A MICROCOMPUTER-BASED INTERACTIVE MODEL FOR COLLECTION AND CODING OF DIETARY DATA

I. Marilyn Buzzard, Nutrition Coding Center, Division of Biometry
Lynda B. M. Ellis, Division of Health Computer Sciences
Elaine D. H. Asp, Department of Food Science and Nutrition
University of Minnesota

INTRODUCTION

Recent technological advances in computer hardware and software have made possible the automation of procedures for collecting and coding dietary data. A model for microcomputer-based interactive programming is being developed at the University of Minnesota Nutrition Coding Center (NCC) to improve the accuracy and efficiency of dietary data collection and coding procedures.

A number of other investigators have successfully implemented interactive systems for dietary data collection and analysis (Witschi et al. 1981; Evans & Gormican, 1973; Slack et al. 1976). However, none of these systems documents dietary intake at the level of precision required by studies using the present NCC system for investigation of diet-disease relationships.

This paper will begin with a brief description of the present NCC system and will then proceed to describe the initial stages of development of the model for computerizing the coding component of the NCC dietary analysis system.

THE PRESENT NCC SYSTEM

The NCC system was developed in 1974 under NHLBI contract to provide standardized procedures for collecting, coding, and calculating nutrients for 24-hour dietary recalls for two collaborative nationwide cardiovascular studies, the Multiple Risk Factor Intervention Trial (MRFIT) and the Lipid Research Clinics (LRC) programs. Standardization of collection and analysis procedures for the 34 clinical centers involved was achieved through centralized training of nutritionist interviewers and centralized coding at the NCC.

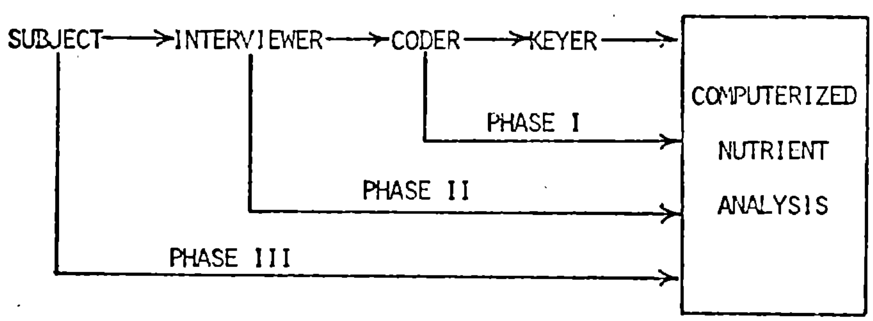
Detailed descriptions of the NCC system have been presented by Dennis et al. (1980) and Tillotson et al. (1981). Briefly, the system provides for detailed documentation of food intake through the use of a 500-page coding manual which includes several hundred coding rules to provide consistency in coding when amounts, preparation methods or food characteristics are documented as unknown (such as amount of creamer used in coffee or unknown cut of beef). The system also includes a Recipe File of over 550 recipes, a Manufacturers' File containing nutrient and ingredient information on thousands of commercial products, and an "Uncodables" File for standardization of infrequent coding decisions. The codebook and nutrient data base are continually updated to reflect changes in the marketplace. Designed specifically to characterize the quality and quantity of dietary fat and cholesterol, the system allows detailed specification of the type and brand of fat used in recipes, fats added in food preparation, and estimates of amount of fat absorbed in various food preparation methods. A comprehensive system of internal and external quality controls has been incorporated into the coding and analysis procedures to ensure accuracy and consistency over time.

In 1977, the NCC system was made available to other studies, and modifications have been made to the individual research needs such as expansion of the system to include processing of multiple-day diet records, diet histories, and quantified food frequencies. In addition to cardiovascular studies, various studies involving cancer, diabetes, hypertension, gastroenterology and other diet-related diseases have used the NCC system since 1977. Specialized coding to accommodate the special research needs of each new study, such as quantification of sodium or fiber intake, have led to ever increasing numbers of new coding rules and the expansion of the nutrient data base. As the system grows in complexity, it has become increasingly difficult and costly to maintain the present level of accuracy and standardization. An automated system would free the food coder from the cumbersome manual coding procedures while maintaining the meticulous attention to coding detail. Thus, a model is being developed for a microcomputer-based interactive system that will incorporate all of the documentation and coding features of the present NCC system, including procedures for verification, error checking, and other quality controls.

MODEL FOR COMPUTERIZATION OF THE NCC SYSTEM

NCC staff members are collaborating with the University of Minnesota Health Computer Sciences Division and the Department of Food Science and Nutrition in developing the computerized model. The Computerized NCC System (CNCC) model involves three phases of development as indicated schematically in Figure 1 below:

Figure 1. Schematic diagram of the steps involved in dietary data collection and analysis. The arrows for phases I, II, and III indicate the successive steps to be eliminated in the development of the microcomputer-based model.



Phase I, the coder-oriented system, will include the development and evaluation of an interactive coding system that will combine coding and keying into a single-step process. Phase II, the interviewer-oriented system, will combine interactive coding with interviewing, allowing improved documentation of food intake while eliminating the need for a food coder. Phase III, the subject-oriented system, will involve modification of the interactive system for use by untrained subjects, thus eliminating the need for the interviewer, as well as for the coder and keyer.

This paper is limited to the development of Phase I, the coder-oriented component of the system. Most of the software to be developed in Phase I will be utilized, with appropriate modifications, in the succeeding phases. Testing at the end of Phase I will allow comparison of the CNCC system with the established NCC system.

Phase I development will include interactive programming that will incorporate the procedures presently used by the NCC food coders. Foods will be identified through an interactive menu-driven branching system based on food acronyms arranged in subsets of food grouping hierarchies. Some 100 to 200 major foods and food groups, designated by a one or two letter acronym (e.g. BF = beef; EG = egg; VG = vegetable; C = coffee; T = tea) will constitute the uppermost level of the hierarchy. Each group will be associated with a menu of its subgroups (either type or form of food or brand name). Each subgroup will be similarly structured using increasingly more specific descriptions and ending with the most precise identification of the food item including method of preparation, fats and salt used in preparation, and fat and salt added at the table.

An illustrative example of interactive coding using such a system is shown in Figures 2-7. Figure 2 is an example of documentation provided on a typical dietary record form. In Figure 3, the item salmon, pink, fresh and be identified either by three successive menu selections or by entering FS.SAL.PK.FR in a single sequence. Figure 4 indicates the amount of fish consumed either by weight, by known volume, or by dimensions for calculating the volume (such as length, width, and thickness of a rectangular serving). Preparation methods are indicated in Figures 5-7. Selection of BST, the present NCC "prep code" for oven cooked and basted, prompts a menu for type of fat used in the preparation as shown in Figure 5. If the fat used is unknown, documentation as to whether the food was consumed at home or in a commercial establishment is used to aid in selecting an appropriate fat. Margarine is an item that must be specified by brand and type, if known, as illustrated in Figure 6. The "Margarine Guide" for the present NCC system indicates which of the 44 margarine entries in the Food Table should be used for each of the several hundred brand name margarines based on type and polyunsaturated to saturated fatty acid (P/S) ratio. This information will be computerized, allowing the coder either to enter the brand name acronym or to request the list of brand name acronyms for margarine. When information is unknown, selections are based on frequency of use in the general population. Coding rules for individual food items regarding amount of salt used in preparation and at the table will be incorporated into the software development. The interactive dialog for entering this information is shown in Figure 7. The system will be flexible enough so that for studies not interested in sodium intake, the salt portions of the interactive dialog can be bypassed.

Figure 2. Example of documentation in a dietary record.

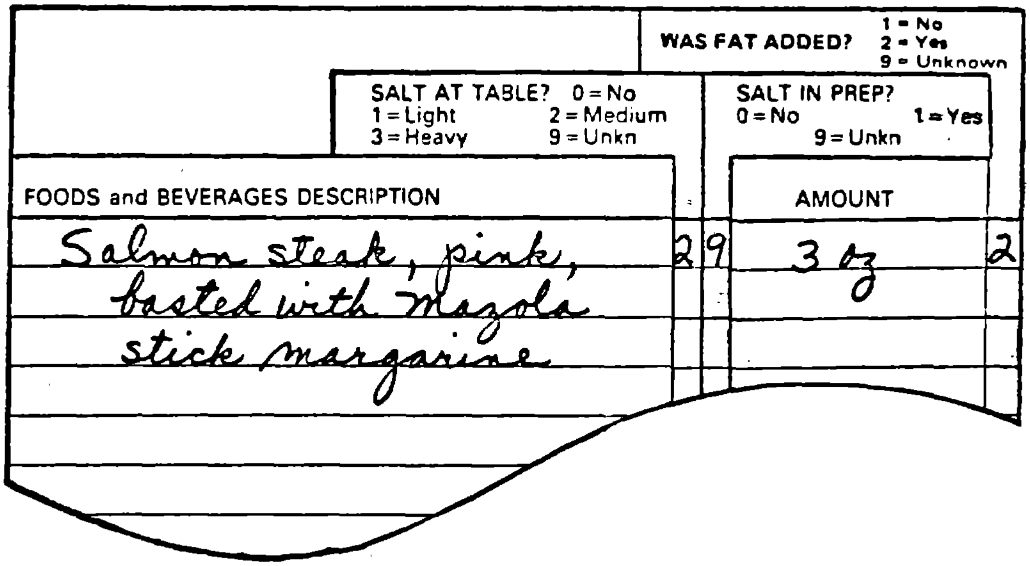
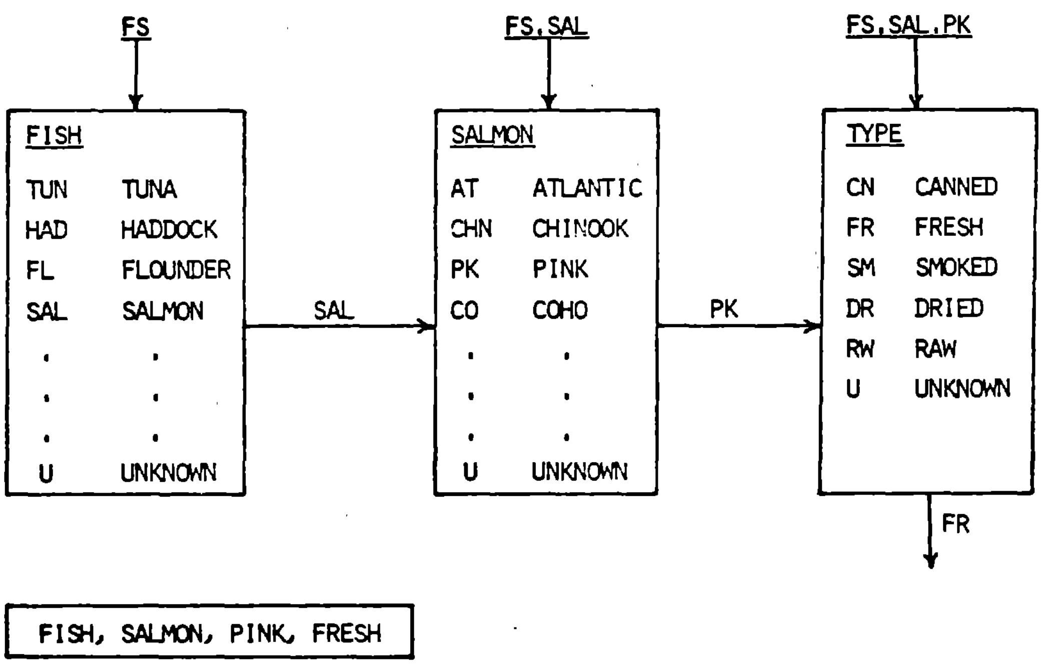


Figure 3. Example of interactive dialog for food item identification.



Note: Information in the boxes appears on the screen in response to information entered by the coder.

Figure 4. Example of interactive dialog for amount consumed.

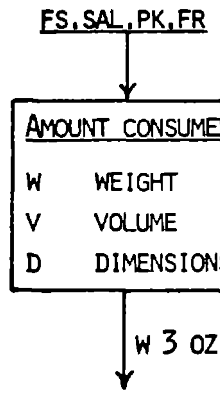


Figure 5. Example of interactive dialog for method of preparation (1).

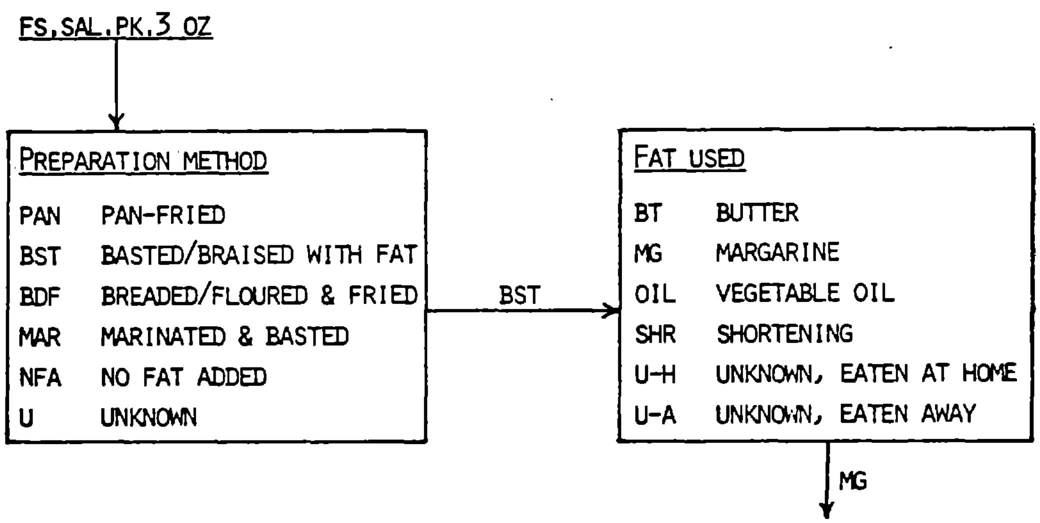


Figure 6. Example of interactive dialog for method of preparation (2).

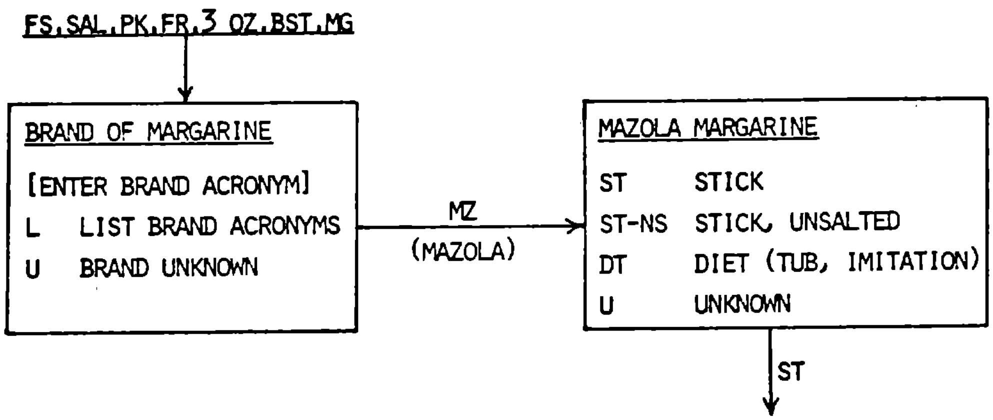
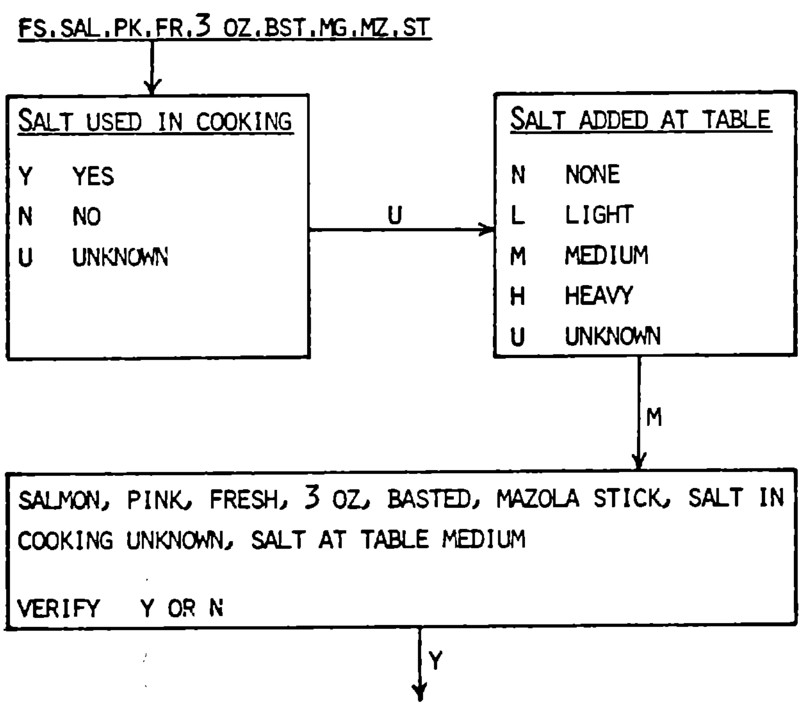


Figure 7. Example of interactive dialog for use of salt.



Once the food item description and preparation have have been specified in detail, the computer will assign the appropriate NCC food codes and quantities. This information will be stored on a disk file for subsequent transmission to a mainframe computer for nutrient calculations and food group analysis.

A feasibility project to demonstrate major aspects of the CNCC model is being developed on the Apple II microcomputer. Although the Apple II does not provide adequate disk storage for the comprehensive system, it is adequate for demonstration purposes. The program for entering the food grouping hierarchies is described in detail by Ellis et all (1983). Software for incorporating preparation codes, salt additions, coding rules, unit conversions, and various edit checks is presently being developed.

EXPANSION-OF THE CNCC SYSTEM TO INCLUDE CODING FOR MULTIPLE NUTRIENT DATA SETS

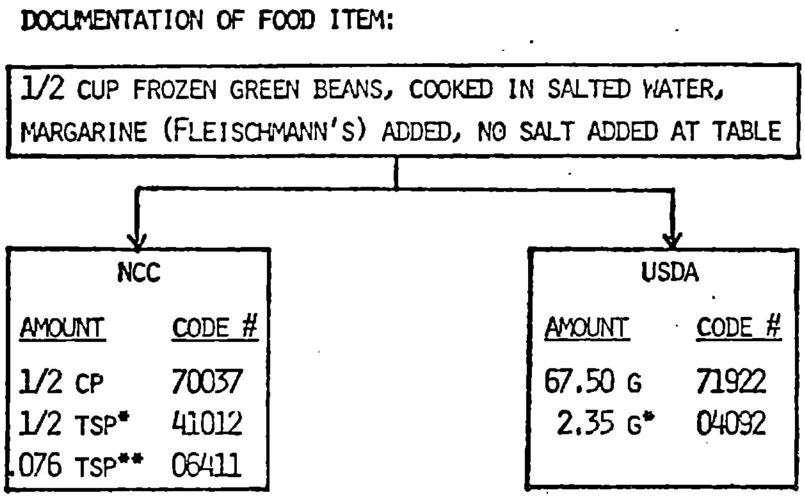
The CNCC model allows for the addition of food codes for any data set desired. Code numbers for each data set will be added at the end of each interactive pathway. The user will then specify which data set(s) is (are) desired for nutrient calculation, and the computer will assign the appropriate code numbers and amount units. Plans are being developed for incorporating the USDA Nutrient Data Base for Standard Reference (USDA, (1982) into the CNCC system along with the NCC nutrient data set. The USDA data base includes approximately 4,500 items compared with approximately 1,500 items in the NCC data base which was constructed to collapse on similarities in fat content. Based on the maximum amount of descriptive detail documented for any food item, the computer will assign the appropriate codes to either or both of the data sets. A one-to-one correspondence between the two data sets is not necessary as can be seen in the example shown below.

Consider an item documented as follows:

> 1/2 cup frozen green beans, cooked in salted water, margarine (Fleischmann's) added, no salt added at table.

The NCC data base includes only two entries for green beans; one entry is for canned green beans and the other entry includes all other forms. The USDA data set includes 17 green bean entries specifying whether canned green beans are drained or not, diet or regular pack, and whether frozen green beans are cooked or raw, French or regular cut, cooked with or without salt, etc. The documented information will be appropriately used by the computer for assigning both the NCC and the USDA code numbers as indicated in Figure 8. Only two code numbers are required for the USDA data set since salt is included in the USDA green bean code, whereas the salt must be added separately in the NCC system. Coding rules built into the software will specify amounts or types whenever maximum detail is not documented, such as the amount of margarine used in the example above. Since the USDA nutrient calculation system requires input in grams, the computer will generate the appropriate conversion when household or other units are entered.

Figure 8. Example of concurrent coding for more than one nutrient data base.



*Amount based on NCC "prep code" rule for unknown amount of fat added to vegetable

**Amount based on NCC coding rule for salt added in preparation

Code # descriptions:

NCC:	70037	beans, green, ckd, fresh or frozen
	41012	margarine, P/S 2.0-2.5, stick or tub
	06411	salt
USDA	71922	beans, snap, green, frz, cut ckd, boiled, drained, w/salt
	04092	margarine, soft, corn, hydr or reg

Providing the capability of calculating nutrients using various data bases will enable studies to compare their results with studies using other data bases without having to recode the dietary records.

SUMMARY

A model for a microcomputer-based interactive system for coding dietary data is described. The computerized model will incorporate all features of the present NCC coding system including coding rules, methods of food preparation, special procedures for sodium coding, and procedures for verification, error checking, and quality control. The system will be expanded to incorporate coding for multiple nutrient data sets, facilitating comparisons among studies using different data bases.

LITERATURE CITED

Dennis B., N. Ernst, M. Hjortland, J. Tillotson and V. Grambsch, 1980. The NHLBI nutrition data system. J. Am. Dietet. Association. 77:641.

Ellis, L.M.B., S. Chow and M. Buzzard, 1983. Microcomputer-based food coding. Proceedings of the Seventh Annual Symposium on Computer Applications in Medical Care, in press.

Evans, S.N. and A. Gormican, 1973. The computer in retrieving dietary history data. I. Designing and evaluating a computerized diabetic dietary history. J. Am. Dietet. Association. 63:397.

Slack, W., B. Porter, J. Witschi, M. Sullivan, R. Buzbaum and F.J. Stare, 1976. Dietary interviewing by computer. J. Am. Dietet. Association. 69:514.

Tillotson, J.L., D.D. Gorder and N. Kassim, 1981. Nutrition data collection in the Multiple Risk Factor Intervention Trial (MRFIT). J. Am. Dietet. Association. 78:235.

USDA, 1982. Nutrient Data Base for Standard Reference, Release 2.

Witschi, J., H. Kowaloff, S. Bloom, and W. Slack, 1981. Analysis of dietary data: an interactive computer method for storage and retrieval. J. Am. Dietet. Association. 78:609.