Development of the Speech Reception Test

Description of the development for use in the Health and Nutrition Examination Survey of 1974-75 of a new test of hearing discrimination ability for speech sounds based on the revised Central Institute for the Deaf Sentences devised by a working group of the Committee on Hearing and Bioacoustics of the National Research Council.

DHEW Publication No. (HRA) 78-1345

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
Public Health Service
Health Resources Administration
Hyattsville, Md. October 1977

Library of Congress Cataloging in Publication Data

Elkins, Earleen.
Development of the speech reception test.

(Vital and health statistics: Series 2, Data evaluation and methods research; no. 71) (DHEW publication; no. (HRA) 78-1345)
Bibliography: p.
1. Audiometry, Speech. I. Causey, G. Donald, joint author. II. Roberts, Jean, joint author. III. Title. IV. Series: United States. National Center for Health Statistics. Vital and health statistics: Series 2, Data evaluation and methods research; no. 71. V. Series: United States. Dept. of Health, Education, and Welfare. DHEW publication; no. (HRA) 78-1345. [DNLM: 1. Hearing tests. 2. Hearing disorders—Diagnosis. W2 A N148vb no. 71]

RA409.U45 no. 71 [RF294.5.S6] 312.'07'23s
ISBN 0-84060088-7 [617.8'9] 77-2800

For sale by the Superintendent of Documents, U.S. Government Printing Office
Washington, D.C. 20402

Stock No. 017-022-00589-1

NATIONAL CENTER FOR HEALTH STATISTICS

DIVISION OF HEALTH EXAMINATION STATISTICS

COOPERATION OF THE U.S. BUREAU OF THE CENSUS

Under the legislation establishing the National Health Survey, the Public Health Service is authorized to use, insofar as possible, the services or facilities of other Federal, State, or private agencies.

In accordance with specifications established by the Division of Health Interview Statistics, the Bureau of the Census, under a contractual arrangement, participated in planning the survey and collecting the data.

Vital and Health Statistics-Series 2-No. 71

DHEW Publication No. (HRA) 78-1345
Library of Congress Catalog Card Number 77-2800

CONTENTS

LIST OF FIGURES

LIST OF TEXT TABLES

SYMBOLS

Data not available	---
Category not applicable	. . .
Quantity zero	-
Quantity more than 0 but less than 0.05	0.0
Figure does not meet standards of reliability or precision	*

DEVELOPMENT OF THE SPEECH RECEPTION TEST

Earleen Elkins, Ph.D., and G. Donald Causey, Ph.D., University of Maryland, and Jean Roberts, *Division of Health Examination Statistics, NCHS*

INTRODUCTION

The hearing component of the Health and Nutrition Examination Survey in 1974-75 was designed to obtain data on the discrimination ability for speech sounds and the relationship of this ability to pure-tone air-conduction hearing thresholds among adults. These data, which will have been obtained for the first time with standardized methods on a national probability sample of civilian noninstitutionalized adults 25-74 years of age in the United States, will make it possible to assess objectively the functional implications of hearing impairment in the adult population. Prior to the start of the Health and Nutrition Examination Survey there was general agreement that one or more measures employing speech stimuli are necessary to quantify an individual's ability to understand suprathreshold speech.[1] Recent research has shown the extent to which specific types of hearing loss affect the intelligibility of speech and that amplification (such as with hearing aids) does not restore full discrimination when there is sensorineural loss.[2]

This new speech reception test provides an objective instrument, not previously available, for the determination within a short period of time of the ability of adults to hear and understand everyday conversational speech. This developmental study further provides some indication of the limitation and precision of this type of survey measurement in comparison with longer clinical tests of speech discrimination.

BACKGROUND

The *ad hoc* advisory committee for the hearing component of the examination had recommended for use in speech discrimination testing the Revised Central Institute for the Deaf (RCID) Sentences, that had been developed by a working group of the Committee on Hearing and Bioacoustics (CHABA) of the National Research Council. The RCID Sentences, while not yet on a suitable recording, were recommended as providing a better indication of functional hearing impairment within the time limits available for this component of the examination than was possible with the test material commercially available.

Recorded materials for speech discrimination testing, then commercially available, consisted of lists of 50 monosyllabic words. Those developed at the Harvard Psychoacoustic Laboratory, known as PB-50 lists,[3] were compiled with strict adherence to phonetic composition. Later, Hirsh and his associates at the Central Institute for the Deaf[4] recorded four 50-word lists compiled with primary emphasis on a restricted range of word familiarity. These recordings are known as the W-22 lists. A considerable number of research projects have employed these two recordings and speech discrimination ability has been clinically assessed with them. No test or set of tests has been generally recognized as acceptable to replace the W-22 and PB-50 commercial recordings which are more than 20 years old. Rerecording and adaptation, including shortening of these tests, would also have been necessary for use in the national survey.

Dr. Hallowell Davis, of the Central Institute for the Deaf, made the lists of RCID Sentences and developmental materials from CHABA available to the Division of Health Examination Statistics, to have adapted for their use.

Under contract with the Division of Health Examination Statistics, Dr. G. Donald Causey and Dr. Earleen Elkins, University of Maryland, recorded the RCID Sentences and developed the speech discrimination test used in the Health and Nutrition Examination Survey during 1974-75. Their methods, findings, and the resultant test based on the RCID Sentences, are described in this report.

DEVELOPMENT OF TEST MATERIALS

Description of Stimuli

The RCID Sentences supplied by Davis consisted of 10 lists of 10 sentences each. The following criteria were employed by the working group of CHABA in the development of these lists:

Vocabulary appropriate to adults.

Words that appear with high frequency in one or more of the well-known word counts of the English language.

Exclusion of proper names and proper nouns.

Free use of common nonslang idioms and constructions.

Avoidance of phonetic loading and tongue twisting.

High redundancy.

Low level of abstraction.

Grammatical structure that varies freely.[5]

Each list of 10 sentences contains 50 key words so that a discrimination score based on 2 percentage points per key word may be obtained. These sentence lists were an attempt to provide speech discrimination stimuli within a construct more closely associated with everyday speech.

Talker

A 24-year-old male, determined by a panel of hearing and speech scientists to have normal speech in relation to fundamental frequency, articulation, general American dialect, rate, and prosodic features, was selected as the talker. He was unsophisticated with regard to monitored live-voice technique for audiometric testing, but was extensively trained in the technique prior to the final recording session. This procedure involved the monitoring of his vocal output on the carrier phrase of the item and allowing the stimuli to occur naturally within the sentence, i.e., the level within each sentence varied as the speech power of its component sounds varied. The carrier phrase provided a reference level around which the elements of the stimulus occurred in their natural speech power relationships. The carrier phrase also served to prepare the listener for the oncoming test item during actual test presentation.

The voice monitoring was accomplished with the phrase "Number__." The talker maintained optimal needle deflection on the first syllable of "number" and on the first or only syllable of the digits 1 through 10.

All RCID Sentence stimuli were recorded during one session in an effort to minimize day-to-day vocal variability of the talker. When a stimulus sentence was judged unacceptable, it was repeated immediately or at the end of a list and subsequently spliced into correct order.

The timing of each item was regulated by an automatic device which triggered a light every 6.5 seconds to alert the talker. This permitted an average response time of 5 seconds which had been found adequate in a preliminary study of men from ages 38 to 78 years.

Recording Equipment

The RCID Sentence recordings were made in an anechoic chamber (Industrial Acoustics Company, Inc.–IAC) with inside dimensions of 7 feet by 7 feet by 7 feet. Four microphones (Bruel & Kjaer, Type 4131) and cathode followers (Bruel & Kjaer, Type 2619) were mounted in free air within the chamber. Microphone No. 1 was 24 inches from the talker's lips at a 30° angle to the right of center; microphone

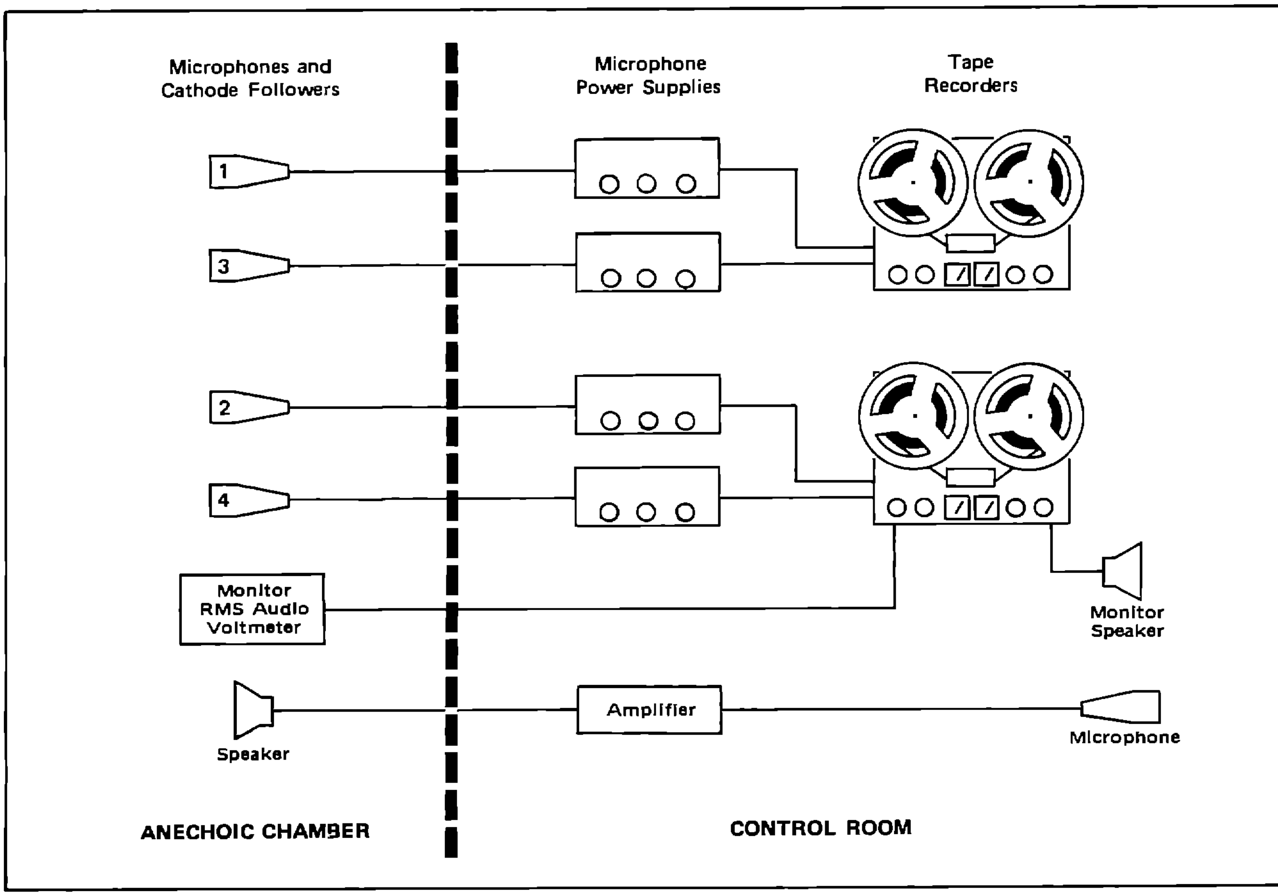

Figure 1-1. Schematic diagram of recording equipment and its placement.

No. 3 was 1 meter from the talker's lips at a 20° angle also on the right side; microphone No. 4 was placed as No. 1 but on the left side; and microphone No. 2 was placed as No. 3 but on the left. All microphones were at a 90° angle of incidence with respect to the talker's lips (figures 1-1 and 1-2).

Outside of the chamber, each microphone complement was electrically coupled to its respective power supply (Bruel & Kjaer, Type 2801) and fed to a separate amplifier and single-channel of two magnetic ½-inch tape recorders (Scully, Type 280). The outside channels of a triple-head tape assembly were employed to minimize the possibility of cross-talk (see figure 1-1).

The talker visually monitored his vocal output on a rms (root-mean-square) audio-voltmeter (Bruel & Kjaer, Type 2410). A talkback system permitted communication between the chamber

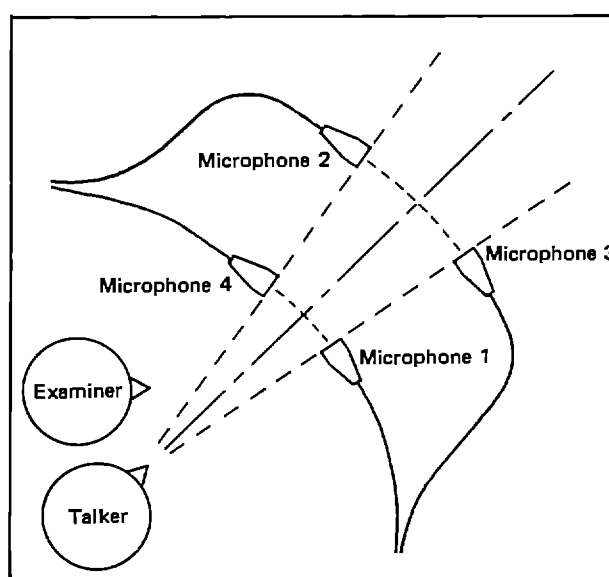

Figure 1-2. Diagram of talker, examiner, and microphones in recording chamber.

and recording technicians. An investigator was present in the chamber with the talker to aid in monitoring and in timing.

The record gain of each channel was adjusted to a -7 dB (decibel) re zero VU (volume units) level to avoid saturation during peak vocal output. All recording was done on low-noise tape (Scotch Brand, Type 208) at 7½ inches per second (ips). The signal-to-noise ratio exceeded 50 dB.

Tape Assembly Procedures

A group of hearing and speech scientists judged the recording on channel 3 as "sounding the most natural." All further procedures were accomplished with this recording. Half-inch tape copies were made of the master recording. This first generation tape was then spliced for the appropriate stimuli, and leader tape was inserted between items to ensure equal intervals of response time.

Tracings of the stimuli of each list were made on a graphic level recorder (General Radio, Type 1521-B) at 20 ips writing speed in order that the average level of the peaks could be calculated. A 1000-Hertz (Hz) tone generated by an oscillator (Hewlett-Packard, Model 200 AB) was placed at the beginning of each list at the average peak level to be used as a calibration tone in administration of the test lists. A second generation ¼-inch tape (Scotch Brand, Type 208) was then made for experimental purposes.

Verification of Stimuli

To determine that each stimulus was the intended one, the 10 lists were presented sound field to 16 normal hearing college students in a semi-sound-treated classroom. The ambient noise level was 60 dB Sound Pressure Level (SPL) as measured on the C-scale of a sound-level meter (Bruel & Kjaer, Type 2205) 15 feet in front of the speaker. The level of the 1000 Hz calibration tone at the beginning of each list was adjusted to obtain a 65 dB SPL reading on the same meter. All lists were presented at this level for the initial session.

Listeners were seated in front of the speaker at a distance of 5 to 25 feet. They were instructed to write the entire sentence on prepared answer sheets. Ample time was provided after each sentence for the write-down task. Two days later, the same 16 listeners performed the same task again, sitting in the same position relative to the speaker as they had for the initial session. Two changes in procedure were made. First, the level of the 1000 Hz calibration tone was adjusted to 60 dB SPL on the sound-level meter, and second, the order of list presentation was changed. Each session took 1 hour and 10 minutes. Following the second session, the listeners were asked to write their reactions to both the task and stimulus material.

The results from both sessions were marked for the number of correct key word responses. None of the listeners had more than one error per list, nor were any of the errors consistent among listeners. It was concluded that the lists all contained the correct stimuli and were not different from each other when perceived by normal-hearing listeners at a comfortable listening level.

The reactions of the listeners were directed toward the repetitive and boring nature of the task and fatigue from writing responses. Comments about the stimulus material were few, and those listeners reported that the content did not affect them one way or another.

Quality of Recording

The quality of a taped copy of this master recording of the RCID Sentences was analyzed by Mr. Edward D. Burnett, Physicist, Sound Section, National Bureau of Standards.

His report shows the broad-band noise to be 47 dB below the reference tone (which is -9 dB re 0 VU) when the tapes are played with a full track head. This is 4 dB better than predicted for a fourth generation tape. The type 208 tape is well suited for this application. The reference tone indicates the peak levels as measured at the 20 inches/second writing speed of a graphic level recorder. The instantaneous peaks, as observed on an oscilloscope, are at least 15 dB higher. The speech waveform is highly asymmetrical as is common for male speech. The level of the signal on the tape is well chosen.

The reference tone does not show any appreciable second-order distortion which indicates that the bias waveform was symmetrical and no heads were magnetized. No indications of overload distortion were found. The lower frequency limit is 100 Hz which is a characteristic of the speaker's voice. Components up to 14,000 Hz were detected, which indicates that heads were in good alignment.

PILOT STUDY

Prior to the experimental study, 10 normal-hearing subjects with normal otological histories were utilized to determine the method and sensation levels necessary to obtain performance-intensity functions for the RCID Sentences. The criterion for normal hearing was a pure-tone threshold of 10 dB or less (ANSI, 1969) at 250, 500, 750, 1000, 1500, 2000, 3000, 4000, and 6000 Hz.

The experimental task was performed with the subject seated in the anechoic chamber described earlier. The test material, reproduced by a tape recorder (Scully, Type 280), was channeled through a speech audiometer (Grason-Stadler, Model 162) to one earphone of a set (Telephonics, TDH-39 in MX-41/AR cushions). The same earphone was used throughout the tests. The 1000 Hz calibration tones on the RCID tape were used to establish 0 dB sound pressure level (SPL) on the speech audiometer. Periodic calibration checks were made of the equipment throughout the period of data collection.

Prior to the test session, subjects were told by the examiner that they would hear lists of sentences at different signal levels and were to repeat all or any of the sentence. The give-back method of response was chosen as a result of the stimuli verification study and the time factor. Subjects also received the instructions, pre-recorded on tape, which consisted of:

> You will hear ten sentences, each preceded by a number. Please listen carefully and repeat only the sentence.

A systematized randomization of the lists was employed so that each list was presented an equal number of times at the same SPL across subjects. The first five subjects in the pilot study received two sentences at each level beginning at 34 dB SPL and decreasing in 2 dB steps. All subjects repeated the first two sentences correctly. The descending method of presentation proceeded in 2 dB steps to 16 dB SPL. At this level all subjects were unable to respond to any portion of the sentences. The remaining eight lists were then presented at sound pressure levels of 18, 20, 22, 24, 26, 28, 30, and 32 dB in an ascending fashion. The examiner recorded all responses, including substitutions, on an answer sheet containing the appropriate list stimuli. All five subjects obtained 0 percent correct at 18 dB SPL and 100 percent correct at 32 dB SPL.

The next five subjects were trained by the same descending method with two lists, but received the following eight lists in a descending mode starting at 32 dB SPL and ending at 18 dB SPL. The percentages of correct responses were similar to those obtained with the ascending method; however, these subjects reported a general frustration with their inability to make correct responses as the signal was attenuated. During the performance of this task, four of the five subjects verbally responded to each stimulus that they could not identify completely with a phrase similar to "I can't make it out."

Based on the results and subjective evaluations of the task, two lists were used for training the subject and the remaining eight lists for obtaining articulation-gain functions by an ascending method.

STUDY AMONG NORMAL-HEARING SUBJECTS

Subjects and Method

One hundred college students—7 males and 93 females between the ages of 18 and 25 years—were selected for participation in this experiment. Each was questioned about his otological history and screened audiometrically as in the pilot study. Only those with normal otological history and audiometric test results were selected for this part of the study. All testing was done monaurally, and an equal number of students were tested in the right and in the left ear. Verbal instructions were given to each subject as in the pilot study, and they also received the recorded instructions (at 50 dB SPL) prior to each list. A systematized presentation of the lists was employed so that each list was presented an equal number of times as training lists and test lists at the same signal level across subjects. Using the equipment described earlier, each subject received the stimuli of the first two lists to accustom them to the task. Four sentences were presented at 34 dB SPL and two sentences each at 32, 30, 28, 26, 24, 22, 20, and 18 dB SPL. Each of the remaining eight lists were presented in full at each presentation level, beginning with 18 dB SPL and increasing in 2 dB increments to 32 dB SPL. All responses were made verbally and recorded by the examiner in the manner described earlier. The percentage of correct key words was calculated for each list. The test session took 35 minutes.

Findings

Table 1 summarizes the data obtained with 100 normal-hearing subjects combined across lists. Each list was heard by 10 subjects at each presentation level. The performance-intensity function derived from these data is shown in figure 2. This curve has the characteristic configuration of traditional performance-intensity functions derived from monosyllabic stimuli in that it exhibits a linear function which reaches a plateau at higher sensation levels. The lower segment of the curve is linear and has a slope of 10-percent increase in correct responses per decibel increase in presentation level. This linear segment terminates at about 24 dB SPL, where discrimination scores are about 76 percent. The upper portion of the function progresses in a curvilinear manner until it approaches asymptosis, and almost perfect discrimination of 99 percent is achieved at 32 dB SPL.

Table 1. Mean, standard deviation, and standard error of the mean obtained with 10 lists of RCID Sentences from 100 normal-hearing subjects at 8 presentation levels

Presentation level (dB SPL)	Mean	Standard deviation	Standard error of mean
18 dB	16.66	17.46	1.74
20 dB	36.78	24.68	2.47
22 dB	58.12	23.27	2.33
24 dB	75.94	20.24	2.02
26 dB	87.24	14.85	1.49
28 dB	94.62	7.36	0.74
30 dB	97.90	3.90	0.39
32 dB	98.80	2.07	0.21

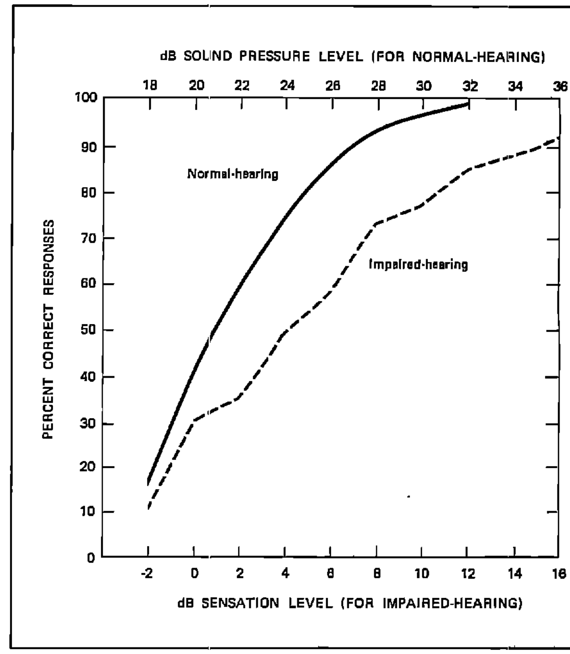

Figure 2. Performance-intensity functions derived from mean discrimination scores on the RCID Sentence Lists by 100 normal-hearing and 55 impaired-hearing subjects.

The measures of score variability, as indicated by the standard deviations in table 1, also compare favorably with the literature on monosyllabic discrimination. Excluding the variability among scores at 18 dB SPL, where the range of scores is truncated at zero, the variability was found to be large within the linear portions of the function. Variability decreased progressively as the signal intensity increased to the point where almost perfect responses were made. The variability present at the highest presentation level probably reflects the subjects' random errors due to physiological factors or lack of attention.

As an indication of subject consistency, Pearson correlation coefficients were calculated to determine the relationship between individuals' scores at the different levels of presentation. The coefficients shown in table 2 reveal that scores obtained under presentation levels adjacent to one another show higher correlations

Table 2. Pearson correlation coefficient from 100 normal-hearing subjects' scores on 10 lists of RCID Sentences at 8 presentation levels

Presentation level (dB SPL)	Presentation level (dB SPL)						
	20	22	24	26	28	30	32
18 dB	[1].629	[1].570	[1].451	[1].450	[1].331	.157	.181
20 dB		[1].613	[1].611	[1].451	[1].387	[1].202	.141
22 dB			[1].656	[1].520	[1].456	.098	[1].212
24 dB				[1].610	[1].568	[1].266	.137
26 dB					[1].656	[1].301	.096
28 dB						[1].521	[1].185
30 dB							.235

[1]Significant difference from zero at .05 level.

than scores obtained under levels 4 to 12 dB apart. This trend would be expected in view of the influence of signal intensity differences. Each column or row follows a logical decrease in the size of the correlation coefficient as the signal intensity difference increases. The general progression of decreasing coefficients as the presentation levels increase probably reflects the truncating effects when scores beyond 100 percent cannot be obtained. These data indicate that subjects maintained systematic relationships among themselves at different sound pressure levels.

In an effort to evaluate interlist equivalency, the scores of 10 subjects for each of the 10 lists A through J at each presentation level were drawn from the data described above. The results are shown in table 3. It is apparent from these data that at the lower levels, particularly those identified in the linear portion of the curve for 100 subjects (through 24 dB SPL), the mean and standard deviation values are quite different. Generally, there is a range of 30 percent from the lowest to the highest mean at these levels. Not until the scores reach 90 percent or better, do the mean and standard deviation values appear similar. Tests for significant differences among means revealed that the lists were different through 28 dB SPL. The means of the 10 lists were not significantly different (.05 level of confidence) at 30 and 32 dB SPL.

STUDY AMONG SENSORINEURAL-HEARING-IMPAIRED SUBJECTS

Subjects

Fifty-five male military veterans served as subjects for this part of the study. The age range was 22 to 63 years with a mean age of 45 years. It was the intent of this part of the study to gather data on all patients seen at the Audiology and Speech Pathology Section of the Veterans Administration Hospital, Washington, D.C. who showed any degree of sensorineural hearing impairment. Consequently, no rigid criteria for subject selection were required. Speech reception thresholds (SRT) could be as low as 2 dB above hearing threshold level (HTL) and pure-tone air conduction thresholds could be within normal range with a drop of 25 dB HTL at only one test frequency. The group mean SRT for the test ears was 17 dB. Generally, the audiometric configurations of the test ears were characterized by normal hearing in the lower frequencies with precipitous losses at various octaves above 1000 Hz. Thirty-one percent of the test population had a flat or gradually sloping configuration. The remaining subjects showed a drop of at least 25 dB starting at the following frequencies: 16 percent at 1500 Hz, 13 percent at 2000 Hz, 35 percent at 3000 Hz, and 5 percent at 4000 Hz.

Table 3. Mean, standard deviation, and standard error of the mean obtained on each of 10 lists of RCID Sentences from 10 normal-hearing subjects at 8 presentation levels

Presentation level (dB SPL)	List of RCID Sentences									
	A	B	C	D	E	F	G	H	I	J
18 dB										
Mean	5.20	17.80	26.00	27.80	8.00	12.00	27.60	5.20	19.20	17.80
Standard deviation	6.55	18.17	21.46	21.18	8.59	15.06	20.39	7.61	16.23	16.04
Standard error	2.07	5.75	6.79	6.70	2.72	4.77	6.45	2.41	5.14	5.08
20 dB										
Mean	39.40	44.60	37.20	48.40	37.20	18.00	44.20	25.60	22.80	46.60
Standard deviation	27.57	27.78	31.00	23.64	22.37	12.58	30.13	19.97	15.67	20.16
Standard error	8.73	8.79	9.81	7.48	7.08	3.98	9.54	6.32	4.96	6.38
22 dB										
Mean	64.80	72.60	67.80	55.00	41.60	55.20	64.40	43.00	67.00	49.80
Standard deviation	22.37	23.04	22.91	25.94	20.08	24.93	16.97	21.61	23.84	14.37
Standard error	7.08	7.29	7.25	8.21	6.35	7.89	5.37	6.84	7.59	4.55
24 dB										
Mean	75.60	90.20	89.80	76.40	62.60	59.40	88.00	60.00	77.60	79.00
Standard deviation	13.98	10.81	8.87	18.52	24.33	24.73	9.57	23.87	17.33	13.57
Standard error	4.42	3.42	2.81	5.86	7.70	7.83	3.03	7.55	5.48	4.29
26 dB										
Mean	93.40	93.00	96.00	94.20	71.60	77.20	88.60	85.80	87.80	84.80
Standard deviation	6.26	8.18	4.52	3.05	15.85	24.66	8.17	21.15	11.72	12.83
Standard error	1.98	2.59	1.43	0.97	5.02	7.80	2.59	6.69	3.71	4.06
28 dB										
Mean	95.40	99.00	98.80	97.20	94.00	90.00	93.20	90.80	96.20	92.20
Standard deviation	5.50	2.16	1.69	3.55	6.93	9.61	8.34	6.27	9.26	10.52
Standard error	1.74	0.68	0.54	1.12	2.19	3.04	2.64	1.98	2.93	3.33
30 dB										
Mean	95.40	99.80	98.40	98.20	98.40	98.00	97.20	97.20	97.80	98.60
Standard deviation	9.05	0.63	2.70	1.75	2.46	2.98	2.86	3.68	3.71	3.78
Standard error	2.86	0.20	0.85	0.55	0.78	0.94	0.91	1.17	1.17	1.20
32 dB										
Mean	99.60	99.00	99.40	98.80	97.40	98.00	99.40	98.60	99.60	98.20
Standard deviation	0.84	2.16	0.97	2.15	3.78	2.67	0.97	1.65	0.84	2.20
Standard error	0.27	0.68	0.31	0.68	1.20	0.84	0.31	0.52	0.27	0.70

Following an otological examination, bilateral pure-tone thresholds by air and bone conduction were obtained for each subject. Only those who had an air-bone gap of 10 dB or less were retained for further testing. This consisted of a determination of their SRT's and discrimination scores on recorded W-22 lists.[4] Performance of the 55 male veterans selected as subjects for this part of the study on air and bone conduction hearing tests and on the W-22 lists are summarized by age in the appendix. Among this study group, the degree of negative association between their scores on the W-22 word lists and their air-conduction hearing level at 1000 Hz

(-0.59) is similar to that with their speech reception threshold (-0.61), i.e., those with lower or better hearing levels attain higher W-22 scores or hear more words correctly.

Method

The ear with the better SRT was then chosen for the monaural RCID Sentence testing. If the SRT's for both ears were the same, the ear with which the subject had obtained the highest W-22 speech discrimination score was selected.

The task was performed with the subject seated in one room of an IAC testing suite. The test material, reproduced by a tape recorder (Scully, Type 280), was channeled through a speech audiometer (Grason-Stadler, Model 162) to one earphone of a set of phones (Telephonics, TDH 49-10Z in MX-41/AR cushions). The same earphone was used throughout. The 1000 Hz calibration tones on the RCID tape were used to establish O VU output level on the VU meter of the speech audiometer. Periodic calibration checks were made of the equipment during the period of data collection.

Prior to the test session, subjects were told by the examiner that they would hear lists of sentences at different signal levels and were to repeat all or any of the sentences. Subjects also received the prerecorded instructions on the test tape. A systematized presentation of the lists was employed so that each list was presented a given number of times as training lists and test lists at the same sensation level across subjects. The reference level for presentation was the SRT of the subject's test ear. For the training portion, two sentences were presented at each of the following sensation levels (SL's): 16, 14, 12, 10, 8, 6, 4, 2, 0, -2 dB. The remaining eight lists were presented in full, beginning with 0 or +2 dB SL and increasing in 2 dB increments to 14 or 16 dB SL. After data on 20 subjects had been obtained, some variation in the initial sensation level was instigated in order to explore the performance intensity function below 0 dB SL and above 16 dB SL. All responses were made verbally and recorded by the examiner in the manner described in the study among the normal-hearing group. The percentage of correct keywords was calculated for each list. The test session took 35 minutes.

Findings

Table 4 summarizes the data obtained from 55 subjects with sensorineural losses. The number of subjects for which the data at any sensation level were obtained is indicated in the last column. The performance-intensity function derived from these data is shown in figure 2 with the performance-intensity function obtained from 100 normal-hearing subjects. The curve for sensorineural subjects exhibits the linear function which appears to terminate at about 8 dB SL, where subjects gave 74 percent correct responses. The slope of this portion of the curve rises at a rate of about 5 percent increase per decibel increase in intensity. The upper portion of the function tends toward curvilinearity and does not approach asymptosis.

The variability of the discrimination scores about the mean values, as shown by the standard deviations in table 4, indicates the large amount of heterogeneity among the impaired-hearing subjects. It is also important to observe that this large variability is maintained from the lower sensation levels up to 14 dB SL. A gradual decrease in subject variability is then observed but a pattern is not apparent. Standard errors of the mean indicate that in other populations with similar sensorineural losses, one would expect about two-thirds of the mean discrimination scores to be plus or minus two key words.

Table 4. Mean, standard deviation, and standard error of the mean obtained with 10 lists of RCID Sentences at 12 sensation levels for the 55 subjects with sensorineural hearing impairments, and the number contributing to each mean.

Sensation level	Mean	Standard deviation	Standard error of the mean	Number of subjects
-4 dB	14.33	24.21	9.88	6
-2 dB	10.89	10.30	3.43	9
0 dB	30.56	28.59	5.72	25
2 dB	35.27	28.10	5.13	30
4 dB	49.26	28.58	4.83	35
6 dB	58.65	32.13	5.51	34
8 dB	73.62	24.20	3.98	37
10 dB	77.82	23.70	4.13	33
12 dB	85.47	15.09	2.76	30
14 dB	86.77	20.15	3.95	26
16 dB	93.53	8.62	2.09	17
18 dB	95.50	4.50	1.59	8

INTERLIST EQUIVALENCY STUDY

In order to control the effect of variable levels of stimuli presentation, another study was conducted to evaluate interlist equivalency—assessing the same subject under different experimental conditions. There would be need in the Health Examination Survey for sentence lists that could be used interchangeably, since different lists of 10 sentences must be available for use at each successive 10 dB intensity level at which the test needed to be presented. The presentation was to be continued until the examinee was no longer able to hear enough of the speech test to pass.

Subjects

Subjects for this experiment were twenty male military veterans with normal or impaired hearing. Their ages ranged from 23 to 69 years, with a mean age of 45.7 years. The subjects were divided into two groups of 10 each. Group 1 had speech reception thresholds (SRT's) in the test ear ranging from 2 to 24 dB with a mean of 9.4 dB. Their lower pure tone averages (PTA's), usually two frequencies (500 and 1000 Hz), ranged from 2 to 30 dB with a mean of 10.2 dB. Group 2 had SRT's from 0 to 34 dB and a mean of 15.2 dB. The lower PTA for these subjects went from 0 to 37 dB with a mean of 16.4 dB.

No restrictions with respect to degree, type, or presence of hearing impairment were placed on subject selection. Five of the subjects could be classified as normal, two had mixed impairments, and the remainder had sensorineural impairments. All but three of the subjects had W-22 discrimination scores of 90 percent or better in the test ear.

Following an otological examination, bilateral pure-tone thresholds by air and bone conduction were obtained for each subject as well as speech reception thresholds and discrimination scores on recorded W-22 lists. The test ear was chosen on the basis of the lower PTA of the better ear for monaural testing with the RCID sentence lists.

Method

The experimental task was performed with the subject seated in one room of an IAC testing suite. The test material, reproduced by a tape recorder (Scully, type 280), was channeled through an audiometer (Allison, Model 22) to one earphone of a set (Telephonics, TDH 39-10Z in MX-41/AR cushions). The same earphone was used throughout the test. The 1000 HZ calibration tones on the RCID tape were used to establish 0 on the VU meter of the audiometer.

Prior to the test session, subjects were told by the examiner that they would hear lists of sentences at a comfortable loudness level and were to repeat all or any of the sentences. Subjects also received the prerecorded instructions on the test tape. A systematized presentation of the lists was employed so that subjects in Group 1 or 2 each heard a different order of the sentence lists. The reference level for presentation was the lower PTA of the subjects' better ear. Group 1 heard all 10 lists at 10 dB SL and Group 2 heard all 10 lists at 20 dB SL. No training lists were administered because all stimuli were presented at a comfortable loudness level and any learning factor would be counterbalanced by the randomization of the lists. All responses were made verbally and recorded by the examiner in the manner described earlier. The percentage of correct keywords was then calculated for each list. The test session took about 35 minutes.

Findings

Table 5 describes the performance of both groups of subjects separately and combined on the 10 RCID sentence lists. The mean correct responses for Group 1 ranged from 88.8 percent for list E to 98.2 percent for list I. The scores of Group 1 were statistically evaluated by analysis of variance. Statistical significance was shown for the main effect of lists ($F=2.67$; $df=9,90$; $p<0.05$). Subsequent comparisons by Duncan's Multiple Range of the mean scores for 10 lists showed that list E was significantly ($p<0.05$) lower than all of the remaining lists with the exception of lists F and H. There were no significant differences among the means of the nine lists, except for list E.

The scores of Group 2 were also evaluated by analysis of variance and no statistical significance was shown (F=0.94; df=9,90; p>0.05). The scores from the two groups were similar at the 10 and 20 dB SL's, and the F-test for homogeneity of variance (Hartley's) was performed on the data to determine the feasibility of combining the data into one group with an N of 20. The results of the F-test indicated that the variance due to experimental error within each of the treatment populations was homogeneous (F=2.22; df=2,9; p>0.05).

Finally, an analysis of variance performed on Groups 1 and 2 combined indicated no significant differences among the scores on the 10 sentence lists (F=1.66; df=9,190; p>0.05). The largest mean difference for Groups 1 and 2 combined is between Lists A and E and constitutes only 4 percent or two keywords.

Table 5. Mean, standard deviation, and standard error of the mean obtained on RCID Sentence lists by 10 subjects at 10 dB SL and 10 subjects at 20 dB SL, and by both groups combined.

RCID Sentences	Group 1 at 10 dB SL	Group 2 at 20 dB SL	Groups 1 and 2 combined
List A			
Mean	97.6	98.6	98.1
Standard deviation	1.744	2.375	2.198
Standard error	0.552	0.752	0.492
List B			
Mean	97.2	98.8	98.0
Standard deviation	2.857	2.040	2.675
Standard error	0.904	0.645	0.599
List C			
Mean	96.2	98.2	97.2
Standard deviaton	3.842	2.272	3.397
Standard error	1.216	0.719	0.760
List D			
Mean	96.0	97.6	96.8
Standard deviation	2.683	3.555	3.334
Standard error	0.849	1.125	0.746
List E			
Mean	88.8	99.4	94.1
Standard deviation	9.474	1.281	8.813
Standard error	2.998	0.405	1.972
List F			
Mean	93.0	96.4	94.7
Standard deviation	6.768	5.499	6.562
Standard error	2.142	1.740	1.468
List G			
Mean	96.2	97.8	97.0
Standard deviation	5.016	2.600	4.180
Standard error	1.587	0.823	0.935
List H			
Mean	93.6	96.6	95.1
Standard deviation	7.526	5.142	6.789
Standard error	2.382	1.627	1.519
List I			
Mean	98.2	97.6	97.9
Standard deviation	2.088	2.154	2.142
Standard error	0.661	0.682	0.479
List J			
Mean	97.2	95.8	96.5
Standard deviation	4.490	5.173	5.021
Standard error	1.421	1.637	1.123

DISCUSSION

Normal Hearing

Several factors probably contribute to the large differences among means and the wide variability among scores at levels from 18 dB to 28 dB SPL. One consideration would be the lack of homogeneity among sentence items due to the linguistic constraints imposed by possible word sequences.[6] Quantification of sentence stimuli for the measurement of discrimination ability has not received much investigation, and some of those who have worked in this area feel that sentence stimuli are "very complicated" and require further study.[1]

Another factor which undoubtedly contributed to the observed differences among lists is related to the experimental method. It has been noted that all subjects heard the lists at specified levels re 0.0002 microbar, and though they had been screened for normal-hearing, it is generally agreed among audiologists that there is marked variability in sensitivity among persons meeting the normal-hearing criteria. The level which may have been the threshold of detectability or intelligibility for one subject could easily have been displaced plus or minus several decibels for another subject. This consideration is simply another way of viewing the marked variability among subjects. However, as the data across all

10 lists for the entire group of 100 subjects indicates, measures of a group of tests are more stable than measures for a single test.

The performance-intensity function for normal-hearing subjects is characteristic of functions obtained with monosyllabic stimuli in that it contains a linear portion which reaches a plateau at higher signal intensities. The linear portion has a slope of 10 percent increase in score per decibel increase in presentation level. This linear segment terminates at a level of about 24 dB SPL, where discrimination scores approximate 76 percent. Asymptosis occurs at a level of 30 dB SPL.

For normal hearing subjects, these findings would indicate that all 10 lists of the RCID Sentences may be used interchangeably for obtaining maximum discrimination scores from normal-hearing persons, provided the signal level under earphone presentation is at least 30 dB SPL.

Hearing Impaired

Comparison of findings among the normal-hearing and impaired-hearing populations show marked differences. First, the slope of the linear portion of the curve for the impaired-hearing subjects was more gradual—approximately 5 percent per decibel compared to 10 percent per decibel for the normal-hearing subjects. The linear portion for both groups of subjects terminated where approximately 76 percent correct responses were obtained. But this occurred at a sensation level of 4 dB for normals and 8 dB for the impaired listener.

The second difference between the two groups of subjects may be observed where the nonlinear portion of the performance-intensity function approached saturation. The normal group attained this level at 32 dB SPL and the impaired group did not achieve a plateau at the sensation levels employed for this study.

The third difference between the normal and impaired-hearing subjects is reflected in the subject variability within each group. As indicated by the standard deviations, the normal-hearing group had less variability throughout the sensation levels examined than the impaired-hearing group. This would indicate that the impaired group was more heterogeneous in the speech discrimination task than was the normal-hearing group. Also, the variability within the normal-hearing group decreased progressively as the signal intensity increased to the point where almost perfect responses were made, but the same type of pattern of intersubject variability was not maintained by the sensorineural group.

Interlist Equivalency

The findings from this study indicate that the University of Maryland recordings of the 10 RCID sentence lists could be used interchangeably for the Health and Nutrition Examination Survey as a measure of the extent of functional hearing impairment for normal conversational levels of speech in the adult population. The discrimination scores obtained from subjects in Group 1 at 10 dB above the PTA of the better ear and scores obtained from subjects in Group 2 at 20 dB above the PTA of the better ear were statistically evaluated. As a result of score similarity between the groups of subjects and a finding of homogeneous variances, the data from Groups 1 and 2 were combined. No significant differences were found among any combination of the 10 means. The standard errors of the means indicate that with other populations, one would expect about two-thirds of the discrimination scores to be within plus or minus one or two keywords of the means obtained at 10 or 20 dB sensation levels. Any real differences among the lists would be apparent only if the sensation levels of presentation were less than 10 dB and representative of the linear portion of a subject's performance-intensity function for these materials.

SUMMARY

These developmental studies demonstrate that the recordings of the RCID sentence lists can readily be used as a discrimination test to separate listeners with normal hearing from those with sensorineural losses. This new test, designed for use in the Health and Nutrition Examination Survey of 1974-75, is capable of providing a rapid assessment of hearing discrimination ability for speech sounds.

REFERENCES

[1]Second Conference on Speech Discrimination, sponsored by the Office of Vocational Rehabilitation. Houston, Texas, 1962.

[2]Personal communication from E. L. R. Corliss on Some Quantitative Consequences of Hearing Loss from the Standpoint of Communication Theory, 1975.

[3]Egan, J. P.: Articulation Testing Methods. *Laryngoscope.* 58:955-91, 1948.

[4]Hirsh, I. J., Davis, H., Silverman, S. R., Reynolds, E. G., Eldert, E., and Benson, R.W.: Development of Materials for Speech Audiometry. *J. Speech Hear Disord.* 17:321-337, 1952.

[5]Davis, H., and Silverman, S. R.: *Hearing and Deafness,* 3d ed. New York. Holt, Rinehart and Winston, Inc., 1970.

[6]Speaks, C., Jerger, J., and Jerger, S.: Performance-intensity characteristics of Synthetic Sentences. *J. Speech Hear Res.* 9:305-12, 1966.

APPENDIX

CONTENTS

APPENDIX TABLE

APPENDIX

HEARING CHARACTERISTICS OF THE SENSORINEURAL HEARING-IMPAIRED SUBJECTS IN THE STUDY

The extent of hearing impairment for pure tone sound in air- and bone-conduction, the sensation levels for speech reception, and the speech discrimination level as determined with the previously available W-22 word lists for the 55 sensorineural-hearing-impaired subjects in this developmental study and the degree of association between the W-22 discrimination scores and the pure tone air-conduction hearing levels for these 55 subjects are shown in Table I. These data are included to give a general idea of the extent of hearing impairment for these subjects, although as previously stated no rigid criteria for their selection were used since all were patients in the Audiology and Speech Section of the Veterans Administration Hospital in Washington, D.C.

Table I. Mean, standard deviation, and standard error of the mean on air- and bone-conduction hearing tests and the W-22 word lists and the correlation between air-conduction levels and W-22 scores for the 55 subjects with sensorineural hearing impairments.

Age	Number of subjects	Air-conduction hearing levels (dB)				Bone-conduction hearing levels (dB)				Speech test	
		500 Hertz	1000 Hertz	2000 Hertz	4000 Hertz	500 Hertz	1000 Hertz	2000 Hertz	4000 Hertz	W-22 scores	SL re SRT
Number of subjects	55	55	55	55	55	15	15	21	26	55	55
					Mean values						
All ages 20-64 years..........	55	17.8	19.5	31.3	53.5	29.7	29.7	36.7	53.3	91.8	40.9
20-29 years..................	8	11.7	12.5	24.4	51.9	-	-	-	42.5	90.5	41.2
30-39 years..................	7	12.9	13.6	25.0	57.1	15.0	12.5	36.2	57.5	94.6	40.9
40-49 years..................	12	19.2	17.9	25.0	52.9	41.7	38.3	32.5	45.0	94.8	39.5
50-54 years..................	20	22.5	26.2	40.2	+63.5	+32.9	+35.0	+42.2	+55.6	89.0	42.3
55-64 years..................	8	14.4	17.5	30.6	65.6	20.0	20.0	28.8	+63.8	91.8	39.5
					Standard deviation						
All ages 20-64 years..........	...	14.03	17.29	21.33	33.92	17.60	20.07	16.30	12.41	9.09	8.13
20-29 years..................	...	9.87	12.99	16.81	19.92	-	-	-	14.36	16.31	6.24
30-39 years..................	...	6.84	6.86	16.69	28.36	-	2.50	12.57	5.60	6.32	0.95
40-49 years..................	...	16.90	18.66	22.63	17.42	12.35	19.35	18.87	11.65	3.88	4.55
50-54 years..................	...	16.47	20.49	22.95	24.35	20.43	21.38	15.53	8.35	13.18	12.23
55-64 years..................	...	6.29	8.26	11.37	20.76	4.08	12.25	14.19	10.52	4.04	2.18
					Standard error of mean						
All ages 20-64 years..........	...	1.89	2.33	2.87	4.57	4.55	5.19	3.56	2.43	1.23	1.09
20-29 years..................	...	3.48	4.58	5.94	7.04	-	-	-	10.18	5.76	2.20
30-39 years..................	...	2.58	2.58	6.29	10.74	-	1.77	6.28	2.80	2.39	0.35
40-49 years..................	...	4.88	5.39	6.54	5.04	7.14	11.18	9.44	4.39	1.12	1.32
50-54 years..................	...	3.68	4.58	5.13	5.45	7.71	8.07	5.18	2.78	2.95	2.74
55-64 years..................	...	2.22	2.92	4.02	7.33	2.35	7.08	7.10	5.26	1.42	0.77

☆U.S. GOVERNMENT PRINTING OFFICE: 1977— 260-937:26

VITAL AND HEALTH STATISTICS PUBLICATIONS SERIES

Formerly Public Health Service Publication No. 1000

Series 1. *Programs and Collection Procedures.*—Reports which describe the general programs of the National Center for Health Statistics and its offices and divisions, data collection methods used, definitions, and other material necessary for understanding the data.

Series 2. *Data Evaluation and Methods Research.*—Studies of new statistical methodology including experimental tests of new survey methods, studies of vital statistics collection methods, new analytical techniques, objective evaluations of reliability of collected data, contributions to statistical theory.

Series 3. *Analytical Studies.*—Reports presenting analytical or interpretive studies based on vital and health statistics, carrying the analysis further than the expository types of reports in the other series.

Series 4. *Documents and Committee Reports.*—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 10. *Data from the Health Interview Survey.*—Statistics on illness; accidental injuries; disability; use of hospital, medical, dental, and other services; and other health-related topics, based on data collected in a continuing national household interview survey.

Series 11. *Data from the Health Examination Survey.*—Data from direct examination, testing, and measurement of national samples of the civilian, noninstitutionalized population provide the basis for two types of reports: (1) estimates of the medically defined prevalence of specific diseases in the United States and the distributions of the population with respect to physical, physiological, and psychological characteristics; and (2) analysis of relationships among the various measurements without reference to an explicit finite universe of persons.

Series 12. *Data from the Institutionalized Population Surveys.*—Discontinued effective 1975. Future reports from these surveys will be in Series 13.

Series 13. *Data on Health Resources Utilization.*—Statistics on the utilization of health manpower and facilities providing long-term care, ambulatory care, hospital care, and family planning services.

Series 14. *Data on Health Resources: Manpower and Facilities.*—Statistics on the numbers, geographic distribution, and characteristics of health resources including physicians, dentists, nurses, other health occupations, hospitals, nursing homes, and outpatient facilities.

Series 20. *Data on Mortality.*—Various statistics on mortality other than as included in regular annual or monthly reports. Special analyses by cause of death, age, and other demographic variables; geographic and time series analyses; and statistics on characteristics of deaths not available from the vital records, based on sample surveys of those records.

Series 21. *Data on Natality, Marriage, and Divorce.*—Various statistics on natality, marriage, and divorce other than as included in regular annual or monthly reports. Special analyses by demographic variables; geographic and time series analyses; studies of fertility; and statistics on characteristics of births not available from the vital records, based on sample surveys of those records.

Series 22. *Data from the National Mortality and Natality Surveys.*—Discontinued effective 1975. Future reports from these sample surveys based on vital records will be included in Series 20 and 21, respectively.

Series 23. *Data from the National Survey of Family Growth.*—Statistics on fertility, family formation and dissolution, family planning, and related maternal and infant health topics derived from a biennial survey of a nationwide probability sample of ever-married women 15-44 years of age.

For a list of titles of reports published in these series, write to:

Scientific and Technical Information Branch
National Center for Health Statistics
Public Health Service, HRA
Hyattsville, Md. 20782

DHEW Publication No. (HRA) 78-1345
Series 2-No. 71

NCHS
U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
Public Health Service
Health Resources Administration
National Center for Health Statistics
Center Building
3700 East-West Highway
Hyattsville, Maryland 20782

For publications in the *Vital and Health Statistics* Series call 301-443-NCHS.