PRINTED BY THE UNIVERSITY OF NORTH DAKOTA PRESS
GRAND FORKS, NORTH DAKOTA

PSYCHOLOGICAL REPORTS

Monograph Supplement 7 1957 price 1.00

# THE HEALTH OPINION SURVEY: TECHNIQUE FOR ESTIMATING PREVALENCE OF PSYCHONEUROTIC AND RELATED TYPES OF DISORDER IN COMMUNITIES

ALLISTER M. MACMILLAN

*Cornell University*

PSYCHOLOGICAL REPORTS

*Psychological Reports* is published quarterly, March, June, September, and December, by Southern Universities Press, Box 11, Grand Forks, North Dakota. The subscription rate is $7.00 per year, with single issues costing $2.50. The journal is cooperative and non-profit. Income beyond publication costs will be used for improvements.

The purpose of this journal is to encourage scientific originality and creativity, in the field of general psychology, for the person who is first a psychologist, then a specialist. It carries experimental, theoretical, and speculative articles; comments; special reviews; and a listing of new books and other material received. Controversial material of scientific merit is welcomed. An attempt is made to balance critical editing by specific suggestions as to changes to meet standards. Articles should be submitted in duplicate to one of the editors or to Southern Universities Press. In the former case, a copy of the covering letter should be sent to Southern Universities Press.

Publication is in order of receipt of proof from authors. Authors pay costs of cuts and special composition, and a page charge. There are three publication arrangements:

(1) *Regular articles.* Page charge is $10.00 per page in multiples of four pages. These are articles which print to from 2 through 12 pages. Authors receive 200 reprints without further charge. Additional reprints can be ordered at cost.

(2) *One-page notes and articles.* This arrangement is useful where the author does not intend to do further work in the area but feels that preliminary findings should be put on record, where it is expected that it will be several years before the final study is completed and reported, where the author submits a summary of a study accompanied by the full report for filing with the American Documentation Institute, or where a particular finding can be reported completely in one page. Page charge is $10.00. Authors receive 50 reprints without further charge. Additional reprints can be ordered at cost.

(3) *Monograph supplements.* Papers justifiably running to more than 12 printed pages will ordinarily be published as monograph supplements. These are distributed to regular subscribers of *Psychological Reports* as part of regular issues, and are also advertised and sold as separates. Page charge is $10.00 per page in multiples of four pages. Authors receive 200 copies of the monograph with covers without further charge. Additional copies can be ordered at cost.

Articles should be prepared carefully, following the form suggested for publications of the American Psychological Association (see *Psychol. Bull.*, 1952, 49, 389-449). Proof is mailed to the author within approximately three weeks of acceptance of his article, and reprints are mailed within five weeks of return of proof by the author.

It is the policy of this journal to file raw data with the American Documentation Institute wherever possible. Authors should submit appropriate tables with their articles. In this way, a permanent file of data can be built up.

*Psychological Reports*, 1957, 3, 325-339. © Southern Universities Press 1957
Monograph Supplement 7

# THE HEALTH OPINION SURVEY: TECHNIQUE FOR ESTIMATING PREVALENCE OF PSYCHONEUROTIC AND RELATED TYPES OF DISORDER IN COMMUNITIES[1]

ALLISTER M. MACMILLAN
*Cornell University*

This paper presents the results of an attempt to devise and standardize a psychological screening test for adults in rural communities. The need for such a test arose in connection with a larger project whose general aim was to study the relationship between social factors of various kinds and the development of psychiatric symptoms (17, 18). This called for some means of discovering what members of the population under investigation showed such symptoms, in short, a case-finding method. Since the test was to be used in a rural county of one of the Atlantic Provinces of Canada, it seemed probable that whatever had already been devised would have to be retailored somewhat to fit this region. The first step, then, was to examine existing tests. This included both a review of the literature and consultation with investigators who were active in the field of psychological screening.

Most such tests had been designed for the selection of persons for positions in academic, business, and military situations, and the elimination of those not qualified according to the various specifications adopted. Such selection and elimination processes, though ancient in application, have only recently been systematized and removed somewhat from the realm of intuition and "hunch." One of the earliest recorded efforts, perhaps, is Gideon's dramatic way of picking 300 warriors to overcome the Midianites (15). Until the beginning of the present century, empirical tests like Gideon's were the only means available for such assessments.

[1]This paper comes from the Stirling County Study which is being conducted by Cornell University in collaboration with the Department of Public Health of the Province of Nova Scotia and with the cooperation of Acadia and Dalhousie Universities. Invaluable help has also been provided by the Faculte des Sciences Sociales, Universite Laval. Within Cornell, the Stirling County Study is attached administratively to the Social Science Research Center and is sponsored by the Department of Sociology and Anthropology and the Department of Psychiatry of the New York Hospital and Cornell Medical College. Financial support is provided by the Carnegie Corporation of New York, the Department of National Health and Welfare of Canada, the Department of Public Health of the Province of Nova Scotia, and the Milbank Memorial Fund.

The present staff of the project consists of the following who are listed according to their functions in the study: Alexander H. Leighton, Director, and Allister M. Macmillan, Deputy Director; Christopher Haffner, Chief of the Psychiatric Clinic; Bruce Dohrenwend, Social Analyst; Bernard Hebert, Clinical Psychologist; Charles C. Hughes, Social Science Coordinator; Ruth O. Kent, Administrative Secretary; Dorothea C. Leighton, Assistant to the Director. Major contributors to the work reported here are too numerous to list in detail. They include Nova Scotia physicians and hospital administrators, community leaders, personnel of Acadia and Dalhousie Universities, statistical assistants, and cooperating respondents. Grateful acknowledgment is made to Oskar Diethelm and Robin M. Williams, Jr., for manuscript review.

In the past 50 years, however, much attention has been devoted to the problem of psychological screening, which includes assessments of individuals in terms of aptitudes, intelligence, personality, temperament, or mental health for the use of various organizations in business and industry, education, and, with World War II, the armed forces. Some use had been made of rough screening devices in World War I. In the interval between wars the extravagance in both money and man power of using men for jobs to which they were not suited became increasingly apparent, as did the high cost of accepting the unfit for service and then supporting them as war casualties forever after. Beginning with the early work of Cattell (2), Heymans and Weirsma (11), Davenport (6), Hoch and Amsden (12), and Wells (34) before World War I, a monumental scientific literature on selection for a wide variety of administrative purposes has grown up. Reviews of this literature, including the period through World War II, by Symonds (31), Snyder (29), Traxler (33), Darley and Anderson (5), Maller (21), and Meehl (24) generally gave faint praise to the questionnaire or inventory method of assessment.

A major difficulty with the earlier questionnaire method appears to have been that the scope was too large. That is, attempts were made to study "personality as a whole" without the benefit of an adequate theory of personality which could be formulated in terms of test items. Personality, with all its ramifications, presented too great a range and complexity of phenomena for a single test to deal with adequately. With the need for conservation of man power that arose with World War II came augmented interest in finding a way of determining which men would be effective in various branches of the armed services and which might better do their part in industry or other civilian occupations. It was of vital importance to develop tests that could be administered quickly, that could be both given and scored by persons with minimal training, and yet would be fairly reliable indicators of a limited set of qualities of the individuals tested. A main consideration in such screening was elimination of those individuals who were potentially or actually neurotic, since it was assumed that such persons were least suited to military service. In reviewing the fruits of these efforts for War II, Morton (25) concludes that the comparative success of the screening tests used by the armed forces depended on the fact that they did not aim to measure polarized traits, but were limited to indicating the extent to which an individual's health problems would be significant in military service. Thus, when the test constructor lowered his level of aspiration from "personality measurement" to a more limited aspect of behavior, the resulting tests were found to be useful. Other reviews of the matter by Ellis (7), Ellis and Conrad (8), Wexler (35), and Zubin (38) tend to support this conclusion.

The variety and range of tests and methods available at the beginning of the Stirling screening field work in 1951 made a field trial of the merits



of each, or even of several, impossible. These included group tests, individual tests, paper-and-pencil tests, personal interviews, inventory check-lists, card sorts, and projective tests. The possibilities are illustrated in papers by Cattell (3), Eysenck (9), Hunt (13, 14), McQuitty (22, 23), Saslow (28), and Wittenborn (36, 37), to mention a few. Reviews of the matter by Rosenzweig (27) and Thorndike (32) indicated that between 1948 and 1951 no noteworthy advance had been made in the field of screening tests as research instruments.

In spite of the assortment to choose from, no test was found which had been standardized on small town and rural adults—the major focus of study in the Stirling County research. Further, it appeared that the tests most closely related to our actual field possibilities were some developed and used in the armed forces. Since these had been administered only to young males, for the most part, and under special circumstances, it became evident that it would be necessary to adapt a test for young and old, applicable to both sexes, and standardize it on the type of population under investigation. This needed to be done outside the Stirling County research area but on a similar population. It seemed clear, also, that the aim should be to obtain a rough measure of neurotic and related symptoms among rural community adults. More specifically, the objective should be to detect those adults whose responses to questions about their health approximated the responses of psychiatric patients, and differed from the responses of controls drawn at random from the community. It did not seem feasible to attempt an elaborate analysis of personality or a precise psychiatric diagnostic appraisal (20).

Accordingly a two-county area with socio-cultural characteristics similar to those of Stirling County was selected as the locale for the standardization operation. Also, permission was obtained to administer the test to patients at the nearest psychiatric hospitals and out-patient clinic. These settings provided us with a community group to screen and a criterion group of psychiatrically disturbed individuals whose illness had been diagnosed.

### Field Procedures

In selecting and arranging the test and its mode of administration, certain known health attitudes and "sets" prevalent in the community population were taken into account. For example, the people were for the most part unfamiliar with social research and questionnaires; they were reticent in discussing personal matters with strangers; they were unaccustomed to doing "paper work." It thus seemed advisable that the test should contain no offensive items (e.g., "Do you believe in God?"), it should be administered confidentially, the interviewer should do the writing, and the respondent should not have to concentrate upon answering questions for more than about 20 min. Further, we gave our operation a name which we felt would

adequately and frankly express our aims in terms acceptable and meaningful to the group to be screened—the Health Opinion Survey (hereafter called the HOS).

The test was built up of items from a number of sources. Its core was 15 questions from the Army's Neuropsychiatric Screening Adjunct (30) with additional questions reported to be useful neurotic discriminators by Eysenck (9), Rimoldi[1], and others. Local general practitioners were consulted by the author in order to establish criteria as to the suitability of the various items. As finally constituted, the test schedule comprised a total of 75 health-oriented queries, the majority being of the psychophysiological-complaint type and the balance dealing with social relationships. In most cases the respondent was required to select one of three possible answers like the often-sometimes-never alternatives of the Army NSA. The more familiar and easily answered sociological questions, such as marital status, religion, and occupation came first followed by some very general questions about health, in order to reassure the respondent and provide some orientation before reaching the main section, the screening items.

While the details of the screening instrument were being worked out and interviewers recruited and trained, a public relations program was also carried on to familiarize the two-county area with our intentions and express our hope for participation. Interviewers, selected from students at Acadia and Dalhousie Universities, attended a three-week special-training session and took part in the final stages of preparing the survey as well as in the field work itself. Six of them interviewed in the two-county area, and two at the psychiatric hospitals. The hospitals afforded facilities for conducting the interviews in private. To insure similar privacy in the community interviewing, without the necessity of bringing respondents to a central place, house trailers were rented and the interiors arranged for the purpose. In the community situation, a slight *quid pro quo* as well as a reinforcement of the medical atmosphere was provided by measuring height and weight and by testing the respondent's eyesight with a telebinocular before completing the questionnaire.

County agricultural agents were consulted for their opinions concerning socio-economic grouping of the residents, and their advice led to the identification of subgroups differentiated as "Good Farm," "Average Farm," "Poor Farm," "Day Labor," and "Urban" (small town, actually). By random cluster sampling methods 559 white adults aged 20 to 59 years were selected from these subgroups and interviewed.

At the same time, 210 hospitalized and out-patient *S*s were interviewed in the clinical setting. The psychiatric diagnostic range included those with

[1]Rimoldi, H. V. A. *Behavior inventory*. Chicago: Univer. of Chicago, by personal communication.



neurosis, psychosis, and character disorder, as well as a number from the medical wards with non-psychiatric diagnoses who were interviewed in order to have some control on the effect of the hospital environment. The total population available, who were diagnosed as having neurosis during the three-month field period, amounted to 78 adults. These interviews were selected as our "ill" criterion group for comparison with the interviews from the community adults.

*Standardization—Discriminant Function Analysis of Good Farm and Psychoneurotic Groups*

The needs of the on-going research program in Stirling County forced a rapid analysis, in at least a preliminary form, to see if the screening test was sufficiently useful to be given a trial as a case-finding instrument. For this reason, initial efforts were limited to comparing the 78 clearly defined neurotics with the county subgroups,[2] findings reported in this paper are based on these 78.

Trials at Guttman scaling of the data failed to show the relationship between responses and personality disturbances such as had been found in the United States Army research with the NSA items (30). Therefore it became necessary to devise a special form of item analysis aimed at constructing a composite health score or index (16).

Chi-square comparisons between the county residents as a whole and the hospital neurotics revealed that 40 of the 75 HOS items discriminated between the two groups at the 1% level. However, technical limitations of the statistical method we were using at the time forced us to reduce this number for computing scoring weights. In addition, there was some variability in responses to certain of the test items among the different socio-economic subgroups ("Good Farm," "Poor Farm," etc.), which we considered important to examine. Therefore chi-square comparisons were made between the hospital neurotics on the one hand and each subgroup separately on the other. By this means, the 40 significant items were reduced to 20 that, for every subgroup, maintained the discrimination at a 1% level. These 20 items are considered "universals" (1).

In carrying out the interviewing program a strong impression developed that the subgroup which the agricultural agent had named "Good Farm" was made up mostly of healthy, well-adjusted, asymptomatic individuals. The impression was corroborated by the findings in the preliminary analysis, that not only the 20 universals but an additional 16 HOS test items distinguished this group from the hospital neurotics at a 1% level. Since the Good Farm group provided the greatest contrast with the hospital neurotics, the data

[2]It may be of interest to note that in a later analysis we found the character disorder group had answered the HOS items very similarly to the neurotics.

TABLE 1

HOS ITEM WEIGHTS DERIVED BY DISCRIMINANT FUNCTION ANALYSIS FROM RESPONSES OF 140 GOOD FARM AND 78 ADULTS WITH DIAGNOSIS OF NEUROSIS

| HOS Items<br>Question Content* | Derived Response Weights<br>"Often" (sick) | <br>"Sometimes" | <br>"Hardly ever" or "Never" (well) |
|---|---|---|---|
| 1. Do you have loss of appetite? | 0 | 12.8 | 25.6 |
| 2. How often are you bothered by having an upset stomach? | 0 | 7.9 | 15.8 |
| 3. Has any ill health affected the amount of work you do? | 0 | 6.7 | 13.4 |
| 4. Have you ever felt that you were going to have a nervous breakdown? | 0 | 6.2 | 12.4 |
| 5. Are you ever troubled by your hands sweating so that they feel damp and clammy? | 0 | 4.8 | 9.6 |
| 6. Do you feel that you are bothered by all sorts (different kinds) of ailments in different parts of your body? | 0 | 3.1 | 6.2 |
| 7. Do you ever have any trouble in getting to sleep and staying asleep? | 0 | 3.0 | 6.0 |
| 8. Do your hands ever tremble enough to bother you? | 0 | 2.7 | 5.4 |
| 9. Do you have any particular physical or health trouble? | 0 | 2.5 | 5.0 |
| 10. Do you ever take weak turns? | 0 | 2.3 | 4.6 |
| 11. Are you ever bothered by having nightmares? (Dreams that frighten or upset you very much?) | 0 | 2.1 | 4.2 |
| 12. Do you smoke a lot? | 0 | 2.0 | 4.0 |
| 13. Have you ever had spells of dizziness? | 0 | 1.4 | 2.8 |
| 14. Have you ever been bothered by your heart beating hard? | 0 | 0.3 | 0.6 |
| 15. Do you tend to lose weight when you have important things bothering you? | 0 | 0.2 | 0.4 |
| 16. Are you ever bothered by nervousness? | 0 | — 1.2 | — 2.4 |
| 17. Have you ever been bothered by shortness of breath when you were not exercising or working hard? | 0 | — 2.0 | — 4.0 |
| 18. Do you tend to feel tired in the mornings? | 0 | — 2.7 | — 5.4 |
| 19. For the most part, do you feel healthy enough to carry out the things that you would like to do? | 0 | — 3.5 | — 7.0 |
| 20. Have you ever been troubled by "cold sweats"? (NOT a Hot sweat—you feel a chill, but you are sweating at the same time.) | 0 | — 5.7 | —11.4 |

*These questions are not presented in order as asked in the interview but in relative rank order by derived weights.



from it were used in a special form of discriminate function analysis to determine the optimal scoring weights to give each of the 20 universal items so that the distribution of scores of the Good Farm group would be maximally distinguished from that of the hospital neurotics. The weights found by this analytic procedure were then applied to the scoring of the 20 universal items in all the protocols in such a fashion as to give a higher score to the non-symptom responses and a lower score to the neurotic-like answers. Table 1 gives the universal items and the derived weights.

## RESULTS

Table 2 shows the distribution of HOS test scores of the Hospital Neurotics, on the right, and of the Good Farm residents on the left. To the right of the Good Farm is the combined total of the other community responses from Average, Poor Farm, Day Labor, and Urban. The next column to the right is the distribution of the two combined counties including the Good Farm. Next to the Hospital Neurotics is the Poor Farm distribution from one of the counties. Limitations of space do not allow showing other subgroup distributions.

The scoring system provides a higher score for fewer of neurotic-like responses, and one can see at a glance that there is a progression from predominantly high scores to predominantly low across the table from left to right, with the "Good Farm" group scoring highest, the Hospital Neurotics lowest. Graphic representations of the distributions shown in columns "A," "B," and "C," each in contrast to column "E" from Table 2, are given in Figs. 1, 2, and 3. These all indicate that the score distributions show some degree of overlap between the community groups and the neurotics so that at certain scores one might misjudge a neurotic person as "well" or a "well" person as neurotic. At the same time, there is a definite tendency for the Hospital Neurotics to be characterized by lower scores than, say, the "Good Farm" group. The problem, then, is to establish a cutting point which will serve as the optimum division line between neurotic and well, with the fewest false negatives on one side and false positives on the other. A single temporary cutting point has been used on the above figures for purposes of explication.

Two possible explanations come to mind for the appearance of county residents on the neurotic side of the cutting point: (a) that the test does not adequately distinguish between neurotic and non-neurotic; and (b) (much more probable) that there are people in the county with neurotic symptoms even though they have not been so diagnosed. During the testing program in the county, the interviewers kept separate notes on behavior and remarks

'Devised by the late Garnet E. McCreary, formerly the Stirling County Study statistician, who supervised the computations of the discriminant function analysis.

Table 2

DISTRIBUTION OF DISCRIMINANT FUNCTION HOS SCORES OF COMMUNITY SUBJECTS AND DIAGNOSED NEUROTICS

| Test Scores | "A" Two County Good Farm | % | "B" Two County Average; Poor; Labor; Urban | % | "C" Two County Total (Combined A & B) | % | "D" Poor Farm County "X" | % | "E" Diagnosed Neurotics | % |
|---|---|---|---|---|---|---|---|---|---|---|
| 10.0-14.9 | | | | | | | | | 1 | 100 |
| 15.0-19.9 | | | | | | | | | 3 | 98 |
| 20.0 | | | 4 | 100 | 4 | 100 | 1 | 100 | 6 | 94 |
| 25.0 | | | 1 | | 1 | | 0 | | 7 | 86 |
| 30.0 | | | 5 | 99 | 5 | 99 | 2 | 98 | 11 | 77 |
| 35.0 | 1 | 100 | 8 | 98 | 9 | 98 | 4 | 95 | 8 | 63 |
| 40.0 | 0 | - | 16 | 96 | 16 | 96 | 4 | 94 | 8 | 53 |
| 45.0 | 3 | 98 | 16 | 92 | 19 | 93 | 3 | 90 | 12 | 43 |
| 50.0 | 3 | 96 | 21 | 88 | 24 | 90 | 5 | 87 | 5 | 26 |
| 55.0 | 9 | 94 | 34 | 83 | 43 | 86 | 10 | 82 | 11 | 22 |
| 60.0 | 17 | 88 | 41 | 75 | 58 | 78 | 13 | 73 | 2 | 8 |
| 65.0 | 20 | 76 | 44 | 65 | 64 | 68 | 11 | 61 | 2 | 5 |
| 70.0 | 23 | 62 | 54 | 55 | 77 | 57 | 13 | 50 | 1 | 2 |
| 75.0 | 25 | 46 | 58 | 42 | 83 | 43 | 18 | 38 | 0 | - |
| 80.0 | 24 | 28 | 61 | 28 | 85 | 28 | 8 | 21 | 0 | - |
| 85.0 | 13 | 11 | 46 | 13 | 59 | 13 | 13 | 11 | 0 | - |
| 90.0 | 2 | 2 | 9 | 2 | 11 | 2 | 1 | 2 | 1 | 1 |
| 95.0-99.9 | 0 | - | 1 | - | 1 | - | 1 | 1 | | |
| TOTAL N | 140 | | 419 | | 559 | | 105 | | 78 | |

Note: The larger scores are associated with good health and smaller scores with poor health. Percentages are cumulative upwards from the bottom of the table.



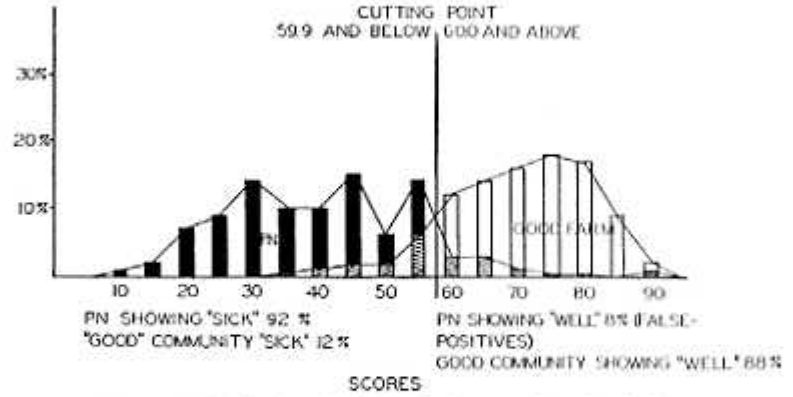


FIG. 1. Distribution Overlap: Psychoneurotic—Good Farm

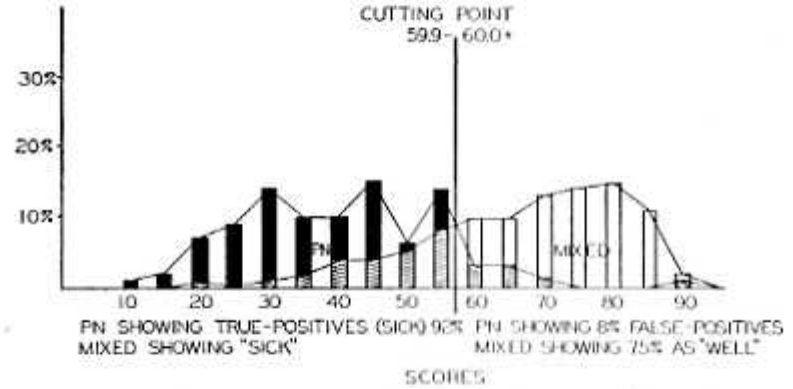


FIG. 2. Distribution Overlap: Psychoneurotic—Average and Poor Farm, Day Labor and Urban.

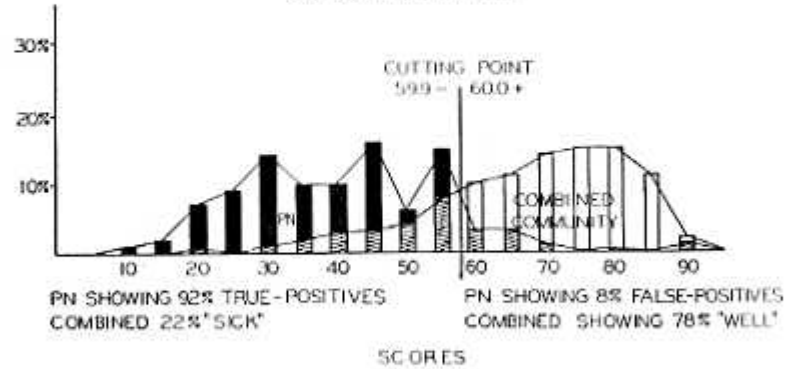


FIG. 3. Distribution Overlap: Psychoneurotic—Two-county Total

of respondents which seemed significant. These were, of course, not utilized in scoring the questions but were assessed independently later. Most of the interviewers' comments turned out to be either describing overt behavior or complaints of the respondents such as are commonly found among clinically recognized neurotics (10, 26), for example, restlessness, irritability and aggression, complaints of fatigue, tremors, excessive perspiration, etc.

After the distribution of HOS scores had been worked out, the interviews of county residents whose scores fell on the neurotic side of the cutting point were examined to see if there were interviewer comments or notations of significant symptoms or complaints. It was found that of those *Ss* scoring in the 50's, 46% had some such indication; of those scoring in the 40's the percent rose to 70; while of those with the most neurotic-like score (20-30) 88% had either noticeable symptoms or complaints. Although obviously not conclusive, these data provide some evidence supporting the test, and indicate that our community groups are certainly not symptom-free. It also suggests that for our screening purposes possibly two cutting points rather than one may be desirable, with a "Doubtful" category between them.

*Establishing Cutting Points*

The worth of a screening instrument depends upon its ability to sort out individuals along some axis, in this case neurotic/non-neurotic. The seldom-achieved ideal would be a clear either/or sort, with no overlap. Few aspects of human behavior can be measured in this neat way, however. If one considers the variety both of the manifestations of neurosis and of the intensity of symptoms, which could play a part in the score respondents would achieve on the HOS, he would hardly expect a simple dichotomy to give adequate separation. At the same time, the more definitely the comparison groups contrast with each other, the nearer one might approach a dichotomy. A further difficulty which had to be dealt with in these data was that, although the neurotic population was quite positively defined both by the individuals themselves in seeking hospital aid and by the physicians who had diagnosed them, there was no way of telling with certainty that the "Good Farm" group which was being used for contrast was, in fact, well and asymptomatic.

There are several possibilities in placing cutting points, depending upon the use to which the data are to be put. One way would be to choose that score which would identify practically every neurotic but would also include some of the "wells" (e.g., 70). The reverse would be to choose the score (e.g., 45), which takes in practically all the "wells" but also includes a good many neurotics. A third way of dealing with the difficulty, and the one which seemed most useful for this data, would be to use twin cutting points: one, the limit below which any score could be considered as showing a neurotic reaction; the other, the limit above which any score could be thought to show absence of neurosis, and the range between being left as doubtful. The desirability of such a scheme will be apparent to all who have tried to classify human beings into either/or categories on almost any basis.



Both single and twin cutting points have been used in this study. For instance, in Figs. 1, 2, and 3 a single cutting point between 59.9 and 60 is used to separate 92% of the hospital neurotics from the county residents. The separation of the Good Farm and Neurosis groups, shown in Fig. 1, is, of course, a function of the discriminant analysis of the responses of these two criterion groups, and the degree of separation, with only 12% and 8% respectively, is to be expected. The real test of the effectiveness of the derived weights, however, is on another community population. This is given in Fig. 2 where the combined two-county sampling of Average Farm, Poor Farm, Day Labor, and Urban score distribution is compared with that of the Neurosis group. Here, we see that 25% of this MIXED group scores like the bulk of the Neurosis group. Fig. 3 shows the distribution of Fig. 2 but with the Good Farm group included. Thus, with a single cutting point between 59.9 and 60.0, we see that one-quarter of the non-Good Farm community adults score on the "sick" side, as do 92% of the Neurosis group. For the latter, there are 8% false-positives but we do not know the proportion among the community group. The use of a double cutting point is demonstrated in the next section.

*Validation by Psychiatric Examination*

In order to push validation a step farther than the evidence from the interviewer's comments, it seemed desirable to get a psychiatric appraisal of a sample of the people tested in the county. To this end a psychiatrist visited 64 (11%) of the individuals tested and interviewed each in the trailer for half an hour, evaluating *S*'s mental health status. This evaluation was done without any knowledge by the psychiatrist of how the individual had scored on the HOS nor of what impression he had made on the HOS interviewer. The results of this evaluation can be seen in Table 3 where the psychiatrist's assessments are compared with the scores achieved by the same individuals on the HOS. Here it is plain that twin cutting points rather than a single one are essential, for under these circumstances it was not possible for the psychiatrist to categorize 23 *Ss* as definitely neurotic or definitely non-neurotic. He simply could not place them with confidence on either the "Well" or the "Sick" side of his rating scale. The psychiatrist estimated his error to be between 10% and 15% of those he rated. He was obviously working under difficulties in that the surroundings were certainly not like the usual clinical situation and, more important, the persons he was interviewing lacked the

TABLE 3

VALIDATION EXPERIMENT

Psychiatrist's Assessment Compared to HOS Score

| Psychiatrist's Assessment | | ◄------- HOS "Well" Range -------► | | | | | *)* | ◄---- HOS "Sick" Range ----► | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 91-86 | 85-80 | 79-73 | 72-68 | 67-62 | 61-59 | 58-53 | 52-47 | 46-41 | 40-35 |
| "Well" | I | N<br>N<br>N | S<br>S<br>S<br>S<br>S | -- | N(Ps?) | N<br>M<br>M<br>M<br>M | M | N(Ps) | N | | |
| "Well" | II | N(Ps?) | - | N<br>Ps<br>Pn | Pn<br>Pn<br>Ps<br>Ps | | Ps<br>Ps?<br>Pn? | | | Pn | |
| "Sick" | IV | - | - | - | - | - | PP<br>PC | - | PN<br>PN | PC<br>PN | PN<br>PN |
| "Sick" | V | - | - | - | - | - | - | PN<br>PN | | PC<br>PN | - |

LEGEND

Psychiatrist's Assessment: Well I - Appears quite well; Well II - Minor symptoms but not disabling; Sick IV - Appears to need help with problems; Sick V - Definitely needs help

Psychiatrist's Diagnosis: N -Normal; Pc -Slight Character Difficulty; Pn -Slightly Psychoneurotic; Pp -Slightly psychotic; Ps -Slight psychosomatic

If a second code letter of any pair is also a "capital" this indicates strong evidence for the diagnosis. A diagnostic code followed by "?" indicates "questionable".

Note: The psychiatrist's "Don't Knows" included 6 in the HOS well range; one in the "?"; and 16 in the sick range, which are not shown in the Table.

familiar motivation of wanting help. Moreover, half an hour is a very short time for making such an appraisal (though considerably longer than is usual in military assessment interviews). In all probability, a proportion of the doubtfuls could have been placed in either the well or the sick group if he had had more time to interview them.

Table 3 shows a high correspondence between the psychiatrist's assessment and the HOS test score; that is, nearly all the persons that he considered well had high HOS scores, and nearly all that he thought sick had low HOS scores. In other words, as far as it goes, the psychiatric clinical assessment tends to validate the HOS as a means of screening the neurotically sick from the well. It tends, further, to indicate that a number of the county residents actually had recognizable neurotic symptoms as well as neurotic HOS scores.[5] There is a total difference of 14% between the psychiatrist's assessment and the designation by the HOS scores. This indicates that, where the picture is reasonably clear from the clinical point of view, there is an excellent chance that the screening test will categorize the person in the same "Sick-Well" way as the clinician.

[5]Four of his "Well" and two of his "Sick" had HOS scores in the doubtful range. (Of his doubtfuls 6 scored from 79 to 62, 1 in the 61-59 range, and 16 were 58-20.) There are no false-negatives. That is, he did not consider "Sick" anyone who had a high (or "Well") HOS score; but there are 3 whom he considered "Well" who made a "Sick" score, while 4 of his "Well" and 2 of his "Sick" appear in the doubtful-score group.



## Discussion

While these results are probably indicative of a usable device, certain cautions must be observed. For example, it may be that the validation by the psychiatrist was a matter of chance and that in another trial he would fail to agree with the test scores because of the unfamiliarity of the situation to a clinician. Furthermore, the test needs a much more extensive try-out in different types of communities and in various cultures in order to see what modifications are required under various socio-cultural conditions. Our experience with and modifications of this test in Stirling County will be reported elsewhere. While its use in Stirling County provides a limited opportunity to test the effect of translation into another language, the full significance of language and cultural differences for the HOS also remains to be explored. Further work is being done with the HOS data with factor analysis, latent structure analysis, and an attempt is being made to derive a psychotic index as a complement to the neurotic index reported here.

Assuming that further trials substantiate the apparent promise of the HOS as a means of screening neurotics from non-neurotics, some of its potential uses might be to: (a) obtain an estimate of the degree of need for psychiatric services in specific communities, industries, or institutions; (b) compare the variation of need across a region (county, province, state, city, etc.) in order to plan the most efficient use of scarce psychiatric resources; (c) measure the effectiveness of mental health programs by redetermining the prevalence of psychiatric symptoms from time to time; (d) designate individuals for psychiatric appraisal in educational, industrial, or military contexts; (e) provide a time-saving general over-view of a patient's symptomatology for the busy physician; and (f) make possible a wide variety of research.

## Summary

A questionnaire-survey test, the Health Opinion Survey, has been developed for screening neurotic individuals from a general population. Results are reported from administration of the test to 559 white adults in a two-county rural area and to 78 diagnosed hospital neurotics in a nearby city. By using a score weighting method based on a discriminant function analysis of a portion of the community responses, 25% of the 419 *Ss* scored low (indicating poor health), as compared to 92% of the Neurosis group. Subsequent independent validation evidence indicates that, at least for this type of population, the test holds considerable promise as a screening instrument.

## References

1. Appel, V., & Kipnis, D. The use of levels of confidence in item analysis. *J. appl. Psychol.*, 1954, 38, 256-259.
2. Cattell, H. McK. Homo-Scientificus Americanus. *Science, U. S.*, 1903, 17, 561-570.