J Clin Epidemiol Vol. 43, No. 11, pp. 1169–1178, 1990
Printed in Great Britain.

0895-4356/90 $3.00 + 0.00

CHARACTERISTICS OF INDIVIDUALS AND LONG TERM REPRODUCIBILITY OF DIETARY REPORTS: THE TECUMSEH DIET METHODOLOGY STUDY

FRANCES E. THOMPSON,[1,2]* HELEN L. METZNER,[1,2] DONALD E. LAMPHIEAR[1] and VICTOR M. HAWTHORNE[1]

[1]Department of Epidemiology and [2]Human Nutrition Program, School of Public Health, University of Michigan, Ann Arbor, MI 48109, U.S.A.

(*Received in revised form* 9 *April* 1990)

Abstract—Food frequency reports in 1967–1969 were compared to frequency reports of the same foods asked retrospectively in 1982–1983 about 1967–1969 for 1184 respondents aged 45–64 years in the Tecumseh Community Health Study. The kappa statistic for concordance of the retrospective and baseline reports was used as a summary measure of the individual's ability to reproduce his or her earlier diet report. Reproducibility was estimated for total diet, represented by 83 foods, and for 9 subsets of foods of epidemiologic interest. In bivariate and multivariate analyses, reproducibility was strongly related to stability of diet; those whose diets changed least over the 15-year period had greatest diet reproducibility. Greater total diet reproducibility was also found among men with higher education, among women of < 110% desirable weight reporting no special diet and among women reporting no medications. Consistent with current models of memory, the retrospective report of diet was strongly related to the current report of diet. Agreement between the retrospective and baseline diet reports was greater than agreement between the current and baseline diet reports. This indicates that, as a proxy for past diet, the retrospective report of diet is superior to the current report. Similar relationships were found for the 9 subset of foods.

Diet methods   Reliability   Food frequency   Retrospective reports

INTRODUCTION

In studies of the dietary antecedents of chronic disease, where low incidence and long exposures prevail, past diet must usually be ascertained retrospectively. Of importance to the researcher are the validity and reliability of the retrospective diet report. For diet history methods, including food frequency instruments, validity has been examined by comparing estimates of diet from the food frequency to those from more quantitative diet records obtained at baseline [1–10]. Reproducibility, as measured by agreement in the results of the same instrument at two different times, has been reported earlier in this study sample for an interval of 15 years [11] and by others for intervals of 2 weeks to 25 years [2, 4, 12–23]. All of this research about retrospective dietary reports has focused on foods and nutrients.

In this paper the attention is on respondents and their ability to reproduce their earlier reports of overall diet rather than specific nutrients, foods or food groups. Epidemiologic studies could be designed more effectively if investigators knew that respondents with certain

*All correspondence should be addressed to: Frances E. Thompson, Martha van Rensselaer Hall, Division of Nutritional Sciences, Cornell University, Ithaca, NY 14853, U.S.A.

characteristics report past diet better than others and to what degree. The study of overall diet is important because findings for a particular food group may not be generalizable to all food groups. Also, food groups and nutrients not of interest today may become so in the near future.

In this report we will consider the following questions: Do subgroups of respondents described by various demographic and health characteristics differ in their ability to reproduce reports of their earlier eating habits? Do those who have greater perceived or actual changes in diet give retrospective diet reports which are less reliable than those who have little perceived or actual changes in diet? Does the ability to report total diet retrospectively differ from the ability to report retrospectively specific food groups of interest?

Our indicator of the ability to report diet retrospectively is the concordance between baseline and retrospective reports, summarized for each individual by a kappa statistic; this score was treated as a characteristic of the individual. Data came from a cohort of 1184 participants in the Tecumseh Community Health Study (TCHS) who were interviewed about their diets in 1967–1969 and again in 1982–1983 about their 1967–1969 diets. Associations between the kappa score for reproducibility and demographic characteristics, health status, adiposity and observed and perceived changes in diet are presented.

METHODS

Study Population

The TCHS is a longitudinal epidemiologic study of health and disease in a total, natural community [24]. The eligible cohort for the 1982–1983 Diet Methodology Study consisted of 1986 participants in the TCHS 1967–1969 round of interviews who had reported their usual dietary pattern by a food frequency instrument and who would have been 45–64 years of age at the time of their scheduled interview in 1982–1983. Of these, 123 (6.2%) had died; 327 (16.5%) had moved outside a 25-mile radius of the study area and so were no longer eligible; and 7 (0.4%) were lost to follow-up. Among the 1529 men and women still living in the study area in 1982–1983, 1387 (90.7%) agreed to participate in the Diet Methodology Study and provided complete data. In general, there were no differences in baseline 1967–1969 characteristics among the respondents, refusals and move-aways/lost to follow-up. However, those who died were more likely to have been men, in the upper age range, and to have smoked cigarettes.

The eligible population was assigned randomly to one of four study groups, each receiving different sets of protocols from the interview, as described previously [11]. Briefly, Groups 1 and 2 received current and retrospective food frequencies; Group 3 received the current food frequency; and Group 4 received the retrospective food frequency. The study sample for most of the analyses reported here consists of 85% of the total Diet Methodology Study sample (from Groups 1, 2 and 4)—1184 participants who gave complete restrospective food frequency information in 1982–1983 about their diets in 1967–1969. When current food frequency data (information about 1982–1983 collected in 1982–1983) are used in conjunction with retrospective data or with baseline data from 1967–1969, the sample size is 998 (from Groups 1 and 2).

Field Procedures

The field procedures and design of the interview are described more fully in Thompson *et al.* [11]. Interviews were conducted in the respondents' homes from March 1982 to March 1983. At baseline, the respondents reported how often they usually ate 110 foods or food groups, including 15 seasonal foods, during the previous year by sorting cards, each labeled with the name of a food or food group, into 1 of 8 categories of frequency of use. In order to shorten the lengthy interview for the current study, respondents were asked to report on 83 of the same foods, including 6 seasonal foods, using the same frequency categories and procedures. The foods omitted were rarely eaten in 1967–1969, or contributed little to fat or vitamin intake; they included, among others, olives, pickles, avocado, dried or smoked fish, lamb and mutton.

Statistical Methods

The kappa statistic for total diet reproducibility was computed for each respondent using all of the 83 foods that were sorted for frequency of use in both the baseline and the retrospective card sorts. Reproducibility is defined here as the item by item agreement of the two reports. This adaptation of the classical problem of interrater agreement between two judges [25] gives a measure of the reproducibility of two reports by the same individual of food use frequency at the time of the baseline study, one at baseline and the other after an interval of *15* years. The retrospective report of frequency of use was made independently of the frequencies reported at baseline since the respondent would have had no memory of the frequency categories selected in the earlier report. Frequency categories were defined for the respondent as never or almost never, once or a few times a year, about 1–3 times a month, about once a week, about twice a week, about 3–5 times a week, about once a day and about twice a day or more often.

The kappa for the individual is computed from an 8 × 8 cross-classification by frequency category of retrospective and baseline frequencies of use, both regarded as subject to random errors of reporting. A weight

$$w_{ij} = 1 - \frac{(i-j)^2}{(k-1)^2},$$

where k is the number of categories, is attached to the ijth cell. The weighted kappa provides partial contributions to kappa for non-matches, according to the degree of agreement, to account for random error in reporting frequency of use and for ambiguities in the definitions of the categories as presented to the respondent. For example, a true frequency of greater than 5 but less than 7 times a week could have been placed into the category of about 3–5 times a week in some sortings, and about once a day in others.

The logarithmic transformation

$$L = 4 + \log_2(f + 0.125)$$

of central frequencies for the 8 categories yields approximately equal intervals. Fleiss and Cohen have shown that when the categories form an ordinal scale, equally spaced, with weights as shown above, kappa is asymptotic to the intraclass correlation coefficient as the number of cases increases [26]. In the study sample, the mean value for the weighted kappa for reproducibility of the total diet is 0.550; the intraclass correlation coefficient is 0.553.

The reproducibility kappas for individuals were found to be normally distributed in the study sample. The kappa statistic is treated here as a score characterizing the individual with respect to the reliability of his dietary report.

Two other kappa scores were defined to characterize the individual with respect to other dietary reports. Similarity of current diet with baseline diet, termed stability of diet, was measured by the kappa for concordance of the baseline food frequency reports and food frequency reports in 1982–1983 for current use. High values of kappa stability scores indicate similar diets; low values, dietary change. The influence of current diet on reporting past diet, termed current influence, was measured by the kappa for concordance of the retrospective and current frequency reports for 1982–1983. The interpretation of this kappa score as an indication of the influence of current dietary practice on the retrospective reporting of diet depends on its magnitude relative to the kappa scores for reproducibility and stability.

Differences in mean kappa scores among subgroups of the population were tested for statistical significance by analysis of variance. Multiple regression, using continuous variables and indicator variables to represent nominal or unscaled categories, was performed in order to examine joint relationships of predictor variables with diet reproducibility.

Kappa scores were also computed for 9 subsets of foods, including food groups for high vitamin C foods, high vitamin A foods, high sodium foods, high cholesterol foods, all cholesterol foods, high fat foods (70–100% calories from fat), medium fat foods (50–70% calories from fat), lower fat foods (30–50% calories from fat) and fried foods. Multiple regression analysis, using kappa reproducibility scores for each of these food groups as the dependent variables and the same predictor variables, was performed in order to examine the generalizability of results from the total diet reproducibility analysis.

Definition of Variables

Health variables were derived from questions about the date of onset and date of last trouble with 9 diet-related chronic disease conditions: ulcer, gallbladder disease, intestinal trouble, heart attack, other heart problems, high blood pressure, diabetes, thyroid trouble and tumor. An illness score summarized this information by assigning for each condition a score of 0 for no history, 1 for history but not within the past 6 months and 2 for history within the past 6 months, then summing across all 9 conditions. Other health variables were derived from questions about medications taken in the last 12 months for 3 diet-related conditions: anemia, high blood pressure and diabetes. A medications index summarized this information as a count of conditions for which medications were used. Other health variables described adherence to special diets and smoking habits.

Current weight was described by an adiposity index, defined as observed weight divided by height squared, scaled separately for each sex to the **1959 Tables of Desirable Weight** from the Metropolitan Life Insurance Company. A value of 100 signifies the midpoint of the range of desirable weight for the medium frame. The distribution of the adiposity index for each sex was divided into quintiles. The index was also used to divide the sample into 3 adiposity categories: <90%, 90–110% and 110% or greater of desirable weight.

Three variables described change over the 15-year period covered by the study. Respondents were coded as having changed their marital status if either the status or partner had changed. Any change in the number of adults or children in the household was coded as change in household composition. Diet change between 1967–1969 and 1982–1983 was indicated by the kappa score for stability; also, the distribution of this score was divided into tertiles with the first tertile representing low stability or high change.

An index of perceived dietary change was constructed for each respondent based on answers to questions about whether the respondents thought there had been a change in how often they were eating particular foods. For each food group, no perceived change was scored 0; some change, that is, more or less often, was scored 1; and much more or much less often was scored 2. These scores were summed across 12 food groups to derive each respondent's index of perceived dietary change.

Reproducibility of the non-dietary variables, marital status, address and household composition, was measured by percent exact agreement, a measure that is not directly comparable to the kappa for diet reproducibility. Observed agreement was used to measure reproducibility of these variables because there was no ambiguity in classifying the true data, because the variables did not have ordinal categories, and because agreement of report of these variables was used to classify respondents into categories according to their ability to remember correctly the context of the earlier time period.

RESULTS

Bivariate Analyses

Demographic characteristics

Total diet reproducibility averaged 0.55 for the total sample of 1184 individuals. Mean reproducibility scores for subsets of individuals defined by sex, age, marital status, employment status, number of jobs held or income quartile did not differ significantly; means for various subgroups (consisting of at least 20 individuals) ranged from 0.53 to 0.57. Men in professional and managerial occupations had a mean reproducibility score of 0.58 (95% CI = 0.56, 0.60), while men in clerical and sales and blue collar occupations had a mean reproducibility score of 0.54 (95% CI = 0.53, 0.56). Men and women who had completed high school had mean reproducibility scores of 0.56 (95% CI = 0.55, 0.57), while those with less education had mean scores of 0.51 (95% CI = 0.50, 0.53).

Health characteristics

In general, respondents who were classified in the more favorable health categories had higher total diet reproducibility scores. Among both men and women, there was a small but significant negative correlation between the summary illness score and diet reproducibility ($r = -0.11$). For men, mean reproducibility scores were similar in subcategories of other health variables. Mean scores for subgroups defined by various health characteristics ranged from 0.49 to 0.56 for individual illnesses, from 0.55 to 0.56 for medications, from 0.54 to 0.56 for special diet, from 0.54 to 0.57 for cigarette smoking status and from 0.54 to 0.56 for adiposity quintile.

For women, small but significant negative associations were found between total diet reproducibility and three sets of health variables: taking blood pressure and diabetes medications and the medications index; report of being on a special diet; and adiposity. Total diet reproducibility scores averaged 0.48 (95% CI = 0.43, 0.52) for women reporting diabetes medications and 0.52 (95% CI = 0.50, 0.54) for women reporting blood pressure medications; scores for women reporting none of the three medications averaged 0.56 (95% CI = 0.55, 0.57). Women reporting two or more special diets had a mean reproducibility score of 0.49 (95% CI = 0.44, 0.54), while women reporting no special diets had a score of 0.55 (95% CI = 0.54, 0.57). Reproducibility was lower with each successively higher adiposity quintile, from a mean score of 0.58 (95% CI = 0.56, 0.61) in the lowest adiposity group to 0.50 (95% CI = 0.47, 0.53) in the highest adiposity group.

Measures of change since 1967–1969

In general, respondents who experienced the least change had the highest total diet reproducibility scores. Men who remained married to the same person had a mean reproducibility score of 0.56 (95% CI = 0.55, 0.57), while men who experienced a change in marital status or spouse had a mean score of 0.52 (95% CI = 0.49, 0.55). However, change in household composition was not significantly associated with diet reproducibility in men or in women; mean reproducibility scores for various change in household composition subgroups ranged from 0.52 to 0.56.

Total diet reproducibility was strongly related to the similarity of the current diet and the baseline diet, measured by the total diet stability kappa score. Men in the top tertile of diet stability had a mean diet reproducibility score of 0.65 (95% CI = 0.64, 0.67), men in the middle tertile a mean score of 0.56 (95% CI = 0.54, 0.57) and men in the lowest tertile a mean score of 0.46 (95% CI = 0.44, 0.48). For women, the respective scores were 0.66 (95% CI = 0.65, 0.68), 0.53 (95% CI = 0.51, 0.55) and 0.43 (95% CI = 0.42, 0.45). Perception of dietary change was inversely related to total diet reproducibility in women but not in men (Table 1). Women who perceived more change in their diet had lower reproducibility scores.

Reproducibility of demographic variables. Exact agreement between retrospective and baseline reports was 99% for marital status, 94% for address and 84% for number of household members. Most of the 16% who reported a different number of household members were discrepant by only one member. Agreement of 89% was obtained for employment status and earlier employer, and 85% for actual job. Many of these differences may be accounted for by the fact that baseline data for these variables were obtained for the day of baseline interview, whereas retrospective questions referred generally to the baseline year.

Men and women whose retrospective reports of their earlier demographic characteristics matched exactly with their baseline data tended to have higher total diet reproducibility scores. Of the four variables analyzed, only name of employer was statistically significant for men, 0.56 (95% CI = 0.55, 0.57) for agree vs 0.51 (95% CI = 0.46, 0.55) for disagree. For women, agreement on household composition and occupation were associated with significantly higher reproducibility scores, 0.55 (95% CI = 0.54, 0.56) vs 0.45 (95% CI = 0.40, 0.51) for household composition and 0.56 (95% CI = 0.55, 0.57) vs 0.52 (95% CI = 0.49, 0.55) for occupation. Mean reproducibility scores for subgroups defined by the other variables ranged from 0.52 to 0.58.

Reproducibility of weight. Recalled weight corresponded to earlier measured weight within 5 lb in 39% of men and women. The difference between retrospective and baseline weight was significantly correlated with the total diet reproducibility scores ($r = -0.14$ for men and -0.19 for women); the greater the discrepancy between the weight reports, the lower the diet reproducibility score.

Table 1. Perceived change in diet and mean current influence, stability and reproducibility scores, by sex; Tecumseh Diet Methodology Study, 1982–1983

Perceived change in diet (in quartiles)	Current influence			Stability			Reproducibility		
		weighted kappa			weighted kappa			weighted kappa	
	n	Mean	95% CI	*n*	Mean	95% CI	*n*	Mean	95% CI
Men									
First—low	87	0.76	(0.74, 0.79)	87	0.55	(0.52, 0.58)	107	0.55	(0.52, 0.58)
Second	126	0.72	(0.70, 0.74)	126	0.53	(0.51, 0.56)	146	0.57	(0.54, 0.59)
Third	122	0.65	(0.63, 0.68)	122	0.49	(0.46, 0.51)	142	0.55	(0.53, 0.58)
Fourth—high	125	0.55	(0.51, 0.58)	125	0.45	(0.42, 0.48)	145	0.54	(0.52, 0.56)
Women									
First—low	102	0.73	(0.71, 0.75)	102	0.55	(0.52, 0.58)	124	0.58	(0.55, 0.60)
Second	129	0.67	(0.64, 0.70)	129	0.53	(0.51, 0.56)	159	0.57	(0.55, 0.59)
Third	136	0.60	(0.57, 0.62)	136	0.48	(0.46, 0.51)	167	0.54	(0.51, 0.56)
Fourth—high	156	0.53	(0.50, 0.56)	156	0.44	(0.42, 0.46)	175	0.50	(0.48, 0.52)

95% CI is the 95% confidence interval for the mean. Current influence is concordance between retrospective and current reports. Stability is concordance between current and baseline reports. Reproducibility is concordance between baseline and retrospective reports.

Multivariate analyses

Results of multiple linear regressions were generally consistent with bivariate results (Table 2). Dietary stability accounted for most of the explained variance in total diet reproducibility. Higher education was associated with higher reproducibility in men. Medications, special diet and the interaction of adiposity and special diet remained statistically significant predictors of diet reproducibility among women. Among both men and women, the illness score was not statistically significant in multivariate analyses, possibly due to its positive correlation with other health variables. In a multiple regression model which omitted dietary stability, the total amount of variability explained decreased dramatically; R^2 dropped from 0.41 to 0.05 for men and from 0.45 to 0.05 for women. The regression coefficients of the remaining predictor variables were uniformly small. In this model, change in marital status was statistically significant among men, and education was statistically significant among women, consistent with bivariate results. Their lack of statistical significance in the model which includes dietary stability may be due to shared variance with that variable.

Table 2. Coefficients for regression of diet reproducibility on predictor variables, by sex; Tecumseh Diet Methodology Study, 1982–1983

Predictor variable	Men	Women
Age	0.0017	−0.0005
Education completed		
≥ 12 yrs	0.0334[b]	0.0090
< 12 yrs		
Adiposity	−0.0002	−0.0001
Special diet		
One or more	0.0110	0.0368[a]
None		
Special diet/adiposity interaction		
Special diet and ≥110% standard weight	−0.0124	−0.0501[a]
No special diet or <110% standard weight		
Illness score	−0.0008	0.0002
Medications	−0.0110	−0.0298[b]
Cigarette smoking		
Never smoked	0.0245	0.0193
Ex-smoker	0.0128	0.0224
Current smoker		
Change in marital status		
Change	−0.0141	0.0006
None		
Report of previous household composition		
Disagree	−0.0022	−0.0201
Agree		
Diet stability kappa	0.5838[b]	0.6165[b]
n	421	463
Adj. R^2	0.4091	0.4496

For categorical variables, regression coefficients are estimated relative to the last category.
[a]Statistically significant at $p < 0.05$ (two-tailed). [b]Statistically significant at $p < 0.01$ (two-tailed).

Comparisons between kappa scores for total diet and subsets of foods

Correlations between the kappa score for reproducibility of the total diet and kappa scores for reproducibility of subsets of foods ranged from 0.31 for fried foods to 0.62 for high fat foods (Table 3).

Mean reproducibility scores ranged from highs of 0.58 for medium fat foods and 0.55 for

Table 3. Correlation between kappa scores for reproducibility of total diet and reproducibility of food groups; Tecumseh Diet Methodology Study, 1982–1983

Food group*	No. of foods	r
Vitamin C	8	0.35
Vitamin A	7	0.34
Sodium	12	0.46
High cholesterol	5	0.43
All cholesterol	15	0.58
High fat	15	0.62
Medium fat	12	0.54
Lower fat	14	0.51
Fried foods	5	0.31

*Vitamin C foods include: lemonade, limeade, fruit punches; apple, pineapple, cranberry juices, cider, fruit nectar; oranges, grapefruit, orange and grapefruit juices; tomatoes; cooked broccoli, Brussel sprouts, cauliflower, cabbage; sweet potatoes, yams; creamed soups; and liver.

Vitamin A foods include: raw vegetables; cooked green beans, spinach, greens, asparagus, summer squash; cooked broccoli, Brussel sprouts, cauliflower, cabbage; cooked beets, carrots, peas, onions, winter squash, turnips; sweet potatoes, yams; clear soups; and liver.

Sodium foods include: nuts; snack foods; potato and corn chips; crackers; cream or cheese sauces; cheese; pizza; canned fish; shellfish; sausage, bacon, salt pork not at breakfast; ham not at breakfast; and meat for breakfast.

High cholesterol foods include: mayonnaise-type salad dressings; butter added to foods; liver; eggs not at breakfast; and eggs for breakfast.

All cholesterol foods include: mayonnaise-type salad dressings; cheese; butter added to foods; shellfish; poultry; liver; hot dogs, luncheon meats; sausage, bacon, salt pork not at breakfast; pork; ham not at breakfast; veal; beef; meat for breakfast; eggs not at breakfast; and eggs for breakfast.

High fat foods include: nuts; whipped cream, topping; mayonnaise-type salad dressings; oil and vinegar dressings; gravy; peanut butter; butter added to foods; margarine added to foods; fried chicken; fried meat; hot dogs, luncheon meats; sausage, bacon, salt pork not at breakfast; cream on cereal; meat for breakfast; and fried eggs.

Medium fat foods include: potato and corn chips; cream or cheese sauces; creamed soups; cheese; pork; ham not at breakfast; veal; beef; sweet rolls; doughnuts; eggs not at breakfast; and eggs for breakfast.

Lower fat foods include: crackers; chocolate, nut, caramel candy; cookies; malts, shakes, sodas; ice cream, ice milk, sherbert; pudding, custard; pie; cake; fried potatoes; pizza; fried fish or shellfish; shellfish; liver; and whole milk to drink.

Fried foods include: fried potatoes; fried fish or shellfish; fried chicken; fried meat; and fried eggs.

Table 4. Mean kappa scores for food groups and total diet; Tecumseh Diet Methodology Study, 1982–1983

	Reproducibility ($n = 1184$)		Stability ($n = 998$)		Current influence ($n = 998$)	
Food group*	Mean	95% CI	Mean	95% CI	Mean	95% CI
Vitamin C	0.48	(0.47, 0.50)	0.45	(0.43, 0.47)	0.64	(0.63, 0.66)
Vitamin A	0.55	(0.54, 0.57)	0.55	(0.54, 0.57)	0.66	(0.65, 0.68)
Sodium	0.49	(0.47, 0.50)	0.43	(0.42, 0.45)	0.56	(0.55, 0.58)
High cholesterol	0.34	(0.32, 0.36)	0.32	(0.29, 0.35)	0.56	(0.53, 0.58)
All cholesterol	0.52	(0.51, 0.53)	0.49	(0.47, 0.51)	0.65	(0.63, 0.66)
High fat	0.44	(0.42, 0.45)	0.40	(0.38, 0.42)	0.58	(0.56, 0.60)
Medium fat	0.58	(0.56, 0.59)	0.54	(0.53, 0.56)	0.64	(0.62, 0.65)
Lower fat	0.51	(0.49, 0.52)	0.39	(0.37, 0.40)	0.53	(0.51, 0.55)
Fried foods	0.42	(0.41, 0.44)	0.35	(0.33, 0.38)	0.43	(0.41, 0.45)
Total diet	0.55	(0.54, 0.56)	0.50	(0.49, 0.51)	0.64	(0.63, 0.65)

*See Table 3 for definitions of food groups.
95% CI is the 95% confidence interval for the mean.

vitamin A foods to lows of 0.44 for high fat foods, 0.42 for fried foods and 0.34 for high cholesterol foods (Table 4). Results for these subsets of foods are representative of values that may be anticipated for food groups tailored to current dietary investigations. As was true for total diet, reproducibility scores for subsets of foods were generally inversely associated with perceived dietary changes in women but not men (data not shown). Those women with the most perceived dietary change had lower dietary reproducibility scores than women with the least perceived dietary change for all food groups except vitamin C and vitamin A. For these 2 food groups, women at the extremes of perceived dietary change had similar reproducibility scores.

In multiple regression analyses, the dietary stability variable was by far the strongest and only consistent predictor of reproducibility for individual subsets of foods (data not shown), similar to findings for the total diet (Table 2). When dietary stability was omitted from the regression model, high educational status was associated with higher reproducibility in 6 of the 9 food groups among men, but in only 1 food group among women (Table 5). Other variables were generally not related to reproducibility. In order to illustrate the effects on predicted kappa reproducibility scores due to differences in total diet stability, we computed predicted kappa reproducibility scores for each food group at two levels of diet stability and for each sex. We used the multiple regression model specified in Table 2, with the kappa reproducibility score for each food group as the dependent variable. Average values on independent variables other than kappa stability are assumed. A low kappa stability score for the total diet is the 10th percentile value for each sex. A high kappa stability score for the total diet is the 90th percentile value for each sex. Results are presented in Table 6, and illustrate the predominant role of diet stability on expected levels of diet reproducibility.

Table 5. Coefficients for regression* of reproducibility of food groups on diet stability and educational status, by sex; Tecumseh Diet Methodology Study, 1982–1983

	Men ($n = 421$)			Women ($n = 463$)		
	Model I: with stability		Model II: without stability	Model I: with stability		Model II: without stability
Food group†	Kappa stability	Education	Education	Kappa stability	Education	Education
Vitamin C	0.4094[b]	0.0687[a]	0.0866[b]	0.4862[b]	0.0771[a]	0.0951[b]
Vitamin A	0.4337[b]	0.1116[b]	0.1306[b]	0.3073[b]	0.0360	0.0473
Sodium	0.5029[b]	0.0333	0.0552[a]	0.6116[b]	−0.0261	−0.0036
High cholesterol	0.7776[b]	−0.0161	0.0178	0.9222[b]	−0.0227	0.0113
All cholesterol	0.5534[b]	0.0441	0.0682[b]	0.7237[b]	0.0129	0.0396
High fat	0.7227[b]	0.0197	0.0512	0.8322[b]	−0.0148	0.0159
Medium fat	0.5370[b]	0.0411	0.0645[b]	0.5973[b]	0.0133	0.0353
Lower fat	0.4635[b]	0.0372	0.0574[a]	0.4733[b]	0.0033	0.0207
Fried foods	0.3838[b]	−0.0214	−0.0046	0.4517[b]	−0.0074	0.0093

*Full set of predictor variables same as in Table 2.
†See Table 3 for definitions of food groups.
[a]Statistically significant at $p < 0.05$ (two-tailed). [b]Statistically significant at $p < 0.01$ (two-tailed).

Table 6. Predicted kappa reproducibility scores* for food groups in individuals with low vs high dietary stability; Tecumseh Diet Methodology Study, 1982–1983

	Predicted kappa reproducibility			
	Men (n = 421)		Women (n = 463)	
Food group†	Low stability	High stability	Low stability	High stability
Vitamin C	0.38	0.54	0.42	0.61
Vitamin A	0.43	0.60	0.52	0.65
Sodium	0.39	0.59	0.36	0.60
High cholesterol	0.22	0.52	0.15	0.51
All cholesterol	0.41	0.63	0.38	0.67
High fat	0.31	0.59	0.26	0.59
Medium fat	0.46	0.67	0.47	0.71
Lower fat	0.43	0.61	0.41	0.59
Fried foods	0.41	0.56	0.30	0.48

*Kappa reproducibility scores are predicted using the multiple regression model specified in Table 2, using the kappa reproducibility score for each food group as the dependent variable. Average values on all variables except diet stability are assumed. A low kappa stability score for the total diet is the 10th percentile value for each sex. A high kappa stability score for the total diet is the 90th percentile value for each sex.
†See Table 3 for definitions of food groups.

DISCUSSION

This report differs from other reports on the reproducibility of retrospective dietary reports by focusing on individuals rather than foods. In a previous report which focused on long term reproducibility of individual food items [11], we used kappa to measure agreement over individuals for each food. The food kappas ranged from 0.05 to 0.74 on a scale of -1.00 to 1.00. In contrast, in this report we focus on characterizing the individual's overall ability to reproduce at follow-up his or her report of diet at baseline; we use kappa to measure agreement over foods for each individual. These individual kappas for total diet reproducibility ranged from 0.14 to 0.88. Thus, the individual kappas, averaged over foods, varied as widely as the food kappas, averaged over individuals.

Even with the large variation in total diet reproducibility scores, means for subgroups defined by various demographic characteristics differed little. Little evidence was found to support the commonly held notions that women are more reliable reporters than men and the young more reliable than the old. In bivariate analyses, sex was not related to diet reproducibility, similar, in sum, to the results of others. Jensen *et al.* [19] found men better retrospective reporters; Byers *et al.* [16] found women better; in a later study Byers *et al.* [22] found women slightly better; van Leeuwen *et al.* [3] and Wu *et al.* [23] found no sex difference. In work examining age effects, no significant differences with age were found by van Leeuwen *et al.* [3] nor by Wu *et al.* [23]; no consistent differences were found by Sobell *et al.* [10]; Byers *et al.* [22] found higher reproducibility for two of three nutrients examined in older respondents as compared to younger respondents. In our analyses, age was marginally associated (positively) with total diet reproducibility in men but was unrelated when controlling for diet stability, indicating that older men may be more reliable than younger men because of their greater dietary stability during the previous 15 years. In our study, only two indicators of socioeconomic status showed the expected relationship with reproducibility: men and women with higher education and men with higher occupational status had higher total diet reproducibility. In multiple regression analyses, educational status remained a significant predictor of total diet reproducibility. When the diet stability variable was included, the regression coefficient decreased, becoming non-significant for women. In the food subset multiple regressions, education was a relatively consistent but minor predictor of reproducibility only among men and only when the diet stability variable was excluded.

Sobell *et al.* [10] found no consistent differences in the validity of retrospective diet reports between those reporting medical problems and those not. Byers *et al.* [22] found higher reproducibility among respondents not currently experiencing a recently diagnosed chronic disease than among those who were. Consistent with his findings, our bivariate analyses indicated that poorer health status, as measured by use of special diet (women), higher adiposity (women), use of at least one medication (women) and higher illness score (men and especially women), were associated with lower total diet reproducibility. In multivariate analyses of these variables, most relationships remained statistically significant. The illness score became non-significant, perhaps because of multicollinearity. For subsets of foods, few relationships between health status variables and reproducibility attained statistical significance. While these findings are equivocal, the importance of examining the relationship between various health indicators and reproducibility warrants further study. If error in estimating past diet is not randomly distributed between cases and controls, then the associations between diet and disease may be either overestimated or underestimated.

In other studies, dietary change and perceptions of dietary change were associated with validity of retrospective report [10] and reproducibility of retrospective report [1, 19], but in another [3], were unrelated to reproducibility over a 4-year period. In our study, measured change in diet was strongly associated with lower reproducibility among men and women, and perceived change in diet was somewhat less strongly associated with lower reproducibility and only in women. The strong relationship between stability of diet and reproducibility may be due to the influence of current eating patterns on retrospective reporting. In this study, retrospective and current reports had a higher intraclass correlation coefficient, 0.64, than did retrospective and baseline reports, 0.55, a relationship also reported by others [14, 16, 17, 19, 22, 23]. This pattern was present for both total diet and subsets of foods.

These interrelationships are consistent with memory theory and applied research on reconstructing past events. Psychologists suggest that events which occur repeatedly and have the same essential characteristics across instances give rise to a generic fused memory for this group of events [27, 28]; a typical rather than a particular event will be reconstructed. Various researchers have found that the present circumstances and self-concept influence reconstruction of past events; what is remembered is compatible with one's present self-knowledge [29]. Similarly, we found that when respondents in our age group were asked about their past diets, their reference was their current diet, and earlier diets were reported as variations. Since memory for current behavior and events is better than for past behavior and events, it is likely that respondents' answers will be highly conditioned on what they believe their present diet is. The high correlation between current and retrospective reports supports this hypothesis. For respondents whose diets have not changed and who perceive no substantial changes from baseline, current *is* baseline, and reproducibility is relatively high. In summary, we suggest that everyone tends to report retrospective diet as a variant of current diet, and when current is close to baseline, reproducibility will be high.

In this study and in others [16, 17, 23], the agreement between restrospective and baseline reports was greater (mean total reproducibility score here of 0.55) than that between current and baseline reports (mean total stability score here of 0.50). At all levels of perceived dietary change, reproducibility was the same or greater than stability. This was true for total diet and generally true for subsets of foods. We conclude that the retrospective report of past diet is superior to the report of current diet in providing estimates of past diet. Retrospective reports are particularly superior to current reports for respondents who have experienced changes or who report changes in their diets.

The mean kappa score for total diet reproducibility of 0.55 in this sample indicates fair to good reliability for group level data. The importance of reliability estimates to statistical power has been reviewed elsewhere [30]. As reliability decreases, statistical power decreases, biasing estimated relationships towards zero, provided that error is random. The mean score of 0.55 considered as an overall intraclass correlation coefficient may represent a fair estimate of the expected reliability of long term retrospective dietary information and can be used in conjunction with expected risk differences to estimate the sample sizes needed in retrospective diet investigations.

Except for stability of diet, no single individual characteristic was important in identifying the more reliable reporters. In bivariate analyses the maximum spread in the mean total diet reproducibility kappa scores among subgroups of individuals was only about 0.10 points. The small differences and the general lack of statistical significance suggest that, in retrospective studies of diet, researchers may not need to be concerned about different levels of reliability of dietary report for different subgroups defined by sex, age or marital status. However, reliability of reports of past dietary practices is likely to be lower for individuals whose diets have changed, and may be slightly lower for lower socioeconomic groups and for those in poor health.

Acknowledgements—This work was supported by National Cancer Institute Grant R18 CA 27979.

The authors wish to thank Dr J. Richard Landis for his invaluable statistical advice, Dr Frances A. Larkin for her nutritional advice and Dr Craig R. Barclay for sharing his expertise in memory research. We also wish to acknowledge gratefully the contributions of Grace Lim to the collection and processing of the nutritional data, Catherine Meade to the design of the questionnaire and supervision of data processing and Dr Mark A. Archer and Mary S. Oh for computer programming.

REFERENCES

1. Jain M, Howe GR, Johnson KC *et al.* Evaluation of a diet history questionnaire for epidemiologic studies. **Am J Epidemiol** 1980; 111: 212–219.
2. Smith-Barbaro P, Darby L, Reddy BS. Reproducibility and accuracy of a food frequency questionnaire used for diet intervention studies. **Nutr Res** 1982; 2: 249–261.
3. van Leeuwen FE, de Vet HCW, Hayes RB *et al.* An assessment of the relative validity of retrospective interviewing for measuring dietary intake. **Am J Epidemiol** 1983; 118: 752–758.
4. Willett WC, Sampson L., Stampfer MJ *et al.* Reproducibility and validity of a semi-quantitative food frequency questionnaire. **Am J Epidemiol** 1985; 122: 51–65.
5. Krall EA, Dwyer JT. Validity of a food frequency questionnaire and a food diary in a short-term recall situation. **J Am Diet Assoc** 1987; 87: 1374–1377.
6. Willett WC, Reynolds RD, Cottrell-Hoehner S *et al.* Validation of a semi-quantitative food frequency questionnaire: comparison with a 1-year diet record. **J Am Diet Assoc** 1987; 87: 43–47.
7. Willett WC, Sampson L., Browne ML *et al.* The use of a self-administered questionnaire to assess diet four years in the past. **Am J Epidemiol** 1988; 127: 188–199.
8. Pietinen P, Hartman AM, Haapa E *et al.* Reproducibility and validity of dietary assessment instruments. II. A qualitative food frequency questionnaire. **Am J Epidemiol** 1988; 128: 667–676.
9. Larkin FA, Metzner HL, Thompson FE *et al.* Comparison of estimated nutrient intakes by food frequency and dietary records in adults. **J Am Diet Assoc** 1989; 89: 215–223.
10. Sobell J, Block G, Koslowe P *et al.* Validation of a retrospective questionnaire assessing diet 10–15 years ago. **Am J Epidemiol** 1989; 130: 173–187.
11. Thompson FE, Lamphiear DE, Metzner HL *et al.* Reproducibility of reports of frequency of food use: the Tecumseh Community Health Study. **Am J Epidemiol** 1987; 125: 658–671.
12. Acheson ED, Doll R. Dietary factors in carcinoma of the stomach: a study of 100 cases and 200 controls. **Gut** 1964; 5: 126–131.
13. Graham S., Lilienfeld AM, Tidings JE. Dietary and purgation factors in the epidemiology of gastric cancer. **Cancer** 1967; 20: 2224–2234.
14. Garland B, Ibrahim M, Grimson R. Assessment of past diet in cancer epidemiology. Presented at the **Annual Meeting of the Society for Epidemiologic Research,** Cincinnati, Ohio; 1982.
15. Hankin JH, Nomura AMY, Lee J *et al.* Reproducibility of a diet history questionnaire in a case-control study of breast cancer. **Am J Clin Nutr** 1983; 37: 981–985.
16. Byers TE, Rosenthal RI, Marshall JR *et al.* Dietary history from the distant past: a methodological study. **Nutr Cancer** 1983; 5: 69–77.
17. Rohan TE, Potter JD. Retrospective assessment of dietary intake. **Am J Epidemiol** 1984; 120: 876–887.
18. Little RE, Worthington-Roberts B, Mann SL *et al.* Test–retest reliability of diet and drinking estimates for pregnancy and post partum. **Am J Epidemiol** 1984; 120: 794–797.
19. Jensen OM, Wahrendorf A, Rosenqvist A *et al.* The reliability of questionnaire-derived historical dietary information and temporal stability of food habits in individuals. **Am J Epidemiol** 1984; 120: 281–290.
20. McKeown-Eyssen GE, Yeung KS, Bright-See E. Assessment of past diet in epidemiologic studies. **Am J Epidemiol** 1986; 124: 94–103.
21. Wu ML, Whittemore AS, Jung DL. Errors in reported dietary intakes. I. Short-term recall. **Am J Epidemiol** 1986; 124: 826–835.
22. Byers T, Marshall J, Anthony E *et al.* The reliability of dietary history from the distant past. **Am J Epidemiol** 1987; 125: 999–1011.
23. Wu ML, Whittemore AS, Jung DL. Errors in reported dietary intakes. II. Long-term recall. **Am J Epidemiol** 1988; 128: 1137–1145.
24. Napier JA, Johnston BC, Epstein FH. The Tecumseh, Michigan Community Health Study. In: Kessler II, Levin ML, Eds. **The Community as an Epidemiologic Laboratory: a Casebook of Community Studies.** Baltimore, Md: The Johns Hopkins University Press; 1970.
25. Fleiss JL. The measurement of interrater agreement. In: **Statistical Methods for Rates and Proportions,** 2nd edn. New York: Wiley; 1981: Chap. 13.
26. Fleiss JL, Cohen J. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. **Educ Psychol Meas** 1973; 33: 613–619.
27. Linton M. Ways of searching and the contents of memory. In: Rubin D, Ed. **Autobiographical Memory.** New York: Cambridge University Press; 1986.
28. Neisser U. Nested structure in autobiographical memory. In: Rubin D, Ed. **Autobiographical Memory.** New York: Cambridge University Press; 1986.
29. Barclay CR. Schematization of autobiographical memory. In: Rubin D, Ed. **Autobiographical Memory.** New York: Cambridge University Press; 1986.
30. Liu K, Stamler J, Dyer A *et al.* Statistical methods to assess and minimize the role of intra-individual variability in obscuring the relationship between dietary lipids and serum cholesterol. **J Chron Dis** 1978; 31: 399–418.