THE USE OF SYMPTOM AND LIFE EVENT STRESS INFORMATION IN PSYCHIATRIC JUDGEMENT AND CASE IDENTIFICATION

BY

LAWRENCE KRASNOFF

A dissertation submitted to the Graduate Faculty in Psychology in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York.

1979

This manuscript has been read and accepted for the Graduate Faculty in Psychology in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Jan. 1, 1979
date

Chairman of Examining Committee

January 2 1979
date

Florence L. Denmark
Executive Officer

Barbara S. Dohrenwend

Laurence Gould

Alan Gross

Supervisory Committee

The City University of New York

Acknowledgements

I would like to thank Barbara Dohrenwend for her great patience and understanding in carefully reviewing a volume of material much larger than the product presented here. More generally, I thank her for being everything an adviser should be.

I would like to thank Alan Gross for his tolerance of my intuitive and less articulated first approaches to the quantitative analysis. I thank Larry Gould for his long and continual support in both this and other endeavors. Thanks to John Antrobus and Harold Wilensky for their thoughtful comments on the wider issues touched on in this work.

I would like to thank Carole Vance, Lou McDonald, Eli Baker and Leon Landovitz for their help and support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

The identification of untreated psychiatric cases is a long standing problem in epidemiological research (Mechanic, 1970; Plunkett & Gordon, 1960). Moreover, the difficulties in the identification of untreated cases are important outside of epidemiology because they bring into question the basic definition of psychopathology. If psychiatric disorder can not be reliably identified independently of treatment, one must question whether those identified as cases in treatment settings have anything else in common besides contact with mental health services.

If research methods used for case identification were consistent with models of psychopathology found in clinical literature, difficulty in the reliable identification of untreated cases would suggest that those identified as treated cases are so classified for situational reasons and would bring into question the models, or theories, of psychopathology used in clinical practice. However research methods used for case identification are often inconsistent with the models of psychopathology found in clinical and research literature and consequently do not provide a fair test of the potential for reliable identification of untreated cases.

Research on the identification of untreated cases will therefore be reviewed and analysed with respect to how well they reflect current theories of psychopathology. A method for case identification will be developed which uses some of the best features of current methods and incorporates an important feature of current models of psychopathology: the belief that when symptoms can be atrributed to temporary stresses, they do not indicate pathology. This belief is generally ignored and at best poorly handled in current reseearch on identification of untreated cases. The efficacy of the proposed method will then be compared to that of a more conventional one in order to see if making the method of case identification more consistent with this feature of current theory improves the reliability of case identification.

A variety of methods, roughly falling into three categories, have been used to identify untreated cases. Historically, the first method was the use of clinicians' judgements based directly on unstandardized interviews (Essen-Moller, 1970; Plunkett & Gordon, 1960; Hagnell, 1966; Lewis, 1961). This method involves judgement during both the data collection and its synthesis. There are many problems with this method including those of questionable comparability of findings, unspecified influence of social factors on judgements, and the limitation of sample size in areas where psychiatric manpower is lacking or expensive (Srole, Langner, Michael, Opler, & Rennie, 1962, p. 141).

The second method involves the use of structured interviews, administered by either psychiatrists or non-psychiatrists, instead of traditional clinical interviews as the source of data for psychiatric judgements. This method greatly restricts or eliminates judgement during data collection but allows it to operate during the sythesis of data involved in making judgements. Such interviews include checklists of standard symptom items thought to be general indicators of psychiatric disorder by the psychiatrists working on these projects (Leighton, Harding, Macklin, Macmillan, & Leighton, 1963; Srole et al., 1962; Wing, 1970).

Although the structured interview method clarifies and standardizes some of the information used as a basis for clinical judgement, it does not standardize how this information is used in generating clinical judgements. This method has been criticized both by the investigators themselves (Srole et al., 1962, pp. 395-407) and by others (Dohrenwend, B. P. & Dohrenwend, B. S., 1969) for the lack of explicit rules for judgements. Without explicit rules a question remains about what information is used and how it is used in the formulation of clinical judgements.

The last method involves using symptom scales composed of symptom items which have been shown to discriminate between patient and "well" groups as direct indicators of psychopathology or status as an undiagnosed case, thereby

eliminating clinical judgements altogether. The use of such screening scales is attractive in that they are quick and easy to administer, do not require the use of expensive psychiatric manpower, and completely specify the basis of the classification. However there are problems in using symptom scales as direct indicators of psychopathology (Dohrenwend, B. P.,Oksenberg, Shrout, Dohrenwend, B. S., & Cook, Notes 3; Tousignant, Denis, & Lachapelle, 1974; Seiler, 1973).

A recurrent criticism is that these scales may identify as cases individuals who are physically sick or those who are reacting to stressful situations (Tousignant et al., 1974; Star, 1949). These problems are generally referred together. However the problem of the influence of recognized physical illness on these scales has been addressed methodologically, whereas no effort has been made with respect to the problem of temporary stress reactions.

A method for dealing with the problem of temporary stress reactions for use with screening scales would be an important improvement since these scales have advantages relative to the other methods of case identification. It is of note that the problem of the effect of environmental stress on symptom items has also been referred to in the literature on structured interviews. Furthermore the remarks of psychiatrists working with structured interviews are informative with respect to how one should deal with the influence of

temporary stress reactions on symptom scales used as direct indicators of psychopathology.

In the Stirling County study a decision was made to classify subjects by symptom pattern instead of diagnosis. In explaining their decision the authors make the following point about diagnosis.

> 4 The attempt to make a diagnosis in the usual clinical sense must be abandoned... Different psychiatrists made different assumptions regarding dynamic process. More than this, they also included in their diagnostic evaluations varying estimates regarding the effects of environmental factors. (Leighton, et al.,1963, p. 48)

This quotation reflects the fact that diagnosis is based on both the presence of symptoms and "estimates regarding the effects of environmental factors." The way these estimates are used in diagnosis is suggested in the following quotation which describes the difficulty the Stirling County psychiatrists encountered when they tried simply to evaluate sypmtom patterns instead of making diagnoses.

> The evaluators frequently found themselves saying, "No wonder so-and-so has such symptoms. Look what he has been through." They had to tell each other repeatedly that a symptom is a symptom and must be rated as to its presence, regardless of explanations or extenuations perceived in the life history. (Leighton, et al., 1963, p.55)

This quote suggests that the psychiatrists typically would discount a symptom if it were 'explained' or 'extenuated' by a stressful situation.

This same process of discounting symptoms in certain contexts is refered to in the Midtown study. In this case

the psychiatrists were trying to rate for "symptom impairment."

> Thus we devised a system of flag items, agreement to which by the respondent precluded a rating of good mental health. But these responses could not be used mechanically as punch-card items, although a "yes" answer indicated the presence of a definite symptom.
> ... Thus, a woman who had agreed that there were "periods of days, weeks, and months when she couldn't get going" (R19-5), a symptom usually considered to be indicative of severe emotional disturbance, was not conceded this item when she volunteered that such occurred only once in her life for a duration of approximately three weeks after the accidental death of her husband. (Srole et al., 1962, p.396)

In this case a decision was made to go along with standard diagnostic practice and discount the symptom for its context. The psychiatrists in both these studies did not regard all instances of these 'symptoms' as pathologically significant and if left to their own diagnostic inclinations would have discounted symptoms in both studies.

An Implicit Model of Psychopathology

Under extreme conditions individuals may exhibit behaviors, commonly called symptoms, which would be judged pathological reactions under normal circumstances, but which are normal responses to extreme conditions. For example, Grinker & Spiegel (1945) in their book on the reactions of Air Force combat personnel to the pressures of war state:

> ... Under sufficient stress any individual may show failure of adaptation, evidenced by neurotic symptoms. Such symptoms then are pathological only in a comparative sense, when contrasted with the symptoms of those still making successful adaptations. (p. vii)

Plunkett & Gordon (1960) raise this issue in a discussion of the problems of "case-finding":

> The competence of the interviewer and a subtlety of method are called to the test when he is faced with distinguishing between bona fide mental illness and transitory symptoms arising from unusual conditions of stress. Without a psychiatric background and an opportunity to probe in depth, it is doubtful whether most interviewers can differentiate between the two. When the contributions of dispositional weakness and environmental stress to symptomatology are complementary, the challenge to an interviewer is even more formidable. (pp. 60-61)

The Diagnostic and Statistical Manual of Mental Disorders of the American Psychiatric Association (Second Edition, 1968) (DSM-II) contains two major classifications which reflect a recognition of the view that symptoms are not pathologically significant when attributable to some identifiable condition or stress:

> 307 Transient situational disturbances
>
> This major category is reserved for more or less transient disorders of any severity (including those of psychotic proportions) that occur in individuals without apparent underlying mental disorders and that represent an acute reaction to overwhelming environmental stress. A diagnosis in this category should specify the cause and manifestations of the disturbance so far as possible. If the patient has good adapative capacity his symptoms usually recede as the stress diminishes. If however, the symptoms persist after the stress is removed, the diagnosis of another mental disorder is indicated. Disorders in this category are classified according to the patient's developmental stage as follows:

307.3 Adjustment reaction of adult life

Example: Resentment with depressive tone associated with an unwanted pregnancy and manifested by hostile complaints and suicidal gestures. Example: Fear associated with military combat and manifested by trembling, running and hiding.

307.4 Adjustment reaction of late life

Example: Feelings of rejection associated with forced retirement and manifested by social withdrawal.

316 Social maladjustment without manifest psychiatric disorder.

This category is for recording the conditions of individuals who are psychiatrically normal but who nevertheless have severe enough problems to warrant examination by a psychiatrist. These conditions may become or precipitate a diagnosable mental disorder.

The 1-15-78 draft of the third edition of the Diagnostic and Statistical Manual of Mental Disorders (Note 1) categorizes the above conditions under a single major classification of Stress Reactions with most cases falling in the subcategory of Adjustment Reaction.

The examples quoted above, in which the psychiatrists of the Midtown and Stirling county studies were tempted to make adjustments for the circumstances of the individual even when they were specifically instructed not to do so, suggest that the tendency to make such adjustments during diagnosis is so strong that it intrudes on the very process of recording clinical data despite specific instructions not to make such adjustments. Moveover, the clinical literature

suggests that clinicians should try to make such adjustments during diagnosis. The suggestion comes from many different sources within this literature and takes the form of a number of similar constructs all of which suggest that an evaluation of how successfully or appropriately a person is functioning in a given context is a part of the diagnostic process.

Within American psychiatry Menninger (1952, pp. 98-105) recommends that "life situation maladjustment" is an important factor that should be evaluated during diagnosis. Other American psychiatrists, Offer & Sabshin (1974) cover the same ground under the heading of "Coping and Adaptation" in their discussion of "The Concept of Normality." Psychoanalytic writings contain essentially the same functionalistic approach in a more intrapsychic construct of "adaptation" which is the cornerstone of the entire field of ego psychology (Hartmann, 1951).

Psychologists with diverse backgrounds suggest that one should consider the context in evaluating the pathognomic significance of particular behaviors. Peterson (1968) in his Radical Behaviorist critique of assessment uses the concept of "functional stimulus response relationships" to argue that psychologists should use information about context more than they do already. The concept of the "etiological equation" is based on the psychometric distinction between stimulus variation and intersubject variation

and is used by Thorne (1961) as a vehicle for separating contextual from individual variation in his work on clinical judgement.

All these concepts suggest that a diagnostic judgement should entail an evaluation of how well or how appropriately a person is responding to his or her environment. These concepts may be thought of as reflecting an implicit model of psychopathology. This model implies that those who are responding well or appropriately be called normal and those who are not be called pathological (cf. Wilson, 1963).

More specifically we can expect that the adjustments during diagnosis take two forms. First, particular symptoms may be explained by specific events, as in the example of the woman who "couldn't get going" in the Midtown quotation. Second, there should be a general discounting of the pathognomic significance of all symptoms under conditions of extreme stress as suggested in the quotes on war time conditions, case finding, and adjustment reactions.

The first type of adjustment can be built into symptom items by specifying that an item not be scored as present under certain conditions. For example, in some scales sypmtom items that often represent physical illness are not scored as present if they can be attributed to a physical illness (Spitzer, Endicott, Fleiss, & Cohen, 1970). In other scales this type of adjustment has been handled by simply dropping from the scale items that often represent physical

illness in a test sample (Dohrenwend, B. P., et al., Note 3). However, the second type of adjustment can not be built into individual items and requires a measure of stress that can be used as an adjustment for all the items.

Stressful Life Events

The examples cited above, especially the diagnostic categories from the DSM-II for adjustment reactions, suggest that stressful life events are an important part of the circumstances to be considered in assessing the pathological significance of gross sypmtomatology. The implicit model of psychopathology cited above suggests that the association between life event measures and disorders reflects, at least in part, a process in which life event stress elicits transitory physical and psychiatric symptoms.

Research on stressful life events has produced measures for the change attributable to reported life events. These measures offer promise as adjustments to symptom scales because they have already been found be to related to a number of physical and psychiatric disorders (Dohrenwend, B. P. & Dohrenwend, B. S., 1974a; Houg, Fairbank, & Garcia, 1976).

Clinical interest in stressful life events can be traced back to the 1930's when Adolf Meyer, influential in shaping the form of American psychiatry, advocated the use of the "life chart" in medical diagnosis (Dohrenwend, B. P. &

Dohrenwend, B. S., 1974b). The life chart is a complete medical history from birth together with "the changes of habitat, of school entrance, graduations or changes, or failures; the various 'jobs'; the dates of possible important births and deaths in the family, and other fundamentally important environmental incidents" (Meyer, 1951, p. 53).

Meyer's influence can be seen in the 1949 conference of the Association for Research in Nervous and Mental Disease: Life Stress and Bodily Disease. Papers were presented at ten panels on the effects of life stress on disorders of growth, development, metabolism, the eye, the airways, stomach, colon, muscles, joints, and periarticular structures as well as headaches and cardiovascular disease (Wolff, Wolf, & Hare, 1950).

Influenced by the work of Meyer, Wolff, and other participants of the 1949 conference, Holmes & Rahe (1967) developed of the Social Readjustment Rating Scale (SRRS), the stressful life event scale most widely used in the last decade (Gunderson & Rahe, 1974; Dohrenwend, B. P. & Dohrenwend, B. S., 1974a). The list of events in the SRRS is based on the systematic study of life charts of over 5000 patients initiated in 1949 (Holmes & Rahe, 1967).

The SRRS is a list of weights, or "life event change ratings," for 43 events. These weights represent relative degree of readjustment in one's usual pattern of life

resulting from each of the events. The list includes events such as marriage, divorce, change in residence, and death of a close family member, as well as business readjustment, minor violations of the law, and vacation.

These ratings or weights are summed for all events reported in a given period of time by subjects for whom there are also measures of pathology. This total represents the "magnitude of life change" according to the authors of the SRRS (Holmes & Rahe, 1967). On the basis of demonstrated associations between these totals and numerous measures of disease Holmes & Masuda concluded: "The greater the magnitude of life change (or life crisis), the greater the probability that life change would be associated with disease onset, and the greater the probability that the population at risk would experience disease" (Holmes & Masuda, 1974, p. 68).

The SRRS represents a convergence of long standing clinical interests with a powerful psychometric technique. The most important influence in making the SRRS a methodological advance over previous work was the adoption of a scaling technique (Hough et al., 1976).

The techniqued used in generating the weights for the 43 events in the SRRS is called magnitude estimation. Holmes & Rahe use this technique to generate a single, fixed weight for each event. These fixed weights, unlike a subject's self-report of the stressfulness of his or her experience,

can not be influenced by a subject's observation of his or her own reaction. These fixed weights are therefore a cleaner measure of the average impact of the event on the individual, because they are not confounded with other determinants of the subject's response, such as his personal strengths and weaknesses, other concurrent events, and sources of emotional and physical support.

However, there is a problem in the way Holmes and Rahe use magnitude estimation. They have used this technique uncritically, as if it answers all the measurement problems posed by stressful life events. Holmes and Rahe (1967) cited only the proponents of magnitude estimation (Stevens & Galanter, 1957; Stevens, 1966; and Sellin & Wolfgang, 1964) and seemed unaware of possible shortcomings that have been pointed out by other investigators working with stressful life events (Brown, 1974a; Paykel, Prusoff, & Uhlenhuth, 1971).

Instead, Stevens, the major proponent of magnitude estimation, suggests that the technique provides weights that are free of the influence of between judge differences in ratings and does not treat the existence of such differences as a serious methodological problem:

> The factors that affect the outcome [of the rating process] are numerous and subtle. Patience and experimental skill can probably clean up part of the variance, but there will always remain irreducible dispersions to set a level below which we sink into uncertainty. (Stevens, 1957, p. 167)

Holmes and Rahe accepted this view and made only minimal attempts to look at differences in the ratings made by raters from contrasting status groups based on sex, age, ethnicity, or class; they showed high rank order correlations across the event list between ratings generated by different groups of judges and concluded that there were no status group differences (Holmes & Masuda, 1974).

Others have argued that this is an insufficient test of group differences; they have proposed a decision model for evaluating life event ratings which tests for group differences using a factorial analysis of variance of life event ratings. They demonstrated the model using the sex, education, and ethnicity of the raters as factors in the analysis of variance (Dohrenwend, B. S., Krasnoff, Askenasy, & Dohrenwend, B. P., 1978; Krasnoff, Askenasy, & Dohrenwend, B. P., Note 4). The analysis was performed on a 102 item event list similar to the SRRS and a number of group differences were found for a variety of events.

These findings raise an issue as to the best method for estimating the total magnitude of life change for an individual. If differences in ratings reflect true differences between groups in the impact of particular events, studies relating stressful life events to other variables should use ratings that differ with the group membership of the individual experiencing the event. If these rating differences are not due to differences in the impact but to other

factors such as response style, a single weight for an event, regardless of group membership, is sufficient.

A number of investigators have criticized additional aspects of Holmes and Rahe's work and the assumptions on which it is based and thereby raised issues concerning the nature of relations between stressful life events and indicators of illness. For example B. P. Dohrenwend (1974) has made the point that certain events in the SRRS and other lists of life events are possibly the result of prior psychopathology and therefore can not be used as independent measures of a person's circumstances. Other investigators have suggested that negatively evaluated change is a better predictor than total change for some or all types of psychopathology (Brown, 1974b; Gersten et al., 1977; Paykel, 1974).

Types of Instruments

There also is debate over the best method for the construction of symptom scales. This debate reflects two distinct traditions in the developement of symptom scale instruments.

One tradition follows from the structured interview approach and uses scales composed of fixed alternative response questions like those used in the Midtown and Stirling County interviews. This type of questionnaire words items so that they can be asked directly of the subject and the interviewer need only record whether the answer was

"Somtimes", "Often" or "Never" for some questions or "Yes" or "No" for others. Such an instrument does not require clinical training on the part of the interviewer and can be administered by non-psychiatrists.

The other tradition is the "standardized" interview and is an outgrowth of efforts to standardize the mental-status interview that is a part of traditional clinical examinations (Spitzer, et al., 1970). A standardized interview only standardizes the areas of questioning and recording, and, unlike a structured interview, allows clinical judgement to operate in the collection of data. Specifically, this type of interview involves the recording of judgements by the interviewer concerning the presence of specific symptoms based largely on standardized open-ended questions and probes. Although the judgements are recorded in terms of fixed categories for each item, the number of questions and probes actually asked of the subject is not fixed. In fact some items have no corresponding question and are based entirely on unspecified observations.

The difference in data collection between these two approaches to scale construction are important with respect to adjusting symptom scales for life event stress. The implict model of psychopathology described earlier suggests that allowing clinical judgement to affect raw scores should result in an adjustment for life events as part of the scoring process. Furthermore, this type of adjustment was

actually described in the quotations on the Stirling County 'evaluators' and Midtown raters. Therefore, one would expect explicit adjustment of symptom scores for stressful life events would improve the validity and utility symptom scales as indicators of psychopathology more in the case of scales from structured interviews than in the case of scales from standardized interviews.

Hypotheses

Four main hypotheses concerning relations between life events and symptom measures may be formulated from the foregoing discussion. (1) Symptom scores increase with life event stress scores. (2) Psychiatric judgements of the probability of being a case increase with symptom scores. These two hypotheses are based on relationships already established in the literature. The primary purpose of these hypotheses is to establish a basis for the test of the next hypothesis, (3) Symptom scores that are adjusted for life event stress are associated more strongly with psychiatric judgements than unadjusted symptom scores. That is, the prediction of psychiatric judgement from both symptom scores and life event stress measures should be better than the prediction from symptom scores alone. This is the central hypothesis in this study.

The next hypothesis is concerned with the distinction between symptom scales based directly on respondents self

reports versus those based on psychiatric judgements. (4) The improvement posited in hypothesis three is greater for symptom measures based directly on respondents self reports than for symptom measures based on clinical judgement. That is, the increase in the ability to predict psychiatric judgements afforded by the addition of life event information to symptom scores is larger when one adds life event information to symptom scores from structured interviews than when life event information is added to symptom scores from standardized interviews. Confirmation of hypothesis (4) would provide further support for the theory that generated hypothesis (3).

METHOD

Subjects

The data come from a methodological study designed "to develop better measures of psychiatric symptomatology and related disability in role functioning than have been used heretofore in epidemiological studies of general populations" (Dohrenwend, B. P. et al., 1970, p.159). There are four samples containing subjects ranging in age from 21 to 64. The sample of major interest is a 'community cross-section' sample (N=257). This was a stratified probability sample of a population previously studied in Washington Heights. The probabilty samples of community respondents were designed to provide roughly equal distributions of educational level, an indicator of class, within each of five ethnic groups: white Protestants of old American ancestry, Jews, Irish, Negroes, and Puerto Ricans.

There is also a sample of patients (N=204) including psychiatric inpatients, outpatients, and prisoners. This sample was selected to have a roughly even distribution of six "behavior types" within each of the five ethnic groups. The 'behavior types' are based on fictious illustrations of different types of disorder developed by Star (Note 5) and used in studies of public attitudes toward mental illness

(e.g. Dohrenwend, B. P., Chin-Shong, 1967). The six illustrations characterize a paranoid schizophrenic, a simple schizophrenic, an anxiety neurotic, a compulsive phobic neurotic, an alcoholic, and a personality disorder of the antisocial type. A patient was selected for inclusion in the study if the quota for the behavior type he most resembled had not been filled for his ethnic group.

Finally there is a sample of community leaders from the same Washington Heights population (N=67). These persons are leaders in political, religious, and civic groups in the community.

Instruments

Both the Structured Interview Schedule (SIS), a structured interview, and the Psychiatric Status Schedule (PSS), a standardized interview were used. Aproximately half the subjects in each sample were interviewed with the SIS and half with the PSS.

Each instrument contained symptom and role impairment scales. The role impairment scales were included on the assumption that serious psychopathology is associated with impaired ability to function in social roles (Sartorious, 1971). Since they may represent alternative measures of psychopathology these role impairment scales will be compared in the same way symptom scales are in the hypotheses.

Interview Procedure

Fifteen psychiatrists conducted the interviews. Each interview lasted from forty-five minutes to two hours and forty-five minutes, with the average being almost one hour and forty-five minutes. The first assignments of respondents were randomized and the interviews conducted by each psychiatrist involved systematic alternation of the two instruments. This procedure was used because it permits controlled comparison of the results for the two instruments despite possible order effects and differences between psychiatrists.

Each interview was composed of three sections. The first section was the interview schedule proper, either the SIS or PSS. The second section was a self-administered questionnaire designed to provide information on the influence of social desirability and acquiesence response styles on answers to questions in the first section. The questions in this section were like those in the SIS. In fact a number items were identical to those on the SIS. This feature allowed for the construction of social desirability and acquiesence scales especially suited for checking for these biases in SIS scales.

The third section was a "stressor inventory." It began with the question, "What was the last major event in your life that, for either better or worse, interupted or changed your usual activities?" , followed by the probe, "For exam-

ple, events affecting your occupation, your physical health, your living arrangments, your relations with other family members, your friends or your personal values or beliefs." There were further probes for details and an open-ended question about other events almost as important since the last major event. This was followed by two identical check-lists, the first one for events that occurred during the last twelve months and the second for those anticipated during the next twelve months.

Psychiatric Judgement

During the second section of the interview, while the subject was filling in the self-administered questionnaire the psychiatrist made two judgements and a diagnosis. The judgements were made on the main rating scales used in the Stirling County and Midtown studies.

The Stirling County rating is described as a four point "caseness" scale representing the "subjective probability" that a subject would be classified as a case if given a "full diagnostic investigation." Each subject was given a rating of A, B, C, or D with average probabilties of being a case being equal to .9, .7, .4, and .1 respectively in accordance with the published descriptions of how the Stirling County staff made their ratings. Some of the Stirling County findings are reported in terms of a dichotomized caseness scale on which all subjects in the two groups with

the higher average probabilities of being a case are considered "cases" and the other two groups are considered "well." This convention of treating subjects from groups with probabilities of being a case of greater than fifty percent as cases and those with probabilities of less than fifty percent as non-cases has been used in subsequent work as well (Dohrenwend, B. P., Egri, & Mendelsohn, 1971), and is based on the underlying probabilistic conception of caseness.

The Midtown rating is a six point scale of symptom "impairment" with categories labeled "well", "mild", "moderate", "marked", "severe", and "incapacitated." The psychiatrists in this study were guided by the published descriptions of the criteria used by the Midtown staff, which included the following. A person could be rated as "well" either if he had no symptoms or only "mild" (e.g. hay fever) and "rare" symptoms that "were not disturbing to the individual." A rating of "mild" was consistent with "mild anxiety or tension [if] the respondent show[s] evidence that he is effective at work, at home and in social and interpersonal relations [and is] not grossly overweight or underweight." "Moderate" is consistent with substantial evidence of psychosomatic interference such as "drinking [or] eating too much, restlessness, poor memory,and not [being] able to get going." "Marked" is similar to "moderate" but with evidence of "interference by these symptoms in the respon-

dent's life adjustment." "Severe" implies "serious symptom formation with some interference in life adjustment." If the "interference in life adjustment" is "great" or incapacitating the respondent is classified as "incapacitated." The Midtown psychiatrists thought "this last category was frequently comparable to patients hospitalized in mental institutions." The Midtown group also used a dichotomized version of their scale with "well" to "moderate" being considered "unimpaired" and "marked" to "incapacitated" being "impaired." (Srole et al., 1962, pp. 395-399)

It may seem that all these judgements could provide criteria against which to test the hypotheses. However, although we are interested in making inferences about a feature of diagnostic practice, the multi-dimensional character of the diagnoses themselves argues against their use as a measure of a single dimension on which we have a more direct measure.

Similarly the Midtown scale attempts to measure a dimension of clinical judgement quite different from "the probability of being a case." A person may be impaired by their symptoms independently of whether the symptoms are attributable to stress. Therefore "impairment" ratings are likely to be unaffected by psychiatrist's estimates of a subject's total stress.

While it is true that the Midtown psychiatrists reported adjusting particular items for the circumstances of indivi-

duals, these adjustments were the exception to the rule (Srole et. al., 1962, p. 396). There is no guarantee that the psychiatrists in the present study will do the same and given the defintion of the scale, "impairment," there is reason to believe that they will not adjust for the general level of stress. Consequently the Stirling County caseness ratings are the only suitable measure of psychiatric judgement against which to compare symptom and role impairment scales and adjusted symptom and role impairment scales.

Symptom and Role Impairment Scales

SIS

The SIS symptom items were drawn mainly from the Army Neuropsychiatric Screening Adjunct, the MMPI, and the Stirling County and Midtown Manhattan studies. New items, not used in previous work, were added to broaden coverage of suspicions, delusions and hallucinations, drinking problems, drug use, and antisocial beliefs and behaviors. Three board certified psychiatrists sorted these items into eleven non-overlapping groups by symptom type.

The reliability of these eleven scales was then examined using coefficient alpha and item-whole correlations. The criterion for a reliable scale was alphas greater than .50 in both the community cross-section and patient-prisoner samples and alphas below .50 in no more than one ethnic, class, or sex group in either or both samples. According to

this criterion six scales proved reliable (Dohrenwend, B. P. et al., Note 3).

High intercorrelations among five of these six scales led investigators to question whether they were really measuring independent aspects of pathology or just measuring one dimension in a number of ways. Therefore they ran principle components analysis on these five scales in the community and in the patient and prisoner samples. Both analyses indicated only one substantial factor with the patient and prisoner analysis suggesting a possible second factor for sadness.

On the basis of these findings the developers of the SIS recommend the use of a single undifferentiated symptom scale which is the sum of the five highly intercorrelated scales as the best measure of psychopathology. The other reliable scale, not included in the undifferentiated scale, is an Alcoholism scale.

The developers of the SIS believe that this scale measures the construct of demoralization as formulated by Jerome Frank (Dohrenwend, B. P. et al., Note 3). They quote Frank's description "... a person becomes demoralized when he finds that he cannot meet the demands placed on him by the environment, and cannot extricate himself from his predicament" (Dohrenwend, B. P. et al., Note 3, p. 30). The sources of the predicament include environment stresses such as wartime experiences, constitutional defects, learned

incapacity, existential despair, physical illness and "crippling [psychiatric] symptoms" (Dohrenwend, B. P. et al., Note 3, p. 30). They think that this concept fits the characteristics of this and other screening scales better than similar concepts such as caseness, impairment, dissatisfaction with self, anxiety, depressed mood, dysthymic states, helplessness and hopelessness, and pseudo neurosis, all of which were also examined. Although the label demoralization (DEMORL) was selected post-hoc and was only subjected to hypothesis testing with scales from a later instrument it will be used to label the undifferentiated scale in this study because it is suggestive of the construct that this scale probably measures.

The SIS also contains three role impairment scales: Job Stability, relevant only to those in the job market, Marriage, relevant only to those currently married, and Housework, relevant only to women.

PSS

The PSS has a longer history than the SIS. Consequently there are 54 scales developed by three different method available for use with this instrument. However, only seven of the fifty-four scales had alphas above .50 for both the community and patient-prisoner samples and alphas below .50 in no more than one ethnic, class, or sex group in either or both samples. Of these seven scales only four are both

relevant to case identification and have non-shared items[1]. These four scales are Delusions and Hallucinations (21 items), Suicidal Tendencies (6 items), Depression-Anxiety (39 items), and Alcoholism (23 items) (Dohrenwend, B. P., Yager, Egri, & Mendelsohn, 1978). However the Alcoholism scale is problematic in that it had an alpha under .50 in one community subgroup and all its alphas are probably inflated because most items are automatically scored false if a key item, 'alcohol may have been a problem during the past month', is false.

Only one PSS role scale, marital role (Marirole), was found to be reliable by the above criteria.

Of the four PSS symptom scales, the one most comparable to the SIS DEMORL scale is the Depression-Anxiety (DeprAnx) scale. The only PSS role scale available for comparison to any of the three SIS role scales is Marriage. Therefore the two Marrige scales will be compared.

There are Alcoholism scales on the two instruments. However, a comparison of the two would be of dubious value because of the questionable reliability of the PSS scale. Although there is no comparable scale on the SIS the results for the PSS DeluHalu scale will be compared with the results for the DeprAnx scale to see whether the nature of the

---

[1]One discarded scale measured interview belligerence and negativism. Another was a less reliable Alcoholism scale, and the other was a depression-suicide scale that shared items with two of the kept scales.

relationship between symptoms, life events, and psychiatric judgements varies with type of pathology.

Life Event Measures

PERI Study

The weights for the life events reported in the third section of the interview were derived from a subsequent study using an instrument based on both the SIS and PSS called the Psychiatric Epidemiology Research Interview (PERI). As part of the PERI study a stratified sample of 124 judges gave magnitude estimations for a list of 101 events excluding Marriage, the modulus. These judges were stratified on ethnicity and class. Ethnicity was defined in terms of three groups; Blacks, Puerto Ricans, and nonPuerto Rican whites. Class was defined in terms of educational level of head of household with three groups; less than high school graduate, high school graduate but less than college graduate, and college graduate or better.

A stratified sample was necessary in order to see whether different groups gave different ratings for these events. A number of events were shown to have ratings that differed with the status group membership of the rater (Dohrenwend, B. S., Krasnoff, Askenasy, & Dohrenwend, B. P., 1978; Krasnoff, Askenasy, & Dohrenwend, B. P., Note 4).

We will test the efficacy of using weights that vary with status group membership as opposed to weights that are

constant. Two total stress scores will be computed, one with and, one without differential weighting according to status group membership, and the question of which method gives better estimates of stress will be tested.

Although the event list in the rating study was designed to cover events reported in the SIS-PSS study, a number of events were described with wording that is different from the wording used to classify the events reported in the SIS-PSS study. The assignment of weights to differently worded events was decided by consensus of two judges. Appendix A contains a list of the events assigned weights in this manner. This list contains the full wording of the events in both the PERI and SIS-PSS studies.

In addition to computing total stress scores both with and without differential weights for status group, totals will be computed excluding events likely to be a result of a persons psychopathology. A comparison of the results using these latter totals with the results using totals containing all events will provide a check on the effect of including events of questionable etiological significance in stress totals. Similarly total stress scores using only negatively evaluated events will be computed. The results using the totals for negative events will be compared to the results using totals containing both positive and negative events and the question of which totals provide a better adjustment for raw symptom scores will be examined.

Procedure

Relationship of variables in hypotheses

Figure 1 shows the relations of the variables in the hypotheses without consideration of other factors that may need to be controlled. The arrow from life events to symptom scales represents the association posited by hypothesis (1), 'symptom scores increase with life event stress scores.' The arrow from symptom scales to psychiatric judgement represent the association posited by hypothesis (2), 'psychiatric judgements of the probability of being a case increase with symptom scores.'

The arrow from life events to psychiatric judgements represents the increase in association of symptom scales with psychiatric judgement due to adjustment for life events posited by hypothesis (3), 'symptom scores that are adjusted for life event stress are associated more strongly with psychiatric judgements than unadjusted symptom scores.'

Hypothesis (4), involves a comparison of the size of the arrow from life events to psychiatric judgements obtained with different symptom scales.

The size of life event adjustments

If hypothesis three is confirmed, a means for evaluating the importance and meaning of the improvement in the prediction of psychiatric judgements will be necessary. Several measures of agreement will be examined in an effort to elucidate how life event information is used in generating

psychiatric judgements. Furthermore a comparison of the magnitude of the adjustment for life events to that expected on an "actuarial" basis will be examined.

The term "actuarial" is used in the sense that Meehl uses it in Clinical vs. Statistical Prediction (1959). According to this usage the term actuarial does not imply scientific validity but simply refers to the fact that the the adjustment for life events should be based solely on the observed association of life events with symptoms.

Comparing the actual adjustments to actuarial ones should indicate whether the adjustment for life event stress is greater or less than that dictated by the observed association of between the two variables. If the actual adjustments are uniformly greater or less than the actuarial ones other factors besides the observed association would be indicated in the judgement process. If the differences between the actual and actuarial adjustments show a great deal of variability the accuracy of the actual adjustments should be questioned.

Factors to be Controlled

The central question is whether adjusting symptom scores for stressful life events increases the association of symptom scores with psychiatric judgements about the likelihood of psychopathology. Since the data available on this question come from uncontrolled real-life sources, the

possible influence of other factors on the relationships among life events, symptom scales and psychiatric judgements must be examined. The effects of these other factors must be either ruled out or, if they are found to be significant, controlled, before the relationships of interest can be tested.

Figure 2 is an elaboration of Figure 1, showing the relation of factors to be controlled to the variables directly involved in the hypotheses. The factors to be controlled are listed under the heading Person Characteristics and represent different status group indicators associated with persons in this study.

The status groups defined by these person characteristics have been shown to differ in response style (Blumenthal, 1967; Fabrega, Rubel, & Wallace, 1967; Holingshead & Redlich, 1958; Phillips & Clancy, 1972). Therefore one check on the presence of differences in response style would be to see if the relations of life events, syptom scales, and psychiatric judgements are constant or vary with these status group indicators. Such a check would detect differences in response style per se as well as other differences in reporting behavior that may effect life events, symptom scales, and psychiatric judgements.

In addition to such differences in reporting there may be real differences in the number and type of life events and symptom scale items experienced by individuals in the

status groups defined by these person characteristics. A further complication is that psychiatrists may try to adjust their judgements of caseness for their subjective estimates of these differences in response style and circumstances. The psychiatrists may adjust either accurately or inaccurately, and the accuracy may vary for the different status groups.

These differences in reporting behavior, real circumstances, and psychiatrists' knowledge may either confound each other or balance out each other. Although it is not feasible to try to separately test and control for all these influences as the first step in the analysis, it is quite possible to test for the net effect of all these factors by seeing if the relation of the variables of interest is constant in the different status groups. If there are such differences the possible reasons for them can be investigated in terms of the individual biases cited above.

Supplemental Rating Study

Central Figure

A number of subjects reported events such as "child enters Armed Services," "spouse stopped working," and "serious injury to loved one (not spouse)." These events have a person other than the subject as the "central figure." All events in the PERI list are non-specific with respect to central figure. It would be a questionable

practice to assign the same PERI weight to an event regardless of central figure without empirical support because these events are clearly different. It also is undesirable to leave these events out of the life event total scores because reducing the number of scoreable events would reduce the reliability of the total scores and increase the number of subjects who had a threshold score of zero for their life event total.

Method

In order to see how ratings elicited under instructions that are specific with respect to central figure might differ from those elicited under the PERI instructions a supplemental rating study was run. The same subjects were given both the original PERI instructions and instructions specifying central figure. The rationale was that if the ratings produced with the PERI instructions were comparable to those in the original PERI study the information in the supplemental study could be used to estimate weights for events with different central figures. Specifically, the ratio of the ratings under the central figure instructions to those under the PERI instructions in the supplemental study could be used as "scaling factors." These scaling factors would then be used as multipliers for the original PERI weights to generate comparable weights for events with central figures other than the respondent.

Recency.

The decision to run a supplemental rating study provided an opportunity to address another concern. The standard method of computing life event totals is to assign the same weights to all events occurring in a given period, generally spanning one or two years prior to the questionnaire date. Our model of psychiatric judgement and some research on life events (Horowitz, Schaefer, & Cooney,1974) suggests that there is a dimunition of effect with time that may be appreciable over a period of a year or two.

Therefore the supplemental rating study included instructions that were specific with respect to recency of occurrence. The feasibility of scaling factors for recency was also examined.

Subjects.

The judges in the supplemental rating study were forty-three college students, ages 18 to 45. These judges came from classes chosen to provide proportions of Blacks and Puerto Ricans comparable to the first rating study.

Procedure.

The supplemental rating study involved the administration of an event list composed of a subset of fifteen events plus the modulus, Marriage, from the PERI study. These fifteen events were chosen to cover events commonly reported with persons other than the respondent as the central figure

and to cover a wide range of event types (e.g. work, school, marriage, illness, injury, and career change). Appendix B contains the rating instructions and a sample questionnaire form, which lists the fifteen events.

The subjects were first instructed to rate the fifteen events according to the instructions in the PERI study. Then they were told to rate the events for the "amount of change in most people's own lives when the event happened to their husband or wife and 500 points still is the amount of change for most people getting married themselves." This process was repeated for "their child" and "loved one other than (their) husband, wife, or child, such as a relative, friend or lover."

The respondents were then told to rate these events for "the amount of change these things cause in most peoples lives now if they occurred more than twelve months ago," "between six and twelve months ago," and "sometime less than six months ago." Finally the subjects were asked to rate the amount of change involved in the expectation that the event would occur in the next twelve months.

Results

The ratings were analysed to see (1) if the ratings in this sample are comparable to those in the PERI study and (2) whether the same weighting for central figure and recency value could be used for all fifteen events and for

subjects in different status groups or if different weights were needed for particular events or status groups. The means for each event under each of the rating instructions for central figure or recency value were found and then the ratio of each mean to the mean for that event under the PERI instructions was computed. These ratios, or scaling factors, are shown in Table 1. Table 1 also shows the ratings under the PERI instructions in this sample and the original PERI weights. The overall range of the ratios is .62 to 1.47 with the range within a given central figure or recency value being much smaller except in the case of "child". These ranges as well as the minimum and maximum values for each column are shown in Table 2. Part of the reason that the range for Child is so large may be that most of the subjects were below 22 years of age. This fact may inflate the variability of these scalings factors because many of these subjects do not have much experience with children.

On the basis of visual inspection the variability of the scaling factors did not seem to be related to event content. The possibility that the variation in scaling factors might be related to the magnitude of the ratings could not be decided by visual inspection and was tested statistically. Correlations of the individual scaling factors for each central figure and recency value with the original PERI weights were computed and tested for significance. Table 3

shows these correlations and the correlation of the mean rating under the PERI instructions in this study with those in the PERI sample.

The ratings under the PERI instructions in this sample are highly correlated to those in the original study, indicating that the two samples are comparable in terms of their ratings of these events. The scaling factors are not correlated. Since the variation in scaling factors within central figure and recency value do not seem to be related to event content or magnitude, the mean scaling factor for each central figure or recency value have been used for all events except for events with Child as central figure.

The exception for Child events is due to the wide range of scaling factors. Although these factors, like the others, show no pattern with respect to event content or magnitude, they do differ to such an extent that using their mean value would be grossly inconsistent with the data. Therefore the individual scaling factors for each Child event were used.

As a further test of the comparability of these data mean ratings across events within central figure and recency value were compared for different sex and ethnic groups. None of the F or t-tests for these mean differences was significant at the .05 level. Therefore the same scaling factors were used for subjects in different sex and ethnic groups.

Since the scaling factors for Spouse were so close to 1.0 unscaled PERI weights were used for all events with Spouse as central figure. The PERI weights were multiplied by .75 for events with Loved One as central figure.

All event weights were also scaled for recency. Weights for all events occurring more than twelve months ago have been multiplied by .84 and the weights for events occurring between six and twelve months ago have been multiplied by .90 .

RESULTS

Hypothesis 1

The first hypothesis, 'symptom scores increase with life event stress', was tested by examining the correlation of stress measures with each of the symptom and role impairment scales. In light of the controversy in the literature concerning the types of events that should be used to measure life event stress, four types of life event measures were examined: (1) the total for all events, (2) the total for all negative events, (3) the total for all events except those likely to be due to a respondents psychopathology, and (4) the total for all positive events. The first three totals were examined because of the controversy over which of them is the best life event stress measure. The total for all events was a reference measure against which the others was compared. This total was compared to the total for negative events in examing the contention of some authors that negative events are most highly associated with psychopathology. Similarly the total for all events except those likely to be due to psychopathology was compared to the total for all events in light of B. P. Dohrenwend's (1974) warnings about the need to separate events of different etiological statuses. The total for positive events is

included as a check on any finding for the total for 'all events except those likely to be the result of psychopathology' since this total has a high proportion of positive events. Each of these four life event measures have been computed both with and without weights specific to the status group membership of the respondent. Table 4 shows the correlations for the life event measures with symptom and role impairment scale in the cross-section sample.

The most striking feature of Table 4 is that the correlations are small for all scales except DEMORL with the only significant correlations being those between the DEMORL scale and the stress totals for all events and for all negative events. The correlations for DELUHALU and the PSS Marirole scale are all close to zero with the only trend being a tendency to be negative in sign. Therefore the first hypothesis is only confirmed for the DEMORL scale and only with the totals for negative and for all events.

The small and predominantly negative correlations for the DELUHALU and PSS Marirole scale are at odds with the literature which suggests that they should be positive and of substantial magnitude. Therefore these essentially negative findings require further examination. Basically there is very little variation in the DELUHALU scale because these symptoms are rarely reported in the sample. Only seven subjects in the community sample had non-zero scores on the DELUHALU scale. Therefore the correlations with

stress measures are close to zero with the sign of these correlations being determined by so few data points that they are unreliable and therefore statistically meaningless.

The situation for the PSS Marirole scale is similar. Although there was a much higher proportion of non-zero scores, the person with the highest score had a very low event total and the one with the next highest score had no events. The relation between the stress totals and Marirole are actually evenly distributed around zero among the other eighty-two married subjects, but all the correlations turn out to be negative because these two individuals were weighted very heavily in the computation of the correlation coefficients.

Another feature of Table 4 should be noted although it does not change the findings for hypothesis one. Despite the small size of most of the correlations there is a rather consistent tendency for the status group adjusted totals to be slightly more associated with all the sypmtom scores than the corresponding non-adjusted totals. This fact is of interest because it suggests that the status group adjusted totals may be more reliable measures of life stress than the corresponding non-status group adjusted totals.

Hypothesis 2

The second hypothesis, 'psychiatric judgements of the probability of being a case increase with symptom scale

scores', was tested by examining the correlation of (Stirling County) caseness ratings with each of the symptom and role impairment scales. However, there is a question as to how to measure these correlations since the caseness ratings have a rather uniform distribution and the item totals for symptom and role scales all have very skewed distributions with a substantial number of subjects having the lowest possible score of zero, a correlation coefficient involving a simple scale total is likely to underestimate the true degree of association. This underestimation is the result of the fact that the maximum possible correlation coefficient will always have an absolute value less than one in cases like this where one variable is highly skewed and the other is not.

One method for dealing with the limitation on the size of the correlation coefficient between variables with distributions that differ in this way is to transform one or both variables to more normally distributed variables (Cohen & Cohen, 1975, p. 59). The most commonly used transformations for variables which are counts of events (as are the symptom and role scales) are logs and square roots (Cohen & Cohen, 1975, p. 252; Tukey, 1977, p. 83). Square roots are used most often with counts of rare events. There is a problem in applying a log transformation to variables that can have a value of zero because the log of zero is undefined. The problem is generally handled by adding a small constant such as 0.1 to each score before taking the log.

The choice of this constant is not trivial because it determines the degree to which differences at the upper end of the scale become reduced relative to equal differences at lower points on the scale. This fact is best shown by example. With a constant of 0.1, a score of zero becomes log(.1) and a score of one becomes log(1.1). Because equal ratios of values become equal intervals on a log scale, the difference between log(.1) and log(1.1) is equal to the difference between log(1.1) and log(12.1). Since log(12.1) corresponds to an original score of twelve, the use of 0.1 as a constant would result in a high correlation with caseness only if the difference in caseness ratings between those having scores of zero and one was equal to the difference in caseness ratings between those having scores of one and twelve.

Compare this to the use of a constant of 1. A score of zero becomes log(1) and a score of one becomes log(2). In this case log(4) is an equal interval further up the log scale. Since log(4) corresponds to an original score of three, the use of this constant is like assuming that the difference in caseness ratings between people with symptom scores of three and one will be equal to the difference in ratings between those having scores of one and zero. As the constant increases in size the differences between scores higher on the scale become discounted less and less, and the log transformation becomes more and more like the original scale.

If one does not want to pick an arbitrary constant, entertaining the use of square root and log transformations results in considering a potentially large number of different measures. As a rational means for selecting transformations that are likely to produce replicable results, a transformation will be used only if it provides a statistically significant increase in the variance accounted for in caseness ratings over and above that accounted for by the original scale. This is the same decision rule used in the polynomial regression approach to the problem of measuring a curvilinear relationship with a measure of linear association. The advantage of using square roots and logs over polynomials is that they are more likely to result in interpretable and meaningful transformations with 'counted' variables such as the symptom and role scales used in this study.

Table 5 shows the correlation of each symptom and role scale with caseness and indicates for each if a log or square root transformation increases the correlation of that scale with caseness. The highest correlation is for DEMORL, and this is the only scale for which transformations significantly increased the correlation. In fact, all log transformations which were tried added significantly to DEMORL (constants of 0.1, 2, 4, 8,... ,24 were tested). Table 5 shows only the correlation for log of DEMORL plus four (D4) because this is the log transformation that produced the

highest zero-order correlation with caseness. Figure 3 shows the joint distribution of caseness and DEMORL along with the predicted values of caseness for DEMORL and D4. In subsequent analyses D4 will be used as the measure of effect of the items in the DEMORL scale, but results using the untransformed total will be reported for comparison.

Although no log transformations added significantly to any of the other scales, an examination of the zero-order correlations of these transformations with caseness sheds light on why log(DEMORL+4) is the best function of DEMORL. On all the symptom scales, the constant that produced the highest correlation with caseness was equal to, or slightly exceeded, the highest score for a person receiving the lowest caseness rating. In effect the constant seems to be a subjective unit, or threshold, for caseness judgements particular to the scale. Specifically, for the DEMORL scale, a subject can have at most four positive items and still be judged well. In addition each doubling of the number of positive answers above this threshold results in a fixed increase in caseness judgements. Therefore hypothesis 2 is confirmed for DEMORL, DeprAnx, and both Marirole scales.

Hypothesis 3

The third hypothesis, 'symptom scores that are adjusted for life event stress are associated more strongly with

psychiatric judgements than unadjusted scores', was tested by examining the increase in the amount of variation in caseness ratings accounted for by the addition of life event stress measures to regression equations predicting caseness from each of the symptom scales. The test of this hypothesis involves questions about the use of transformations for stress measures similar to those used for symptom measures in the test of hypothesis two. Although the stress totals are not 'counted' variables like the symptom and role scales they have similarly skewed distributions. Therefore square root and log transformations of the stress scales were considered as part of this test. However such transformations all resulted in poorer adjustments to symptom scales than the original stress totals. Therefore log transformations have only been applied to the symptom measures and not to stress measures.

The failure of log and square root transformations to improve the adjustment provided by stress measures may be related to the fact that these transformations reduce the size of a given difference in stress scores at a higher point on the scale relative to the same difference at a lower point. This sort of transformation would increase a correlation if there were a ceiling effect, that is, if differences in scores above a certain point on the original scale were associated less strongly with changes in the other variable than differences below the point. An oppo-

site effect is likely to be true for stress measures, that is there may be a threshold effect, in that the psychiatrists may weight differences above a point (considered the limit of normal or insignificant stress) more heavily in discounting symptoms than differences below the point. Therefore transformations that weight differences between high scores more heavily than equal differences between low scores should be considered. The usual squared and cubed terms in a polynomial regression are transformations with precisely this effect of stretching differences at the upper end of the scale relative to those at the lower end.

The findings for hypothesis three are presented in terms of the partial correlation of stress measures with caseness ratings, partialling out symptoms. These partial correlations are significant whenever the increase in the amount of variation in caseness ratings due to the addition of life events to symptoms is significant. Partial correlations are reported because they are more informative than the simple increase in R or R^2 since they show how large the simple increase is relative to the remaining variation in caseness ratings. Moreover, the partial correlations can be interpreted as the correlation of life event stress measures with caseness ratings controlling for symptoms.

Table 6 shows the partial correlation of stress totals with caseness (partialling out symptoms). Only totals that added significantly to the symptom scales in predicting caseness ratings are shown.

Table 6 only contains columns for log(DEMORL+4) and DEMORL because these are the only symptom measures for which any life event stress adjustments were significant. There are no rows for positive event totals or non-status group adjusted negative event totals because these totals never added significantly to any symptom measure in predicting caseness ratings.

In general, one would expect a truely unrelated second variable to be less likely to provide a significant increase in R^2 (or have a significant partial correlation) as the amount of variation accounted for by the first variable increases. This is because there is less variation left in the dependent measure (restriction in range) with which a second variable can correlate. With respect to the results in Table 6, D4 should provide a more conservative test for the contribution of life event stress measures than does DEMORL in that it leaves less residual variation in the caseness ratings withwhich the life event measures can correlate. However the results using D4 are almost identical to those found using the untransformed total, with the only difference being a tendency to find slightly larger effects with D4.

The fact that the findings are at least as firm under the more conservative test strenghtens their credibility. That is, stress measures add to the symptom measures in predicting caseness, not only when using the simpler and

possibly inadequate measure of symptom effects but add as much, if not more, when using a maximal estimate of the effect of symptoms.

Hypothesis 4

A separate test of hypothesis four, 'the improvement posited in hypothesis 3 is greater for symptom measures based directly on respondents' self reports than for symptom measures based on clinical judgement', is not necessary since only one symptom scale, DEMORL, shows the improvement posited in hypothesis 3. These results are consistent with hypothesis four, however, since the statistically significant improvement for this scale is clearly greater than the non-significant results for DeprAnx, the corresponding scale from a standardized interview.

Different life event measures

The confirmation of hypothesis three raised a number of questions. The first of which is whether certain life event measures were better than others as adjustments to symptom information.

Since hypothesis three was confirmed for a number of different life event measures the relative size of the partial correlations presented in Table 6 must be considered. The most important feature with respect to differences between life event measures based on different classes

of events is that the partial correlations show a pattern just the opposite of that for the correlations of stress totals with symptom measures. A review of Table 4 shows that negative event totals were most highly correlated with the SIS scales, followed by the totals for all events, and then the totals for all events except those likely to be due to psychopathology. The totals for all events except those likely to be due to psychopathology has the highest partial correlation with caseness, followed by the totals for all events, and then the totals for all negative events.

There is a tendency for the partial correlations for status group adjusted totals to be slightly higher than those for non-status group totals just as there was with the zero-order correlations with symptom measures. This common feature of generally higher correlations with other variables suggests that status group adjustment provides slightly more reliable measures of life event stress.

The best adjustment for either symptom measure was the squared status group adjusted total for all events except those likely to be due to psychopathology. This total accounts for about seven percent of the remaining variation in caseness ratings with either symptom measure. This is twice the amount accounted for by either the status group adjusted or non-status group adjusted total for all events.

Evaluating the magnitude of life event adjustments

The practical importance of the adjustment afforded by life event measures can be examined by seeing how well predictions made with an equation containing both life event and symptom information match, or model, psychiatric judgements and comparing these results to those for an equation that contains only symptoms. Since the judgements were made on a four point scale with integer values from one to four, the most relevant measure of agreement would be the proportion of cases for which the equation matches the actual rating. Such a measure can be generated in the following way. The first step is to use (1) the regression equation which predicts caseness ratings from a symptom measure alone and (2) the equation which predicts from a symptom and a stress measure together, to generate predicted values. The second step is to assign to each individual predicted caseness ratings which are the whole number value, from 1 to 4, closest to the predicted values from each equation. Then the proportion of agreement with actual caseness ratings can be compared for the predicted ratings generated by the two equations. This process can be performed for either the original four point caseness ratings or for the simpler case of dichotomized caseness ratings in which both the actual rating and predicted ratings contain just two possible values, one for cases and another for non-cases. This simpler case allows one to see how useful symptom and life

event information might be in providing a check on clinical judgement in situations where there is a dichotomous choice such as whether to recommend someone for treatment.

Table 7 shows a number of measures of association between caseness and either symptom measures alone or symptom measures combined with (adjusted for) life event stress measures. There are columns for both D4 and DEMORL so that comparisons can be made between the optimal measure of symptoms and the simplest. Similarly, the non-status group adjusted total for all events (S) and the squared status group adjusted total for all events except those likely to be due to psychopathology (SGX^2) are the stress measures used as adjustments: S represents the simpliest adjustment and SGX^2 is the optimal adjustment in terms of increasing R^2 or maximum partial correlation. The first row show R^2s for the regression equation predicting caseness from the predictors which label each column. The second row shows Sum(D^2), which is the sum of the squared differences between the actual and predicted four point integer caseness ratings (e.g. a quadractic error function). This quantity is the error sum of squares for the predicted ratings after they have been converted to whole numbers on the same four point scale as the actual values for caseness. Row three contains the percent agreement between four point integer predicted caseness ratings and actual ratings. The last row shows the percent agreement between predicted dichotomous ratings and actual dichotomized ratings.

It can be seen that all measures of agreement increase going from D4 to D4 & S and again from D4 & S to D4 & SGX^2, with SGX^2 providing about twice as much of an improvement over D4 as S. This is not the case for DEMORL, DEMORL & S, and DEMORL & SGX^2 however. Despite the fact that the R^2 increases steadily going from DEMORL to DEMORL & S to DEMORL & SGX^2, all other measures of agreement are poorer for DEMORL & SGX^2 than for DEMORL & S. In addition, the percent agreement for four point predicted caseness is actually lower with DEMORL & SGX^2 than that for DEMORL alone.

It would seem that although adding adjustments to a less than optimal measure of the effect of symptoms (on casenesss ratings) does account for significantly more variation in caseness ratings (as measured by R^2), it does so in such a way that it does not add much to the ability to predict in a practical sense. The values for Sum(D^2) are important in this respect. If the percent agreement measures were the only measures available, one could argue that the poorer results for DEMORL & SGX^2 compared DEMORL & S are simply an accidental artifact of using measures of agreement that are important from a practical point of view but rather arbitrary and based on only part of the information about agreement. Specifically, one could imagine that, although the number of exact agreements did not increase, the proportion of near misses increased relative to predictions that were far from the actual caseness value. The size of Sum(D^2) is

informative in this respect because it is based not only on the number of disagreements (as are the percent agreement measures) but also on the the extent of the disagreements. The relatively small reductions in Sum(D^2) with adjustments to DEMORL as compared to the reduction from adjusting D4 suggest that although the adjustments to DEMORL increase one's ability to account for variation in caseness ratings, this increase is not due as much to, assigning more individuals to a correct, or a more correct, caseness category, as to other refinements, such as, predicting closer to the whole number rating value for those who would be assigned to the right category on the basis of symptom information alone or predicting closer to the actual ratings of outliers.

Clinical and actuarial judgement

As suggested earlier, one way of gaining perspective on the meaning of these results is to compare the size of the regression coefficient for stress measures to the value one would expect if the psychiatrists adjusted for life events on a purely 'actuarial' basis. Actuarial prediction is based entirely on observed frequencies and associations between variables. If the psychiatrists were operating as actuaries they should discount symptoms for life events to the degree to which symptoms are associated with those life event measures.

The regression equation relating symptoms to life events shows how an actuary would adjust symptoms for life events. The correlation coefficient in this equation determines (as in any regression equation) how much to increase or decrease the predicted value for changes in the predictor variable. A version of the regression equation that highlights this fact is (Y- Mean of Y) / Standard deviation of Y= Correlation * (X- Mean of X)/ Standard deviation of X + Error. From this equation it can be seen that the correlation coefficient determines (is equal to) the difference in the predicted value, in standard deviation units, that results from a standard deviation change in the predictor variable. For standardized variables the correlation coefficient is equal to the change in the predicted value for each unit change in the predictor variable, and the equation reduces to Z_y= correlation * Z_x. For example if the correlation is .4 the predicted value increases by .4 standard deviations for each standard deviation increase in the predictor. If the correlation is -.3 the prediction would go down .3 standard deviations for each standard deviation increase in the predictor.

Since there are fewer terms in the equations when using standardized variables the expected value will be developed for the regression coefficients for the standardized variables. However, the arguement can be extended to the regression coefficient for the non-standardized variables and would generate the same results.

If we assume that the correlation of life event stress measures with symptom measures for this sample is representative of the psychiatrists' past experience (similar to correlations observed in the past), we should expect that the adjustment to symptoms for life events should equal the correlation of symptoms with life events. That is, the actuarial estimate of the part of a person's symptom score that is due to life event stress is equal to the person's life event score times the correlation of life event stress with symptoms. According to the psychiatrists' model of psychopathology the portion of the symptoms that can be identified as due to life events should not be treated as indicative of pathology and therefore should not be multiplied by the correlation of symptoms with caseness in the prediction of caseness ratings. Since the regression equation predicting caseness from both symptoms and life events is CASENESS= B1 * SYMPTOMS + B2 * LIFE EVENTS, B2 should be equal to -r(life events_symptoms) * r(symptoms_caseness), and B1 should be equal to r(symptoms_caseness), if the psychiatrists are adjusting according to the model of psychopathology and are actuarially accurate.

This rationale for generating an expected value for the beta weight for life events can be thought of in the following way. Supposed an actuary wanted to simulate psychiatric judgements of caseness and did not have the actual judgements but only the symptom scores, S, the life event

measures, LE, correlation of life events with symptoms, r(LE_S), and the correlation of caseness judgements with symptoms, r(C_S). If she or he knew that psychiatrists discounted symptoms for life events, her or his first step would be to generate an adjusted symptom score, S', representing the psychiatrists' estimate of true symptoms. Being an actuary she or he would set S' equal to observed symptoms, S, minus r(LE_S) times life events, LE, that is she or he would use the equation S'=S - r(LE_S)*LE. The actuary could then predict judgements with the equation C=r(C_S)*S'. By substituting the equation for S' into the equation for C she or he would get an equation in terms of observed symptoms and life events: C=r(C_S)*S + r(C_S)* -r(LE_S)*LE. This equation has the same form as the equation developed on the actual caseness judgements and specifies the actuary's expectations for the beta weights in terms of the correlations she or he was given.

Table 8 shows the correlation of each life event stress total and squared total with D4 and DEMORL. Following each correlation is the expected, actuarially determined, value of B_life events for the regression equation predicting caseness from both symptoms and life events. Following each expected value is the actual B_life events.

The ratio of the size of the actual beta weights relative to the corresponding expected values varies with type of event that compose the totals but are about the same for

totals and squared totals composed of events of the same type. The total and squared total for all events have actual betas which are approxiamately equal to the corresponding expected values. The actual betas for negative events and negative events squared are about half the size of the actuarial values, and those for all events except those likely to be due to psychopathology are about twice the expected value. The size of the actual betas for positive event totals and squared totals are all negative while the signs of all the expected values for these measures are positive. The absolute size (magnitude) of the actual betas for positive event measures are all smaller than the corresponding expected value, but the ratio of the actual to expected value is not consistent as with other stress measures. Rather, the magnitude of the actual betas decrease as the magnitude of the expected values increase.

This relationship suggests that the positive event totals measure two distinct components. One may be called good fortune and is inversely related to symptoms. The second is stress which increases symptoms. To the extent that a particular measure based on positive events does measure stress, it functions as other life event measures and the beta for this measure will have a negative sign like that for other life event stress measures. On the other hand the size of the negative zero-order correlation with symptoms reflects the degree to which the positive event

measure also reflects the component of good fortune. Since the adjustment for good fortune should be the opposite of that for stress, larger negative zero-order correlations tend to force the beta weight in a more positive (i.e. less negative) direction. This tendency results in smaller negative beta weights for those measures which have larger negative zero-order correlations with symptoms, such measures being those which reflect good fortune more.

The fact that the totals for all events except those likely to be due to psychopathology have betas twice as large as the expected values suggests either that (1) these totals are weak measures of the underlying factor and the psychiatrists are adjusting accurately or that (2) the psychiatrists overadjust for reported events (perhaps because they think this can correct for underreporting of events). Since there are other reasons for believing that these life event stress measures are less reliable than they should be (e.g. the high proportion of zero scores on all the measures suggests that the enquiry for events could have been more thorough) the first interpretation is preferred although they both could be true to some extent.

Influence of other factors

As stated previously, the question of whether the relationship posited in hypothesis three is the same in different groups of subjects is especially important since the

data come from uncontrolled real life sources. This question was cast in terms of the influence of group membership variables. These variables, or factors, were referred to as person characteristics and have been measured as Age, Mrd, (married vs. other), Ed (high, middle and low), sex, and ethnicity (Black, Puerto Rican, and others).

There are two ways person characteristics can affect the relation between caseness ratings and symptoms and life events. The first is that individuals with different person characteristics may be allowed a different number of (life event adjusted) symptoms before they are judged to be cases. This type of difference may be called a threshold difference. The statistical test for differences in thresholds is similar to the test for life event adjustments. If a difference in threshold for a person characteristic is substantial a variable representing that characteristic will significantly increase the R^2 in the regression equation predicting caseness from symptoms and life events. Of all the person characteristic variables only Age and Mrd add significantly to the equation using DEMORL and SGX^2 as preditors. Only Age adds significantly to the equation predicting from D4 and SGX^2 with Mrd having a large, but not significant F ratio. The beta weights for Age and Mrd are both negative. This indicates that older people are allowed more adjusted symptoms before they are judged cases than are younger people. Similarly individuals who are married are

judged less likely to be a case than are unmarried individuals with the same amount of adjusted symptoms. However, the difference for Mrd is much smaller than that for age and may not be a reliable finding in light of the fact that using a stronger measure of symptoms, D4, results in this difference being statistical insignificant.

The second way person characteristics may affect the relation of caseness to symptoms and life events is by amplifying or reducing the importance of differences in symptoms or life events in determining caseness. The differential weighting of symptoms in different groups defined by person characteristics can be referred to as the interaction of person characteristics with symptoms. The test for interactions involves generating additional variables called cross-product terms.

There is one cross-product term for the interaction of each person characteristic with symptoms. Each cross-product term is the result of multiplying the person characteristic variable by the symptom score. For all the person characteristics except Age the cross-product term contains the value of the symptom measure when a person is in the appropriate group and a zero otherwise. With Age the cross-product term contains the subjects age times the symptom score. These cross-product terms, in effect, allow the beta weight for symptoms to vary with the person characteristic. Therefore if any of these cross-product terms

adds significantly to the R^2 of equations predicting caseness from symptoms, life events and significant person characteristics, then the beta weights for symptoms is significantly different for groups defined by the person characteristic involved in the interaction term. This is precisely what is meant by saying that there is an interaction between a person characteristic and symptoms.

Similarly life events may be weighted differently in different groups. Cross-product terms representing the interaction with each person characteristic can be generated by multiplying the life event score by the person characteristic variable. A significant increase in R^2 upon addition of such a cross-product term would indicate that the beta weight for life events is significantly different in groups defined by the person characteristic involved in that cross-product term.

The only cross-product term that added significantly to either the equation predicting caseness from D4, SGX^2, & Age or the equation predicting from DEMORL, SGX^2, Age, & Mrd was that for the interaction of low education with symptoms (DEMORL) in the latter equation. The beta weight for this term was negative indicating that psychiatrists did not weight high scores on DEMORL as heavily in imputing caseness for low education respondents as for others. The cross-product term for high education and DEMORL was large but not statisticallly significant. This suggests that the tendency

to underweight the pathognomic significance of symptom reports is a direct function of the lack of education of the respondent. However, since neither of these cross-product terms added significantly to the equation predicting from D4, SGX^2, & Age, which uses the stronger measure of symptom impact, and less educated subjects score higher on this scale, it is likely that this apparent relationship between education and symptom weighting is an artifact of using a linear measure of symptom impact and having respondents with lower education score higher on the scale. The only cross-product term that approaches statistical significance for the equation using D4, SGX^2, & Age is that for the interaction of Mrd with symptoms (D4). The large size of this term could simply be an artifact of not including in the equation the previously noted almost significant contribution of Mrd. The truth of this interpretation was verified by adding Mrd to the equation and then checking that the F-test for the cross-product term representing the interaction of Mrd and D4 did in fact drop below one.

Deciding between equations

The final equations starting from D4 and DEMORL are:

$$\text{Caseness} = 1.73(\text{D4}) - .00195(\text{SGX}^2) - .0211(\text{Age}) - .290 \quad (1)$$

and

$$\text{Caseness} = .184(\text{DEMORL}) - .00183(\text{SGX}^2) - .0191(\text{Age}) - .330(\text{Mrd}) - .0470(\text{DEMORL})(\text{Low\_ed}) + 2.60. \quad (2)$$

The coefficients for SGX^2 are those for scaled squared totals, in which each event rating is divided by 100.

These equations represent different pictures of the relation of symptoms and life events to psychiatric judgements of caseness. The first equation suggests that the psychiatrists use a non-linear function of the number of symptoms to generate their caseness ratings and adjust only the threshold number of symptoms for two factors: (1) a non-linear function of the amount of life event stress and (2) the person's age. The particular non-linear function of symptoms used suggests that there is a declining impact for each symptom in raising the caseness ratings as the number of symptoms increases, with the first four symptoms increasing the probability of being a case by about as much as the next eight. The particular nonlinear function of life events used suggests that a given amount of life event stress is used to discount symptoms more when it occur with other stresses. That is, the amount of discounting per unit

of stress increases with the total amount of stress. The adjustment for age suggests that the psychiatrists expect older people to report more symptoms at any given level of functioning. The second equation suggests that each increase in the number of reported symptoms increases the likelihood of being a case by the same amount. The second equation also suggests that psychiatrists adjust the threshold number of symptom for life events and age as in the first equation but that the weighting of symptoms varies with the respondents education.

The R^2s for the two equations are .602 and .582 respectively and provide no basis for choosing between the two. As further information in evaluating the meaning and importance of the difference between these two pictures of the judgement process the ability of each equation to predict on a practical level will be examined as was done for earlier equations.

Table 9 shows that the predictions of caseness ratings using Equation 1 are slightly better than the predictions of Equation 2. The first equation is only marginally better on R^2, Sum(D^2), and percent agreement for dichotomized ratings, but is clearly better on percent agreement for four point integer ratings. Since the actual ratings are on the four point scale, this measure is the most important in deciding how well an equation models the judgement process. Therefore, although the equations are about equally good at

predicting caseness according to a number of measures of agreement (the predicted values have a correlation of .944), the first equation is a better model for psychiatric judgements.

Inter-rater reliability

One way of evaluating the level of agreement between equations and psychiatrists is to compare it to the agreement between the ratings of different psychiatrists. A number of different checks on the reliability of the caseness ratings have been reported for this study (Dohrenwend, B. P., et al., 1971). The best measure, of those reported, for comparison with our data is the agreement between original interview ratings and those made by a reviewing psychiatrist from tape recordings of 35 community sample interviews (Dohrenwend, B. P., Egri, & Mendelsohn, Note 2). Two factors make this measure less than ideal as an estimate of the upper limit of possible agreement measures. Firstly, tape recordings provide less information than the original interview, and secondly, the reviewing psychiatrist received special training by the Stirling County study staff which the other psychiatrists did not receive. However this measure is the closest to inter-judge reliability of those available for these ratings. Futhermore, these two factors may actually counterbalance each other, since the tape recording may reduce agreement and the special training may increase the reliability of the second rater.

Sixty percent of the ratings for the original and review psychiatist were in complete agreement, and 28.6 percent were one step disagreements. The proportion of the remaining two and three step disagreements were not reported. By comparison the four point scale percent agreement rate of 58.1 for the best equation is about as good as that obtainable by another psychiatrist. Similarly, the equation's agreement rate on the dichotomized scale of 86.3 is about equal to the sum of the agreements and one step disagreements found with another psychiatrist. This sum of 88.6 is a high estimate of the psychiatrists' agreement on the dichotomized scale since it is based on the assumption that all one step disagreements are between scores of 1 and 2 or between scores of 3 and 4. These results suggest that our equations are about as reliable as another psychiatrist. This means that, although the predictions using the best equation are far from perfect, they are about as good as one can expect considering the reliability of the criterion against which they were developed. Therefore, Equation 1 should be useful as a check on clinical judgement since it is about as accurate as a second clinician's opinion.

Reliability of similar measures

In order to provide perspective on the agreement results in this study it is desirable to have a single measure of overall degree of agreement on the four point scale that is

comparable to measures available in reliability studies of other psychiatric judgements and ratings scales. The statistics kappa and weighted kappa have been used in a number of studies and reviews of the reliability of psychiatric diagnosis (Helzer, Clayton, Pambakian, Reich, Woodruff, & Reveley, 1977; Spitzer & Fleiss, 1974). Kappa and weighted kappa measure the proportion of agreement over and above that expected by chance given the distribution of scores for each rater. These statistics vary from +1 for perfect agreement, through 0 for chance agreement, to negative values for less than chance agreement. Weighted kappa is appropriate for situations in which the seriousness of disagreement varies with different combinations of ratings. Kappa is a special case appropriate for situations in which all disagreements are treated as equally severe. Since disagreements on the caseness scale can by ranked in terms of seriousnes as one, two, and three steps apart the weighted version kappa will be reported for the equations.

The most important feature of weighted kappa with respect to providing a perspective on the findings in this study is that under certain conditions weighted kappa is equivalent to the intraclass correlation coefficient. This means that weighted kappa can provide a measure of agreement that can be compared to agreement figures for both diagnostic judgements and numerical scales of psychopathology, and be interpreted as a ratio of variances. When the degree of

disagreement for a given combination of ratings is set equal to the squared difference in the two values, weighted kappa is equivalent to the intraclass correlation coefficient. This condition implies that a two step disagreement between predicted and actual ratings will be treated as four times as serious as a one step disagreement, and a three step disagreement will be treated as nine times as serious as a one step disagreement. This particular weighting scheme is in some sense arbitrary, but it is not inconsistent with conventional methods for scaling errors (Fleiss & Cohen, 1973).

The work on inter-rater reliability of caseness ratings for this study does not report sufficient information to compute weighted kappa. However, weighted kappa can be computed for the agreement between equations and psychiatrists. Weighted kappa (using disagreement weights which make it comparable to intraclass correlation coefficients) is .767 for Equation 1 and .686 for Equation 2. Kappas for diagnoses, as normally practiced, vary greatly with diagnostic category and typically average around .40 or .50 (Fleiss, Spitzer, Cohen, & Endicott, 1972). On the other hand, kappas for diagnoses made with structured interviews and explicit, specific criteria typically are somewhat higher having kappas around .60 or .70 (Helzer, et al., 1977). Finally, intraclass correlation coefficients for scales of psychopathology or specific dimensions such as

depression, phobic anxiety, delusions and halucinations, and mania are frequently in the .80s and .90s (Spitzer, Fleiss, & Endicott, 1978). The kappas for the two equations are therefore equivalent to those found in studies of psychiatric diagnosis when diagnosis is based on structured interviews and explicit diagnostic criteria. This is consistent with the previous finding that the equations are about as reliable as another psychiatrist since these ratings are made from structured interviews and are based on written descriptions of the criteria used by the Stirling County staff. Furthermore, the kappa for Equation 1 indicates that only 23 percent of the variance of the caseness ratings is attributable to measurement error and differences between the equation and the psychiatrists.

Cross-validation

The fact that the first equation contains only three variables and predicts slightly better than the other which contains five suggests that the first equation should predict better on a cross-validation sample as well. The community leader sample can be used for such a test although it is not ideal since it was selected differently from the cross-section sample. However, differences between two samples would generally make a cross-validation test more severe. If this is kept in mind, the results may still be useful.

Table 10 shows the agreement results for the leader sample using the two equations developed on the cross-section sample. The percent agreement for both four point and dichotomous ratings for both equations are about the same or slightly better than in the community cross-section sample. This finding is unexpected. The results for R^2, Sum(D^2), and kappa are more consistent with expectations. They show almost as good agreement as in the original sample. These findings are encouraging in that the level of agreement is high on all measures, but the results for the percent agreement measures require explanation.

One possibility is that since the leaders are closer to the psychiatrists in terms of class and values, the psychiatrists feel more comfortable using their habitual judgement processes, and rely less on idiocyncratic information in makings judgements. Such a response set could make the judgements more independent of variables outside the equations and consequently closer to the predictions based on the variables in the equations. This increased dependence of judgements on the variables in the equations may be strong enough to counter the tendency of agreement measures to drop across samples.

Criterion oriented validity

Psychiatric judgements are not a perfect or ultimate criterion for caseness. Psychiatric judgements are not

perfectly reliable and even if they were one could argue that if some individuals who were judged cases did not subsequently act like cases then the psychiatrists were wrong in these instances. However, since this study was not a longitudinal work, psychiatric judgement seemed clearly the best indicator of caseness, and the question of other criteria was not discussed.

It seems at this point however, that some additional information on criterion oriented validity may be gained by examining the predictions of these two equations in the inpatient and outpatient samples. Presummably inpatients and outpatients groups will contain higher proportions of true cases than the community cross-section and leader samples. Furthermore, inpatients can be expected to receive the highest rating, "probably a case," more often than outpatients because the pathology is more severe among hospitalized patients and therefore should be detected more often[2].

Table 11 shows the distribution of actual and predicted four point integer caseness ratings for each sample and the same agreement statistics for the patient samples as were presented earlier for the leader sample. The hospitalized sample does not have entries for R^2 or kappa because all

---

[2]This argument assumes that there are equal proportions of those who are actually recovered but still in treatment in the two patient groups.

inpatients received the highest rating, "probably a case," which makes kappa meaningless and the computation of a correlation coefficient impossible. To an unknown extent the fact that all inpatients received the highest caseness rating is a result of being seen in a psychiatric hospital. The implications of this fact will have to be considered in interpreting the validity results.

In the cross-section, leader, and outpatient samples the distribution of predictions for the first equation is closer to that of the actual values than the distribution for the second equation. In the hospitalized sample the second equation predicts closer to the actual ratings simply because it produces more predictions of 4, "probably a case". However it is unknown whether the psychiatrists would rate all inpatients so extremely if they were seen outside the hospital as were the other samples. Part of the reason the first equation's predictions are distributed more closely to the actual rating than those for the second equation in all samples except inpatients is that there are more extreme scores (ones and fours) predicted by the first equation. This is interesting because the second equation is better in the hospitalized sample because it predicts more scores of 4. This result is in part due to the fact that the second equation contains a term for Mrd and the hospitalized sample contains a smaller proportion of married subjects. Therefore, it would seem that the one instance in

which the second equation predicts better is an artifact of a combination of sample differences and judgement conditions. The rest of the evidence, especially the great drop in kappa for the Equation 2, suggests the first equation is superior in terms of criterion validity.

A check on the construct of caseness

The psychiatrists in this study also made judgements of symptom impairment as in the Midtown study. These Midtown ratings represent a construct different from caseness. The Midtown ratings are described as ratings of symptom severity and impairment and should be solely a function of the psychiatrist's estimate of total symptomatology.

It is possible that the psychiatrists adjusted for life event measures because they were using a construct of caseness or because this adjustment was simply a part of how they perceived total symptomatology. These possibilities can be tested by regressing the Midtown ratings on symptoms and life events and comparing the results to those for caseness. If the adjustment for life events is only a part of how psychiatrists estimate total symtoms, the same relationship found for caseness ratings with symptoms and life events should be found for the Midtown ratings of symptom impairment. If we are observing an adjustment process that is essentially connected to the psychiatrists' model of psychopathology, or caseness, then life events should not be used to discount the effect of symptoms on the Midtown ratings.

The use of square root and log transformations for the symptom and role impairment scales was considered as it was for caseness ratings. Table 12 shows the correlation of the Midtown ratings with each symptom or role scale total and any transformation of those totals that significantly increased R^2. These correlations are almost identical to those in Table 5 for the correlations with caseness ratings. log(DEMORL+5) was the log function with the highest correlation with Midtown ratings, although a number of others added significantly to DEMORL in predicting Midtown ratings. The interpretation of the constant is the same, in that five was the highest DEMORL score for subjects with the lowest Midtown rating. (The correlation for D4 is almost identical and there was only one person with DEMORL=5 and Midtown=0).

The only scale for which life event measures improved the prediction of Midtown ratings was the SIS Marirole scale. The totals for negative events, both status group adjusted and non-adjusted, as well as squared and cubed totals for these events added significantly to the prediction of Midtown ratings when predicting from SIS Marirole. Also the squared status group adjusted total for all events added siginficantly to the amount of variance explained by Marirole. The largest improvement among both totals and squared totals was obtained using the non-status group adjusted total for negative events, with the squared total providing the largest overall improvement in R^2. The strik-

ing feature in this finding is that all these event measures had positive beta weights, meaning that higher life event scores result in higher Midtown ratings when these measures are added to the equation predicting from Marirole. This relationship is the opposite of that found when life event measures were added to D4 or DEMORL in predicting caseness ratings. Instead of discounting high Marirole scores for high life events the psychiatrists treated high scores on both measures as indicating high symptom impairment.

These findings demonstrate that the discounting of symptoms for life events is not simply a reflection of how psychiatrists estimate total symptoms. The fact that negative event totals, and to a lesser extent totals for all events, actually add to Marirole scores in predicting Midtown ratings suggests that the Marirole scale and these event totals are treated as weak but independent indicators of total symptoms by the psychiatrists.

As a check that this finding is reliable a parallel analysis was performed on the leader sample. An analysis of the hospitalized sample is not feasible because there are only six married subjects in this sample. An analysis of the clinic sample would not be comparible to the cross section analysis because the correlation of the Marirole scale with the Midtown ratings is not statistically significant in the clinic sample.

In the leader sample the beta weights for negative event totals and squared totals are essentially equal to zero. The beta weights for the squared status group adjusted total for all events is negative with an F-ratio equal to one. The largest beta weights are for squared positive event totals and they are negative but not statistically significant at the .05 level. These results indicate that the additive relationship for Marirole and event measures is not reliable across samples.

Beta weights are zero for negative event totals, become negative for all events, and then become even more negative for positive events. This trend suggests that, although the psychiatrists treat Marirole as a weak measure of symptoms in the leader sample, as they do in the community cross-section sample, they use life event information differently. Specifically, they use life events only as indicators of good fortune and disregard any other of information in these measures. The psychiatrist use positive events, which clearly indicate good fortune, to discount Marirole scores. They do not use negative events to discount scores because these event clearly do not measure good fortune. Lastly, the totals for all events seem to be used to discount scores slightly, and this is understandable since they measure a number of factors, only one of which is good fortune.

The failure to replicate the additive relation between Marirole and life event measures in the leader sample may be

related to the higher level of social functioning in the leader sample. It may be that the weak indication of pathology implicit in high life event scores is simply outweighed by the evidence of success in members of the leader sample.

An overview

The major finding in this study is that psychiatrists discount the pathognomic significance of symptom reports for information they gathered during the interview and that this information is related to a number of measures based on the stressful life events. This relationship was demonstrated only for an objectively scored symptom scale. The life event measures showing the strongest relation to the symptom adjustment process were the squared totals for all events except those likely to be due to psychopathology. It was further demonstrated that this discounting for life event measures is particular to caseness ratings in that it does not occur in the prediction of a related but different judgement, that of symptom impairment.

These results are, on the whole, consistent with the hypotheses proposed. However, these findings and other details of the results reported suggest a number of interesting things about the judgement process which are only tangentially addressed by the hypotheses tested.

Judgement in the scoring process

First there is the fact that an objectively scored symptom scale, DEMORL, showed an improved ability to predict caseness ratings when combined with life event measures whereas a corresponding scale, DeprAnx, that allows judgement to influence the scoring process did not. This suggests that the tendency to discount symptoms for context is so strong that symptoms are, in a practical sense, discounted for context during the very recording of symptom reports. The hypotheses allowed for the operation of such a process during the recording of symptom reports but did not suggest that it would be so strong.

The interpretation that the reason life event measures do not add to the ability of the DeprAnx scale to predict caseness is that psychiatrists adjust for context in the scoring of the scale is supported by the fact that the DeprAnx scale is uniformly less correlated with all life event measures than is DEMORL. These smaller correlations can be conceived of as being the result of prior discounting for (removal of correlation with) life event information.

Symptom vs role scales

The fact that neither marital role scale is improved as a predictor of caseness by the adjustment for life events suggests that although these scales measure an aspect of social functioning closely related to symptom measures, they

are too different from symptom measures in content to be treated as strong indicators of symptomatology. Since these measures are only weakly related to caseness it is not surprising that no discounting of their small affect can be demonstrated. The finding that life event measures actually add to the SIS Marirole in predicting impairment lends credence to the interpretation that these role scales are treated, at best, as only circumstantial evidence of symptomatology and are essentially perceived as measures of social functioning.

Another possible reason for not discounting marital role scales for life events is that the psychiatrists may not think that such measures reflect the influence of stressful life events. This possibility finds some support in the extremely low correlations between stress measures and the marital role scales. That is, the psychiatrists would be actuarially correct in not adjusting role scales for factors which show no correlation to them. However, the correlation for negative and total events with the SIS Marirole scale are about the same size as those between DEMORL and the totals for all events except those likely to be due to psychopathology. Since the totals and squared totals for all events except those likely to be due to psychopathology provided the largest adjustment to DEMORL, the size of the correlations between life event measures and Marirole do not, by themselves, provide a full explanation for why

psychiatrists do not discount marital role scores for life event measures.

DISCUSSION

Test conditions

In order to interpret the results it is necessary to understand the conditions under which the hypotheses were tested. In this regard the most important fact is that the psychiatrists made their judgements of caseness and symptom impairment before eliciting the life event reports. The information that psychiatrists had at the time they made their judgements were only spontaneous reports of life events volunterred by subjects and signs of the effects of life events on the subjects. These reports and signs are at best imperfectly correlated with the life event reports upon which our measures were based. Additionally, the number of such signs and reports available to the psychiatrists are fewer than would be the case in an ordinary clinical interview: the use of structured interviews severely restricted any spontaneous tendency of the psychiatrists to probe for the context of a particular symptom or the subjects general background. All these factors tend to reduce the amount of information about life events available to the psychiatrists and therefore make the test of these hypotheses extremely severe.

DEMORL scale

Only one scale shows the relationship to judgements and life events postulated in hypothesis 3 --the relation of symptoms to judgement increases when adjusted for life events. Explanations for the negative results with the other scales have already been presented. However, considering the severity of the test of hypothesis 3 some speculation about the reasons for positive findings with the DEMORL scale is in order.

Frank's conception of demoralization may be described as a psychological equivalent of pain. He says that people come to psychotherapy because of their demoralization, regardless of its cause, and that the common feature of diverse types of therapy is that they attempt to alleviate this condition. Such a statement is like saying that people go to physicians because they are in pain and that doctors simple try to relieve pain by a number of means. Let us ignore for the moment any implications one may be tempted to draw from this statement about the scientific status of psychotherapy and medicine. Instead let us concentrate on the interesting point that despite the reason for a person's demoralization, it is this feeling, state, or condition that actually brings one into therapy.

One possibility for why DEMORL works in a model of psychiatric judgement is that the DEMORL scale may contain the sort of complaints that people seeking psychotherapy

present. One would expect psychiatrists to be especially sensitive and reliable evaluators of the significance of these complaints simply as a function of exposure.

Another possibility is that DEMORL scale represents a concept similar to caseness, except that demoralization is a scale of personal or subjective caseness, showing a persons's own attributions about his or her adequacy of functioning and need for help. However, a person's own attributions about his or her adequacy of functioning are likely to be untempered by situational explanations of the distress: the fact that there are stresses which explain the difficulty does not reduce the suffering. It may be that psychiatrists use the same information as the sufferer to estimate adequacy of functioning but have the perspective to discount these impressions for situational explanations of the distress when making judgements about caseness.

Practical importance

The high level of agreement for the better equation in the clinic sample is important; kappa is as high as that found for ordinary diagnostic judgements, and the dichotomous percent agreement is as good as in the cross section sample. Despite changes in reporting style and meaning of terms presummably associated with the assumption of patient status, the agreement results are almost as good as in the cross section sample. The fact that the high level of

agreement may be due to the higher proportion of clear cut cases in the clinic sample does not detract from the implication that the use of an objective procedure for checking psychiatric judgements could be useful in practice since the population of individuals who are seeking or refered to psychiatric help can be expected to be similar to either the cross section or clinic sample.

In practical terms the improvement in the percent agreement measures for Equation 1 represent a 20 percent reduction in error on the original four point scale when life events and Age are added as compared with predictions from DEMORL or D4 alone and seven percent when only life events are added. There is 33 percent reduction in the errors on the dichotomized scale when life events and Age are added to symptom measures. This same degree of improvement is obtainable with the addition of life events alone. This degree of improvement is impressive. It seems unlikely that the dichotomous rating percent agreement could go above ninety percent using only three predictor variables no matter how good they are simply because there will always be rare indicators of pathology which account for a few high ratings. It is unlikely that reliable scales will contain many of these rare indicators since they tend to reduce reliability unless their error of measurement is very small. The absence from the DEMORL scale of items for some rare symptoms may be the reason that the equations developed

always predicted fewer of the highest caseness values than actually exist in any of the samples.

One can even argue that the equations, especially the better equation, is a bettter indicator of psychopathology than psychiatric judgements. The equations predict more non-cases in the leader sample and at least as many cases in the clinic sample as the psychiatrists. This means that the predictions of the equations correlate more highly with group membership, another indicator of the presence of psychopathology, than do the judgements on which they were developed. In light of the fact that psychiatric judgements are not a perfect indicator of true caseness, one can argue that since the equations show a stronger relation to other indicators of the underlying state, they are better indicator of psychopathology than are psychiatric judgements.

It should be noted however that this argument implies that sample membership is as valid an indicator of psychopathology as psychiatric judgement and that more leaders should be classified as well and more clinic patients as cases than are classified by the psychiatrists. These assumptions are questionable but the argument is worth consideration nonetheless because it reminds one that the criteria of psychiatric judgement is imperfect and that case identification remains a problem of convergent validity.

A model of the judgement process

Since the equations provide a fairly good model of psychiatric judgements the question of whether the model fits the judgement process can be turned around and one can ask whether the the judgement process fits the model as embodied in the equations. In this respect the fact that age is the only person characteristic that adds significantly to the better equation suggests that the psychiatrists are relatively unbiased in applying the model of psychopathology implicit in clinical literature. That is, the psychiatrists do not use different thresholds or weight symptom or life event information differently for different sex, ethnic, or educational groups. The fact that Mrd and an education by symptom interaction do add to the equation using the simpler measure of symptoms points up the need for care in measuring the impact of sypmtoms since it would seem that the psychiatrists are biased if we only had the equation based on the less elaborate measure of symptoms.

Possible biases

It is noteworthy that age has a stronger effect than Mrd. The fact that someone is married is often cited as evidence of a higher level of social functioning. However, age seems more important in the symptom discounting process. One possible explanation for this seeming paradox may be found by considering that the question of the level of

social functioning is generally explicitly raised only when considering the diagnosis of schizophrenia or other psychoses. It may be that DEMORL does not tap enough indicators in the domain of psychotic symptoms to make an indicator of social functioning like Mrd relevant as a basis for discounting scores.

It may also be true that psychiatrists overadjust for age. Additional analysis has shown that the discounting for age is truely linear, that is, psychiatrists not only discount symptoms for respondents who are quite old, in the upper third of the age range, but discount symptoms of those in the middle of the age range relative to the respondents in the lowest third. There seems to be no support in the clinical literature for such a general discounting for age. To the extent that the effect of age is due to this linear discounting of symptoms rather than a discounting for old age and ill health, it should be considered a bias.

Independence of life events and age

The relative size of the adjustments for life events and age are about equal. Futhermore, the size of the partial correlation of life events with caseness after age is entered into the equation is the same as before age is added. Similarly, the size of the partial correlation of age with caseness remains unchanged by entering life events. This means that psychiatrists treat age and information

about life events as independent and equally important reasons for discounting symptom information.

Controversy over different life event measures

The opposite ranking of stress totals representing different sets of events according to their correlation with DEMORL and their partial correlation with caseness (partialling out DEMORL) is revealing with respect to the controversy about which life event measure is best as an adjustment for symptoms. It seems that the advocates of negative events are correct only about the fact that negative events are the most highly correlated with symptom scales. The event totals that provide the best adjustment to symptoms in predicting caseness ratings are those for all events except those likely to be due to psychopathology. It would seem that the psychiatrists use the etiologically correct set of events (in terms of their model or theory of psychopathology) to adjust symptoms despite the lower correlation with symptoms. It is impressive that the event totals for all events except those likely to be due to psychopathology are actually weighted slightly more heavily in the adjustment of symptoms than the totals for all events since the totals for all events are based on many more events and therefore can be expected to have higher correlations with any other variables simply because they are more reliable measures. It should be noted that all correlations for the totals for

positive events are close to zero, indicating that the findings for the totals for all events except those likely to be due to psychopathology can not be attributed to the high proportion of positive events in this set.

Implications

In one sense this entire chapter and part of the previous one are concerned with the implications of the findings of this study. However, previous sections have been devoted only to implications that involved issues that needed substantial explication or interpretation. It seems useful to summarize these implications along with others that do not require much elaboration in this section.

Firstly it is possible to improve upon current methods for the identification of untreated cases by using information about stressful life events in conjunction with symptom information. The increase in agreement with psychiatric judgement provided by the addition of life event information is substantial. Furthermore, the high overall level of agreement of predictions based on both symptom and life event information with caseness judgements suggests that the equations developed represent a useful model of the actual judgement process.

The relative size of the increase in association with casseness judgements afforded by life event measures based on different classes of events shows that psychiatrists

weight events that may themselves be due to psychopathology less heavily in discounting symptoms than events that are clearly not the result of psychopathology. This implies that the psychiatrists make important and rather sophisticated distinctions in applying life event information to the symptom discounting process.

The fact that status group adjusted life event measures are uniformly slightly more highly correlated with both symptom measures and judgements implies that they are more reliable measures than non-status group adjusted life event measures.

The lack of improvement in the association with judgements when life event measures are combined with symptom scales from standardized interviews implies that psychiatrists adjust for life events in the scoring process. The lack of improvement for the addition of life events to role scales suggests that role scale differ from symptom scales to such a degree that psychiatrists treat them only as measures of social functioning.

The higher association of log transformations of symptom scales with psychiatric judgements than that for for simple item totals suggests that the perception of total symptomatology is similar to other judgemental and sensory continua for which a declining sensitivity with increasing magnitude has been demonstrated. The higher association of squared and cubed life event totals with the symptom discounting

process implies a threshold for the discounting of symptoms for life events.

The lack of symptom discounting for life events in the generation of impairment ratings suggests that the discounting of symptoms is an intregal part of the construct of caseness and is not simply a part of the perception of symptoms.

The fact that age is the only group membership variable that clearly influences the relationship of symptoms and life events with caseness judgements suggests the the psychiatrists were relatively unbiased in using symptom and life event information to make their judgements.

Finally, the high level of agreement for the equation using symptoms, life event information, and age with psychiatric judgements of caseness in samples from different populations implies that the model it represents is applied consistently to subjects at different levels of social functioning and both within and outside of treatment settings. This suggests that the applicaton of screeening procedures based on this model or equation, can be useful both as a research tool in community populations and as a source of a second opinion in clinical practice with individuals seeking or referred to treatment.

Appendix A

Table A.1

Event numbers and descriptions from the SIS-PSS study with the weights used in computing life event totals. Frequency of events in all SIS-PSS samples and PERI rating numbers are shown for each event. Event type information indicates special event totals in which the event was included and whether the event was subject to differential weighting in different status groups.

SIS-PSS Event number and description	Weight used	N	PERI #	Event type¤
11 Starts school, training program, etc.	344	37	90	X P
12 Graduates from school, training program, etc.	326	9	72	X P
13 Fails school, training programme, etc.	324	6	94	N
19 Positive educational event other than above (Ph.D. orals, etc.)	358	1	78	X P
21 Move to better neighborhood	389	66	77	X P
22 Moves to worse neighborhood	422	25	63	N
23 Changes neighborhood, but cannot ascertain whether new neighborhood is better or worse	262	14	82	X
24 Move forced because of urban renewal, slum clearance, transportation routing, etc.	422	1	63	N X
25 Move forced because of eviction for other reason	422	1	63	N

26	Moves to continental U.S.	262	7	82		X		
27	Moves from continental U.S.	262	1	82		X		
28	Buys house	470	2	24		X	P	G
29	Travel to foreign lands or throughout U.S.	258	4	57		X	P	
30	Broken love relationship with "steady"	294	24	5	N			G
31	Going steady, incl falling in love	335	4	50		X	P	
32	Engaged	353	10	55		X	P	
33	Married	500	41	M		X	P	
34	Remarried	500	11	M		X	P	
35	Widowed	823	24	96	N			
36	Divorced	609	13	69	N			G
37	Separated	516	38	41	N			
39	Affective tie with spouse broken	568	6	100	N			
40	Respondent adopts child	261	3	40		X	P	
41	Pregnancy	451	7	6		X	P	
42	Birth of first child	608	25	75		X	P	
43	Birth of child other than first	488	23	61		X	P	G
44	Child starts Kindergarten or grammar school	301	3	90*		X	P	
45	Child goes away to college	300	1	13		X	P	
47	Child marries (assume approval if unmentioned)	522	20	m*		X	P	
48	Child marries without approval	522	4	m*	N	X		
49	Respondent becomes grandparent	596	11	75*		X	P	
4x	Respondent adopts child	481	4	35		X	P	

50	Child drops out of school, or training program, etc.	351	2	37*	N X	
51	Child enters Armed Services (or goes off to war)	459	3	1*	X	
52	Child in employment, legal, or financial difficulty	236	2	15*	N X	
54	Child in social (sex related) difficulty (e.g., out of wedlock child)	504	2	19*	N X	G
55	Child having/had marital difficulty or problem	454	1	100*	N X	
59	Child graduates from school or college	336	1	72*	N X	
60	Drug or alcoholic event to spouse	625	2	67	N X	G
61	Serious physical illness to self	625	105	67	N X	G
61	Serious injury to self	556	26	85	N X	
63	Serious injury to loved one (not spouse)	417	12	85*	N X	
64	Serious physical illness of loved one (not spouse)	469	47	67*	N X	G
65	Death of loved one (not spouse)	489	115	73	N X	
66	Death, illness, or injury to "significant other" (e.g., boss)	417	8	85*	N X	
67	Injury to spouse	556	7	85	N X	
68	Illness to spouse	625	24	67	N X	G
69	Drug or alcoholic event to self	625	14	67	N	G
70	Entered Armed Services	373	12	1	X	
71	Started to work on job for the first time	395	58	42	X	G
72	Expanded business	437	18	28	X P	G

73	Promoted or moved to more responsible job	384	60	56		X	P	
74	Changed to more secure job	391	17	10		X	P	
75	Business failed	462	10	2	N			
76	Demoted or changed to less responsible job	379	10	44	N			
77	Laid off	344	26	98	N			
78	Fired	405	18	89	N			
79	Changed job, but quality of change not ascertainable	277	12	87		X		
7x	Stopped working (usually as a sequallae to illness etc.)	387	12	22	N			
81	Spouse started to work on job for the first time	395	1	42		X		G
83	Spouse promoted or moved to more responsible job	384	2	56		X	P	
85	Spouse's business failed	462	1	2	N	X		
87	Spouse laid off	344	2	98	N	X		
88	Spouse fired	405	1	89	N	X		
8x	Spouse stopped working (usually as a sequallae to illness etc.)	387	1	22	N	X		
94	Accident (no deaths)	253	8	71	N	X		
95	Accident involving death(s)	253	1	71	N	X		
96	Natural disasters, Acts of God or the Queen's enemies, etc.	556	1	45	N	X		G
99	Severe financial, legal difficulty (perhaps inprisonment)	518	12	3	N			
x1	Self/spouse conflict (non-terminal, emotional tie not broken)	568	5	100	N			
x4	Self/family conflict	258	7	79	N			

y1	Seemingly insignificant positive event concerning self (e.g., buying a new car)	261	11	36	X	P
y2	Seemingly insignificant positive event concerning others	196	4	36*	X	P
y3	Seemingly insignificant negative event concerning self	261	8	58	N	
y5	Positive event to self	343	3	40	X	P

nN = included in negative event total.
X = included in total for all events except those likely to be due to psychopathology.
P = included in total for positive evnts.
G = event assigned different weights in staus group adjusted totals (see Table A.2 for group specific weights).

* Weight for this event is the product of PERI weight and scaling factor from supplemental rating study.

Table A.2

PERI rating number, event number (used in other work published on the PERI scale), event descriptions, overall ratings, and status group specific ratings.

PERI rating	Event number and description	Overall mean	Group means Status	Levels		
			Class	High	Middle	Low
			Ethnicity	Black	Puerto Rican	Other
			Sex	Male	Female	
90	1 Started school or a training program after not going for a long time	344				
72	3 Graduated from school or training program	326				
94	5 Failed school, training program	324				
37	6 Did not graduate from school, training program when expected	323				
42	7 Started work for the first time	395	Ethn.	495	428	337
10	9 Changed jobs for a better one	391				
87	11 Changed jobs for one that was no better and no worse than the last one	277				
15	12 Had trouble with boss	310				
44	13 Demoted at work	379				
56	16 Promoted	384				
78	17 Had significant success at work	358				
98	19 Laid off	344				
89	20 Fired	405				

28	22	Expanded business or professional practice	437	Class	528	360	405
2	2	Took on a greatly increased work load.	462	Ethn.	251	332	213
22	27	Stopped working, <u>not</u> retirement, foran extended period	387				
55	28	Became engaged	353				
5	29	Engagement was broken	294	Sex	248	333	
m	30	Married	500				
50	31	Started a love affair	335				
100	32	Relations with spouse change for worse, without separation or divorce	568				
41	33	Married couple separated	516				
69	34	Divorce	609	Class	518	707	601
19	37	Marital infidelity	514	Sex	586	449	
96	39	Spouse died	823				
6	40	Became pregnant	451				
75	41	Birth of a first child	608				
61	42	Birth of a second or later child	488	Ethn.	397	507	456
35	47	Adopted a child	481				
13	50	Person moved out of household	300				
79	52	Serious family arguement other than with spouse	258				
73	54	Family member other than spouse or child died	489				
77	55	Moved to a better residence or neighborhood	389				

63	56	Moved to a worse residence or neighborhood	422				
82	57	Moved to a residence or neighborhood no better or worse than the last on	262				
24	59	Built a home, or had a home built	470	Class Ethn.	542 522	447 531	386 406
45	61	Lost a home through fire, flood or other disaster	556	Ethn.	688	538	526
71	64	Accident in which there were no injuries	253				
3	66	Accused of something for which a person could be sent to jail	518				
36	76	Started buying a car, furniture, or other large purchase on the installment plan	261				
58	80	Suffered a financial loss or loss of property not related to work	383				
40	83	Got a substantial increase in wage or salar without a promotion	343				
1	96	Entered the Armed Services	373				
57	98	Took a trip other than a vacation	258				
67	100	Physical illness	625	Ethn.	696	512	685
85	101	Injury	556				

Appendix B

Part I: Rating instructions for supplemental study.

This is not a test of intelligence or ability. There are no right or wrong answers. We want to know your opinion, what you think. We are concerned with just how much people agree or disagree in their opinions. Therefore each person's own opinion is important. We can examine such differences if each of you gives your own opinion even if we don't have your names. That is why I am asking all of you to help me and why I do not want your names on these papers.

After you give me your opinions I will give a talk on the research for which I am collecting this information. It will probably be more interesting than an ordinary talk about such research because you will know what you thought about when you gave me your opinions as well as what I was trying to find out. I would like to thank you in advance for your help and ask you to give me your opinions without talking to each other. HAND OUT SHEETS

Please check off your sex and ethnic group and enter your age in the spaces provided at the top of the sheet.

WAIT

This is a list of certain things that happen in people's lives. Each one of these things changes a person's life, causes a person a certain amount of stress and strain.

What I'd like you to do is to compare each of the things on the list to 'getting married'. For each thing on the list, think to yourself -- is this thing 'less' or 'more' of a change in someone's life than getting married is? Does it take a longer or shorter time for a person to get used to this change than it does to get used to being married?

Look at the first column, the one marked X. In this column 'getting married' is equal to 500 points, as you can see. If you think something else on the list is more of a change than getting married, you put down more than 500 points in column X. You put down as many points as you want, depending on how much more of a change it is than getting married. If you think something is less of a change than getting married, you put down less than 500 points. It's as simply as that.

In all of these things, tell us how much of a change you think these things are for 'most people', not just you.

Before you begin, I'd like to remind you of one thing. It doesn't matter whether you think something is better or worse than getting married. That's 'not' what we're interested in. It's whether you think it is less or more of a 'change' than getting married is.

Now are there any questions?

WAIT

You can start now.

IF R ASKS FOR ANY EVENT: SHOULD I ASSUME THAT THIS WAS THE 'FIRST TIME' (EVENT) HAPPENED. ANSWER: Please assume whatever you think is likely to be the case for most people.

Now what I'd like you to do is to write down on the bottom of the sheet who you thought of as the person experiencing the events and when you thought the events happened.

WAIT

Look at the next column, the one marked S. In this column put a number for the amount of change in most people's own lives when the event happened to their husband or wife and 500 points still is the amount of change for most people in getting married themselves. Do not change any numbers in column X from now on. For example, if you think that most people have less change in their own lives as a result of their husband or wife starting school than they do from getting married put a number less than 500 points showing how much less of a change you think it is. If you think a person's husband or wife's starting school causes more change than getting married put a number larger than

500. Note that you do not need to give a number in this column for getting married, marital difficulty, or become engaged.

WAIT

Now look at the next column, the one marked C. In this column put a number for the amount of change in most people's own lives when the event happened to their child. Remember to compare each event to marriage which is 500 points when it happens to most people themselves. For example if you think the graduation of a person's child causes most people more change than their own marriage put a number larger than 500 showing how much more of a change it is. If you think it causes less change put a number less than 500. You should fill in a number for a person's child getting married in this column comparing it to 500 points for most people getting married themselves. Note also to fill in a number for marital difficulty and become engaged in the spaces provided.

WAIT

Now look at the next column, the one marked L. In this column put a number for the amount of change in most people's lives when the event happened to a loved one other than their husband, wife or child, such as a relative, friend, or lover. Remember to compare each event to marriage

which is 500 points when it happens to most people themselves. Please fill in a number for a person's loved one getting married comparing it to 500 points for most people getting married themselves.

WAIT

Now look at the next column, the one marked 12+. In this column put a number for the amount of change thses things cause in most people's lives now if they occured more than 12 months ago. Compare each event to getting married as 500 points assuming it occured at the time you used in column X. For example let's say you assumed in column X that these things happened one week ago. If you think that dropping out of school more than 12 months ago causes more change in most people's lives now than getting married one week ago put a number larger than 500 showing how much more of a change it is. If you think that dropping out of school more than 12 months ago causes less change than getting married a week ago put a number less than 500. Please fill in a number for getting married in this column also. It should be equal to the amount of change for most people in getting married sometime longer than twelve months ago compared to getting married being 500 if it occured at the time you assumed when you filled in column X.

WAIT

Please fill in the column marked 6-12 with a number equal to the amount of change for the event when it happened sometime between 6 and 12 months ago and knowing that getting married is 500 points if it occured at the time you assumed for column X. Please fill in a number for most people getting married sometime between six and twelve months ago compared to getting married being 500 if it occured at the time you assumed for column X.

WAIT

Please fill in the next column, the one marked 6 with a number equal to the amount of change for the event when it happened sometime less than 6 months ago and knowing that getting married is 500 points if it occured at the time you assumed for column X. Please fill in a number for most people getting married sometime less than six months ago compared to getting married being 500 if it occured at the time you used when you filled in column X.

WAIT

Please fill in the last column, the one marked F, with a number equal to the amount of change for the event when most people expect it to happen to them some time in the next 12 months. Compare each event to getting married as 500 points at the time you assumed for column X. Please fill in a number for getting married in this column also. It should

be equal to the amount of change for expecting that one is getting married sometime in the next twelve months compared to getting married being 500 at whatever time you assumed when you filled in column X. Note that you do not have to gave a number for Serious physical illness or for Serious injury.

SEX: MALE ( ) FEMALE ( )
AGE: _____

ETHNIC GROUP: BLACK ( ) PUERTO RICAN ( )
WHITE NON-PUERTO RICAN ( ) OTHER ( )_____
SPECIFY

	X	S	C	L	12+	6-12	G	F
Getting Married	500	XXX						
1. Started school or training program after not going to school for a long time								
2. Graduated from school or training program								
3. Drop out of (stops) school or training program								
4. Enter armed services								
5. Marital difficulty		XXX						
6. Become engaged		XXX						
7. Serious physical illness								XXX
8. Serious injury								XXX
9. Promotion or move to more responsible job								
10. Changed to more secure job								
11. Business failed								
12. Laid off								
13. Fired								
14. Change of job-quality not known								

Table 1

PERI weights, mean rating under PERI instructions in supplemental rating sample, and ratio of mean rating under other instructions to those under PERI instructions in supplemental rating sample. (N=43)

EVENTS	PERI Weight	PERI Instruction Mean	. . Ratio of ratings under other instructions . . to those obtained using PERI instructions . .					
			Spouse	Child	Loved one	12+	6-12	6-
Marriage	500			1.04	0.87	0.92	1.03	1.13
Start school	334	395	1.12	0.89	0.69	0.82	0.92	1.03
Graduate school	326	406	1.02	1.03	0.81	0.85	0.93	1.00
Drop out of school	286	376	1.11	1.23	0.75	0.93	0.96	1.01
Enter Armed Services	373	506	1.21	1.47	0.91	0.77	0.92	0.98
Martial difficulty	568	567		0.80	0.69	0.84	0.83	0.84
Engaged	353	365		1.10	0.94	0.94	0.92	1.03
Illness	625	690	1.01	1.01	0.82	0.82	0.86	0.94
Injury	556	660	1.04	1.09	0.86	0.86	0.91	0.96
Promoted	384	455	1.00	0.83	0.74	0.87	0.87	0.89
More secure job	391	437	0.97	0.85	0.67	0.88	0.90	0.92
Business failed	462	606	1.02	0.76	0.65	0.73	0.85	0.89
Laid off	344	548	1.04	0.82	0.62	0.82	0.94	0.98
Fired	405	615	1.03	0.81	0.65	0.80	0.89	0.92
Change job quality unknown	277	410	1.06	1.14	0.69	0.72	0.79	0.87
Stopped work	387	508	1.11	0.85	0.69	0.83	0.90	0.94

Table 2

Summary information across all events.

Statistic	PERI Weight	PERI Instruc-tions	Spouse . . .	Child	Loved one 12+ Ratios	6-12	6- . . .
Mean	411	503	1.06	0.98	0.75	0.84	0.90	0.96
Minimum	277	365	0.97	0.76	0.62	0.72	0.79	0.84
Maximun	625	690	1.21	1.47	0.94	0.94	1.03	1.13
Range	348	325	0.24	0.70	0.32	0.22	0.24	0.29

Table 3

Correlations of event means under different instructions in supplemental sample with PERI weights from original study.

	PERI	Spouse	Child	Loved one	12+	6-12	6-
Correlation with original PERI rating	.832	-.331	-.216	.190	.033	-.095	-.171
Number of events	15	13	16	16	16	16	16
P(rho=0)	.001	.135	.211	.240	.452	.364	.263

Table 4

Correlations of life event totals with symptom and role impairment scales in community cross section samples (N=256).

	SIS		PSS		
	DEMORL N=124	Marirole N=77	DeprAnx N=132	DeluHalu N=132	Marirole N=84
All events	.222*	.087	.128	-.048	-.101
All events adjusted for status groups	.237**	.096	.132	-.058	-.108
Negative events	.371**	.118	.114	-.087	-.116
Negative events adjusted for status groups	.378**	.121	.114	-.097	-.119
Positive events	-.152	-.017	.067	.064	-.021
Positive events adjusted for status groups	-.153	-.015	.072	.066	-.017
All events except those likely to be due to psychopathology	.128	.050	.098	-.074	-.102
All events except... adjusted for status groups	.143	.061	.106	-.085	-.104

* $p < .05$
** $p < .01$

Table 5

Correlations of symptom and role impairment scale totals and other functions with caseness in community cross section samples (N=256).

Function of scale total	SIS		PSS		
	DEMORL N=124	Marirole N=77	DeprAnx N=132	DeluHalu N=132	Marirole N=84
Item total	.687	.340	.590	.145	.441
Square root of total*	.705				
Log of (total + 4)*	.725				

*There is an entry in this row only if the funcion provides a significant increase in the correlation of the scale with caseness ratings (over that of item total). Alpha = .05.

Table 6

Statistically significant partial correlations of stress measures with caseness ratings, controlling for symptoms in the community cross section sample (N=124). Alpha = .05.

	Symptom measure	
	log(DEMORL+4)	DEMORL
All events		
Total		
Unadjusted for status group	-.183	-.178
Status group adjusted	-.188	-.183
Squared total		
Unadjusted for status group	-.245	-.244
Status group adjusted	-.252	-.248
Cubed total		
Unadjusted for status group	-.251	-.252
Status group adjusted	-.253	-.249
Negative events		
Total		
Status group adjusted	-.178	N. S.
Squared total		
Status group adjusted	-.189	N. S.
Cubed total		
Status group adjusted	N. S.	N. S.
All events except those likely to be due to psychopathology		
Total		
Unadjusted for status group	-.211	-.223
Status group adjusted	-.215	-.223
Squared total		
Unadjusted for status group	-.260	-.262
Status group adjusted	-.265	-.263
Cubed total		
Unadjusted for status group	-.247	-.238
Status group adjusted	-.246	-.233

Table 7

Measures of agreement between equations and psychiatrists' caseness ratings in community cross section sample (N=124).

	Predictor variables for equations predicting casenss ratings.					
	D4*	D4* & S*	D4* & SGX²*	DEMORL	DEMORL & S*	DEMORL & SGX²*
R^2	.525	.541	.558	.472	.488	.500
Sum(D^2), quadratic error function	91	83	80	97	90	91
Four point scale % agreement	47.6	49.2	51.6	47.6	48.4	46.8
Dichotomous scale % agreement	80.6	83.9	87.1	82.3	85.5	83.1

*D4= log(DEMORL+4). S = Life event stress total for all events, unadjusted for status group. SGX^2 = Squared status group adjusted total for all events except those likely to be due to psychopathology.

Table 8

Correlations of stress measures with symptoms and corresponding expected and actual beta weights for stress measures in equations predicting caseness ratings from both symptoms and life events. Community cross section sample (N=124).

	Equation using log(DEMORL+4) as the symptom measure.			Equation using DEMORL as the symptom measure.		
	Correlation with events	Expected beta	Actual beta	Correlation with events	Expected beta	Actual beta
All events						
Total						
Unadjusted for status group	.206	-.150	-.129	.222	-.152	-.132
Status group adjusted	.221	-.160	-.133	.237	-.163	-.137
Squared total						
Unadjusted for status group	.227	-.124	-.174	.250	-.172	-.183
Status group adjusted	.245	-.132	-.179	.266	-.182	-.187
Negative events						
Total						
Unadjusted for status group	.382	-.276	-.125	.371	-.255	-.100
Status group adjusted	.386	-.280	-.133	.378	-.260	-.109
Squared total						
Unadjusted for status group	.317	-.230	-.128	.312	-.214	-.110
Status group adjusted	.324	-.235	-.138	.322	-.221	-.122
Positive events						
Total						
Unadjusted for status group	-.183	.133	-.033	-.152	.104	-.061
Status group adjusted	-.184	.133	-.030	-.153	.105	-.058
Squared total						
Unadjusted for status group	-.179	.130	-.045	-.140	.096	-.079
Status group adjusted	-.180	.130	-.042	-.140	.096	-.076
All events except those likely to be due to psychopathology						
Total						
Unadjusted for status group	.100	-.072	-.146	.128	-.088	-.164
Status group adjusted	.115	-.083	-.150	.143	-.098	-.166
Squared total						
Unadjusted for status group	.114	-.083	-.181	.136	-.094	-.192
Status group adjusted	.138	-.129	-.184	.257	-.108	-.194

Table 9

Agreement measures between equations and actual caseness ratings for cross section sample. (N=124)

	Equation□ 1	2
R^2	.602	.582
Sum(D^2) quadratic error function	75	78
Four point scale % agreement	58.1	49.2
Dichotomous scale % agreement	86.1	85.5

□Equation 1

Caseness = 1.73(D4) −.00195(SGX^2) −.0211(Age) −.290

Equation 2

Caseness = .184(DEMORL) −.00183(SGX^2) −.0191(Age) −.330(Mrd) −.0470(DEMORL)(Low_ed) + 2.60

Table 10

Agreement measures between equations and actual caseness ratings for leader sample. (N=41)

	Equation¤ 1	2
R^2	.569	.570
Sum(D^2)* quadractic error function	85	88
Four point scale % agreement	61.0	58.5
Dichotomous scale % agreement	87.8	87.8
Kappa	.520	.492

*Prorated to be comparable to value in community cross-section sample:
Sum(D^2)*= Sum(D^2) times 124/41.

¤Equation 1

Caseness = 1.73(D4) -.00195(SGX^2) -.0211(Age) -.290

Equation 2

Caseness = .184(DEMORL) -.00183(SGX^2) -.0191(Age) -.330(Mrd) -.0470(DEMORL)(Low_ed) + 2.60

Table 11

(a)

Agreement measures between equations and actual caseness ratings for clinic sample. (N=56)

	Equation¤ 1	2
R^2	.414	.316
Sum(D^2)* quadratic error function	80	115
Four point scale % agreement	55.4	50.0
Dichotomous scale % agreement	85.7	78.6
Kappa	.426	.263

(b)

Agreement measures between equations and actual caseness ratings for inpatient sample. (N=33)

	Equation¤ 1	2
Sum(D^2)* quadratic error function	147	187
Four point scale % agreement	39.4	42.4
Dichotomous scale % agreement	90.9	78.8

Prorated to be comparable to value in community cross-section sample: Sum(D^2)= Sum(D^2) times 124/56 for clinic sample and =Sum(D^2) times 124/33 for inpatient sample.

(c)

Distribution of caseness values for equations with actual values in parentheses (Total N=254).

	Cross section sample N=124			Leader sample N=41		
Value	Eq. 1¤	Actual	Eq. 2¤	Eq. 1¤	Actual	Eq. 2¤
1	21.0	(33.1)	14.5	56.1	(56.1)	48.8
2	46.8	(32.3)	56.5	36.6	(29.3)	43.9
3	18.5	(12.1)	17.7	7.3	(2.4)	7.3
4	13.7	(22.6)	11.3	0.0	(12.2)	0.0

	Clinic sample N=56			Inpatient sample N=33		
Value	Eq. 1¤	Actual	Eq. 2¤	Eq. 1¤	Actual	Eq. 2¤
1	1.8	(1.8)	0.0	6.1	(0.0)	6.1
2	10.7	(14.3)	16.1	3.0	(0.0)	15.2
3	37.5	(5.3)	39.3	51.5	(0.0)	36.4
4	50.0	(78.6)	44.6	39.4	(100.0)	42.4

¤Equation 1

Caseness = 1.73(D4) -.00195(SGX²) -.0211(Age) -.290

Equation 2

Caseness = .184(DEMORL) -.00183(SGX²) -.0191(Age) -.330(Mrd) -.0470(DEMORL)(Low_ed) + 2.60

Table 12

Correlations of symptom and role impairment scale totals and other functions with Midtown impairment ratings. Community cross section sample (N=124)

	SIS		PSS		
Function of scale total	DEMORL	Marirole	DeprAnx	DeluHalu	Marirole
Item total	.688	.358	.614	.144	.452
Log of (total + 5)*	.710				

*There is an entry in this row only if the funcion provides a significant increase in the correlation of the scale with Midtown ratings (over that of item total). Alpha = .05.

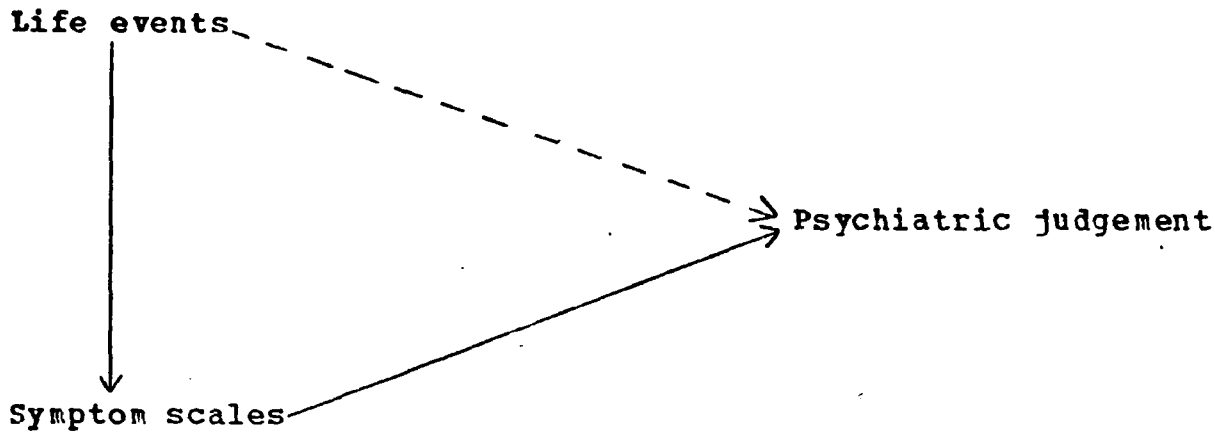

Figure 1: Relation of variables in hypotheses.

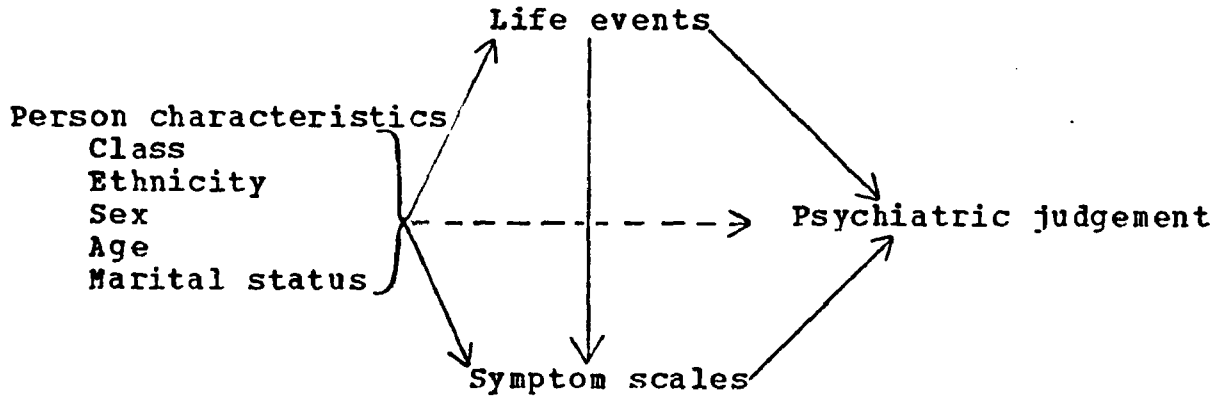

Figure 2: Relation of variables in hypotheses and those to be controlled.

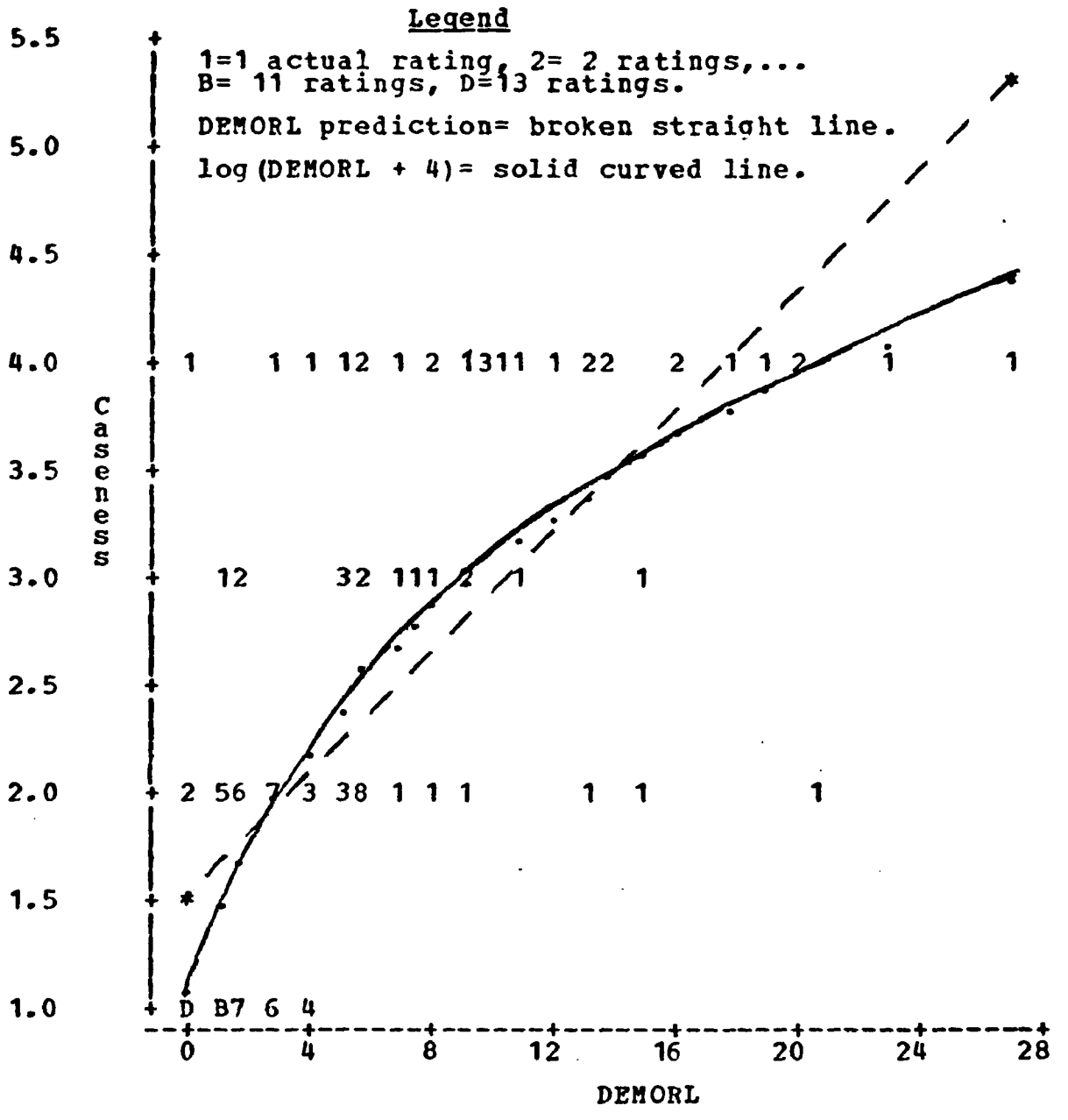

Figure 3: Distribution of actual caseness ratings and predictions using DEMORL and log(DEMORL + 4).

REFERENCE NOTES

1. American Psychiatric Association Task Force on Nomenclature and Statistics. DSM-III Draft: Diagnostic and Statistical Manual of Mental Disorders (Third ed.), 1/15/78 draft, first printing. (Available from American Psychiatric Association, 1700 Eighteen Street N.W., Washington, D. C.)

2. Dohrenwend, B. P., Egri, G., & Mendelsohn, F. S. Psychiatric disorder in general populations: a study of the problem of clinical judgment. Longer version of article with the same name published in American Journal of Psychiatry, 1971, 127, 1304-1312. Available on request from the authors. The authors are with the College of Physicians and Surgeons of Columbia University, 128 Fort Washington Ave., New York, N.Y. 10032, Dr. Dohrenwend is Professor of Social Science, Dr. Egri is Instructor in Psychiatry, and Dr. Mendelsohn is Research Associate in Psychiatry.

3. Dohrenwend, B. P., Oksenberg, L., Shrout, P. E., Dohrenwend, B. S., & Cook, D. What psychiatric screening scales measure in the general population I: Jerome Frank's concept of demoralization. Manuscript submitted for publication, 1978.

4. Krasnoff, L., Askenasy, A. R. , & Dohrenwend, B. P. A Decision Model for evaluating reliability of life event change ratings. Paper presented at the meeting of the Eastern Psychological Association, Boston, April 1977.

5. Star, S. A. The public's ideas about mental illness. Chicago: National Opinion Research Center, University of Chicago, 1955 (mimeographed).

REFERENCES

American Psychiatric Association Committee on Nomenclature and Statistics. DSM-II: Diagnostic and Statistical Manual of Mental Disorders (2nd ed.). Washington, D. C.: American Psychiatric Association, 1968.

Blumenthal, M. D. Sex as a source of heterogeneity in a mental health survey. Journal of Psychiatric Research, 1967, 5, 75-87.

Brown, G. W. Meaning, measurement, and stress of life events. In B. S. Dohrenwend & B. P. Dohrenwend (Eds.), Stressful Life Events: Their Nature and Effects. New York: John Wiley, 1974a.

Brown, G. W. Life-events and the onset of depressive and schizophrenic conditions. In E. K. E. Gunderson & R. H. Rahe (Eds.), Life Stress and Illness. Springfield, Ill.: Charles C Thomas, 1974.

Cohen, J. & Cohen, P. Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences. New York: John Wiley & Sons, 1975.

Dohrenwend, B. P. Problems in defining and sampling the relevant population in stressful life events. In Dohrenwend, B. P. & Dohrewend, B. S. (Eds.), Stressful Life Events: Their Nature and Effects. New York: John Wiley, 1974.

Dohrenwend, B. P., Chin-Shong, E. T. Social status and attitudes towards psychological disorder: The problem of tolerance of deviance. American Sociological Review, 1967, 32, 417-433.

Dohrenwend, B. P., Chin-Shong, E. T., Kagi, G., Mendelsohn, F. S., & Stokes, J. Measures of psychiatric disorder in contrasting class and ethnic groups. In E. H. Hare & J. K. Wing (Eds.), Psychiatric Epidemiology: An International Symposium, London: Oxford University Press, 1970.

Dohrenwend, B. P. & Dohrenwend, B. S. Social Status and Psychological Disorder: A Causal Inquiry. New York: Wiley, 1969.

Dohrenwend, B. S. & Dohrenwend, B. P. (Eds.), Stressful Life Events: Their Nature and Effects. New York: John Wiley, 1974a.

Dohrenwend, B. S. & Dohrenwend, B. P. A Brief Historical Introduction to Research on Stressful Life Events. In B. S. Dohrenwend & B. P. Dohrenwend (Eds.), Stressful Life Events: Their Nature and Effects. New York: John Wiley, 1974b.

Dohrenwend, B. P., Egri, G., & Mendelsohn, F. S. Psychiatric disorder in general populations: a study of the problem of clinical judgment. American Journal of Psychiatry, 1971, 127, 1304-1312.

Dohrenwend, B. S., Krasnoff, L., Askenasy, A. R., & Dohrenwend, Exemplification of a method for scaling life events. Journal of Health and Social Behavior, 1978, 19, 205-299.

Dohrenwend, B. P., Yager, T. J., Egri, G., & Mendelsohn, F. S. The Psychiatric Status Schedule (PSS) as a measure of psychopathology in the general population. Archives of General Psychiatry, 1978, 35, 731-737.

Essen-Moller, E. Discussion (of "Problems and perspectives in psychiatric epidemiology"). In E. H. Hare & J. K. Wing (Eds.), Psychiatric Epidemiology: An International Symposium, London: Oxford University Press, 1970.

Fabrega, H., Rubel, A. J., & Wallace, C. A. Some social and cultural features of working class Mexican psychiatric outpatients. Archives of General Psychiatry, 1967, 16, 704-712.

Fleiss, J. L. & Cohen, J. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. Educational and Psychological Measurement, 1973, 33, 613-619.

Fleiss, J. L., Spitzer, R. L., Cohen, J., & Endicott, J. Three computer diagnosis methods compared. Archives of General Psychiatry, 1972, 27, 643-649.

Gersten, J. C., Langner, T. S., Eisenberg, J. G., & Simcha-Fagan, O. An evaluation of the etiologic role of stressful life-change events in psychologocal disorders. Journal of Health and Social Behavior, 1977, 18, 228-244.

Grinker, R. R., & Spiegel, J, P. Men Under Stress. Philadelphia, Blakiston, 1945.

Hagnell, O. A Prospective Study of the Incidence of Mental Disorder. Stockholm: Svenska Bokforlaget Norstedts-Bonniers, 1966.

Hartmann, H. Ego psychology and the problem of adaptation. In D. Rapaport (Ed.), Organization and Pathology of Thought. New York: Columbia University Press, 1951.

Helzer, J. E., Clayton, P. J., Pambakian R., Reich, T., Woodruff, R. A., & Reveley, M. A. Reliability of psychiatric diagnosis 2: The test/retest reliability of diagnostic calssification. Archives of General Psychiatry, 1977, 34, 136-141.

Hollingshead, A. B., & Redlich, F. C. Social Class and Mental Illness New York: Wiley, 1958.

Holmes,T. H. & Masuda, M. Life change and illness suspectibility. In B. S. Dohrenwend & B. P. Dohrenwend (Eds.), Stressful Life Events: Their Nature and Effects. New York: John Wiley, 1974.

Holmes, T. H. & Rahe. R. H. The Social Readjustment Rating Scale. Journal of Psychosomatic Research, 1967, 11, 213-218.

Horowitz, M. J., Schaefer, C. & Cooney, P. Life event scaling for recency of experience. In E. K. E. Gunderson & R. H. Rahe (Eds.), Life Stress and Illness. Springfield, Ill.: Charles C Thomas, 1974.

Hough, R. L., Fairbank, D. T., & Garcia, A. M. Problems in the ratio measurement of life stress. Journal of Health and Social Behavior, 1976, 17, 70-82.

Komaroff, A. F., Masuda, M., & Holmes, T. H. The Social Readjustment Rating Scale: A comparative study of Negro, Mexican and white Americans. Journal of Psychosomatic Research, 1968, 12, 121-128.

Leighton, D. C., Harding, J. S., Macklin, D. B., MacMillan, A. M., & Leighton, A. H. The Character of Danger. New York: Basic Books, 1963.

Lewis, A. The Samuel Hamilton Lecture: Field studies in mental disorders in Britain. In P. H. Hoch & J. Zubin (Eds.), Comparative Epidemiology of the Mental Disorders: The Proceedings of the American Psychopathological Association, Held in New York City February, 1959. New york: Grune & Stratton, 1961.

Mechanic, D. Problems and prospectives in psychiatric epidemiology. In E. H. Hare & J. K. Wing (Eds.), Psychiatric Epidemiology: An International Symposium, London: Oxford University Press, 1970.

Meehl, P. E. Clinical vs. Statistical Prediction. Minneapolis: University of Minnesota Press, 1959.

Menninger, K. A. A Manual for Psychiatric Case Study. New York: Grune & Stratton, 1952.

Meyer, A. The Life Chart and the obligation of specifying positive data in psychopathological diagnosis. In P. E. Winters (Ed.), The Collected Papers of Adolp Meyer, Vol. III, Medical Teaching. Baltimore: The John Hopkins Press, 1951.

Offer, D., & Sabshin, M. The concept of normality. In S. Arieti (Ed.), American Handbook of Psychiatry (2nd ed.), Vol. 1. New York: Basic Books, 1974.

Paykel, E. S. Recent life events and clinical depression. In E. K. E. Gunderson & R. H. Rahe (Eds.), Life Stress and Illness. Springfield, Ill.: Charles C Thomas, 1974.

Paykel, E. S., Prusoff, B. A., & Uhlenhuth, E. H. Scaling life events. Archives of General Psychiatry, 1971, 25, 237-247.

Peterson, D. R. The Clinical Study of Social Behavior. New York: Appleton-Century-Crofts, 1968.

Plunkett, R. J. & Gordon, J. E. Epidemiology and mental illness. New York; Basic Books, 1960.

Phillips, D. L., & Clancy, K. J. Some effects of "social desirability" in survey studies. American Journal of Sociology, 1972, 77, 921-940

Rashevsky, N. Some Medical Aspects of Mathematical Biology. Springfield, Ill.: Charles C Thomas, 1964.

Sartorious, N. An epidemiological approach to psychiatric disorders. Acta Paedopsychiatr, 1971, 38, 335-345.

Sellin, T. & Wolfgang, M. E. The Measurement of Delinquency. New York: John Wiley, 1964.

Seiler, L. H. The 22-Item scale used in field studies of mental illness. Journal of Health and Social Behavior, 1973, 14, 252-264.

Spitzer, R. L., Endicott, J., Fleiss, J. L., & Cohen, J. The Psychiatric Status Schedule. Archives of General Psychiatry, 1970, 23, 41-55.

Spitzer, R. L. & Fleiss, J. L. A re-analysis of the reliability of psychiatric diagnosis. British Journal of Psychiatry, 1974, 125, 341-347.

Spitzer, R. L., Fleiss, J. L., & Endicott, J. Problems of classification: Reliability and validity. In M. A. Lipton, A. DiMascio, & K. F. Killam (Eds.), Psychopharmacology: A Generation of Progress. New York: Raven Press, 1978.

Srole, L., Langner, T. S., Michael, S. T., Opler, M. K., & Rennie, T. A. C. Mental Health in the Metropolis: The Midtown Study, Volume I. New York: McGraw-Hill Book Company, 1962.

Star, S. A. Psychoneurotic symptoms in the army. In S. A. Stouffer, L. Guttman, E. A. Suchman, P. F. Lazarsfeld, S. A. Star, & J. A. Clausen (Eds.), Studies in Social Psychology in World War II. The American Soldier: Combat and Its Aftermath. Princeton, N. J.: Princeton University Press, 1949.

Stevens, S. S. On the Psychophysical Law. Psychological Review, 1957, 64, 153-181.

Stevens, S. S. A metric for the social consensus. Science, 1966, 151, 530-541.

Stevens, S. S. & Galanter, E. H. Ratio scales and category scales for a dozen perceptual continua. Journal of Experimental Psychology, 1957, 54, 377-411.

Thorne, F. C. Clinical Judgement. Brandon, Vermont: Journal of Clinical Psychology, 1961.

Tousignant, M., Denis, G., & Lachapelle, R. Some considerations concerning the validity and use of the Health Opinion Survey. Journal of Health and Social Behavior, 1974, 15, 241-252.

Tukey, J. W. Exploratory Data Analysis. Reading, Massachusetts: Addison-Wesley Company, 1977.

Wilson, R. S. On behavior pathology. Psychological Bulletin, 1963, 60, 130-146.

Wing, J. K. A standard form of psychiatric present state examination. In E. H. Hare & J. K. Wing (Eds.), Psychiatric Epidemiology: An International Symposium, London: Oxford University Press, 1970.

Wolff, H. G. Wolf, S. G. Jr., & Hare, C. C. (Eds.), Life Stress and Bodily Disease. Baltimore: The Williams and Wilkins Company, 1950.

INDEX