt to assess whether and to what degree the assignment of standardized small, medium, and large portion sizes would be capable of enhancing the agreement between the two methods. That is, if people could accurately report that their usual portion size was large with respect to the stated medium portion, how much would such a small, medium, and large portion quantitation improve the estimates? This approach can identify the degree of potential improvement as a result of this simplified portion size assessment.

Nutrients were summed by the diet record method calculated in the usual way, and by portion size and nutrient composition quantitations developed for the self-administered diet history method, for each individual. The paired *t* test was used to test the hypothesis of no difference in the means between each of the two self-administered diet history quantitation methods (single median and small, medium, large) and the diary reference. Pearson correlation coefficients (r) were used to estimate the strength of the linear relationship between each self-administered diet history scoring method and the diary method (16). The data for vitamin A and vitamin C were logarithmically transformed so as to reduce skewness and improve normality, as re-

quired for the proper usage of parametric statistics and hypothesis testing.

Kappa (17), a measure of agreement for categoric data corrected for chance, was used to evaluate how well the self-administered diet history quantitation methods could categorize individuals into quartiles of the nutrient distributions when compared with the food record categorization. Quartiles were determined from each method's own nutrient distribution. In addition to Kappa for exact agreement, weighted Kappa was also calculated, in which agreement plus or minus one category was considered as full agreement. Weighted in this fashion, Kappa allows for less than perfect agreement and is more realistic for most epidemiologic purposes.

In addition to the developmental evaluations described above, data are also presented on mean nutrient estimates obtained from actual field use of the questionnaire by approximately 1,000 volunteers or healthy participants in cancer prevention studies (see "Acknowledgements"). After administration of the questionnaire, nutrient estimates were calculated using the nutrient composition and portion-size data base developed as described above, and a computer program developed by the authors. For comparison, sample-weighted mean nutrient values from the NHANES II data are presented for the relevant age and sex groups. Standard errors of the means for the NHANES II data were estimated using SESUDAAN (18, 19), a program which takes complexity of the sample design into account.

The questionnaire described here is in use, in both self-administered and interviewer-administered formats. It takes approximately 20–25 minutes to self-administer, although some respondents take longer. Coding takes less than 15 minutes, and requires no special expertise. The nutrient analysis program contains features that permit the study of several methodological issues in dietary assessment, in addition to performance of the nutrient analysis. The nutrient analysis program and the associated nutrient composition and portion-size array may be obtained from the authors upon request.

## RESULTS

Tables 3-5 present results of the first-stage evaluations for several nutrients which served as the focus of the early developmental process. They are nutrients o: current research interest, are distribute differently in the diet, and are nutrients or which questionnaires typically perform dissimilarly. That is, food lists which assess vitamins A and C adequately often do less well for calories and fat, and vice versa.

At the group level, table 3 presents mean nutrient amounts as estimated by the actual diet records of the 50 healthy individuals and as estimated when the food list and quantitation developed for it are used to score the diet records. (In a few cases, a comparison of mean and median suggests that outliers may have had an important effect on the mean value, a common observation in dietary data (20). ) Use of a standard medium portion appears to produce accurate estimates only for females. For males, it produces statistically significant underestimates for energy and the nutrients examined. When even three levels of portion size (small, medium, large) are used, however, the quantitation produces values which are not statistically different from the diet record, for nine of the ten age-sex-specific nutrient estimates.

At the individual level, use of the single median portion to quantify the foods produced correlation coefficients ranging from 0.54 for fat to 0.90 for vitamin A (table 4), while use of the quantitation based on three levels of portion size strengthened the correlation coefficients to the range of 0.73 for fat and 0.94 for vitamin A. Spearman rank order correlations were also examined and produced similar results. Table 5 indicates that when the self-administered diet history food list and the quantitation based on three levels of portion size were used to

TABLE 3

*Mean energy and nutrient values estimated by one-day diet record and two methods of quantitation using the self-administered diet history (SADH) food list to score the diet records*

| Energy or nutrient, by sex | One-day diet record (mean) | Quantitation method for SADH food list | |
|---|---|---|---|
| | | Single median portion size† (mean) | Small, medium, and large portion sizes‡ (mean) |
| Energy (Kcal) | | | |
| Males | 2,260 | 1,773** | 2,103 |
| Females | 1,644 | 1,510 | 1,508 |
| Total fat (g) | | | |
| Males | 88.5 | 71.7* | 85.1 |
| Females | 73.6 | 66.8 | 67.4 |
| % Kcal from fat | | | |
| Males | 35.2 | 36.8 | 36.4 |
| Females | 39.7 | 39.7 | 40.4 |
| Vitamin A§ (IU) | | | |
| Males | 8,290 | 6,527* | 7,288 |
| Females | 5,224 | 5,131 | 4,855 |
| Vitamin C§ (mg) | | | |
| Males | 97.8 | 64.6* | 71.3* |
| Females | 65.9 | 58.9 | 58.8 |

† For SADH, servings mentioned on diet record were assigned a standard portion size based on NHANES II dietary data.

‡ For SADH, servings mentioned on diet record were assigned small, medium, or large portion sizes based on distributions observed in NHANES II data.

§ Log-transformed to improve normality.

* $p \leq 0.05$ for difference between SADH and diet record.

** $p < 0.01$ for difference between SADH and diet record.

TABLE 4

*Correlations between nutrients estimated by one-day diet record and nutrients estimated by two methods of quantitation using the self-administered diet history (SADH) food list to score the diet records*

| Energy or nutrient | Quantitation method for SADH food list | |
|---|---|---|
| | Single median portion size* ($r$)‡ | Small, medium, and large portion sizes† ($r$)‡ |
| Energy (Kcal) | 0.61 | 0.82 |
| Fat | 0.54 | 0.73 |
| % Kcal from fat | 0.74 | 0.81 |
| Vitamin A§ | 0.90 | 0.94 |
| Vitamin C§ | 0.77 | 0.86 |

* Based on NHANES II dietary data.

† Based on distributions observed in NHANES II dietary data.

‡ Pearson product-moment correlation.

§ Log-transformed to improve normality.

Note: $p < 0.01$ for all values shown.

score the diet record, at least 80 per cent were placed in the same or an adjacent quartile of the distribution, and none were grossly misclassified. These developmental evaluations resulted in the addition of a few foods to the food list or open-ended section, and in the review and revision of the portion-size assumptions to improve their accuracy.

The above results address the adequacy of the self-administered diet history food list and its associated nutrient and portion size assumptions, but do not evaluate the potential error introduced by the respondent himself, since the questionnaire was not actually administered to the respondent. In tables 6 and 7, on the other hand, mean values resulting from actual administrations of the questionnaire are compared with data observed in the NHANES II sur-

TABLE 5

*Per cent of nutrients correctly classified by quartiles, based on one-day diet record and two methods of quantitation using the self-administered diet history food list to score the diet records*

| Energy or nutrient, by quartiles included | Quantitation method for food list | | | |
|---|---|---|---|---|
| | Single median portion size† | | Small, medium, and large portion sizes‡ | |
| | Exact quartile (%) | Exact ± 1 category (%) | Exact quartile (%) | Exact ± 1 category (%) |
| All quartiles | | | | |
| Energy | 36§ | 82 | 64 | 92 |
| Fat | 44 | 82 | 58 | 94 |
| % Kcal from fat | 48 | 88 | 54 | 90 |
| Vitamin A | 62 | 96 | 70 | 94 |
| Vitamin C | 54 | 90 | 58 | 94 |
| Highest or lowest quartile on record | | | | |
| Energy | 44 | 84 | 72 | 96 |
| Fat | 52 | 80 | 68 | 92 |
| % Kcal from fat | 60 | 96 | 64 | 96 |
| Vitamin A | 80 | 92 | 84 | 96 |
| Vitamin C | 64 | 92 | 68 | 96 |

† Based on NHANES II dietary data.

‡ Based on distributions observed in NHANES II dietary data.

§ Kappa = 0.15, $p = 0.308$. For all other values, significance of Kappa statistic $p < 0.01$.

Note: Expectations by chance alone: All quartiles, exact agreement, 25%; all quartiles, exact ± 1 category, 56%; highest or lowest quartile, exact agreement, 25%; highest or lowest quartile, exact ± 1 category, 50%.

vey, for several age and sex groups. The questionnaire, in actual use by respondents, produces mean nutrient estimates which are quite similar to those found in national data, and is able to detect and reproduce the major differences by age and sex. The young consume considerably more calories than the old, for example, and men consume more than women at every age, and this is clearly reflected in the estimates produced by the questionnaire and its associated quantitation. None of the calorie estimates is statistically significantly different from the corresponding national mean value for that age and sex.

For many of the specific nutrients (including the fats, iron, and the B vitamins), few or none of the estimates differ significantly from the national values, either in absolute terms or on a per 1,000 kilocalorie basis. Some nutrient estimates in tables 6 and 7 are significantly different, however. These include higher estimates of calcium intake among women and older men, lower reported cholesterol intakes in older men, and higher intakes of vitamins A and C in all age and sex groups. One cannot be certain how much of these differences may be due to under- or overestimates by the questionnaire, and how much may be due to the different points in time and populations studied. These particular nutrients have been the subjects of considerable popular attention in the 6–10 years since the NHANES II data were collected, and the observed differences are in directions consistent with educational campaigns and popular trends (down for cholesterol, up for calcium and vitamins A and C). Furthermore, study participants tend to be more affluent and more health conscious than the general population (21), an observation consistent with the differences noted above. The apparently elevated vitamin A values are also consistent with those found by Willett et al. (6) and Hinds et al. (22) and others (23-25), who found mean values ranging from 8,000 to 12,000 IUs with diet

TABLE 6

*Energy and nutrient estimates resulting from administration of the self-administered diet history (small, medium, and large portions quantitation method) ($SADH_{SML}$) to healthy clinical trial participants and volunteers (1984–1985), compared with relevant national values from the NHANES II survey (1976–1980)*

| Sex and age (years) | Energy (Kcal) | Protein (gm) | Fat (gm) | Carbohydrate (gm) | Calcium (mg) | Phosphorus (mg) | Iron (mg) | Sodium (mg) | Potassium (mg) | Vitamin A (IU) | Vitamin $B_1$ (mg) | Vitamin $B_2$ (mg) | Niacin (mg) | Vitamin C (mg) | Saturated fatty acids (gm) | Oleic acid (gm) | Linoleic acid (gm) | Cholesterol (mg) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Males | | | | | | | | | | | | | | | | | | |
| 25–34 | | | | | | | | | | | | | | | | | | |
| $SADH_{SML}$ | 2,643 | 103 | 109 | 282 | 973 | 1,664 | 17 | 3,651 | 3,395 | 8,560 | 1.5 | 2.3 | 27 | 157 | 42 | 41 | 15 | 484 |
| ($n$ = 35) | (137)* | (5) | (6) | (16) | (67) | (88) | (1) | (194) | (189) | (887) | (0.1) | (0.1) | (2) | (21) | (3) | (3) | (1) | (40) |
| NHANES | 2,746 | 108 | 112 | 291 | 978 | 1,783 | 17 | 3,702 | 3,210 | 5,812 | 1.7 | 2.4 | 29 | 107 | 41 | 42 | 16 | 454 |
| ($n$ = 1,034) | (52) | (2) | (3) | (5) | (32) | (38) | (0.4) | (76) | (67) | (398) | (0.1) | (0.1) | (0.5) | (4) | (1) | (1) | (1) | (11) |
| 35–44 | | | | | | | | | | | | | | | | | | |
| $SADH_{SML}$ | 2,351 | 93 | 98 | 242 | 806 | 1,446 | 16 | 3,260 | 3,068 | 7,286 | 1.4 | 2.1 | 25 | 116 | 37 | 37 | 14 | 461 |
| ($n$ = 116) | (66) | (3) | (3) | (8) | (35) | (42) | (0.5) | (103) | (93) | (365) | (0.1) | (0.1) | (1) | (9) | (1) | (1) | (1) | (20) |
| NHANES | 2,434 | 95 | 104 | 250 | 867 | 1,536 | 16 | 3,410 | 3,002 | 5,548 | 1.6 | 2.1 | 25 | 96 | 38 | 39 | 14 | 431 |
| ($n$ = 711) | (47) | (2) | (3) | (6) | (27) | (29) | (0.3) | (93) | (64) | (321) | (0.1) | (0.1) | (0.5) | (4) | (1) | (1) | (1) | (14) |
| 45–54 | | | | | | | | | | | | | | | | | | |
| $SADH_{SML}$ | 2,249 | 88 | 95 | 239 | 818 | 1,411 | 16 | 3,442 | 3,023 | 8,071 | 1.4 | 2.1 | 24 | 113 | 34 | 36 | 14 | 450 |
| ($n$ = 135) | (73) | (3) | (4) | (8) | (33) | (45) | (0.5) | (118) | (97) | (484) | (0.1) | (0.1) | (1) | (6) | (1) | (1) | (1) | (22) |
| NHANES | 2,371 | 93 | 99 | 245 | 848 | 1,506 | 16 | 3,566 | 2,892 | 6,154 | 1.5 | 2.1 | 25 | 101 | 36 | 37 | 13 | 449 |
| ($n$ = 668) | (48) | (2) | (2) | (7) | (24) | (30) | (0.3) | (106) | (60) | (359) | (0.1) | (0.1) | (0.5) | (5) | (1) | (1) | (1) | (15) |
| 55–64 | | | | | | | | | | | | | | | | | | |
| $SADH_{SML}$ | 2,139 | 84 | 88 | 232 | 919 | 1,422 | 16 | 3,351 | 3,155 | 9,061 | 1.5 | 2.1 | 22 | 153 | 32 | 33 | 13 | 402 |
| ($n$ = 154) | (62) | (2) | (3) | (7) | (35) | (40) | (0.5) | (103) | (85) | (355) | (0.1) | (0.1) | (1) | (8) | (1) | (1) | (1) | (17) |
| NHANES | 2,076 | 84 | 86 | 217 | 804 | 1,368 | 15 | 3,056 | 2,674 | 5,821 | 1.4 | 1.9 | 22 | 104 | 31 | 33 | 12 | 427 |
| ($n$ = 1,185) | (26) | (2) | (1) | (4) | (26) | (24) | (0.3) | (49) | (48) | (240) | (0.1) | (0.1) | (0.4) | (3) | (1) | (1) | (1) | (10) |
| 65+ | | | | | | | | | | | | | | | | | | |
| $SADH_{SML}$ | 1,883 | 72 | 77 | 204 | 869 | 1,271 | 14 | 2,957 | 2,949 | 8,786 | 1.3 | 1.9 | 18 | 154 | 28 | 28 | 11 | 351 |
| ($n$ = 132) | (69) | (3) | (3) | (7) | (42) | (48) | (0.5) | (118) | (114) | (477) | (0.1) | (0.1) | (1) | (11) | (1) | (1) | (1) | (19) |
| NHANES | 1,818 | 73 | 75 | 201 | 698 | 1,194 | 14 | 2,872 | 2,399 | 6,588 | 1.3 | 1.8 | 20 | 100 | 27 | 28 | 10 | 383 |
| ($n$ = 1,133) | (24) | (1) | (1) | (3) | (14) | (15) | (0.2) | (48) | (31) | (372) | (0.1) | (0.1) | (0.3) | (4) | (1) | (1) | (1) | (9) |
| Females | | | | | | | | | | | | | | | | | | |
| 25–34 | | | | | | | | | | | | | | | | | | |
| $SADH_{SML}$ | 1,672 | 71 | 61 | 194 | 1,078 | 1,370 | 12 | 2,712 | 2,969 | 8,529 | 1.2 | 2.1 | 17 | 169 | 22 | 22 | 9 | 260 |
| ($n$ = 104) | (54) | (2) | (3) | (7) | (45) | (44) | (0.4) | (80) | (98) | (358) | (0.1) | (0.1) | (1) | (8) | (1) | (1) | (1) | (13) |
| NHANES | 1,633 | 63 | 67 | 185 | 631 | 1,053 | 11 | 2,392 | 2,085 | 4,705 | 1.1 | 1.5 | 16 | 91 | 24 | 25 | 10 | 275 |
| ($n$ = 1,121) | (25) | (1) | (1) | (3) | (22) | (21) | (0.2) | (52) | (46) | (318) | (0.1) | (0.1) | (0.4) | (4) | (1) | (1) | (1) | (8) |
| 35–44 | | | | | | | | | | | | | | | | | | |
| $SADH_{SML}$ | 1,573 | 67 | 62 | 172 | 1,005 | 1,277 | 11 | 2,623 | 2,926 | 9,098 | 1.1 | 1.9 | 16 | 158 | 21 | 22 | 10 | 277 |
| ($n$ = 44) | (77) | (3) | (4) | (10) | (73) | (72) | (1) | (136) | (161) | (756) | (0.1) | (0.1) | (1) | (14) | (1) | (1) | (1) | (23) |
| NHANES | 1,584 | 62 | 67 | 172 | 601 | 1,023 | 11 | 2,378 | 2,093 | 4,839 | 1.0 | 1.4 | 17 | 84 | 23 | 24 | 10 | 294 |
| ($n$ = 814) | (29) | (2) | (2) | (3) | (21) | (24) | (0.3) | (54) | (43) | (413) | (0.1) | (0.1) | (0.4) | (3) | (1) | (1) | (1) | (9) |

| | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 45–54 | | | | | | | | | | | | | | | | | | | | | |
| $SADH_{SML}$ | 1,569 | 65 | 61 | 172 | 860 | 1,217 | 13 | 2,633 | 2,698 | 7,964 | 1.2 | 1.9 | 17 | 145 | 22 | 22 | 8 | 270 |
| ($n$ = 40) | (100) | (5) | (4) | (14) | (89) | (104) | (1) | (209) | (167) | (588) | (0.1) | (0.2) | (2) | (13) | (2) | (1) | (1) | (35) |
| NHANES | 1,443 | 58 | 60 | 159 | 587 | 958 | 10 | 2,202 | 2,044 | 5,370 | 1.0 | 1.4 | 16 | 98 | 21 | 22 | 9 | 286 |
| ($n$ = 728) | (27) | (1) | (1) | (3) | (19) | (22) | (0.1) | (53) | (48) | (326) | (0.1) | (0.1) | (0.3) | (5) | (1) | (1) | (1) | (12) |
| 55–64 | | | | | | | | | | | | | | | | | | | | | |
| $SADH_{SML}$ | 1,472 | 66 | 56 | 164 | 850 | 1,194 | 12 | 2,388 | 2,752 | 8,670 | 1.1 | 1.9 | 17 | 164 | 19 | 20 | 9 | 306 |
| ($n$ = 64) | (61) | (3) | (3) | (7) | (51) | (62) | (1) | (113) | (114) | (527) | (0.1) | (0.1) | (1) | (11) | (1) | (1) | (1) | (43) |
| NHANES | 1,402 | 55 | 57 | 160 | 564 | 920 | 11 | 2,183 | 2,034 | 5,518 | 1.0 | 1.4 | 15 | 106 | 20 | 21 | 8 | 265 |
| ($n$ = 1,277) | (19) | (1) | (1) | (2) | (12) | (12) | (0.2) | (35) | (30) | (238) | (0.1) | (0.1) | (0.3) | (3) | (1) | (1) | (1) | (8) |
| 65+ | | | | | | | | | | | | | | | | | | | | | |
| $SADH_{SML}$ | 1,437 | 63 | 54 | 164 | 995 | 1,286 | 12 | 2,399 | 2,845 | 9,066 | 1.2 | 2.1 | 16 | 149 | 19 | 19 | 8 | 257 |
| ($n$ = 35) | (109) | (5) | (4) | (16) | (148) | (158) | (1) | (176) | (252) | (896) | (0.1) | (0.3) | (2) | (12) | (2) | (2) | (1) | (20) |
| NHANES | 1,301 | 51 | 51 | 159 | 543 | 884 | 10 | 1,995 | 1,975 | 5,559 | 1.0 | 1.4 | 15 | 106 | 17 | 19 | 8 | 240 |
| ($n$ = 1,355) | (18) | (1) | (1) | (2) | (12) | (13) | (0.2) | (38) | (32) | (246) | (0.1) | (0.1) | (0.3) | (3) | (1) | (1) | (1) | (6) |

* Standard error of the mean.

history or frequency methods. Final conclusions with regard to the degree to which this questionnaire may over- or underestimate some nutrients await further validation efforts, some of which are reported elsewhere (13, 14).

## Discussion

The purpose of this paper has been to describe a methodological approach which may make it possible to improve questionnaire design by making use of large population data bases. Several differences between this diet assessment instrument and previous questionnaires will be discussed below. However, it should be noted that, with one or two exceptions, all of the features have been used before, by different investigators, in different combinations, or on a different scale. Thus, this instrument has much in common with and owes a great deal to the work of such investigators as Hankin et al. (4, 5), Willett et al. (6, 7), Jain et al. (8), and Morgan et al. (3).

The following developmental and design features are notable.

*Targeted toward assessment of the whole diet.* The developmental work was aimed at assessing not one or a few similar nutrients, but the whole spectrum of micro- and macronutrients. This approach, used with a food list in which individual foods are quite distinguishable, will make it possible to reanalyze data sets if new nutrients become of research interest in the future.

*Selection of the food list based on population data on nutrient contributors.* This approach has resulted in the inclusion of a few foods sometimes overlooked as potentially important contributors of a given nutrient, and the exclusion of a few foods whose infrequent consumption is not sufficiently outweighed by nutrient composition or portion size to warrant their inclusion in a compact assessment instrument. The use of population data on nutrient contribution to select food items for inclusion has been used by other investigators (5, 26). Hankin et al. (5), for example, se-

TABLE 7

*Nutrients per 1,000 kilocalories resulting from administration of the self-administered diet history (small, medium, and large portions quantitation method) ($SADH_{SML}$) to healthy clinical trial participants and volunteers (1984–1985), compared with relevant national values from the NHANES II survey (1976–1980)*

| Sex and age (years) | Protein (gm/Kcal) | Fat (gm/Kcal) | Carbohydrate (gm/Kcal) | Calcium (mg/Kcal) | Phosphorus (mg/Kcal) | Iron (mg/Kcal) | Sodium (mg/Kcal) | Potassium (mg/Kcal) | Vitamin A (IU/Kcal) | Vitamin $B_1$ (mg/Kcal) | Vitamin $B_2$ (mg/Kcal) | Niacin (mg/Kcal) | Vitamin C (mg/Kcal) | Saturated fatty acids (gm/Kcal) | Oleic acid (gm/Kcal) | Linoleic acid (gm/Kcal) | Cholesterol (mg/Kcal) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Males | | | | | | | | | | | | | | | | | |
| 25–34 | | | | | | | | | | | | | | | | | |
| $SADH_{SML}$ | 39 | 41 | 108 | 388 | 639 | 6 | 1,413 | 1,302 | 3,289 | 0.6 | 0.9 | 10 | 60 | 16 | 15 | 6 | 182 |
| ($n$ = 35) | (1)* | (1) | (3) | (32) | (23) | (0.2) | (57) | (47) | (284) | (0.1) | (0.1) | (0.3) | (7) | (0.5) | (0.5) | (0.3) | (12) |
| NHANES | 40 | 40 | 107 | 358 | 652 | 6 | 1,371 | 1,200 | 2,192 | 0.6 | 0.9 | 11 | 42 | 15 | 15 | 6 | 170 |
| ($n$ = 1,304) | (0.5) | (0.4) | (0.8) | (12) | (10) | (0.1) | (23) | (19) | (132) | (0.1) | (0.1) | (0.2) | (2) | (0.2) | (0.2) | (0.1) | (3) |
| 35–44 | | | | | | | | | | | | | | | | | |
| $SADH_{SML}$ | 40 | 42 | 103 | 351 | 623 | 7 | 1,397 | 1,339 | 3,180 | 0.6 | 0.9 | 11 | 50 | 16 | 16 | 6 | 197 |
| ($n$ = 116) | (1) | (1) | (2) | (12) | (10) | (0.1) | (25) | (32) | (166) | (0.1) | (0.1) | (0.2) | (3) | (0.3) | (0.3) | (0.1) | (7) |
| NHANES | 40 | 42 | 105 | 364 | 650 | 7 | 1,428 | 1,279 | 2,442 | 0.7 | 0.9 | 11 | 42 | 15 | 15 | 6 | 181 |
| ($n$ = 711) | (0.7) | (0.5) | (1.2) | (9) | (9) | (0.2) | (27) | (18) | (157) | (0.1) | (0.1) | (0.5) | (2) | (0.2) | (0.2) | (0.1) | (6) |
| 45–54 | | | | | | | | | | | | | | | | | |
| $SADH_{SML}$ | 40 | 42 | 107 | 373 | 638 | 7 | 1,550 | 1,375 | 3,660 | 0.6 | 0.9 | 11 | 52 | 15 | 16 | 6 | 202 |
| ($n$ = 135) | (1) | (1) | (2) | (12) | (10) | (0.1) | (26) | (25) | (167) | (0.1) | (0.1) | (0.2) | (2) | (0.3) | (0.3) | (0.1) | (8) |
| NHANES | 40 | 41 | 105 | 369 | 653 | 7 | 1,561 | 1,273 | 2,908 | 0.7 | 0.9 | 11 | 48 | 15 | 15 | 5 | 197 |
| ($n$ = 668) | (0.5) | (0.4) | (1.1) | (9) | (9) | (0.1) | (31) | (19) | (193) | (0.1) | (0.1) | (0.2) | (2) | (0.2) | (0.2) | (0.1) | (6) |
| 55–64 | | | | | | | | | | | | | | | | | |
| $SADH_{SML}$ | 40 | 41 | 109 | 437 | 679 | 7 | 1,579 | 1,524 | 4,422 | 0.7 | 1.0 | 11 | 76 | 15 | 15 | 6 | 188 |
| ($n$ = 154) | (1) | (1) | (2) | (12) | (10) | (0.1) | (23) | (28) | (158) | (0.1) | (0.1) | (0.2) | (4) | (0.3) | (0.3) | (0.2) | (6) |
| NHANES | 41 | 41 | 106 | 393 | 674 | 7 | 1,514 | 1,328 | 2,949 | 0.7 | 0.9 | 11 | 52 | 15 | 15 | 6 | 214 |
| ($n$ = 1,185) | (0.5) | (0.4) | (0.9) | (9) | (7) | (0.1) | (19) | (15) | (119) | (0.1) | (0.1) | (0.1) | (1) | (0.1) | (0.2) | (0.1) | (5) |
| 65+ | | | | | | | | | | | | | | | | | |
| $SADH_{SML}$ | 39 | 40 | 110 | 464 | 681 | 7 | 1,574 | 1,586 | 4,591 | 0.7 | 1.0 | 10 | 82 | 15 | 15 | 6 | 188 |
| ($n$ = 132) | (1) | (1) | (2) | (13) | (10) | (0.1) | (26) | (27) | (145) | (0.1) | (0.1) | (0.2) | (4) | (0.3) | (0.3) | (0.2) | (7) |
| NHANES | 41 | 41 | 112 | 390 | 665 | 8 | 1,626 | 1,349 | 3,787 | 0.7 | 1.0 | 11 | 59 | 14 | 15 | 6 | 222 |
| ($n$ = 1,133) | (0.5) | (0.3) | (1) | (7) | (5) | (0.1) | (24) | (13) | (202) | (0.1) | (0.1) | (0.2) | (2) | (0.2) | (0.2) | (0.1) | (5) |
| Females | | | | | | | | | | | | | | | | | |
| 25–34 | | | | | | | | | | | | | | | | | |
| $SADH_{SML}$ | 44 | 36 | 116 | 649 | 832 | 7 | 1,656 | 1,812 | 5,261 | 0.7 | 1.2 | 10 | 103 | 13 | 13 | 5 | 157 |
| ($n$ = 104) | (1) | (1) | (2) | (21) | (17) | (0.1) | (28) | (39) | (202) | (0.1) | (0.1) | (0.2) | (4) | (0.3) | (0.3) | (0.2) | (7) |
| NHANES | 40 | 40 | 114 | 390 | 656 | 7 | 1,536 | 1,337 | 3,056 | 0.7 | 0.9 | 11 | 59 | 14 | 15 | 6 | 172 |
| ($n$ = 1,121) | (0.4) | (0.4) | (0.9) | (11) | (9) | (0.1) | (44) | (26) | (161) | (0.1) | (0.1) | (0.2) | (2) | (0.2) | (0.2) | (0.1) | (4) |
| 35–44 | | | | | | | | | | | | | | | | | |
| $SADH_{SML}$ | 43 | 39 | 108 | 624 | 807 | 7 | 1,669 | 1,875 | 5,877 | 0.7 | 1.2 | 10 | 102 | 13 | 14 | 6 | 180 |
| ($n$ = 44) | (1) | (1) | (3) | (32) | (24) | (0.2) | (45) | (69) | (408) | (0.1) | (0.1) | (0.4) | (8) | (0.5) | (0.5) | (0.3) | (16) |
| NHANES | 40 | 41 | 110 | 384 | 662 | 7 | 1,521 | 1,398 | 3,170 | 0.7 | 0.9 | 12 | 56 | 14 | 15 | 6 | 191 |
| ($n$ = 814) | (0.5) | (0.3) | (1) | (11) | (11) | (0.2) | (25) | (34) | (246) | (0.1) | (0.1) | (0.3) | (2) | (0.2) | (0.2) | (0.1) | (4) |

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 45–54 | | | | | | | | | | | | | | | | | | | |
| $SADH_{SML}$ | 41 | 40 | 109 | 537 | 754 | 8 | 1,649 | 1,744 | 5,291 | 0.7 | 1.1 | 11 | 98 | 14 | 15 | 6 | 176 |
| ($n$ = 40) | (1) | (1) | (3) | (34) | (27) | (0.3) | (46) | (57) | (342) | (0.1) | (0.1) | (0.5) | (9) | (0.6) | (0.7) | (0.3) | (15) |
| NHANES | 41 | 41 | 112 | 406 | 674 | 8 | 1,559 | 1,491 | 3,938 | 0.7 | 1.0 | 11 | 74 | 14 | 15 | 6 | 200 |
| ($n$ = 728) | (0.6) | (0.5) | (1.2) | (10) | (10) | (0.1) | (33) | (36) | (251) | (0.1) | (0.1) | (0.2) | (4) | (0.2) | (0.3) | (0.1) | (7) |
| 55–64 | | | | | | | | | | | | | | | | | | | |
| $SADH_{SML}$ | 45 | 38 | 112 | 575 | 808 | 8 | 1,641 | 1,920 | 6,161 | 0.8 | 1.3 | 12 | 115 | 13 | 14 | 6 | 196 |
| ($n$ = 64) | (1) | (1) | (3) | (24) | (21) | (0.3) | (44) | (52) | (357) | (0.1) | (0.1) | (0.4) | (7) | (0.5) | (0.5) | (0.3) | (16) |
| NHANES | 40 | 40 | 116 | 405 | 669 | 8 | 1,591 | 1,525 | 4,143 | 0.7 | 1.0 | 11 | 82 | 14 | 14 | 6 | 194 |
| ($n$ = 1,277) | (0.4) | (0.3) | (0.9) | (9) | (7) | (0.1) | (21) | (22) | (177) | (0.1) | (0.1) | (0.2) | (3) | (0.1) | (0.2) | (0.1) | (5) |
| 65+ | | | | | | | | | | | | | | | | | | | |
| $SADH_{SML}$ | 44 | 38 | 111 | 652 | 864 | 8 | 1,681 | 1,995 | 6,418 | 0.8 | 1.4 | 11 | 113 | 13 | 14 | 6 | 189 |
| ($n$ = 35) | (2) | (1) | (3) | (57) | (51) | (0.3) | (46) | (75) | (450) | (0.1) | (0.1) | (0.4) | (8) | (0.6) | (0.6) | (0.4) | (15) |
| NHANES | 40 | 38 | 123 | 425 | 696 | 8 | 1,580 | 1,569 | 4,560 | 0.8 | 1.1 | 12 | 86 | 13 | 14 | 6 | 192 |
| ($n$ = 1,355) | (0.4) | (0.4) | (1) | (9) | (8) | (0.1) | (24) | (21) | (188) | (0.1) | (0.1) | (0.2) | (2) | (0.2) | (0.2) | (0.1) | (5) |

* Standard error of the mean.

lected foods for a cholesterol-fat-protein questionnaire on the basis of their contribution of those nutrients in 100 four-day food records of Hawaiian Americans.

*Development of the nutrient composition array based on population data.* The use of population data permits the assignment of nutrient composition values which reflect the frequency of consumption of each of the varieties of food within that item. This approach, together with the use of medians rather than means, prevents an infrequent but extremely rich subcategory of food from unduly influencing the nutrient value which is used.

*Development of portion sizes based on population data.* "Standard" portion sizes, such as one-half cup, reflect measurement convenience and approximation rather than any inherent behavioral truth about the portions people actually consume. Portion sizes calculated from large-population data collected with the use of three-dimensional models permit the questionnaire to more closely approximate the behavioral truth. This approach is similar to that of Hankin et al. (5), who based her portion-size attributions on food-record data from 100 Hawaiian Americans. In addition, the use of a very large data base, such as the 11,658 adult respondents in NHANES II, permits the further step of identifying age-sex-specific portion sizes for foods on the list. Since age and sex are important determinants of portion sizes (27–30), the use of population-based age-sex-specific data enhances the accuracy of nutrient estimates from questionnaires.

The two features mentioned above, the nutrient composition and portion size arrays, constitute the quantitation of the food list. The particular quantitation approach is of central importance in the performance of this, or any other, dietary assessment instrument. The food list constitutes only part of any questionnaire, and must always be evaluated in connection with the nutrient composition and portion size assumptions which accompany it. The food list presented here was designed to perform

optimally only with the values developed for it from large population data, and not simply with US Department of Agriculture Handbook no. 456 data (31).

*Capable of being self-administered.* For many epidemiologic and some clinical situations, this is desirable or essential. This capability may make possible the collection of dietary data in research situations where such collection has previously been important but not feasible. Other investigators have developed or used questionnaires designed to be self-administered, including Willett et al. (6, 7), Jain et al. (8), and Bjelke (32), and have thereby considerably advanced our ability to understand the relationship of diet and disease.

*Use of portion sizes identified by the respondent.* In this instrument, the respondent identifies his usual portion of a given food as small, medium or large with respect to a stated medium portion. This approach adds personal variability in food preferences and quantity to the age-sex component addressed above.

Finally, developmental evaluations (tables 3-5) used a different data set to evaluate the food list and nutrient attributions, so that these aspects of the instrument could be optimized before the ambiguities of the human response to it were added. In one respect, this evaluation overestimates the agreement likely to be observed in a field validation, since respondents' ability to estimate frequency of consumption or portion size will not be perfect. This drawback is not, of course, unique to this instrument but is common to any self-report of diet, even dietary records (33). Thus, field validations and psychometric improvements are the logical next step, once the potential performance of the instrument and its values has been maximized.

Conversely, in some respects the developmental evaluation is somewhat conservative. First, tables 3–5 do not include any of the foods normally captured in the open-ended section in actual use. Second, they do not include the nutrient adjustments resulting from supplemental questions on the use of added fat or the consumption of fat on meat or skin on poultry. Finally, it should be noted that the two methods compared here, the food diary and the frequency food list quantitation, did not use identical data bases or portion size assumptions. The agreement which nevertheless resulted suggests the robustness of the quantitation developed for the self-administered diet history food list.

These evaluations, and the comparison with national values in tables 6 and 7, provide support for this food list and its quantitation in assessing nutrient intake. Quite apart from the particular instrument, however, it is hoped that the development methods presented here can be useful to other investigators in maximizing the scope and accuracy of dietary assessment instruments. In addition, further work is needed to improve quantitation, to determine whether some foods could be omitted, and to assess the applicability of this food list to geographic or demographic subgroups. Several studies are planned or are underway to address these issues.

## References


1. Burk MC, Pao EM. Methodology for large-scale surveys of household and individual diets. Home economics research report no. 40. Consumer and Food Economics Institute, Agricultural Research Service, US Department of Agriculture. Washington, DC: US GPO, 1976. (US GPO publication no. 001-000-03620-3).
2. Block G. A review of validations of dietary assessment methods. Am J Epidemiol 1982;115:492-505.
3. Morgan RW, Jain M, Miller AB, et al. A comparison of dietary methods in epidemiologic studies. Am J Epidemiol 1978; 107:488-98.
4. Hankin JH, Rawlings V, Nomura A. Assessment of a short dietary method for a prospective study on cancer. Am J Clin Nutr 1978;31:355–9.
5. Hankin JH, Nomura AM, Lee J, et al. Reproducibility of a diet history questionnaire in a case-control study of breast cancer. Am J Clin Nutr 1983; 37:981-5.
6. Willett WC, Sampson L, Stampfer MJ, et al. Reproducibility and validity of a semiquantitative food frequency questionnaire. Am J Epidemiol 1985; 122:51-65.
7. Willett WC, Stampfer MJ, Underwood BA, et al. Validation of a dietary questionnaire with plasma carotenoid and alpha-tocopherol levels. Am J Clin Nutr 1983; 38:631-9.
8. Jain MG, Harrison L, Howe GR, et al. Evaluation of a self-administered dietary questionnaire for

use in a cohort study. Am J Clin Nutr 1982; 36:931-5.
9. National Center for Health Statistics, Office of Health Research, Statistics and Technology. Plan and operation of the Second National Health and Nutrition Examination Survey, 1976-80. Vital and Health Statistics. Series 1, no. 15. Washington, DC: US GPO, 1981. (DHEW publication no. (PHS) 81-1317).
10. Block G, Dresser CM, Hartman AM, et al. Nutrient sources in the American diet: quantitative data from the NHANES II Survey. I. Vitamins and minerals. Am J Epidemiol 1985; 122:13-26.
11. Block G, Dresser CM, Hartman AM, et al. Nutrient sources in the American diet: quantitative data from the NHANES II Survey. H. Macronutrients and fats. Am J Epidemiol 1985; 122:27-40.
12. Block G, Hartman A. The role of portion size in dietary assessment instruments. Manuscript in preparation.
13. Cummings SR, McHenry K, Block G, et al. Validation and comparison of two food frequency methods of measuring dietary calcium intake. Unpublished manuscript.
14. Block G, Bowen P, Burgess M, et al. A self-administered diet questionnaire validated against serum carotenoids. Manuscript in preparation.
15. Nutriquest™ Software to access the USDA Food Consu-mption Survey data base. Washington, DC: Capital Systems Group Inc.
16. Snedecor GW, Cochran WG. Statistical methods. Ames, IA: Iowa State University Press, 1967.
17. Fleiss JL. Statistical methods for rates and proportions. New York: John Wiley and Sons, 1981.
18. SAS Institute, Inc. SAS user's guide, 1982. Cary, NC: SAS Institute, Inc., 1982.
19. Shah BV. SESUDAAN: Standard error program for computing of standardized rates from sample survey data. Unpublished report from Research Triangle Institute, Research Triangle Park, NC, 1980.
20. Lee J. Alternate approaches for quantifying aggregate and individual agreements between two methods for assessing dietary intakes. Am J Clin Nutr 1980; 33:956-64.
21. Mettlin C. Cancer control research issues in chemoprevention. Paper presented at Second International Conference on the Modulation and Mediation of Cancer by Vitamins and Micronutrients, Tucson, AZ, February 10-13, 1985.
22. Hinds MW, Kolonel LN, Hankin JH, et al. Dietary vitamin A, carotene, vitamin C and risk of lung cancer in Hawaii. Am J Epidemiol 1984; 119:227-37.
23. Gregor A, Lee PN, Roe FJC, et al. Comparison of dietary histories in lung cancer cases and controls with specific reference to vitamin A. Nutr Cancer 1980; 2:93-7.
24. Jain M, Howe GR, Johnson KC, et al. Evaluation of a diet history questionnaire for epidemiologic studies. Am J Epidemiol 1980; 111:212-19.
25. Paul O, Lepper MH, Phelan WH, et al. A longitudinal study of coronary heart disease. Circulation 1963; 28:20-3.
26. Batcher OM, Nichols JM. Identifying important food sources of nutrients. J Nutrition Education 1984; 16:177-81.
27. Beaton GH, Milner BA, Corey V, et al. Sources of variance in 24-hour dietary recall data: implications for nutrition study design and interpretation. Am J Clin Nutr 1979; 32:2546-59.
28. Pennington JAT. Revision of the total diet study food list and diets. J Am Diet Assoc 1983;82:166-73.
29. National Center for Health Statistics. Dietary intake source data: United States, 1976–80. Series 11, no. 231. Washington, DC: US GPO, 1983. (DHHS publication no. (PHS) 83-1681).
30. US Department of Agriculture, Human Nutrition Information Service. Food intakes: individuals in 48 states, year 1977-78. NFCS 1977-78 report no. I-1. Washington, DC: US GPO, 1983.
31. US Department of Agriculture. Nutritive value of American foods, in common units. Agriculture handbook no. 456. Washington, DC: US GPO, 1975.
32. Bjelke E. Dietary vitamin A and human lung cancer. Int J Cancer 1975; 15:561-5.
33. Kim WW, Mertz W, Judd JT, et al. Effect of making duplicate food collections on nutrient intakes calculated from diet records. Am J Clin Nutr 1984; 40:1333-7.