Letters to the Editor

Lifetime Psychotic Symptoms Assessed With the DIS

To the Editor:

The article by Pulver and Carpenter (1983) attempts to test the reliability of a structured instrument, the National Institute of Mental Health Diagnostic Interview Schedule (DIS), in eliciting lifetime psychotic symptoms. The symptoms were first detected approximately 11 years previously through the use of another structured instrument, the Present State Examination (PSE) (Wing, Cooper, and Sartorius 1974), as part of the World Health Organization's International Pilot Study of Schizophrenia (WHO-IPSS) at the Washington, D.C. study center. The authors report that the DIS failed to detect significant psychotic symptoms previously elicited by the PSE, and therefore conclude that the DIS/lay interviewer approach underestimates the lifetime occurrence of psychotic symptoms and diagnoses. Since the DIS is currently in use in large-scale epidemiological surveys in the U.S. (the Epidemiological Catchment Area Project), underreporting of psychotic symptoms could produce substantial error in detecting the true prevalence of disorders such as schizophrenia.

Unfortunately, Pulver and Carpenter's research design leaves unaddressed important factors which could account for discrepancies between the PSE, the Psychiatric Assessment Interview (PAI) (Carpenter et al. 1976), and the DIS symptom reports. The most obvious of these, independent of differences between the DIS and the PSE/PAI, is the fact that the PSE/PAI asks about current symptoms and the DIS asks if a symptom has ever occurred. False-negative information in the retrospective reporting of major psychiatric symptoms, even in repeated interviews with the same, familiar, research psychiatrist over time, has been reported by Barbara Fish (1982) regarding subjects followed since birth for almost 30 years and judged to be at risk for the development of schizophrenia:

To the extent that subjects and their families turned to us in times of crisis, they told of their experiences with all the intensity of the initial impact. We found that in the later retrospective histories, these same events were often forgotten, blurred, or completely distorted to fit a more comfortable self-concept. Symptoms were often revealed at times of stress and then later forgotten or denied during calmer periods. We became aware of the considerable false negative information provided during the regular, standardized interviews. [Fish 1982, pp. 226-227; emphasis added]

The unreliability of self reports of past experiences has also been demonstrated in regard to life changes (Uhlenhuth et al. 1977; Jenkins, Hurst, and Rose 1979), the use of prescription medication (Parry, Baiter, and Cisin 1970-71), and other health-related events such as hospitalizations, auto accidents, and episodes of illness. The general conclusion that can be drawn from such studies is that recall declines over time, with a significant decline detectable after a 6- to 9-month period, even for dramatic and serious experiences. In view of this expected pattern of forgetting past experiences, the overall sensitivity of the DIS to events reported 11 years earlier is quite impressive. Until further studies are made, however, it is misleading to suggest that the sensitivity levels reported for the DIS in this study reflect inadequacies of the lay relative to the clinician interview or inadequacies of highly structured versus more subjective methods of eliciting and evaluating behavior.

Another factor which could account for the study findings is the different wording used in the PSE/PAI and the DIS in the case of
several symptom items. For example, the very low agreement (20 percent) between the DIS and the PSE/PAI ratings for olfactory hallucinations may reflect wording differences in the two instruments. This seems particularly possible considering the much briefer and more general questioning on this item as represented in the PSE/PAI. The low agreement on tactile hallucinations might similarly be influenced by the difference in the DIS and PSE/PAI items. Additionally, the PSE permits psychiatrist-interviewers latitude and judgment in the wording of specific inquiry on these and other symptoms. This creates a larger “net” to elicit symptoms than is found in the DIS.

The wording that would more correctly identify symptoms of schizophrenia in a population is not known. Pulver and Carpenter apparently take the view that the PSE generates the truth. This position is one we would question given the lack of an ultimate standard for determining “true” symptoms or even “true” diagnoses in the field. Questions which cast a broader net and are inclusive of a wider range of experiences such as those in the PSE/PAI are also more easily subject to false-positive errors; but Pulver and Carpenter have chosen to interpret each instance of disagreement between the PSE/PAI and the DIS as false-negative error on the part of the DIS rather than false-positive error obtained by the PSE/PAI.

This last point is closely related to a further problem with the study. Only one aspect of valid diagnostic assessment, sensitivity, is considered by Pulver and Carpenter. A research instrument’s value must also be judged in light of its specificity, particularly if it is to be used in community epidemiologic studies in which the risk of identifying false positives is greater than in a treated population. Since most of Pulver and Carpenter’s sample were cases of schizophrenia, the specificity of PSE/PAI or DIS items could not be examined. One cannot conclude that the DIS would overestimate or underestimate symptoms relative to the PSE without a full examination of both sensitivity and specificity.

Psychiatric nosology has been in a state of flux for the last 10 years and studies like the WHO-IPSS represented pioneering work in development of the more current operational systems—e.g., DSM-III (American Psychiatric Association 1980). Indeed, WHO-IPSS reports pointed to marked diagnostic variation in the evaluation of schizophrenic patients despite uniform training in the use of a standardized, structured interview by research psychiatrists. The difficulty of arriving at diagnostic agreement is illustrated by the wide range of concordance in WHO/IPSS schizophrenia diagnoses derived from PSE, Catego program, and McKeon cluster analyses. The percentage of concordant diagnoses ranged from a high of 60 percent (Cali, Colombia) to a low of 16 percent (Washington, D.C.) (World Health Organization 1973). Spitzer and Fleiss (1974) and Spitzer, Forman, and Nee (1979) have shown that even though operational systems like the Research Diagnostic Criteria (Spitzer, Endicott, and Robins 1978) and DSM-III do significantly better than their predecessors (DSM-II) (American Psychiatric Association 1968) in terms of improving diagnostic reliability, levels of concordance are by no means perfect. Thus, even clinical assessments by fully trained interviewers would be expected to miss a substantial number of “true” schizophrenic cases, if one considers the “kappas” reported in the literature. In view of the above, it is surprising to see that the authors attributed the discrepant findings entirely to a methodological deficiency of the novel lay-administered instrument.

The data presented by Pulver and Carpenter, even when we acknowledge the points raised above, are in our view more supportive of the sensitivity of the DIS in detecting psychotic symptoms which had taken place many years before than the conclusions drawn by the authors would suggest. In spite of the passage of time and the fact that the PSE and DIS differed in specific questions to elicit symptoms, a majority of subjects reported hallucinations and delusions (85 percent of the total sample) in the DIS. And since a DSM-III diagnosis of schizophrenia requires only one hallucination or delusion to have occurred, loss of sensitivity at the level of any single symptom may have negligible effect on the sensitivity of the DIS to estimate lifetime prevalence of the disorder. The overall sensitivity of the DIS is particularly striking in view of the fact that the “standard” against which the DIS was being compared comprised a cumulative total of data obtained at three different time periods by highly trained clinicians. It would be useful to know what the sensitivity of a PSE/PAI clinical interview assessing lifetime symptoms would be when administered at the same 11-year followup time as this study’s DIS administration. One could then distinguish effects of faulty recall from the effects of using different instruments.

References

American Psychiatric Association. DSM-II: Diagnostic and Statistical
The Authors Reply:

Karno et al. (1984) call attention to many reasons why the DIS/lay interviewer approach may produce false-negative assessment of lifetime psychotic features, but then complain that we conclude that the DIS/lay interviewer approach underestimates the magnitude of various psychiatric disorders inaccurately. Data are not yet available to judge the extent to which the false positives for psychosis would counterbalance the false negatives in large-scale epidemiologic studies. However, the false negative problems will seriously flaw research relying on accurate classification of individual cases (e.g., family pedigree studies). [Pulver and Carpenter 1983, pp. 381-382]

Those who believe our concern is valid, as Karno et al. (1984) apparently do, will surely welcome data bearing on the extent, circumstances, and solutions to this issue. We agree that “unfortunately Pulver and Carpenter’s research design leaves unaddressed important factors which could account for discrepancies between the Present State Examination (PSE) (Wing, Cooper, and Sartorius 1974), the Psychiatric Assessment Interview (PAI) (Carpenter et al. 1976), and the DIS symptoms.” However, we would like to point out to the readers that as far as we know this was the first test of the ability of the DIS/lay interviewer approach to ascertain lifetime histories of symptoms where prior assessment during illness episodes could serve as validating criteria. It was, in fact, the absence of such data that prompted our inquiry before applying the DIS method in family pedigree studies. Whether similar