Vital and Health Statistics

Social Cognition Approach to Reporting Chronic Conditions in Health Surveys

Series 6:
Cognition and Survey Measurement
No. 3

This report describes a study, using methods derived from social cognition, designed to gain a better understanding of how the lay person organizes illnesses into natural categories and how these categories are linked to other knowledge about self and others.

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
Public Health Service
Centers for Disease Control
National Center for Health Statistics

Hyattsville, Maryland
October 1989
DHHS Publication No. (PHS) 89-1078
Reprinted November 1992

Copyright Information

Suggested Citation

Brewer MB, Dull VT, Jobe JB. Social cognition approach to reporting chronic conditions in health surveys National Center for Health Statistics. Vital Health Stat 6(3). 1989.

Library of Congress Cataloging-in-Publication Data

Brewer, Marilynn B.
Social cognition approach to reporting chronic conditions in health surveys / Marilynn B. Brewer, Valerie T. Dull, Jared B. Jobe.
p. cm – (Vital & health statistics. Series 6, Cognition and survey measurement ; no. 3) (DHHS publication ; no. (PHS) 89-1078)
"This research was a part of a larger project entitled "Laboratory-based Research on the Cognition Aspects of Survey Methodology" being conducted by the National Center for Health Statistics under a grant #SES 84-03415 from the National Science Foundation."
"DHHS contract no. 282-2123."
Bibliography: p. ISBN 0-8406-0414-9
1. Health surveys. 2. National Health Interview Surveys (U.S.) 3. Chronic diseases–Reporting. 4. Cognition. 5. Memory. I. Dull, Valerie T. II. Jobe Jared B. III. National Center for Health Statistics (U.S.) IV. National Science Foundation (U.S.) V. Title. VI. Series. VII. Series: DHHS publication ; no. (PHS) 89-1078.
[DNLM: 1. Chronic Disease–epidemiology 2. Cognition. 3. Data Collection–methods. 4 Health Surveys–United States. WT 30 B847s]
RA408.5.B74 1989
614.4'2–dc19
DNLM/DLC
for Library of Congress 89-600021
CIP

This research was part of a larger project entitled, "Laboratory-based Research on the Cognitive Aspects of Survey Methodology," being conducted by the National Center for Health Statistics under grant SES 84-03415 from the National Science Foundation (NSF). Monroe G. Sirken, Ph.D., was the principal investigator for this grant. Murray Aborn, Ph.D., was the NSF program director.

This research was performed under DHHS contract 282-85-2123 from NCHS to the University of California, Los Angeles. Marilynn Brewer, Ph.D., was the principal investigator. Patricia Royston was the NCHS project officer. Jared B. Jobe, Ph.D., is the technical editor and reviewer for Series 6.

The authors thank MAXICARE for making patient medical files available and gratefully acknowledge the work of research assistants Sherry Schneider and Jane Asawa.

Contents

Appendixes

List of tables

List of figures

Social Cognition Approach to Reporting Chronic Conditions in Health Surveys

by Marilynn B. Brewer, Ph.D., and Valerie T. Dull, Ph.D., Institute for Social Science Research, University of California, Los Angeles, and Jared B. Jobe, Ph.D., Office of Research and Methodology

Introduction

Since 1957 the National Center for Health Statistics has annually conducted the National Health Interview Survey (NHIS), with a sample of 40,000–50,000 households. The core content of the NHIS survey instrument is used to elicit information on the frequency and impact of acute health problems, the presence of chronic illnesses and disabilities, and experiences with the health care system for respondents and other members of the household. The specific content of the interview questionnaire varies somewhat from year to year as supplemental sections on specific health problems change (National Center for Health Statistics, 1967, 1973, 1987).

The NHIS interview is conducted in the respondent's home. Each adult responds for himself or herself; adults are permitted to give responses for children and for all relatives in the household who are either unavailable at the time of the interview or are unable or unwilling to participate. Thus, information on chronic conditions is obtained about all members of the household, either through self-report or by proxy responses from a related member of the same household.

One of the purposes of NHIS is to produce statistics on the prevalence of chronic illness and disability in the U.S. population. In attempting to maximize the accuracy and completeness of reports of chronic conditions, many changes have been made to NHIS since the survey began. The changes have primarily been of two types: (1) Changing from a "condition approach" to a "person approach" in January 1969, after an experimental comparison was made of the two approaches, and (2) changing the chronic disease checklists used in the condition approach.

The condition approach utilizes a short series of direct questions to obtain information from respondents about recent occurrences of illnesses and injuries, followed by a chronic condition checklist. Examples are: "Last week or the week before, did you have any accidents or injuries, either at home or away from home?" and "At the present time, do you have any ailments or conditions that have continued for a long time?" The questions were structured to elicit information about any departure from a state of physical or mental well-being resulting from disease or injury (National Center for Health Statistics, 1972).

The current person approach, on the other hand, utilizes probe questions on health-related actions, such as restricting activities and making health care visits, with followup questions on the illnesses and injuries that caused these actions. An example is: "During the past 2 weeks, did ______ stay in bed all or most of the day because of any illness or injury?" Neither the condition approach nor the person approach is a pure condition or person approach but a hybrid approach, employing some aspects of the other approach.

The major change in the condition checklists was made in January 1985, when 100 chronic conditions were divided into six separate lists based on a general differentiation into body systems or parts. One checklist is randomly assigned to each household. Several minor modifications to this approach occurred in January 1982. A more detailed account of the changes made in NHIS is contained in work by Gleeson (National Center for Health Statistics, 1972), and a summary of NHIS methods and instruments is contained in work by Jabine (National Center for Health Statistics, 1987). The organization of the six chronic condition checklists is contained in table 1.

Among other things, the NHIS interview is designed to tap into the respondent's episodic memory—in this case, recall of specific events involving illness or contact with health care facilities within a specified time period. Whenever a particular health-related event is reported using the person approach, the respondent is probed for further information concerning the health-related event that may reveal the presence of a chronic health condition. The respondent, therefore, is required first to retrieve information about specific events or experiences in the recent past and then to recall information relevant to the diagnosis of existing chronic conditions.

As part of the person approach, the NHIS interview also contains items used to tap into the respondent's general (semantic) knowledge about his or her chronic illnesses and personal health status using one of the six

Table 1. Chronic conditions, by organization in National Health Interview Survey checklist

Group 1
Arthritis or rheumatism
Gout
Lumbago
Sciatica
Slipped disk
Bursitis
Skin cancer
Eczema or psoriasis
Acne
Dermatitis
Bunions
Group 2
Cataracts
Glaucoma
Cleft palate
Cerebral palsy
Curvature of spine
Clubfoot
Group 3
Gallstones
Cirrhosis
Hepatitis
Jaundice
Ulcer
Hernia
Gastritis
Enteritis
Diverticulitis
Colitis
Cancer of stomach, etc.
Group 4
Goiter
Diabetes
Anemia
Epilepsy
Multiple sclerosis
Group 4—Con.
Migraine
Neuralgia or neuritis
Nephritis
Kidney stones
Breast cancer
Cancer of uterus
Cancer of prostate
Herpes
Syphilis
Group 5
Rheumatic fever
Hardening arteries
Coronary heart disease
Hypertension
Stroke
Brain hemorrhage
Angina pectoris
Myocardial infarction-heart attack
Damaged heart valves
Tachycardia
Heart murmur
Aneurysm
Blood clots
Varicose veins
Leukemia
Hemorrhoids
Phlebitis
Group 6
Bronchitis
Asthma
Hay fever
Sinusitis
Tonsillitis
Lung cancer
Emphysema
Pleurisy
Tuberculosis

highly structured chronic condition checklists described earlier. In this method respondents are read a list of specific chronic conditions and asked to indicate, for each, whether they (and related adults and children for whom they serve as proxy respondents) have that condition. With this method, the name of the chronic condition is used as a cue and the respondents are asked to recognize whether that cue is applicable to themselves or related household members.

Cognitive analysis of the NHIS checklist procedure

In cognitive psychology, the survey respondent is viewed as a processor of information, much the same as a computer is a processor of information. Cognitive psychology focuses on the respondent's thought processes through four stages in the question-answering process. These stages are comprehension, information retrieval, judgment, and response.

The success with which information about health conditions is elicited depends in part on how that information is encoded and organized in the respondent's long-term memory and what search strategies are employed to retrieve that information from memory. From this perspective, the current checklist procedure being used in the NHIS interview has two potential drawbacks.

First, the organization of chronic conditions into six subsets (checklists) is based on methodological considerations (grouping by disease type), and the organization of some checklists may not necessarily reflect the knowledge of the lay public. As a result, the context in which the chronic conditions occur may or may not match the way they are represented in the individual respondent's hierarchical memory structure. Therefore, the knowledge that oneself or a family member has a particular condition may be less likely to be retrieved from memory.

Second, the retrieval strategy associated with the checklist procedure relies on "bottom up" processing, starting from the condition name as a retrieval cue and working up through various associative pathways to respondents' general knowledge about themselves and other members of their households. From what is known about the retrieval process involved in such recognition memory, it appears that the NHIS approach does not take advantage of associative connections or higher order categorizations that may cue memory for multiple conditions at the same time.

A simplified schematic representation of the issues involved in retrieving information from a general knowledge structure is provided in figure 1. Suppose, for example, that the individual has organized his or her knowledge of chronic illnesses into general categories (conditions that involve chronic pain, stomach troubles, and so on). Such categories are represented as A, B, and C in the figure. Any particular chronic condition (c_i) is stored as a member of one of these categories, with associative links to other members of the same category and to the category as a whole. In turn, the categories are linked to other knowledge about self, friends, and relatives. In figure 1, only one of the categories (A) is associated with the respondent's

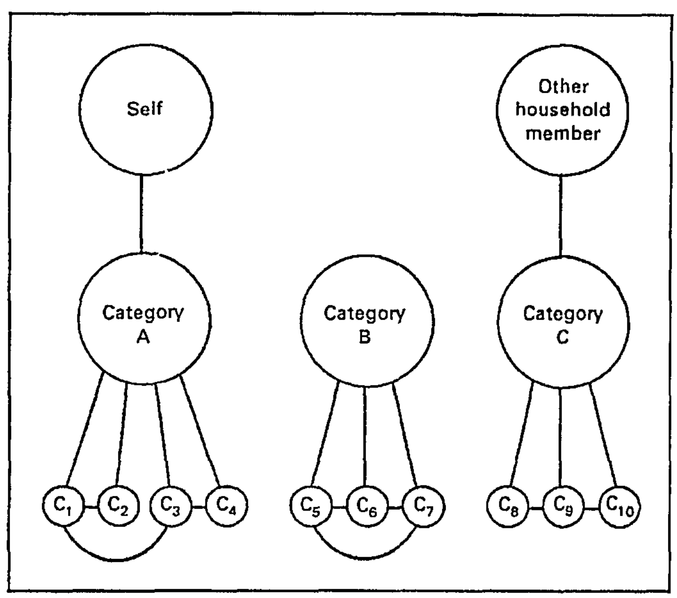

Figure 1. Associative network model of the representation of chronic conditions in memory

self. For instance, the respondent may think of himself or herself as a person who has stomach problems but no heart condition (represented as Category B). Category C in the figure is not associated with the self but is linked to some other household member (such as the knowledge that "my husband has a lung disease").

If figure 1 is an accurate representation of one way in which memory about chronic diseases is structured, then it has implications for how such knowledge will be retrieved from memory. When retrieval starts from a specific cue (for example, c_2), successful recognition depends on whether all of the associative paths that link the cue to the target (in this case, self) are activated during the retrieval process. In terms of the model shown here, this would involve linking up through the system from the cue to the category and from the category to the self. If either of these linkages failed to be activated, recognition would not occur and the condition would not be reported. On the other hand, if the retrieval process started from the more general categories (such as endocrine disorders or heart and blood disorders), direct linkages would be available both down to specific conditions and up to the self, thus possibly increasing the chances that complete linkage between self and any specific condition would be made.

If this model is correct, the efficiency and accuracy of reporting of chronic conditions could be improved by knowledge of how the lay public tends to organize illnesses into natural categories and how these categories are linked to knowledge about self and others. Specifically, organizing the condition categories for NHIS in the same way that respondents typically organize them in memory would increase the probability of respondents' recognizing and reporting conditions of self and others. Because research in the area of social cognition is particularly concerned with how information about self and others is structured in memory, it seems that the insights and methods of social cognition research would be applicable to understanding the problems associated with reporting chronic conditions.

Theory of natural categories

Outside of medical anthropology (for example, D'Andrade et al., 1972), relatively little research has been conducted on how health-related conditions are categorized by the lay population as part of shared social knowledge. Recent advances in the theory of natural object categories (Rosch and Lloyd, 1978; Smith and Medin, 1981) and its extension to social categories (Cantor and Mischel, 1979; Brewer, Dull, and Lui, 1981) provide a framework for studying lay knowledge of chronic health conditions. From this perspective, natural categories differ from formal category systems in that they are made up of "fuzzy sets" (that is, groups lacking sharp boundaries or formal rules of inclusion), with categories being defined in terms of multiple overlapping resemblances among category members rather than in terms of single defining features. Thus, the categorization of health conditions that individuals develop from personal experience and social communication may be quite different from the organizational structure represented in formal classification systems such as *International Classification of Diseases, 9th Revision* (ICD-9) diagnostic coding system of the World Health Organization (1977). For example, whereas diagnostic codes are organized around physiological systems, lay categories may be organized around more behaviorally relevant features, such as degree of debility, embarrassment, or prevalence. The degree of correspondence between lay categorizations and the ICD-9 system has obvious relevance to the future design of health survey questionnaires.

Knowledge about an individual's implicit system of categorization within a particular domain can be derived from judgments of similarity among representative items from that domain. One method used in social cognition research to obtain such similarity judgments is the free-sort procedure. Subjects are given a large set of items representing the domain under investigation and asked to sort these items into groupings in any way they see fit. This technique has the advantage of imposing no a priori structure on subjects' judgments, thus permitting spontaneous associations and clusterings to be represented in the final sortings. When several individual subjects sort the same set of items, the frequency with which specific items are grouped together in the same category becomes a measure of perceived similarity among those items. Analysis techniques such as multidimensional scaling and cluster analysis can then be applied to these similarity judgments to determine the structure underlying the shared perceptions.

The research reported here is part of a larger programmatic effort to apply laboratory-based research methods from cognitive psychology to issues of survey research. Within that larger framework, the objective of the present study was to assess the utility of procedures derived from social cognition research on natural categories to achieving a better understanding of how individuals store and retrieve information about chronic health conditions. The results could be applied to a redesign of the NHIS checklist categories to more closely reflect the manner in which respondents organize the categories in their memory, thereby producing more accurate retrieval and better estimates of the prevalence of chronic conditions. In the study, standard laboratory methods for the assessment of natural categories were applied to the domain of chronic diseases.

Method

Subjects

Participants were recruited for the study by sampling from a list of subscribers to MAXICARE, a health maintenance organization that services the Greater Los Angeles area. A health maintenance organization was selected because the medical records of participants could be used to validate their responses. Letters with return cards enclosed were mailed to 750 subscribers selected at random who matched the following criteria.

- Were 45–64 years of age.
- Lived in the Los Angeles area.
- Had telephone access.
- Had insurance coverage for multiple persons.
- Had medical records that indicated a history of a chronic condition from at least one of these categories: Neoplasms, metabolic disease, blood or circulatory disease, respiratory problems, digestive or genitourinary disease, and musculoskeletal problems.

On the return cards, sample persons were asked to provide a printed name, signature, and phone number of the person in the household interested in volunteering for the study. After the cards had been returned, each household was called so that further screening could be conducted and a session scheduled for the laboratory. (See appendix I for a detailed explanation of recruitment and return rates.)

This recruitment procedure resulted in participation by 70 individuals (34 males and 36 females). In the initial design, all participants were to be interviewed in pairs, which required that two adult members from the same household participate. (The reason for this requirement was to obtain validation data.) However, because of difficulties encountered in recruiting couples, some participants were interviewed individually. The final group of 70 was composed of 62 paired participants (31 pairs) and 8 singles. All experimental sessions were conducted in the Institute for Social Science Research (ISSR) laboratory at the University of California, Los Angeles, by trained graduate student experimenters, with the exception of two sessions conducted in participants' homes. Respondents were paid $20.00 for their participation and reimbursed for parking.

Procedure

Participants arrived at the ISSR laboratory for their scheduled session and were greeted by the two research assistants. For those sessions involving pairs of participants, the two experimenters were randomly assigned to the two family members, who were interviewed in separate rooms of the laboratory to ensure independent responding. At the beginning of the session, each participant was asked to read and sign an informed consent form, which described the nature of the experiment and its general purpose. (See appendix II). An experimental session consisted of three tasks, administered in a fixed order.

Free-sort task

Participants were given a stack of 68 index cards, each with the name of a specific chronic condition printed on it. The 64 conditions were selected from the 1985 NHIS checklist, with approximately an equal number of conditions selected from each of the six checklist categories. All nonspecific diseases were omitted. In addition to conditions taken directly from the interview checklist, the experiment included four conditions (leukemia, cancer of the uterus, herpes, and syphilis) that were added to increase the range of conditions represented in the set to be studied. (Appendix III contains the complete list of conditions used in the experiment, including ICD-9 codings and descriptions for each.)

After participants read through the entire set of cards, they were asked whether they were familiar with the term used to describe each condition. If participants indicated that they had not heard of the term or were not sure what it was, a standardized definition was read to them from a prepared list of descriptions of all conditions included in the study (appendix III).

After they had read the entire set of cards for initial familiarity, participants were given a new set of shuffled cards and were asked to sort those into "piles based on how you think they should go together." No constraints were placed on the number of piles to be generated or the basis on which those sortings should be made. Participants were encouraged to use their own judgment in deciding what conditions should be grouped together and were assured that there were no right or wrong answers for this task.

These procedures were designed to impose minimal structure on the participants' judgments of similarity among the chronic conditions.

The participants were given an unlimited amount of time to complete their sortings. When they indicated that they had finished, they were asked to go through the stacks one last time to evaluate whether each condition had been placed where it belonged. When the participant had verified the final sorting, the experimenter went on to the next phase.

Dimensional-sorting task

Following the free sort, participants were asked to sort the 68 conditions again in a set of structured-sorting tasks. This phase of the experiment consisted of seven different sortings along preselected rating scales. The scales were selected to represent dimensions judged by the research team (in consultation with National Center for Health Statistics staff) to be relevant to the perception of chronic conditions. The seven scales, administered in a fixed order, were:

1. A 4-point scale of perceived incidence, ranging from "extremely common" to "extremely rare"
2. A 4-point scale of perceived painfulness, from "very painful" to "not at all painful"
3. A 4-point scale of perceived seriousness, from "major" to "minor"
4. A 4-point scale of perceived severity, from "disabling" to "not at all disabling"
5. A 3-point scale of embarrassment, from "definitely embarrassing" to "not at all embarrassing"
6. A 3-point scale of perceived controllability, from "definitely controllable" to "definitely not controllable"
7. A 4-point scale of subjective likelihood of the respondent contracting the condition, from "very unlikely" to "already has the condition (or has had it in the past)." This last scale constituted our primary measure of the accuracy of self-reporting with respect to each condition. When participants assigned a card to level 4 on this scale, they were recorded as having that specific condition.

For each of these structured sortings, participants were given a new deck of shuffled index cards with the 68 different conditions and were instructed to place each condition at the appropriate point along the rating scale. Plywood boxes partitioned into labeled sections were provided for this purpose. Each box also had a slot labeled "don't know" for placement of those conditions that the participant found difficult to evaluate on that particular scale. As each sorting was completed, the box was set aside for later recording.

Postsorting interview

Following the eight sorting tasks, a brief, structured, free-response interview was conducted with each participant to allow him or her to talk about personal experience with chronic illness. This interview included a section in which participants were asked to indicate which of the conditions in the study their partners now had or had in the past. Thus, for each pair of interviewees in the experiment, we obtained both a self-report of chronic conditions and a proxy report from the partner.

At the close of the session, participants were thanked for their participation and were requested to sign two forms authorizing the release of their medical record information to the project. They were then given the remuneration agreed on and parking reimbursements. After the participants' departure, the research assistants recorded the results from all of the sortings, entering the data directly on an IBM PC/AT microcomputer.

Results

As a result of the selection and recruitment procedures, the participants in the laboratory study constituted an older than average and more middle-class sample of the Los Angeles area population. The 36 female and 34 male participants ranged in age from 41 to 67 years, with a mean age of 56 years. They came from households containing from one to six persons, with a mean household size of three. Median annual family income of our participant group was $46,809.

Initial familiarity

Table 2 shows the chronic condition terms by the percent of participants who initially indicated that they were unfamiliar with them. (As indicated in the Method section, when this response occurred, the participant was provided with a standardized description of the specific condition.) Because of the selection procedures used to recruit participants for this study, all had had direct or indirect experience with chronic illness and had received some form of medical treatment within the last year. Thus, our participant sample might not be representative of the population at large, in that they were relatively experienced with illness conditions in general. Nevertheless, a few of the terms on our list of 68 were familiar to only one-half of our participants, and two items (enteritis and nephritis) were unfamiliar to more than 75 percent of the participants. By far the majority of terms, however, had unfamiliarity responses of less than 20 percent.

In general, participants who were initially unfamiliar with a particular term were satisfied with the descriptive information provided and did not show any reluctance to incorporate those conditions in the sorting task. (See table 3 and appendix IV for further analyses of "don't know" responses to the dimensional sorting task.)

Table 2. Chronic conditions, by percent of respondents indicating initial unfamiliarity with condition

[Number of respondents = 70]

High unfamiliarity	*Percent*	*Moderate unfamiliarity*	*Percent*	*Low unfamiliarity*	*Percent*
Enteritis	87	Gastritis	27	Cleft palate	9
Nephritis	76	Multiple sclerosis	27	Hernia	9
Diverticulitis	54	Goiter	24	Cirrhosis	7
Phlebitis	49	Hepatitis	21	Curvature of spine	7
Sciatica	49	Sinusitis	21	Gallstones	7
Lumbago	46	Cerebral palsy	20	Hardening arteries	7
Neuralgia or		Glaucoma	20	Cataracts	6
neuritis	43	Jaundice	20	Brain hemorrhage	6
Angina		Rheumatic fever	20	Tuberculosis	6
pectoris	39	Gout	17	Arthritis	
Dermatitis	37	Myocardial infarction-	17	or rheumatism	4
Aneurysm	34	heart attack	16	Cancer of	4
Pleurisy	33	Tachycardia	14	stomach, etc	4
Colitis	30	Bursitis	14	Emphysema	4
		Eczema or psoriasis	14	Stroke	3
		Leukemia	11	Acne	3
		Heart murmur	10	Asthma	
		Anemia	10	Bronchitis	3
		Bunions	10	Clubfoot	3
		Cancer of prostate	10	Epilepsy	3
		Coronary heart		Hypertension	3
		disease	10	Skin cancer	3
		Damaged heart		Syphilis	3
		valves	10	Cancer of uterus	1
		Herpes	10	Diabetes	1
				Hay fever	1
				Hemorrhoids	1
				Varicose veins	1
				Blood clots	0
				Breast cancer	0
				Kidney stones	0
				Lung cancer	0
				Migraine	0
				Slipped disk	0
				Tonsillitis	0
				Ulcer	0

Free-sort analyses

The average number of groupings produced by sortings of the 68 items was 19.54, with a standard deviation of 7.64 and a range of 6–38. A measure of perceived similarity was derived by computing the frequency with which chronic conditions co-occurred with others in the participant sortings. If a condition was put in the same category as another condition, the pairing received a co-occurrence score of 1, and the co-occurrence scores were tabulated for each of the 70 participants. When all sortings had been scored in this way, the result was a unique 68 × 68 co-occurrence matrix for each of the 70 sortings. These co-occurrence scores were then summed across all participants using matrix addition, which produced an aggregate symmetric 68 × 68 co-occurrence matrix with numbers of participants (70) on the diagonal. The data in the aggregate matrix represented the extent to which participants agreed that pairs of conditions were similar to each other. A high co-occurrence value in the matrix meant that the chronic conditions were placed in the same category with high frequency.

Table 3. Means for each of seven rating scales, by cluster

Cluster and condition	Common–rare (1–4)	Painful–not painful (1–4)	Major–minor (1–4)	Disabling–not disabling (1–4)	Embarrassing–not embarrassing (1–3)	Controllable–not controllable (1–3)	Unlikely to get it oneself–have it (1–4)
Arthritis or rheumatism	1.4	1.3	2.7	2.0	2.6	2.1	2.9
Lumbago	2.2	1.7	3.2	2.6	2.8	1.9	2.0
Sciatica	2.4	1.6	3.2	2.5	2.8	1.9	2.0
Bursitis	1.7	1.7	3.4	2.5	2.8	1.8	2.7
Migraine	1.6	1.2	3.2	2.4	2.8	1.9	2.3
Neuralgia or neuritis	2.1	1.9	3.4	2.8	2.9	1.8	2.2
Slipped disk	2.0	1.2	2.6	1.6	2.8	1.9	2.2
Bunions	1.8	2.3	3.9	3.3	2.6	1.4	2.3
Cleft palate	3.4	3.8	3.6	3.3	1.7	2.0	1.0
Curvature of spine	2.6	2.7	3.1	2.6	2.4	2.3	1.8
Clubfoot	3.5	3.6	3.6	2.9	1.8	2.3	1.1
Eczema or psoriasis	2.0	3.3	3.7	3.5	2.1	1.8	2.0
Acne	1.4	3.7	3.9	3.8	2.1	1.6	2.2
Dermatitis	1.7	3.2	3.8	3.6	2.3	1.5	2.0
Herpes	2.3	2.7	2.8	2.9	1.4	2.1	1.4
Syphilis	2.5	2.9	2.1	2.4	1.1	1.6	1.2
Gallstones	1.8	1.4	2.4	2.3	2.9	1.6	2.2
Cirrhosis	2.2	2.5	1.4	1.8	2.0	2.3	1.5
Hepatitis	2.4	2.8	2.2	2.2	2.2	1.6	1.7
Jaundice	2.7	3.7	2.6	2.8	2.4	1.6	1.9
Nephritis	2.7	2.3	2.6	2.6	2.8	1.8	1.7
Kidney stones	2.1	1.2	2.3	2.3	2.8	1.6	2.0
Ulcer	1.5	1.5	2.3	2.4	2.8	1.5	2.3
Hernia	1.8	2.0	2.9	2.5	2.6	1.4	2.1
Gastritis	1.8	1.8	3.0	2.8	2.5	1.6	2.5
Enteritis	2.6	2.3	3.0	2.7	2.7	1.7	1.8
Diverticulitis	2.5	2.1	2.8	2.6	2.5	1.6	2.0
Colitis	2.1	2.0	2.8	2.5	2.4	1.6	2.0
Hemorrhoids	1.4	1.9	3.3	2.9	1.8	1.5	3.1
Bronchitis	1.6	2.5	3.0	2.8	2.8	1.5	2.9
Asthma	1.7	2.5	2.6	2.4	2.7	1.8	1.8
Hay fever	1.3	3.2	3.7	3.3	2.9	1.6	2.4
Sinusitis	1.4	2.4	3.6	3.2	2.8	1.6	2.9
Tonsillitis	1.4	2.1	3.6	3.1	2.9	1.4	2.9
Emphysema	2.0	2.2	1.7	1.6	2.4	2.2	1.7
Pleurisy	2.4	2.0	2.8	2.6	2.7	1.7	2.2
Tuberculosis	2.9	2.7	1.8	1.8	2.1	1.6	1.5
Anemia	2.0	3.8	2.7	2.7	2.9	1.4	2.2
Blood clots	2.2	2.5	1.5	1.7	2.7	1.9	2.1
Varicose veins	1.7	2.6	3.3	3.0	2.4	1.9	2.5
Phlebitis	2.6	1.9	2.3	2.1	2.6	1.8	2.0
Cataracts	1.8	3.6	3.1	2.1	2.8	1.6	2.1
Glaucoma	2.1	3.3	2.8	2.1	2.8	1.8	2.0
Cerebral palsy	2.9	2.8	2.1	1.6	2.0	2.5	1.3
Diabetes	1.9	3.5	1.8	2.3	2.8	1.6	2.0
Epilepsy	2.8	3.2	2.4	2.0	1.5	2.0	1.3
Multiple sclerosis	2.7	2.6	1.6	1.4	2.0	2.6	1.4
Rheumatic fever	2.8	2.5	2.0	2.0	2.8	1.9	1.5
Hardening arteries	1.6	3.0	1.7	1.9	2.7	2.2	2.2
Coronary heart disease	1.6	1.8	1.2	1.4	2.6	2.0	2.2
Hypertension	1.3	3.3	1.7	2.2	2.9	1.5	2.6
Stroke	1.8	1.9	1.2	1.2	2.4	2.5	2.1
Brain hemorrhage	2.6	1.6	1.0	1.1	2.4	2.7	2.0
Angina pectoris	2.0	1.6	1.8	1.9	2.7	2.0	2.1
Myocardial infarction-heart attack	1.7	1.2	1.1	1.1	2.6	2.3	2.2
Damaged heart valves	2.2	2.2	1.4	1.6	2.7	2.0	1.9
Tachycardia	2.1	3.2	2.3	2.3	2.8	2.0	2.1
Heart murmur	2.0	3.5	2.6	2.6	2.8	2.1	2.0
Aneurysm	2.3	2.0	1.3	1.4	2.7	2.3	1.9
Skin cancer	1.8	3.2	2.4	3.0	2.3	1.6	2.2
Cancer of stomach, etc.	2.0	1.3	1.1	1.3	2.1	2.2	2.0
Breast cancer	1.9	2.5	1.2	2.0	2.0	2.0	1.6
Cancer of prostate	2.1	2.0	1.3	1.8	1.9	2.0	1.6
Lung cancer	1.9	1.8	1.0	1.3	2.5	2.3	1.7
Cancer of uterus	1.9	2.1	1.2	1.7	2.1	2.0	1.4
Leukemia	2.5	2.7	1.1	1.3	2.6	2.4	1.6
Gout	2.7	1.7	3.2	2.6	2.6	1.7	1.7
Goiter	2.9	3.1	2.8	2.9	2.4	1.6	1.7

The aggregate co-occurrence matrix was analyzed using several nonparametric techniques to determine what underlying structure could be derived from participants' similarity judgments. The co-occurrence matrix was first subjected to a complete-link hierarchical cluster analysis using the CLUSTAN computer program package, version 2.1. (See appendix IV for a detailed explanation of this computer package.) In this method all objects in the stimulus set begin as separate, single-member clusters. Then, based on similarity values, the most similar items are joined to each other in various stages. The final stage occurs when all objects are contained within a single large cluster.

When the co-occurrence matrix for 68 chronic conditions was subjected to this analysis, the result was a solution of 12 separate clusters, shown in table 4. Descriptively, the final cluster solution had a distinct grouping of the chronic conditions on the basis of body location, with some consideration of symptomatology as well. The most cohesive large cluster groups pertained to cancers, heart disease, and respiratory ailments. Less cohesive clusters involved conditions pertaining to stomach-digestive ailments, liver-kidney problems, and musculoskeletal conditions.

Table 4. Chronic conditions, by organization resulting from hierarchical cluster analysis—12-cluster solution

Cluster 1 Arthritis or rheumatism Lumbago Sciatica Bursitis Migraine Neuralgia or neuritis	**Cluster 7** Rheumatic fever Hardening arteries Coronary heart disease Hypertension Stroke Brain hemorrhage Angina pectoris Myocardial infarction-heart attack Damaged heart valves Tachycardia Heart murmur Aneurysm
Cluster 2 Slipped disk Bunions Curvature of spine Cleft palate Clubfoot	**Cluster 8** Anemia Blood clots Varicose veins Phlebitis
Cluster 3 Eczema or psoriasis Acne Dermatitis Herpes Syphilis	**Cluster 9** Cataracts Glaucoma
Cluster 4 Gallstones Cirrhosis Hepatitis Jaundice Nephritis Kidney stones	**Cluster 10** Cerebral palsy Diabetes Epilepsy Multiple sclerosis
Cluster 5 Ulcer Hernia Gastritis Enteritis Diverticulitis Colitis Hemorrhoids	**Cluster 11** Skin cancer Cancer of the stomach, etc. Breast cancer Cancer of prostate Cancer of uterus Lung cancer Leukemia
Cluster 6 Bronchitis Asthma Hay fever Sinusitis Tonsillitis Emphysema Pleurisy Tuberculosis	**Cluster 12** Gout Goiter

Symptomatology evidently contributed to the aggregate sorting solution for Clusters 1, 3, and 10. Pain and nerve-related conditions were grouped in Cluster 1, which would explain the presence of migraine in this cluster. Although herpes and syphilis joined Cluster 3 rather late in the linkage process, they probably co-occurred in some subjects' sortings because of their associated skin symptomatology. Similarly, Cluster 10 includes conditions that cause tremors and, in the case of diabetes and epilepsy, loss of consciousness.

The two small (two-item) clusters in the solution varied widely in their level of similarity. Cataracts and glaucoma were among the first conditions to link, with a co-occurrence score of 60 (that is, 60 of 70 participants placed these conditions in the same category). On the other hand, the pairing of gout and goiter represents a co-occurrence score of only 14. Both gout and goiter had insufficiently high overall similarities with other conditions in previously formed clusters to be able to link them with those preexisting groups. This cluster probably indicates a confusion on the part of participants as to the proper body system and symptoms to use in assigning gout and goiter.

Reliability of aggregate cluster solution

Although the cluster analysis of the aggregate co-occurrence matrix resulted in meaningful clusterings of the 68 chronic conditions, it is always possible that solutions obtained from aggregated matrixes bear little resemblance to the structure of the sorting produced at the individual level. To test whether our aggregate cluster solution was representative of the participants' own groupings, a correlation was computed between the 12-cluster structure and the individual co-occurrence matrix obtained from each respondent's free sort. The majority of participants produced more than 12 groupings of the stimulus items in their free sorts, indicating that individuals made somewhat finer differentiations than were reflected in the aggregate clustering. Overall, however, correspondence between individual and aggregate matrixes was quite high.

The mean correlation (using Fisher's z-transformation) was .454, with individual correlations ranging from .09 to .71. All the correlations were positive and significantly higher than .00. Only seven participants produced sortings that had correlations of less than .25 with the aggregate cluster solution. An examination of these seven sortings indicated that they were deviant from the typical sorting in that they had either a large number of single-item groupings or a small number of categories in which almost all of the conditions were lumped together. For those participants who sorted the conditions more systematically, there was generally a close correspondence between individual sorting and the aggregated data structure. No indication of meaningful individual differences in the nature of the underlying categorization of chronic conditions was evident from this correlation analysis.

Comparative analyses

For purposes of better understanding the nature of our participants' subjective categorization of chronic conditions, the results of the sorting analysis were compared with the categorizations derived from other methods. First, the natural clusterings were compared with formal categorizations based on physiological subsystems. Second, the results of cluster analyses from the free-sort data were compared with clusterings derived from similarities among the diseases in their placement on the dimensional sortings. (A third comparison, based on multidimensional scaling, is reported in appendix IV.)

Comparing natural clusterings with formal classification systems

A nonparametric analysis known as the quadratic assignment paradigm (QAP) was employed to assess the match between our participants' clustering of the 68 chronic conditions and those generated from more formal classification systems. (See appendix IV for a technical description.) Two comparisons were conducted, one involving classifications based on ICD-9 codes and the second involving the disease classifications represented in the NHIS checklists.

In order to conduct these analyses, two new similarity matrixes were constructed to represent the clustering of conditions within the ICD-9 and checklist systems. The checklist matrix was constructed by assigning a similarity score of 1 to those pairs of conditions that fell within the same checklist subcategory in the NHIS interview schedule and assigning a 0 to pairings not in the same subcategory. A similar procedure was followed to generate a similarity matrix representing the ICD-9 categories. All conditions that are grouped together at the two-digit level in the ICD-9 coding scheme were assigned pair similarity scores of 1, and conditions that do not share codes at this level were assigned similarity scores of 0. The NHIS and ICD groupings used to generate these two formal similarity matrixes are displayed in tables 1 and 5, respectively.

The QAP analysis is done by performing a cross multiplication of the data matrix (the 68 × 68 aggregate co-occurrence matrix) and the comparison similarity matrix. If the comparison matrix mirrors the data matrix, then the sum of these cross-products should be high; the large values of co-occurrence would remain in the summation, but the small values of co-occurrence would drop out as a result of being multiplied by 0. This summed cross-product can be thought of as an unnormalized correlation coefficient. To test against randomness, the expected value and variance of this index was calculated (Hubert and Levin, 1976) and a z-score derived. This score was then compared with a normal distribution in the usual manner to find the appropriate p-value for the obtained index.

Table 5. Chronic conditions, by organization in *International Classification of Diseases, 9th Revision* (ICD-9) and National Health Interview Survey (NHIS) checklist group

ICD-9 category	*NHIS group*
Category I	
Arthritis or rheumatism	1
Gout	1
Lumbago	1
Sciatica	1
Slipped disk	1
Bursitis	1
Bunions	1
Cleft palate	2
Curvature of spine	2
Clubfoot	2
Neuralgia or Neuritis	4
Category II	
Eczema or psoriasis	1
Acne	1
Dermatitis	1
Category III	
Nephritis	4
Kidney stones	4
Category IV	
Gallstones	3
Cirrhosis	3
Hepatitis	3
Ulcer	3
Hernia	3
Gastritis	3
Enteritis	3
Diverticulitis	3
Colitis	3
Category V	
Bronchitis	6
Asthma	6
Hay fever	6
Sinusitis	6
Tonsillitis	6
Emphysema	6
Pleurisy	6
Category VI	
Rheumatic fever	5
Hardening arteries	5
Coronary heart disease	5
Hypertension	5
Stroke	5
Brain hemorrhage	5
Angina pectoris	5
Myocardial infarction-heart attack	5
Damaged heart valves	5
Tachycardia	5
Heart murmur	5
Aneurysm	5
Blood clots	5
Varicose veins	5
Hemorrhoids	5
Phlebitis	5
Category VII	
Cataracts	2
Glaucoma	2
Cerebral palsy	2
Epilepsy	4
Multiple sclerosis	4
Migraine	4
Category VIII	
Jaundice	3
Goiter	4
Diabetes	4
Anemia	4
Category IX	
Leukemia	5
Skin cancer	1
Cancer of stomach, etc.	3
Breast cancer	4
Cancer of prostate	4
Lung cancer	6
Cancer of uterus	4
Category X	
Tuberculosis	5
Herpes	4
Syphilis	4

Results of these analyses are shown in table 6. Although both of the formal classification matrixes fit the obtained co-occurrences significantly better than chance, the ICD-9 codes show a better fit than the NHIS checklist categories, probably because NHIS has only 6 categories to simplify administration, compared with 10 categories for

Table 6. Statistics from quadratic assignment paradigm analysis comparing natural clustering with *International Classification of Diseases, 9th Revision (ICD-9) and National Health Interview Survey (NHIS) groupings of chronic diseases*

Statistic checklist	*ICD-9*	*NHIS checklist*
Similarity index	13,110.0	12,796.0
r	.690	.513
z	31.88	24.57
p	.00001	.00001

the ICD-9. The checklist organizational structure differs from the ICD-9 system in the placement of the various cancers. In the ICD-9 system, all cancers (neoplasms) are coded together at the two-digit level, but in the checklist grouping, the various cancers are separated according to the body part affected. In our participant sorting, the cancers were grouped together in a tight cohesive cluster, thus resembling the ICD-9 classification system more closely.

Clearly the physiological systems that form the basis of the ICD-9 categories are also a significant component of the participants' natural categorization. Although participant sortings did match the ICD-9 categories quite closely, there were several exceptions with interesting implications. For example, diabetes was placed by our participants with nervous system conditions, although it is classified in the ICD-9 categories as a metabolic condition. Migraine is a nervous system condition, but participants placed it with musculoskeletal-connective tissue problems such as arthritis. If participants had strictly adhered to the etiology underlying the ICD-9 organization, herpes and syphilis would have been sorted with tuberculosis because they are all from the infectious-parasitic disease category. The sorting cluster that violates the ICD-9 categorization to the greatest degree is the cluster including gallstones, cirrhosis, and kidney stones (Cluster 4). Three different ICD-9 categories are represented in this cluster—digestive, endocrine-metabolic, and genitourinary—indicating the extent to which participant categories crosscut ICD-9 classifications.

Where participant sortings deviate from ICD-9 codes, the differences tend to reflect differences between the external symptomatic characteristics of a condition and its underlying systemic base, with participants focusing on external symptomatic characteristics. Herpes and syphilis, for instance, are seen by participants as most similar to other skin conditions, and tuberculosis is grouped with lung disorders and congestive conditions. Hemorrhoids are seen by our research participants as more similar to diseases of the intestine than to circulatory conditions. Diabetes is classified by participants among the nervous system conditions, in comparison to its classification as an endocrinologic condition in the ICD-9 coding. Research participants also drew more distinction between spinal and skeletal disorders than are reflected in the ICD-9 classification, as well as somewhat different distinctions between heart and circulatory conditions. In general, however, the aggregate clustering of respondent sortings corresponds remarkably closely to ICD-9 classification at the two-digit level.

Profile analyses

In analyses of the free-sort data, an index of similarity between chronic conditions is derived from the frequency with which they are sorted into the same category (co-occurrence). Another method for calculating similarity scores is to assess the degree to which pairs of conditions have similar patterns of ratings on relevant dimensions. Such "profile" similarity measures have been used in earlier research on the natural classification of diseases, such as work by D'Andrade et al. (1972). After mean ratings have been obtained on selected rating dimensions, the differences between conditions in their ratings across the dimension can be calculated, and this index can be subjected to cluster analysis in a manner parallel to analysis of free-sort co-occurrences. The results of the dimensional sortings in our study were subjected to this type of profile analysis for purposes of comparing the structure of the free-sort clusterings with the structure of groupings derived more indirectly through the rating scales, based on their placements in the structured sorting task. The ratings obtained for each condition on each scale are reported in table 3 (along with the frequency with which participants were unfamiliar with each condition). These ratings reflect our participants' general knowledgeability regarding chronic disorders, particularly with respect to painfulness, severity, and relative frequency of occurrence. In general, participants rated most chronic disorders as moderately controllable, with the exception of various cancers, multiple sclerosis, heart attacks, strokes, and birth defects. Chronic diseases were also generally evaluated as not particularly embarrassing (with the exception of epilepsy, syphilis, herpes, and some deformities, such as cleft palate and clubfoot). Among our research participants, although syphilis was rated as the most embarrassing condition, epilepsy was almost as high in embarrassment rating.

The mean ratings on the seven dimensions were treated as a profile for each condition in the study. A squared Euclidean distance measure was computed as an index of similarity between the profiles of each pair of conditions. The 68 × 68 matrix of similarity (distance) scores was subjected to cluster analysis for purposes of comparison with the clusters derived from the free-sort co-occurrences.

In general, the profile similarity scores did not produce as good a cluster solution as the free-sort measure. The optimal profile clustering was found to be 4 groupings (as opposed to 12 for the aggregate free-sort matrix), and the correlation between the two solutions was not high (appendix IV). The groupings generated by the four-cluster profile solution are presented in table 7. Cluster 1 is characterized by relatively low ratings on the severity and seriousness scales and high ratings on relative frequency. Cluster 4, on the other hand, is characterized by high ratings on seriousness and Cluster 3 by relatively high embarrassment ratings. Cluster 2 was distinguished primarily by moderate ratings on most of the scales. Although these groupings have some intuitive appeal in terms of single dimensions of distinction among conditions, they collapse across many other interesting differences and do not produce an optimal categorization in terms of category size or intercategory differentiation. Only the free-sort procedure produced meaningful cohesive clusterings that appear to correspond to natural categories.

Table 7. Chronic conditions, by organization resulting from profile cluster analysis—Four-cluster solution

Cluster 1: Minor, frequent

Arthritis
Bunions
Bursitis
Cataracts
Skin cancer
Ulcer
Acne
Hernia
Dermatitis
Gastritis
Anemia
Migraine
Hypertension
Varicose veins
Hemorrhoids
Bronchitis
Hay fever
Sinusitis
Tonsillitis
Eczema or psoriasis

Cluster 2: Moderate

Gout
Lumbago
Sciatica
Slipped disk
Glaucoma
Prostate cancer
Curvature of spine
Gallstones
Hepatitis
Jaundice
Enteritis
Diverticulitis
Colitis
Goiter
Diabetes
Pleurisy
Tuberculosis
Neuralgia or neuritis
Nephritis
Kidney stones
Rheumatic fever
Tachycardia
Heart murmur
Phlebitis

Cluster 3: Embarrassing

Cleft palate
Clubfoot
Epilepsy
Herpes
Syphilis

Cluster 4: Major

Cerebral palsy
Cirrhosis
Cancer of stomach, etc.
Mutiple sclerosis
Breast cancer
Emphysema
Leukemia
Cancer of uterus
Lung cancer
Arteriosclerosis
Coronary heart disease
Stroke
Brain hemorrhage
Angina pectoris
Myocardial infarction-heart attack
Damaged heart valves
Aneurysm
Blood clots

Discussion

The present study served two general purposes. The main contribution of this study is methodological; it represents a unique application of social cognition research methods to health problems. Specifically, classical social cognition methods of free sort and multidimensional scaling were used to gain insights into the natural categorization of chronic health conditions by the lay public. The results of the cluster analyses of this study suggest that respondents with different personal medical histories may be quite similar in their subjective categorization of chronic illnesses. If this preliminary finding is confirmed by subsequent research, such conditions could be grouped into cohesive natural categories that have considerable generality across individuals in this culture. Even with the small sample size used here, the lay categories show overlap with the systemic-based classification of chronic conditions represented in ICD-9 codes, although the natural categorizations crosscut ICD-9 coding categories in a number of cases.

The fact that the laboratory methods revealed a coherent natural category system makes it possible to test whether a category-based approach would be useful in surveys designed to elicit information on the prevalence of chronic conditions. The checklist recognition procedure currently employed in NHIS requires respondents to answer "yes" or "no" as to whether they or household members have a series of conditions read from a checklist. This method may not be the most efficient for eliciting such information because it does not make use of subjective categorizations to guide memory retrieval. The results of a small unpublished pilot study demonstrated the feasibility of an alternative interview procedure involving open-ended responses to a series of hierarchically structured questions for obtaining self and proxy reports on specific chronic conditions. A larger scale field study would be needed to determine whether the reports obtained from the hierarchical interview are more accurate, when cross-validated against medical records, than are reports obtained from the checklist procedure.

Another important methodological feature of this study is the use of health maintenance organization patients as respondents. The medical records of these patients are readily accessible. Although the records themselves are subject to inaccuracies, respondent-reported conditions can be validated against their medical records to the extent that conditions have been professionally diagnosed. In this way, reasonably accurate estimates of underreporting or overreporting can be determined, and the effects of cognitive or other independent variables can be measured.

Finally, the free-sort methodology may also apply to other kinds of questions. These include questions on food groups in dietary intake studies and kinds of physical activity in lifetime exercise studies. Further research will determine if this methodology is applicable to these and other areas.

References

Aldenderfer, M. S., and R. K. Blashfield. 1984. *Cluster Analysis.* Beverly Hills, California: Sage Publications.

Baker, F. B., L. J. Hubert, and J. V. Schultz. 1977. *Quadratic Assignment Program (QAP) Technical Manual.* Laboratory of Experimental Design, University of Wisconsin.

Brewer, M. B., V. Dull, and L. Lui. 1981. Perceptions of the elderly: Stereotypes as prototypes. *J. Pers. Soc. Psychol.* 41:656–670.

Cantor, N., and W. Mischel. 1979. Categorization Processes in the Perception of People. In *Advances in Experimental Social Psychology,* edited by L. Berkowitz. Vol. 10. New York: Academic Press.

D'Andrade, R., N. Quinn, S. Nerlove, and A. K. Romney. 1972. Categories of Disease in American-English and Mexican-Spanish. In A. K. Romney, R. Shepard, and S. Nerlove (eds.) *Multidimensional Scaling: Theories and Application in the Behavioral Sciences.* Vol. 2. New York: Seminar Press.

Hubert, L., and J. Levin. 1976. Evaluating object set partitions: Free-sort analysis and some generalizations. *Journal of Verbal Learning and Verbal Behavior* 15:459–470.

National Center for Health Statistics, W. G. Madow. 1967. Interview data on chronic conditions compared with information derived from medical records. *Vital and Health Statistics.* Series 2, No. 23. PHS Pub. No. 1000. Public Health Service. Washington: U.S. Government Printing Office.

National Center for Health Statistics, G. A. Gleeson. 1972. Interviewing methods in the Health Interview Survey. *Vital and Health Statistics.* Series 2, No. 48. DHEW Pub. No. (HSM) 72-1048. Health Services and Mental Health Administration. Washington: U.S. Government Printing Office.

National Center for Health Statistics, W. G. Madow. 1973. Net differences in interview data on chronic conditions and information derived from medical records. *Vital and Health Statistics.* Series 2, No. 57. DHEW Pub. No. (HSM) 73-1331. Washington: U.S. Government Printing Office.

National Center for Health Statistics, T. B. Jabine. 1987. Reporting chronic conditions in the National Health Interview Survey, a review of tendencies from evaluation studies and methodological test. *Vital and Health Statistics.* Series 2, No. 105. DHHS Pub. No. (PHS) 87-1379. Public Health Service. Washington: U.S. Government Printing Office.

Rosch, E., and B. B. Lloyd. 1978. *Cognition and Categorization.* Hillsdale, New Jersey: Erlbaum.

Smith, E. E., and D. L. Medin. 1981. *Categories and Concepts.* Cambridge, Massachusetts.: Harvard University Press.

Wishart, D. 1982. *CLUSTAN User Manual.* Release 2.1. Interuniversity/Research Councils Series. Report No. 47. Edinburgh: Edinburgh University.

World Health Organization. 1977. *Manual of the International Statistical Classification of Diseases, Injuries, and Causes of Death.* Based on the Recommendations of the Ninth Revision Conference, 1975. Geneva: World Health Organization.

Young, F. W., and R. Lewyckyj. 1981. *A Guide for Users of ALSCAL-4: A Nonmetric Multidimensional Scaling and Unfolding Program with Several Individual Differences Options.* Chapel Hill, N.C.: Psychometric Laboratory, University of North Carolina.

Appendixes

Contents

Appendix figure

Appendix I
Recruitment procedures

Summary of recruitment outcomes

At the beginning of the study, MAXICARE, a health care provider, referred the Institute for Social Science Research (ISSR) to the Hawthorne Community Medical Group (HCMG), their data management firm. Recruitment for the sample was subsequently performed through this firm.

Recruitment was initially designed to target 35 individuals who satisfied the following requirements:

- Were 45–64 years of age
- Had insurance coverage for multiple persons (including one other person in the 45–64 age range)
- Lived in the immediate Los Angeles area
- Had access to a telephone
- Had a history of one of the following types of conditions or had a partner with such a history
  - Neoplasms
  - Metabolic diseases
  - Blood and circulatory problems
  - Respiratory ailments
  - Digestive or gastrourinary disease
  - Musculoskeletal problems

Subjects were recruited from multiple-person households to participate in pairs in order that each member of the pair could act as a self-respondent and proxy respondent. The 35 recruited subjects with their partners would result in 70 participants.

On January 27, 1986, HCMG mailed 500 letters to health care subscribers in the Greater Los Angeles area. Their names and addresses had been selected randomly from the HCMG medical data files of individuals who met the outlined criteria. By returning a card enclosed with the letter, subscribers could volunteer for a University of California, Los Angeles (UCLA) study called HealthInfo. These cards were addressed directly to ISSR.

In the last week of March, it became apparent that the goal of 35 pairs of subjects would not be met by the initial mailing of 500 letters. As of April 1, 63 cards had been returned (a 13-percent return rate). Of these 63, 42 had already been successfully contacted. Of these 42, 14 pairs agreed to participate, and 28 had refused or had not met the criteria. In summary, 33 percent of those contacted prior to April 1 had agreed to participate.

Based on the rate of return and participation from the first 500 cards, an additional 250 cards were deemed sufficient to complete the study. The second mailing was completed March 27, 1986.

A total of 750 letters were mailed for recruitment purposes. All subsequent figures reported are based on a mailing of 750 and a return of 101 cards (a 14-percent return rate).

After the cards had been received at ISSR, the volunteers were called and screened, using a standardized form. It required approximately 257 telephone calls to recruit the number of participants necessary for the study. This is an average of 2.5 calls per participant, or an average of 6.4 calls for every scheduled session (40 sessions in all). These calls took place predominantly on weekdays from 7:00 p.m. to 9:30 p.m. (Los Angeles is a city of commuters; calls made earlier than 7:00 p.m. were usually picked up by an answering machine.)

A total of 101 cards were returned. Screening resulted in the following breakdown:

- 28 single females
- 12 single males
- 58 couples
- 3 persons of unknown status (no telephone number given or telephone out of service)

Of the 58 couples:

- In 10 cases, one partner refused to participate
- Five lived too far away (perceived or actual)
- Two were too busy to participate
- Two were in the medical profession
- One spouse was out of town for the study
- In one case, the spouse did not speak English
- In six cases, miscellaneous problems or excuses occurred

Thus, 27 couples were unable to participate. This left 31 eligible participants.

To obtain these 31 eligible pairs, the screening criteria were, of necessity, relaxed such that one partner could be out of the age range by plus or minus 5 years.

A pervasive problem throughout the screening process was the lack of cooperation from the spouse who did not volunteer for the study. The project director phoned 12 couples (10 "partner refused" and 2 "too busy") to attempt refusal conversions; in all cases the spouse remained unpersuaded. It seemed that the major obstacle that prevented these couples from participating was the time required to travel to UCLA, find parking, and participate in a 2-hour session. (Unfortunately, UCLA has a reputation for its lack of parking and confusing campus structure.)

To overcome this problem and to complete the study in a timely manner, the study was expanded to include several individuals who would participate without partners. The final breakdown in participants in the experiment was as follows:

31 couples	=	62 participants
8 singles	=	8 participants
Total	=	70 participants

During the course of the study, most sessions were conducted after 5:00 p.m. and on weekends to accommodate participants' work schedules. Several steps were taken to reduce the number of cancellations and persons who did not arrive for scheduled appointments (no shows). Despite these efforts, there were 5 no shows and 11 last-minute cancellations. One couple was scheduled three times before actually participating. Participants were generally scheduled 1–4 weeks in advance. Those scheduled more than 1 week in advance were sent campus maps with detailed parking instructions. Those scheduled 3–4 weeks in advance were usually telephoned to confirm their appointments. Those couples who canceled or missed their appointments were called immediately and asked to reschedule. Of 5 no shows, 3 couples subsequently rescheduled, and of 11 last-minute cancellations, 6 were rescheduled.

Even with written and oral instructions, participants reported that they found negotiating an unfamiliar college campus a somewhat frustrating experience (as any college freshman can attest). For this reason, two sessions were conducted in the households of the volunteers. One couple was unable to negotiate the UCLA campus because of the husband's recent knee surgery, and another volunteer was a single male who could not travel to UCLA because of some recent surgery that had arisen from complications with his diabetes. Conducting the interview in the home of the single person went smoothly. However, finding adequate space and privacy to interview each member of the pair independently was more difficult in a small house than it would have been in the laboratory.

Appendix II
Experimental instructions and instruments

Informed consent form

HEALTHINFO STUDY
Date: August 20, 1985

INFORMED CONSENT FORM

By signing this form, I agree to participate in the research called "HealthInfo" under the following conditions:

I understand that during this 1 1/2 to two-hour session, I will be asked to provide my ideas about the nature and causes of health problems. I also understand that I will be paid $20.00 for my participation.

I understand that my participation is purely voluntary and that I am free not to answer any particular questions.

I understand that information gathered from any of the tasks performed during this experimental session will be held in strict confidence by the researchers at ISSR and the National Center for Health Statistics.

I further understand that the information gathered in this session will be used only for the purposes of evaluating techniques of health surveys and will not involve risk or discomfort.

I understand that I may refuse to participate or may withdraw from this study at any time without any negative consequences. Also, the investigator may stop the study at any time. I also understand that no information which identifies me will be released without my separate consent, and that all identifiable information will be protected to the limits allowed by law. If the study design or the use of the data is to be changed, I will be so informed and my consent reobtained. I understand that if I have any questions, comments, or concerns about the study or the informed consent process, I may write or call the office of the Vice Chancellor-Research Programs, 3134 Murphy Hall, UCLA, Los Angeles, CA 90024, (213) 825-8714. I acknowledge that I have received a copy of this form.

Signed: ______________________ Date: ______________

We are sure that you will find the session interesting. Your participation will be a contribution to a valuable research project. You may request further information about this study by contacting Dr. Marilynn B. Brewer or Dr. Valerie T. Dull at (213) 825-0711, Institute for Social Science Research, Bunche Hall, UCLA, Los Angeles, CA 90024.

Medical record release form

AUTHORIZATION TO RECEIVE OR RELEASE MEDICAL INFORMATION

The signing of this form allows my health maintenance organization to release medical information, under terms authorized by Section 306 of the Public Health Service Act (42 U.S.C. 242k).

I authorize my health maintenance organization to release to representatives of the Institute for Social Science Research at UCLA records pertaining to my medical history and any medical treatments I have received, for the purpose of research.

This authorization becomes effective immediately and expires on July 31, 1986.

I understand that the Institute for Social Science Research at UCLA will not release any of my medical information to any other person or group, and that the same restrictions for receipt or release of medical information apply to the Institute for Social Science Research as to my health maintenance organization and that no further authorization is indicated in this form.

I understand that no permanent records will be kept in a form that would permit individual identification and that all information will be used solely for statistical research and report purposes.

NAME:

SIGNATURE:

DATE:

WITNESS:

DATE:

Experimental instructions and script

HEALTHINFO STUDY SCRIPT

[EXPERIMENTERS WELCOME BOTH SUBJECTS TO THE SESSION AND TAKE THEM TO SEPARATE ROOMS OF THE LAB WHERE THEY BEGIN THE SESSION.]

Thank you for coming to UCLA today. To start this session, I'd like you to read this form and if you have any questions about it, please do not hesitate to ask me.

[ALLOW AMPLE TIME FOR SUBJECT TO READ THE INFORMED CONSENT FORM.]

Now, do you have any questions? At this point, I'd like you to sign this form if you understand the nature of your participation during this session.

[LET SUBJECT SIGN FORM AND GIVE THEM A COPY.]

To start off, I want to tell you what this study is about. Today we will be concerned with your ideas about illness. We will be asking you to perform a variety of tasks during this time in order to find out about your thoughts regarding disease and illness. We want to stress that during this whole time, there will be no right or wrong answers. This is not at all a test of any kind. If there is a point at which you do not understand the instructions, please feel free to ask.

To begin, here are some index cards. Each one has a name of a disease or illness typed on it. Often medical terms are not familiar. Since we'll be using these for the rest of the session, I'd like you to first read each term. Then, tell me if you've heard of the term and definitely know what it is, you have heard of the term and are pretty sure you know what it is, or have not heard of the term or you are not sure you know what it is.

[E GOES THROUGH EACH CARD WITH SUBJECT AND RECORDS FAMILIARITY RATINGS FOR EACH ITEM. (SEE RECORDING SHEET) IF SUBJECT INDICATES THEY HAVE NOT HEARD OF TERM, E SHOULD READ FROM DEFINITION LIST.]

These are a set with the same conditions written on them but they have been shuffled. I'd like you to sort these cards into piles based on how you think they should go together. Each pile will represent those illnesses that you think are similar in some way. You can use as many different piles as you think are necessary. In other words, I'd like you to freely sort these cards into groups based on how you think the illnesses should be grouped together. Again, keep in mind that we are interested in what you think these illnesses are like. There are no in correct answers. Do you have any questions about this task? Please begin your sorting. (IF NECESSARY READ: We would like you to put these illnesses into groups based on something that they have in common. In other words, you will put them in the same pile if they are similar in some way.)

WHEN SUBJECT IS FINISHED: Do you want to go through each stack to see that everything belongs there?

[WHEN THE SORT TASK IS FINISHED, RUBBER BAND EACH STACK TOGETHER AND PLACE ON THE SIDE FOR RECORDING AT END OF SESSION.]

Now, we are going to ask you to go through these same cards again. This time we want you to tell us some specific things about each illness. First, we are interested in how common or rare you think each of these illnesses are. You may feel that some of these diseases are ones that are extremely common and others are extremely rare. If you feel that the illness is extremely common, put the card here. If you feel it is extremely rare place it here. You may feel that others of these illneses are only somewhat common or somewhat rare, in which case you would place the cards here or here. If you feel you do not know enough about a particular illness to make a guess about whether it is common or rare, use this don't know slot. But remember, there are no right or wrong answers and we would prefer that you indicate what you think even if you are not sure.

[THROUGHOUT THIS PORTION OF THE SCRIPT, THE E SHOULD POINT TO THE APPARATUS AND SHOW THE PHYSICAL LOCATIONS OF THE SCALE AS THEY TALK. DEMONSTRATE THIS PROCEDURE BY WALKING THROUGH IT STEP-BY-STEP.]

Now, we'd like you to think about how physically painful these illnesses might be. Some of these conditions are physically very painful while others are not at all painful. Using these slots again, please place the illness here if you think it is very painful or here if it is not at all painful. If you think the illness is somewhat painful put it here or if you think that it is slightly painful put it here. Here is the don't know slot but please try to make a guess about what you think, even if you are not sure.

Some of these illnesses might cause life-threatening situations for certain people. This time, please think about how life-threatening these diseases are for people that have them. If you think they are major illnesses put them here. If they are minor illnesses place them here. Then, if they are somewhat major put them here or somewhat minor place them here. (IF NECESSARY: Here is the don't know slot but remember there are no right or wrong answers and we are interested in what you think, even if you are not sure.)

These illnesses could also cause problems in the way people lead their lives. This time, please think about how disabling each of these illnesses are. Then, if they are very disabling put them in this slot. If they are not at all disabling put them in this slot. If the illnesses could be disabling put them here. If they are not likely to be disabling put them here. (IF NECESSARY: Here is the don't know slot but remember there are no right or wrong...

Please think about how embarassing some of these illnesses might be. Some could definitely be seen as embarassing, others could be seen as not at all embarassing. Still others may be somewhat embarassing. Please take these cards and once more put them in the slot where you think they should go. (IF NECESSARY: Here is the don't know slot but remember there are no right or ...)

Certain of these illnesses could definitely be controllable or some of these could definitely not be controllable. So, if someone has the illness, it may or may not be possible to control ~~the~~ its occurrence or flare-up. Other illnesses might only be partly controllable. Once again, please take these cards and put them in the slot where you believe they should go. (IF NECESSARY: Here is the...)

Lastly, I would like for you to think about your own chances of getting each of these illnesses. If you think it is very unlikely that you will get this illness or there is no chance at all, put the card in this slot. If you think you could get this illness put it in here. If you think it is very likely that you will get this illness put the card here. If the illness is something you already have at this time, put the card here. Remember, this slot is for illneses you know you have.

Thank you. I see the conditions in this pile are conditions you have indicated you have. We would like to be able to ask you some further questions about these particular illnesses later if that's ok with you. We have now completed all the sorting tasks. The last task involves some questions we would like to ask you. These questions have to do with your feelings regarding illness and the impact it has had on your life and your partner's life as well. This information is of course completely confidential.

[ADMINISTER POSTSESSION INTERVIEW QUESTIONNAIRE. E RECORDS ALL RESPONSES VERBATIM (OR AS CLOSE AS POSSIBLE)]

We would like to thank you very much for your participation. As a final request, may we ask you to read this authorization to release medical information? As part of this study, we need to have access to the formal code numbers for any illnesses for which you have been treated. It will help us to determine how accurate our findings about illness are.

(IF NECESSARY READ: We are interested in only the official code numbers that the doctor writes. We are not interested in any background information or history. We need to know only the code numbers for the precision of this research project. It will be used for research purposes only.)

[PAY AND THANK SUBJECTS. ESCORT OUTSIDE. RECORD DATA IMMEDIATELY.]

I.D. Number ___________

OMB NO. ___________
Exp. Date ________

POSTSESSION INTERVIEW QUESTIONNAIRE

NOTICE: Information contained on this form which would permit identification of any individual or establishment has been collected with a guarantee that it will be held in strict confidence, will be used only for purposed stated for this study, and will not be disclosed or released to others without the consent of the individual or the establishment in accordance with section 308(d) of the Public Health Service Act (42 USC 242k).

Background information:

1. Sex (Record without asking) _______

2. What is your date of birth? __/__/__

3. What is your relationship to (Name of partner*)? _______________

4. What is the highest grade in school that you completed?

5. How many individuals are living in your household? This number includes you, your spouse/partner, any children or other financially dependent relatives.

   Number ______

6. What is the gross annual income, before taxes, for your household? This would include wages and salaries before deductions, as well as income from sources such as Social Security, unemployment insurance, pensions, welfare, alimony or child support.

   $____________

7. Do you consider your general health to be:

   __ Excellent
   __ Good
   __ Fair
   __ Poor

"Partner" refers to the paired household member

Target Condition of the Self: (If no apriori target condition, skip this section.)

Note: Subject must have disclosed this target condition during the session. If not disclosed, then replace with a condition that has been disclosed.

1. Do you have an idea or theory about how you developed (target condition?) Where it came from? Why?

   [open ended]

2. If you had to say a specific thing that is responsible for your illness, which ONE would you choose?

   __Yourself (own actions)
   __Someone else (specific person, actions)
   __Environment (pollution, stress, etc.)
   __Heredity
   __Chance (God, fate, etc.)

3. Do you think you can control the episodes, attacks or spells that you experience with your illness?

   __Not at all (Skip 3a)
   __Slightly
   __Somewhat
   __Very much

   3a. If so, how do you control it?
   [open ended]

4. When your illness flares up, what do you have to do?
   [open ended]

5. Rate how much effect you've experienced in these different aspects of your life because of (target condition).

		No Effect	Slight Effect	Some Effect	Alot of Effect	N/A or DK
a.	Athletic Activity/ Exercise	1	2	3	4	8
b.	Reliance on Medication	1	2	3	4	8
c.	Amount of Sleep/Rest	1	2	3	4	8
d.	Personality	1	2	3	4	8
e.	Personal appearance	1	2	3	4	8
f.	Work/Career Activities	1	2	3	4	8

	No Effect	Slight Effect	Some Effect	Alot of Effect	N/A or DK
g. Family Life	1	2	3	4	8
h. Social Life	1	2	3	4	8

Proxy Interview Section: (Subject now answers for partner)

6. Of the conditions we have been discussing, are there any that (partner) has?

   [open ended]

NOTE: The following questions should be asked for partner's target condition (if any AND if disclosed in the question above.)

7. Do you have an idea or theory about how (partner) developed (target condition)? Where it came from? Why?

   [open ended]

8. If you had to say a specific thing that is responsible for (partner's) illness, which ONE would you choose?

   __Partner (their own actions)
   __Someone else (specific person, actions)
   __Environment (pollution, stress, etc.)
   __Heredity
   __Chance (God, fate, etc.)

9. Do you think (partner) can control the episodes, attacks or spells that they experience with their illness?

   __Not at all (Skip 9a)
   __Slightly
   __Somewhat
   __Very much

   9a. If so, how do they control it?
   [open ended]

10. When their illness flares up, what do they have to do?

   [open ended]

11. Rate how much effect (partner) has experienced in these different aspects of their life because of (partner's target condition).

	No Effect	Slight Effect	Some Effect	Alot of Effect	N/A or DK
a. Athletic Activity/ Exercise	1	2	3	4	8
b. Reliance on Medication	1	2	3	4	8
c. Amount of Sleep/ Rest	1	2	3	4	8
d. Personality	1	2	3	4	8
e. Personal appearance	1	2	3	4	8
f. Work/Career Activities	1	2	3	4	8
g. Family life	1	2	3	4	8
h. Social life	1	2	3	4	8

Appendix III Definitions of chronic conditions included in study

01: Arthritis of any kind or rheumatism

A general term for inflammation of a joint which causes pain, swelling, stiffness and often changes the shape of the joint. (714, 715) (729.0)

02: Gout

A disease caused by build up of a waste deposited in and around joints. Symptoms are pain, inflammation, redness and swelling around the joints, especially the big toe. (712, 274)

03: Lumbago

A dull ache in the lower back—may be caused by overworking muscles in the area or by problems with one of the disks between the vertebrae. (724.2)

04: Sciatica

Severe pain in the back of the thigh along the length of the large sciatic nerve running down the leg. It is caused by inflammation or damage to the nerve. Symptoms are a sharp pain running down the back of the thigh, numbness, and tingling. (724.3)

05: A slipped or ruptured disk

Damage to the cartilage disks between the vertebrae of the backbone due to too much pressure put on it. If severe pressure is put on a disk suddenly, the disk may break open or be moved from its correct position. It may cause severe pain because of pressure on nerves running between vertebrae. (722)

06: Bursitis

An inflammation of the fluid sac found around joints and over bones near the skin. It results in pain and swelling of the area. Locations often affected are shoulder, elbow, knee, and big toe. (726, 727.2, 727.3)

07: Skin cancer

An abnormal growth on the skin. Often caused by overexposure to harmful sun rays. (172.8, 172.9, 173)

08: Eczema or psoriasis

Skin disorder in which there may be inflammation, itching, swelling, blistering, scaling, a rash, or discharge. The cause is usually unknown but in some cases it may be caused by an allergy or fungal infection. Psoriasis is a skin irritation with redness and scaling of the outer surfaces of arms and legs which lasts a long time. (696)

09: Acne

A disease of the skin which results in small red bumps caused by an inflammation in the gland at the bottom of a hair in the skin. (706)

10: Dermatitis

A general term for inflammation of the skin with itching, redness, and perhaps bumps. (681.8, 692)

11: Bunions

Inflammation of the joint of the big toe causing pain and swelling. Sometimes caused by wearing shoes too tight for many years. (727.1)

12: Cataracts

A clouding of the crystalline lens of the eye so that light rays can't pass through, causing poor vision and perhaps blindness. Cataracts may occur in older people as part of aging or in younger people because of disease. (366)

13: Glaucoma

Disease of the eye in which there is an increase in the pressure inside the eye causing the eye to become hard and perhaps causing blindness. A person with glaucoma may have pain in the eyes, headaches, and fuzzy vision. (365)

14: A cleft palate or harelip

A tuck or split of the upper lip or roof of mouth caused by a birth defect. (749)

15: Palsy or cerebral palsy

Abnormal development of the nervous system because of damage to the brain, sometimes at birth. The symptoms include poor muscle control and coordination. (343)

16: Curvature of the spine

A bending of the spine forward, backward, or to either side. (737)

17: A clubfoot

A condition caused by contraction of one or more muscles of the foot causing an abnormal foot position. (754.7)

18: Gallstones

A small stone formed in the gallbladder from the concentrated bile stored there. They can plug the duct going to the small intestine, causing pain and digestion problems. (574)

19: Cirrhosis of the liver

Chronic disease of the liver in which the cells are destroyed, scarring occurs and the organ becomes hard and doesn't work well. (571.2, 571.5, 571.6)

20: Hepatitis

Inflammation of the liver caused by a viral infection or poisoning. A person with hepatitis will usually have yellow colored skin, an enlarged sore liver, a fever, be very weak, nauseous, and lose weight. (571.4) (573.3)

21: Jaundice

Yellowness of the skin, eyes, and body fluids caused by too much of a yellow substance in the blood and tissues. This substance is made from worn-out red blood cells. (277.4)

22: Ulcer

An open sore on the skin or on an internal organ. Ulcers are areas of dead tissues which heal slowly and sometimes bleed and make pus. The most common are those in the digestive system caused by digestive enzymes and acids attacking and digesting the lining of the esophagus, stomach, or small intestine. (531, 532, 533)

23: Hernia

The pushing of an organ or part of an organ through the wall of the area it belongs in; a rupture. Hernias are sometimes caused by straining, as when lifting something heavy. (550) (552, 553)

24: Gastritis

Inflammation of the stomach. Symptoms are pain or tenderness, nausea, vomiting, or a coated tongue. (535)

25: Enteritis

Inflammation of the intestines. (555)

26: Diverticulitis

Inflammation of the sacs and pouches of the large intestine. (562)

27: Colitis

Inflammation of colon (large intestine) usually causing pain and diarrhea. (556, 558)

28: Cancer of the stomach, intestines, colon, or rectum

An abnormal, malignant cell growth found in the stomach, intestines, colon, or rectum. Cancer cells don't work correctly, use up the body's energy, and replace the normal cells in these areas so that they don't work properly. (151, 152, 153, 154)

29: Goiter

An enlargement of thyroid which causes a swelling at the front of the throat. Some persons with goiter may also be nervous, tremble, sweat a lot, be skinny, and have a high heart rate. (240, 241, 242) (246)

30: Diabetes

A disorder of carbohydrate metabolism, due to an absence or lack of insulin produced from the pancreas. It may be more common in certain families or be developed later in life. A person with diabetes has extra sugar in the blood and urine, needs to urinate often, is thirsty, loses weight, and is nervous and weak. (250)

33: Anemia

A shortage of red blood cells in the body or not enough hemoglobin inside the red blood cells carrying oxygen to tissues. Symptoms are paleness, weakness, and shortness of breath. (280, 282, 283, 284)

34: Epilepsy

A disease in which the brain doesn't work correctly, at times resulting in a partial or complete loss of consciousness and in some cases convulsions. (345)

35: Multiple sclerosis

A disease of the nervous system which slowly progresses. It affects a person by causing scars to form on the nerves of the brain and spinal cord. Usually affects young adults with symptoms of double vision, slow, difficult speech, weakness, and trembling. (340)

36: Migraine

A very severe throbbing headache which affects a person from time to time. The headache will usually be on the side of the head and the person may also have vision problems, be unable to stand light, have an upset stomach, and be sweating. (346)

37: Neuralgia or neuritis

Sudden attacks of pain along the course of one or more nerves and inflammation of a nerve. (729.2)

38: Nephritis

Inflammation of the kidneys. Symptoms may be fever, dull pain in the back, water retention, rapid pulse, vomiting, and discoloration of urine. (582)

39: Kidney stones

A hard stone-like object made from minerals in the urine that forms in the kidneys and may stop the flow of urine if it gets into a ureter and plugs it. (592.0)

40: Breast cancer

An abnormal, malignant growth found in the breast. Cancer cells don't work correctly, use up the body's energy and replace the normal cells in these areas so that they don't work properly. (174)

41: Cancer of the prostate

An abnormal, malignant cell growth found in the prostate gland. This gland is found in men and is responsible for making a milky fluid that is put into semen. Cancer cells don't work correctly, use up the body's energy and replace the normal cells in these areas so that they don't work properly. (185)

42: Rheumatic fever

Rheumatic fever can result from a strep throat infection. The person will suddenly feel joint pain and fever and inflammation of the heart may result. At this point the valves of the heart are commonly damaged. (390, 392, 393, 398)

41: Hardening of the arteries or arteriosclerosis
Thickening and hardening of the walls of blood vessels, especially the arteries, which make blood flow more difficult. (440)

42: Coronary heart disease
Disease of the artery that goes around or encircles the heart. (411, 414.0)

43: Hypertension, sometimes called high blood pressure
Higher than normal force of the blood pushing against the walls of the arteries; high pressure symptoms may be dizziness, headaches, bleeding from small blood vessels, vision problems, and kidney inflammation. (401, 402, 403, 404, 405)

44: A stroke
A sudden, severe attack, especially one caused by the bursting of a blood vessel or a clot in the brain. Symptoms of a brain stroke are sudden unconsciousness, loud and difficult breathing, paralysis on one side of the body, a cold sweat on the body, a below-normal temperature, and speech problems. (436)

45: A hemorrhage of the brain
Massive bleeding in brain tissue. (431, 432)

46: Angina pectoris
A severe pain in the chest, usually around the heart, which may spread to the arms, neck, or jaw. The person will have a feeling of pressure around the heart, will not move, may be sweating and very afraid. It's caused by the heart not getting enough blood. (413)

47: A myocardial infarction, heart attack
The dying of a part of the heart muscle because the blood flow through an artery bringing fresh blood to the heart is decreased or stopped. The person will have a severe squeezing pain in the chest, irregular heartbeats, may go into shock and the heart may stop beating. (410, 412, 414.8)

48: Damaged heart valves
Injury of valves in the heart between the different chambers on each side that keep the blood flowing in the correct direction as the heart muscle contracts. (394, 395, 396, 397) (424)

49: Tachycardia or rapid heart
Abnormal rapidity of heartbeat or pulse rate above 100 per minute. (427.0, 427.1, 427.2) (785.0)

50: A heart murmur
An abnormal, soft blowing sound which may be heard when listening to the heart or the large blood vessels around it. Murmurs are heard as the blood moves through a valve in the heart not working properly, an abnormal narrowing or enlargement of one of the vessels near the heart, or openings in the heart where they should not be. (785.2)

51: An aneurysm
An enlargement in one spot of a blood vessel, usually the large artery leading from the heart, caused by a weakness in the vessel's wall; it may cause a rupture. (414.1, 441, 442) (437.3)

52: Blood clots
Jelly-like mass of blood which doesn't flow normally through blood vessels. (453)

53: Varicose veins
Swollen, twisted, and knotted veins. Most common in legs. Veins have little valves along their length that allow blood to move along them in only one direction, preventing it from backing up. If a vein enlarges, the valves won't close together and blood can then flow back and pool in the vein. (454) (456)

54: Hemorrhoids or piles
A group of enlarged veins around the anus and rectum. These veins may bleed, be painful or itchy. (455)

55: Phlebitis or thrombophlebitis
Inflammation of a vein. Symptoms are pain along the vein, swelling below the inflammation, and a change in the color of the skin. (451)

56: Bronchitis
An inflammation of the lining of the tubes which take air into the lungs. Symptoms are pain in the chest, a cough, and fever. (491)

57: Asthma
A condition in which the person coughs a great deal and has trouble breathing because the tubes taking air from the main branch of the lungs tighten and swell. (493)

58: Hay fever
An allergy to pollen or dust, causing the nose, eyes, and upper airways to become irritated. A person with this condition will sneeze, have a runny nose, and sore, red, watery eyes during certain times of the year, especially spring and fall. (477)

59: Sinusitis
Inflammation of a sinus membrane in the nose. May result in difficulty in breathing through nose, pain, or fever. (473)

60: Tonsillitis
Inflammation of the tonsils. Symptoms are a sore throat, fever, headache, chills, and swollen lymph nodes in the neck. (474)

61: Lung cancer
An abnormal, malignant cell growth found in the lungs. Cancer cells don't work correctly, use up the body's energy

and replace the normal cells in these areas so that they don't work properly. (162)

62: Emphysema

A chronic lung disease in which the air spaces in the lungs become enlarged and the tissue between them becomes stiff. The person will have difficulty breathing and a cough. The lungs will become larger than normal, causing a barrel-shaped chest. (492)

63: Pleurisy

Inflammation of the thin membrane covering the outside of the lungs, the inside of the chest, and the top of the diaphragm. There are several different kinds of pleurisy but symptoms are usually temperature, painful breathing, and maybe a collection of pus, blood, or clear fluid in the chest. (511)

64: Tuberculosis

Infectious disease caused by bacteria which usually affects the lungs but can affect other areas of the body. Symptoms vary, depending on type. Small bumps usually form in the infected area and there is fever, sweating, tissue death, scarring, weight loss, and weakness. If it is affecting the lungs, the person will usually have a cough, chest pain, and blood in their mucus and spit. (011, 012)

65: Leukemia

Disease which affects blood-making organs, causing too many white cells to be made. Symptoms may include poor appetite, weakness, enlarged spleen and liver, swollen lymph nodes, anemia, pain in bones, bleeding, and an inability to fight off infection. (204, 205, 206, 207, 208)

66: Cancer of the uterus

An abnormal, malignant cell growth found in the uterus. Cancer cells don't work correctly, use up the body's energy and replace the normal cells in these areas so that they don't work properly. (179, 180, 182)

67: Herpes

An infectious skin disease caused by one of the herpes viruses. The virus causes eruptions of the skin or mucous membranes, painful blisters usually occur on the head or face or genitals. If it returns, it's usually in the same spot as before. (054)

68: Syphilis

A severe, contagious venereal disease caused by bacteria, spread mostly by sexual contact or through infected blood. If the disease is not treated in the early stages, it may cause sores on the skin, bones, and liver and may cause heart damage, paralysis, or insanity. (091.0)

Appendix IV
Analysis techniques

Cluster analysis and the CLUSTAN computer package

Cluster analysis is a method of classifying objects into groupings based on some similarity measure. In this study, the similarity measures were the co-occurrence scores from the aggregate free-sort matrix. The hierarchy procedure from the CLUSTAN computer package was used to analyze these data. In a hierarchical clustering procedure, all the objects begin as single entities unlinked to any others. Then, based on similarity scores, the most similar objects are linked together in a step-by-step process until all objects end up in one large cluster. For every step in the analysis, one linkage occurs. Once an object has joined a particular cluster, it remains there for the rest of the analysis. This sequence of linkages, or fusions, can be represented in a tree diagram that graphically displays which objects have joined each other for every step.

The procedure began with 68 objects—each chronic condition in a cluster by itself. In each of the subsequent fusions, similarity was judged by the frequency of co-occurrence in the free-sort task, with high values indicating greater similarity between chronic conditions. For this study, a complete linkage option was selected as the method for evaluating similarity among the conditions. Basically, every object in a cluster, using this analysis, is completely linked to every other member of the cluster. (See Aldenderfer and Blashfield (1984) for a detailed explanation of the fusion criterion.)

The output is a sequence of linkages with their associated fusion coefficients (the numerical co-occurrence values at which various clusters merge). Based on these output results, one is able to see a history of the cluster linkages from 68 single-member clusters to 1 cluster with 68 members. From these results, a dendrogram figure in which the chronic conditions were graphed by their fusion values, was plotted using the tree procedure.

To determine the optimal number of clusters in this analysis, the fusion values were plotted against the number of clusters. This heuristic is commonly used as a method to show how many clusters best represent the data. When the graph appears to flatten, it suggests that no new information is being contributed to the linkages that follow. However, researchers also suggest that this heuristic be used in conjunction with "subjective inspection" (Aldenderfer and Blashfield, 1984).

Note on sortings of unfamiliar conditions

For the most part, participants seemed comfortable sorting the diseases with which they were initially unfamiliar based on the definitions read to them earlier in the experiment. These participants generally sorted the unfamiliar diseases into clusters similar to clusters formed by those who said they knew of those diseases initially. In other words, participants did not seem to sort all of their unknown diseases into one pile or to assign them haphazardly to groupings. Those diseases that many participants were unfamiliar with (enteritis, diverticulitis, and nephritis) were frequently sorted into single card categories and, therefore, did not affect the co-occurrence cluster solution. In conclusion, the participants were able to place initially unfamiliar diseases into meaningful groups based on a brief description; therefore, their sortings of these unfamiliar diseases probably did not affect the integrity of the co-occurrence cluster solution.

Comparison with multidimensional scaling

Multidimensional scaling (MDS) was used as another nonparametric method of analysis to represent the similarity data. A classical nonparametric MDS analysis was performed on the co-occurrence matrix. Only the lower half of the matrix was used because similarity between pairs was transitive. The diagonal values of the matrix were removed because they were not necessary for the calculation of inter-point distances.

Using this method, the co-occurrences were transformed into "proximities," which were represented in a dimensional space. The proximities, or distances, were derived from a ranking of the interobject similarities. They were then represented graphically as configurations of points in an n-dimensional space, thus showing the similarity or dissimilarity among the chronic conditions.

In order to evaluate whether a dimensional solution might be appropriate for the data being analyzed, two factors should be considered. The first is whether the elements are arrayed in a dimensional manner, having unique values, and with points spread rather evenly along dimensions. This is especially true for solutions using many elements, such as

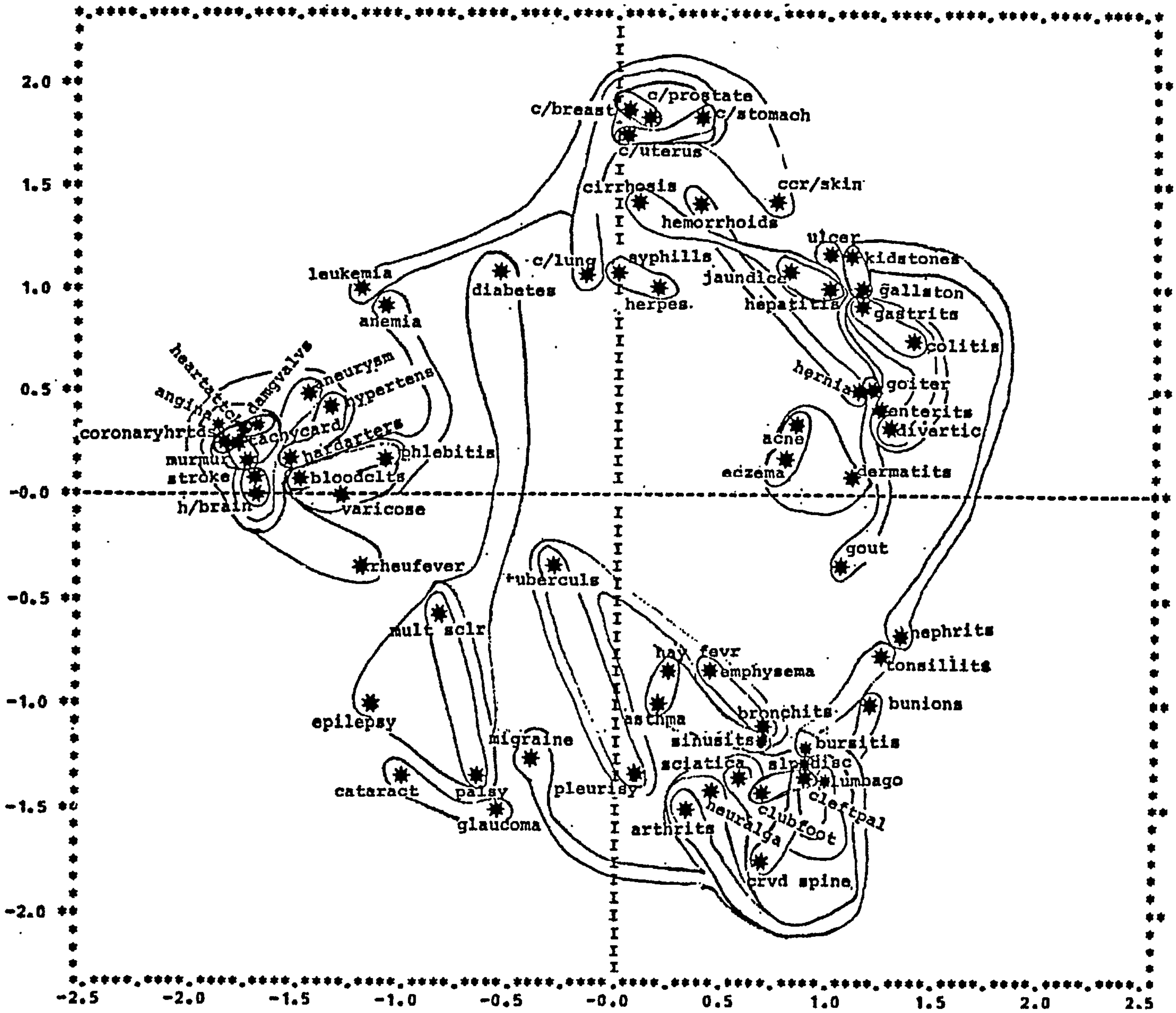

Figure I. Chronic condition proximities—nonmetric analysis

the 68 chronic conditions used in this study. The second factor to consider when analyzing the appropriateness of MDS is the actual interpretability of the dimensions within the solution.

The ALSCAL procedure within the SAS statistical package was used as the data analysis computer package (Young and Lewyckyj, 1981). The result was a two-dimensional MDS solution, which is shown in figure I. Most of the conditions are spread along the two dimensions. However, there is some "clumping" for the heart conditions, such as heart attack, tachycardia, and angina, as well as for musculoskeletal conditions, such as bursitis and lumbago. Certain of these conditions are so close together that they were not printed by the computer program because they had identical proximity values.

In addition, although many of the points have been fitted along the dimensions, it is not clear just what concepts are being represented in the data solution. Coronary heart disease has the most extreme *x*-axis negative value (–1.83) in the solution; colitis has the most extreme positive value (1.41). Potential substantive labels that come to mind, such as "inherited-contracted" or "major-minor," are readily dismissed when noting that many cancers are neutral on this dimension, as are tuberculosis, pleurisy, and asthma. Similarly, the *y* axis does not appear to be interpretable. Cancers of the prostate, breast, and stomach are the most positively extreme (coordinates around 1.85); glaucoma, arthritis, and curvature of the spine are the most negative (with coordinates ranging from –1.47 to –1.71). Although "severity" might seem an appropriate label at first, it is an unlikely choice because brain hemorrhage and heart attack are neutral with respect to this dimension.

A simple way to compare the adequacy of the similarity solution for the MDS and the clustering analysis is to embed the clusters for the free-sort task into the MDS solution obtained for the same data matrix. If both solutions are representing the data with high agreement, then the clusters should form neat concentric circles around the conditions that are near each other when superimposed in the two-dimensional space. Such a pattern would indicate that the cluster and MDS solutions matched in their degree of association among chronic conditions. A poor match would lead to messy elongations within the two-dimensional space when the clusters are embedded. Taken together with the other characteristics of the MDS solution, such a poor match would be further confirmation that MDS is not a very good means of representing the underlying similarity structure of the obtained data.

Figure I displays the clusters as they appear when superimposed in the two-dimensional space. Elongations in the figure are clearly evident, with the exception of two strong clusters—heart conditions and certain cancers—both of which lie near poles of the two dimensions.

Based on this comparison, it would seem that MDS does not do a very effective job of reflecting the inherent structure of the similarity data. Although one might argue that increasing the number of dimensions in the solution would help the interpretability of the solution, such was not the case in our analysis. Unfortunately, MDS is limited as an analysis tool after about three dimensions because it is difficult to conceptualize elements in a space with four or more dimensions. For these reasons, MDS was not performed with more than three dimensions.

The cluster analysis solution would appear to be the best means of representing the similarity structure in these data because of its emphasis on grouping similar entities rather than distancing dissimilarities, as is the case in MDS. In addition, clusters were formed on the basis of subjective similarity, without regard to any particular dimensions. MDS, on the other hand, entailed fitting the similarities to a few (2–4) meaningful dimensions. Using this type of evaluative process to judge similarity does not appear to be the strategy employed by respondents in the free-sort task. Further analyses, therefore, were focused on interpreting the cluster analysis solution as the better representation of similarity judgments.

Quadratic assignment paradigm

The quadratic assignment paradigm was the technique used to examine the organizational structure that subjects mapped onto the chronic conditions used in the study. The paradigm assumes that there is some independent way to organize or structure the set of objects a priori. (In this study, the set of objects was the 68 chronic conditions typed on individual cards.) This analysis enables an investigator to test several known organizational structures against the data actually obtained. Some possible a priori organizations of these chronic conditions are the National Health Interview Survey (NHIS) checklist groups and the 9th Revision International Classification of Diseases (ICD-9) coding categories. A brief explanation of this analysis is discussed next.

For this study, subjects were asked to sort 68 cards into groups or piles according to "how they should go together." The task was not limited in any way, so subjects had no restrictions as to the number of groupings created or the number of cards within each grouping. Following this free-sort task, the number of groupings were counted and the conditions that had been placed within them were recorded. Scoring was performed by giving every pair of conditions within a group a co-occurrence score of 1. Every subject's free-sort solution was scored in this way, according to what cards had been placed together within the same pile. If a pile contained a single member, the chronic condition within it received a score of 0 with all other conditions because it had not co-occurred with others within that particular subject's free-sort solution.

When all sort solutions had been scored in this way, the result was a unique 68 × 68 co-occurrence matrix for each of the 70 free-sort solutions. These co-occurrence scores were then summed across all subjects, using matrix addition, which produced an aggregate and symmetric co-occurrence matrix with the same dimensions and N (70) on the diagonal. This data matrix represented the degree to which conditions were seen as being similar to other conditions. Thus, the higher the co-occurrence values, the greater the degree of similarity as judged by subjects who placed them in the same category with high regularity.

In addition to this matrix, two structure matrixes were specified such that the checklist groupings and the ICD-9 categories for these conditions could be tested against the obtained data matrix. The checklist structure matrix was constructed by giving a score of 1 to the pairings of conditions within each of the six groupings of the checklist and a 0 to pairings not in the same grouping. The same procedure was followed for a second structure matrix constructed in the same manner but with 0, 1 pairings based on the ICD-9 categories. The analysis issue, then, involved the extent to which these two structure matrixes fit the data matrix.

The analysis in this paradigm was done by performing a cross-multiplication of the data matrix and the relevant structure matrix. If the structure matrix mirrors the data matrix, then the sum of these cross-products should be high; the large values of co-occurrence would remain in the summation, but the small values of co-occurrence would drop out as a result of being multiplied by 0. This summed cross-product can be thought of as an unnormalized correlation coefficient. For an approximate test against randomness, the expected value and variance of this index can be obtained by formulation (rather than unwieldy permutation) and a z-score calculated. This score can then be compared with a normal distribution in the usual manner to find the appropriate p-value for the obtained index. (See Hubert and Levin, (1976) for expected value and variance formulas for the index.)

This analysis was performed using a customized computer package called the Quadratic Assignment Program (Baker, Hubert, and Schultz, 1977) available at the University of California, Santa Barbara.

The results of this analysis and the degree of association between each of the two a priori groupings and the obtained data matrix are reported in the body of this report.

Profile analysis procedures

The 70 participants were asked to rate the 68 diseases (on scales from 1–3 or 1–4) on seven dimensions: common-rare, painful-not painful, major-minor, disabling-not disabling, embarrassing-not embarrassing, controllable-not controllable, and unlikely to get the disease-have it already. To cluster the 68 diseases using these attributes, the mean rating for each disease on each of the seven attributes was used to create a profile matrix.

	Attribute						
	1	2	3	4	5	6	7
Disease 1							
2							
.							
.							
.							
68							

The CLUSTAN statistical package was used to perform the cluster analysis. Squared Euclidean distance was the similarity coefficient chosen. Procedure CORREL was used to create the similarity matrix between diseases, and procedure HIERARCHY (Wishart, 1982, p. 31) created the fusion hierarchy of the similarity matrix. Procedure TREE (Wishart, 1982, p. 37) was used to plot the dendrogram produced by HIERARCHY.

To find the best cluster solution, several agglomerative hierarchical algorithms were explored. Procedure COMPARE (Wishart, 1982, p. 18) was used to compare the hierarchical classifications produced by CLUSTAN with the original similarity matrix. The resulting cophenetic correlations were:

Complete linkage	.56
Single linkage	.55
Average linkage	.65
Ward's method	.52

(The similarity coefficient was sums of squares instead of Euclidean distance.)

The higher the cophenetic correlation, the greater is the agreement between the original data and the solution. Because the average linkage solution had the highest cophenetic correlation, its solution was utilized for further analyses.

Procedure RULES in CLUSTAN (Wishart, 1982, p. 14) was used to test for the significant number of clusters in the hierarchical clustering sequence. The "upper tail" rule (commonly referred to as Mojena's Rule) in RULES was used to compute z-scores for each level of hierarchy. The first level to reach a z-score greater than 1.96 is considered a significant number of clusters. For the 68-disease profile solution, $n = 4$ clusters was significant.

In order to find the optimal four-cluster solution, the k-means procedure was performed on the four-cluster average-linkage solution. In the k-means procedure, the similarity of an object to other objects in its cluster and to all other clusters is tested, and the object is relocated if it is more similar to another cluster. Program RELOCATE (Wishart, 1982, p. 43), which performs the k-means analysis, relocates until all objects remain stable in their current clusters.

To verify the k-means relocation of the average-linkage solution, we also had k-means relocate the diseases from a random four-cluster start. (The computer arbitrarily assigned objects to the initial four-clusters.) The random-start solution was identical to the k-means average-linkage solution, verifying that the diseases were being meaningfully relocated to each of the four clusters.

The k-means average-linkage solution is the profile solution that is discussed in the body of this report. It yielded four clusters that have intuitive appeal: the minor disease profile, the moderate profile, the major profile, and the embarrassing profile.

☆ U.S. GOVERNMENT PRINTING OFFICE: 1992—3 4 2 -3 2 7 / 6 0 0 1 7

Vital and Health Statistics series descriptions

SERIES 1. **Programs and Collection Procedures** – These reports describe the data collection programs of the National Center for Health Statistics. They include descriptions of the methods used to collect and process the data, definitions, and other material necessary for understanding the data.

SERIES 2. **Data Evaluation and Methods Research** – These reports are studies of new statistical methods and include analytical techniques, objective evaluations of reliability of collected data, and contributions to statistical theory. These studies also include experimental tests of new survey methods and comparisons of U.S. methodology with those of other countries.

SERIES 3. **Analytical and Epidemiological Studies** – These reports present analytical or interpretive studies based on vital and health statistics. These reports carry the analyses further than the expository types of reports in the other series.

SERIES 4. **Documents and Committee Reports** – These are final reports of major committees concerned with vital and health statistics and documents such as recommended model vital registration laws and revised birth and death certificates.

SERIES 5. **International Vital and Health Statistics Reports** – These reports are analytical or descriptive reports that compare U.S. vital and health statistics with those of other countries or present other international data of relevance to the health statistics system of the United States.

SERIES 6. **Cognition and Survey Measurement** – These reports are from the National Laboratory for Collaborative Research in Cognition and Survey Measurement. They use methods of cognitive science to design, evaluate, and test survey instruments.

SERIES 10. **Data From the National Health Interview Survey** – These reports contain statistics on illness, unintentional injuries, disability; use of hospital, medical, and other health services, and a wide range of special current health topics covering many aspects of health behaviors, health status, and health care utilization. They are based on data collected in a continuing national household interview survey.

SERIES 11. **Data From the National Health Examination Survey, the National Health and Nutrition Examination Surveys, and the Hispanic Health and Nutrition Examination Survey** – Data from direct examination, testing, and measurement on representative samples of the civilian noninstitutionalized population provide the basis for (1) medically defined total prevalence of specific diseases or conditions in the United States and the distributions of the population with respect to physical, physiological, and psychological characteristics, and (2) analyses of trends and relationships among various measurements and between survey periods.

SERIES 12. **Data From the Institutionalized Population Surveys** – Discontinued in 1975. Reports from these surveys are included in Series 13.

SERIES 13. **Data From the National Health Care Survey** – These reports contain statistics on health resources and the public's use of health care resources including ambulatory, hospital, and long-term care services based on data collected directly from health care providers and provider records.

SERIES 14. **Data on Health Resources: Manpower and Facilities** – Discontinued in 1990. Reports on the numbers, geographic distribution, and characteristics of health resources are now included in Series 13.

SERIES 15. **Data From Special Surveys** – These reports contain statistics on health and health-related topics collected in special surveys that are not part of the continuing data systems of the National Center for Health Statistics.

SERIES 16. **Compilations of Advance Data From Vital and Health Statistics** – Advance Data Reports provide early release of information from the National Center for Health Statistics' health and demographic surveys. They are compiled in the order in which they are published. Some of these releases may be followed by detailed reports in Series 10–13.

SERIES 20. **Data on Mortality** – These reports contain statistics on mortality that are not included in regular, annual, or monthly reports. Special analyses by cause of death, age, other demographic variables, and geographic and trend analyses are included.

SERIES 21. **Data on Natality, Marriage, and Divorce** – These reports contain statistics on natality, marriage, and divorce that are not included in regular, annual, or monthly reports. Special analyses by health and demographic variables and geographic and trend analyses are included.

SERIES 22. **Data From the National Mortality and Natality Surveys** – Discontinued in 1975. Reports from these sample surveys, based on vital records, are now published in Series 20 or 21.

SERIES 23. **Data From the National Survey of Family Growth** – These reports contain statistics on factors that affect birth rates, including contraception, infertility, cohabitation, marriage, divorce, and remarriage, adoption; use of medical care for family planning and infertility, and related maternal and infant health topics. These statistics are based on national surveys of childbearing age.

SERIES 24. **Compilations of Data on Natality, Mortality, Marriage, Divorce, and Induced Terminations of Pregnancy** – These include advance reports of births, deaths, marriages, and divorces based on final data from the National Vital Statistics System that were published as supplements to the *Monthly Vital Statistics Report* (MVSR). These reports provide highlights and summaries of detailed data subsequently published in *Vital Statistics of the United States*. Other supplements to the MVSR published here provide selected findings based on final data from the National Vital Statistics System and may be followed by detailed reports in Series 20 or 21.

For answers to questions about this report or for a list of reports published in these series, contact

Scientific and Technical Information Branch
National Center for Health Statistics
Centers for Disease Control
Public Health Service
6525 Belcrest Road, Room 1064
Hyattsville, MD 20782
(301) 436-8500

DEPARTMENT OF
HEALTH & HUMAN SERVICES

Public Health Service
Centers for Disease Control
National Center for Health Statistics
6525 Belcrest Road
Hyattsville, Maryland 20782

DHHS Publication No. (PHS) 89-1078, Series 6, No. 3