# Limitations of Diagnostic Criteria and Assessment Instruments for Mental Disorders

## *Implications for Research and Policy*

*Darrel A. Regier, MD, MPH; Charles T. Kaelber, MD, DrPH; Donald S. Rae, MA; Mary E. Farmer, MD, MPH; Barbel Knauper, PhD; Ronald C. Kessler, PhD; Grayson S. Norquist, MD, MSPH*

During the past 2 decades, psychiatric epidemiological studies have contributed a rapidly growing body of scientific knowledge on the scope and risk factors associated with mental disorders in communities. Technological advances in diagnostic criteria specificity and community case-identification interview methods, which made such progress feasible, now face new challenges. Standardized methods are needed to reduce apparent discrepancies in prevalence rates between similar population surveys and to differentiate clinically important disorders in need of treatment from less severe syndromes. Reports of some significant differences in mental disorder rates from 2 large community surveys conducted in the United States—the Epidemiologic Catchment Area study and the National Comorbidity Survey—provide the basis for examining the stability of methods in this field. We discuss the health policy implications of discrepant and/or high prevalence rates for determining treatment need in the context of managed care definitions of "medical necessity." *Arch Gen Psychiatry. 1998;55:109-115*

During the past 20 years, the research field of psychiatric epidemiology has had a remarkably productive period due to conceptual and technological advances over the previous generation of studies. These advances began with a conceptual paper in 1970 by Robins and Guze[1] on an approach for establishing the validity of psychiatric diagnoses, an approach that led to the publication of the criteria by Feighner et al[2] in 1972, the research diagnostic criteria,[3] and the *DSM-III*.[4] The availability of a widely accepted diagnostic system that relied on more explicit criteria, was relatively free of etiologic assumptions, and was subject to direct observation and empirical measurement allowed psychiatric epidemiologists to create the measurement technology to obtain prevalence data for specific disorders.[5]

Highly structured research interviews incorporating these criteria were developed that could be administered by lay interviewers to large population groups and scored by computers. The National Institute of Mental Health (NIMH) Diagnostic Interview Schedule (DIS)[6] was initially used as the case-identification instrument in the NIMH Epidemiologic Catchment Area (ECA) study and became a prototype for subsequent epidemiological diagnostic methods. The DIS

*See also pages 119 and 120*

was used as the basis for a diagnostic instrument sponsored by the World Health Organization/National Institutes of Health (WHO/NIH)–the Composite International Diagnostic Interview (CIDI).[7,8] The CIDI initially incorporated the *DSM-III-R*,[9] and then added the *DSM-IV*[10] and the *International Statistical Classification of Diseases, 10th Revision (ICD-10)*[11] criteria more recently (CIDI 2.1).

*From the office of the Associate Director for Epidemiology and Health Policy Research, National Institute of Mental Health (NIMH), National Institutes of Health (NIH), Bethesda, Md (Dr Regier); the Division of Intervention and Services Research, NIMH, NIH (Drs Kaelber and Norquist and Mr Rae), and the Division of Basic and Clinical Neuroscience Research, NIMH, NIH (Dr Farmer), Rockville, Md; the Institute for Social Research, University of Michigan, Ann Arbor (Drs Knauper and Kessler); and the Department of Health Care Policy, Harvard University Medical School, Boston, Mass (Dr Kessler).*

Many international epidemiological surveys have incorporated these advances in producing large-scale prevalence, incidence, and service use data. In addition, substantial information has been collected on the sociodemographic risk factors and correlates of disorders and service use to identify the groups most likely to have disorders and to use services. Over time, however, the modifications in both diagnostic criteria and assessment instruments have revealed their sensitivity to seemingly small changes and their possible limitations in defining the need for mental health services.

Once community-based rates of specific syndromes, defined by explicit research criteria, were shown to be obtainable, relatively small changes in diagnostic criteria (eg, *DSM-III* to *DSM-III-R*) and methods of ascertainment (eg, DIS to the University of Michigan [UM], Ann Arbor, CIDI) have produced substantialy different results. The relatively higher prevalence rates that more recent studies such as the National Comorbidity Survey (NCS) have produced have served to heighten concerns about the clinical (and possible insurance reimbursement) significance of some of the syndromes identified in this study and by the earlier ECA study. Hence, the new challenges to this field will be to standardize assessment methods and to specify the scope of clinically significant disorders that are in need of treatment. The background and current recommendations to address these issues are provided in this discussion.

Based on a 1-year longitudinal study (between 1980 and 1985) of more than 18 000 adults in the community and 2500 in institutions in 5 sites, standardized to the US census population, the ECA reported initial overall prevalence estimates from the single-wave-I interview of all assessed mental and addictive disorders. About 16% (15.7%) of the adult population was found to have 1 or more such disorders in 1 month,[12] with 20% of this population meeting diagnostic criteria during the past year,[13] and almost a third (32%) reporting a history of these disorders in their lifetime.[13] More recent longitudinal data analyses of 2 diagnostic assessments during the 1-year time frame of the study revealed a substantial increase over the single-wave findings in both 1-year and lifetime rates when several prospective assessments were available.[14] Based on this longitudinal analysis, the 1-year estimate increased from 20% to 28% of the population when diagnostic information from both waves of interviews were added together. Asking subjects at 2 points in time about lifetime diagnostic criteria also increased the lifetime prevalence estimates from about 32% to 44% of the population. Although the courses of illness and the lifetime findings from the 2-wave ECA data were presented in scientific meetings,[15,16] investigators at the NIMH attempted to obtain additional information on the clinical significance of such high rates of identified psychopathologic disorders before publishing the 2-wave lifetime rates.

A partial replication of the ECA was requested by the NIMH in the early 1990s and funded as the NCS.[17] This single-wave-interview (cross-sectional) study of a national probability sample of 8098 respondents focused on a younger age group (15-54 years) to address the scientific and public health policy issue of the relationship between comorbid mental and addictive disorders—more specifically, on the level of the co-occurrence of multiple disorders and on the sequence of the beginning and the end of each disorder. A request for applications was announced with the anticipation that the methods were sufficiently standardized that it would be possible to support this activity under an investigator-initiated (R-01) grant rather than under the contract or cooperative agreement mechanism previously used for the ECA, in which a broader scientific participation of the field would be involved. Our confidence in the standardizing of assessment methods was based on more than a decade of DIS instrument refinements that had been undertaken by a panel of international experts under WHO to produce the WHO CIDI. Further modifications in the format of the CIDI were made at the UM where investigators used the new version (the UM-CIDI) to carry out the NCS.

## CONCERNS ABOUT HIGH PREVALENCE ESTIMATES AND THEIR VARIATION

When the first findings of the NCS were released, the major emphasis of the reported data analysis was that 48% of the adult population (age 15-54 years) could be identified a having had a lifetime disorder wit higher overall rates of depression an anxiety disorders than had been re ported previously in the ECA. Als noted were a substantially highe rate of comorbid disorders and tha 29% met diagnostic criteria for a least 1 of the disorders assesse within the past year.[17] Both the sci entific and political implications c these high prevalence rates wer highlighted by the timing of this re lease during the national debate o health care reform. Major polic questions were raised about the nee for mental health services that wer implied by these high rates, alon with concern about possible insu ance cost-benefit consequences Some major media commentator identified such high rates as indi cating a bottomless pit of possibl demand for mental health service

### Published ECA and NCS Comparison Results

To address the issue of apparent dis crepancies between the ECA and NC findings, we have compared previou analyses of the ECA single-wave an 2-wave annual and lifetime prevalenc estimates with those published fro the NCS (**Table 1**). Note that the 29 of the US population sample (aged 1 54 years) in the NCS who met *DSM-III-R* diagnostic criteria in the past ye was comparable to the 28% rate fro the 2-wave ECA study population (ag ≥18 years). The coverage of disorde included in these 2 surveys was n completely concordant because th ECA included somatization disorde obsessive-compulsive disorder, a orexia, and severe cognitive impai ment, and the NCS did not. On th other hand, the NCS included gene alized anxiety disorder and posttrau matic stress disorder, which were i cluded in only a few of the ECA site and were not part of the core data se

## FURTHER ECA AND NCS COMPARATIVE ANALYSES

### Limited Age-Range Comparison Results

Some of the difference in rates of spe cific and total disorders between th

**Table 1. Community Prevalence Rates of Mental and Addictive Disorders in the ECA Studies and the NCS***

| Mental and Addictive Disorders | ECA 1-Wave *DSM-III* (n=20 206) | ECA 2-Wave *DSM-III* (n=15 816)† | NCS 1-Wave *DSM-III-R* (n=8098) |
|---|---|---|---|
| Any 12-mo disorder | 21.8 (0.4) | 28.1 (0.5) | 29.5 (1.0) |
| Any lifetime disorder | 32.7 (0.5) | 43.7 (0.6) | 48.0 (1.1) |

**The Epidemiologic Catchment Area (ECA) data include persons 18 years of age or older; the National Comorbidity Survey (NCS) data include persons aged 15-54 years. The 2 waves of the ECA are explained in the text. The numbers in parentheses are SE.*

*†The actual number of persons reinterviewed at the second wave. Diagnostic data for those not reinterviewed were imputed using a "hot deck" method.*

**Table 2. One-Year and Lifetime Prevalence Rates Limited to Common Age Range of 18-54 Years in the ECA Studies and the NCS***

| Mental and Addictive Disorders | ECA 1-Wave *DSM-III* (n=11 432) | ECA 2-Wave *DSM-III* (n=9045) | NCS 1-Wave *DSM-III-R* (n=7599) |
|---|---|---|---|
| Any 12-mo disorder | 24.1 (0.5) | 29.8 (0.6) | 28.5 (1.0) |
| Any lifetime disorder | 36.1 (0.6) | 46.9 (0.7) | 48.0 (1.1) |
| Any substance abuse or dependence | | | |
| 12 mo | 9.5 (0.4) | 10.5 (0.4) | 11.5 (0.5) |
| Lifetime | 19.9 (0.5) | 24.3 (0.6) | 28.1 (1.0) |
| Alcohol dependence | | | |
| 12 mo | 4.1 (0.3) | 4.4 (0.2) | 7.4 (0.5) |
| Lifetime | 8.6 (0.4) | 11.3 (0.4) | 14.9 (0.7) |
| Drug dependence | | | |
| 12 mo | 2.0 (0.2) | 2.4 (0.2) | 2.9 (0.3) |
| Lifetime | 4.8 (0.3) | 6.4 (0.3) | 7.9 (0.5) |
| Any affective disorder | | | |
| 12 mo | 7.0 (0.3) | 10.1 (0.4) | 11.1 (0.7) |
| Lifetime | 9.3 (0.4) | 14.9 (0.4) | 19.7 (0.7) |
| Major depressive episode | | | |
| 12 mo | 4.2 (0.3) | 6.4 (0.3) | 10.1 (0.7) |
| Lifetime | 7.2 (0.3) | 12.5 (0.4) | 17.3 (0.7) |
| Dysthymia–lifetime | 3.6 (0.2) | 5.5 (0.3) | 6.7 (0.4) |
| Any anxiety disorder | | | |
| 12 mo | 9.9 (0.4) | 11.8 (0.4) | 15.3 (0.7) |
| Lifetime | 14.2 (0.4) | 19.2 (0.5) | 22.8 (0.8) |
| Panic disorder | | | |
| 12 mo | 1.1 (0.1) | 1.5 (0.1) | 2.2 (0.3) |
| Lifetime | 1.9 (0.2) | 2.8 (0.2) | 3.6 (0.3) |
| Social phobia | | | |
| 12 mo | 1.6 (0.1) | 2.1 (0.2) | 7.4 (0.4) |
| Lifetime | 2.5 (0.2) | 3.7 (0.3) | 13.3 (0.7) |

**ECA indicates Epidemiologic Catchment Area; NCS, National Comorbidity Survey. The numbers in parentheses are SE.*

NCS and ECA results were initially thought to be attributed to the younger NCS age group (15-54 years). The NCS omitted persons 55 years and older, who were shown in the ECA to have lower rates of disorder compared with younger subjects. Alternatively, other contributions to the variance could be differences between the nationally representative sample of the NCS and the ECA samples from 5 sites, even though the latter were standardized by age, sex, and ethnic status to the US population. Although the 1-year rates for any disorder obtained from the single-wave NCS interview were roughly equivalent to the rates obtained from the ECA in 2 interviews, they were considerably in excess of those obtained from a single interview with the DIS.

The NIMH received many inquiries to identify the best estimates of prevalence for mental disorders because in particular the ECA and NCS were noted to have substantial discrepancies in their rates of anxiety and depressive disorders. Drawing on published NCS data,[17,18] the next level of analyses limited comparisons to the 18- to 54-year age group for selected disorders. The overall annual and lifetime prevalence estimates varied only slightly, as shown in **Table 2**, but marked discrepancies occurred for 12-month prevalence rates of specific disorders such as major depressive episode for both the single-wave ECA (4.2%) and the 2-wave ECA (6.4%) when compared with the NCS (10.1%) rates.

Similar discrepancies appeared for any 1-year anxiety disorder (1-wave ECA, 9.9%; 2-wave ECA, 11.8%; and NCS, 15.3%) and specifically for social phobia (1-wave ECA, 1.6%; 2-wave ECA, 2.1%; and NCS, 7.4%). Rates for any substance abuse or dependence were comparable, although those for alcohol dependence were substantially higher for the NCS in the 12-month interval (ECA, 4.4%; and NCS, 7.4%).

### Variations in Instrument Construction

Differences in the actual instrument construction were considered next. These instrument differences included the number and placement of "stem" screening questions. Stem questions for the diagnoses of mood and anxiety disorders were placed at the beginning of the NCS instrument. In contrast, comparable stem questions were placed at the beginning of each diagnostic module in the DIS and the WHO version of the CIDI. An additional difference in mood disorders was the inclusion of 3 stem questions in the NCS whereas the DIS included only 1. Although there was only 1 stem question for panic disorder in each instrument, there were differences in wording and placement, as shown in the footnotes to **Table 3**.

**Table 3. Positive Responses to Lifetime Stem Questions in the ECA Study and the NCS, Controlling for Common Demographic Variables***

| | ECA 1-Wave *DSM-III* | | NCS 1-Wave *DSM-III-R* | |
|---|---|---|---|---|
| | Prevalence | % of Stem Pos w/Dx | Prevalence | % of Stem Pos w/Dx |
| Major depressive episode stem questions† | 33.7 (0.7) | ... | 56.8 (1.4) | ... |
| 12-mo diagnosis | 4.5 (0.3) | 13.4 (0.8) | 10.4 (0.7) | 18.3 (0.1) |
| Lifetime diagnosis | 7.7 (0.4) | 22.8 (1.0) | 18.4 (0.7) | 31.7 (1.0) |
| Panic disorder stem questions† | 6.5 (0.3) | ... | 16.4 (0.9) | ... |
| 12-mo diagnosis | 1.1 (0.1) | 16.9 (2.0) | 2.3 (0.4) | 13.8 (2.0) |
| Lifetime diagnosis | 1.9 (0.2) | 29.2 (2.2) | 3.6 (0.4) | 21.9 (2.0) |

*ECA indicates Epidemiologic Catchment Area; NCS, National Comorbidity Survey; Pos, positive; w/Dx, with diagnosis; and ellipses, not applicable. The numbers in parentheses are SE.

†Stem questions: Major Depressive Episode: ECA (question DIS072), "Have you ever had 2 weeks or more during which you felt sad, blue, depressed, or when you lost all interest and pleasure in things that you usually cared about or enjoyed?" NCS (question B4), "In your lifetime, have you ever had 2 weeks or more when nearly every day you felt sad, blue, or depressed?" (question B4a), "Have you ever had 2 weeks or more when nearly every day you felt down in the dumps, low, or gloomy?" (question B5), "Has there ever been 2 weeks or more when you lost interest in most things like work, hobbies, or things you usually liked to do for fun?" Panic Disorder: ECA (question DIS062), "Have you ever had a spell or an attack when all of a sudden you felt frightened, anxious, or very uneasy in situations when most people would not be afraid?" NCS (question B1), "Have you ever in your life had a spell or an attack when all of a sudden you felt frightened, anxious, or very uneasy in situations when most people would not be afraid or anxious?"

The relative prevalence rates of positive responses are shown in Table 3, with the stem questions for major depressive episode resulting in prevalence estimates of 33.7% and 56.8% for the ECA and NCS, respectively. The subsequent 12-month rates of major depressive disorder were 4.5% for the ECA and 10.4% for the NCS, indicating that fairly similar proportions of those who screened positive (13% and 18%, respectively) were later identified as meeting full criteria for the disorder.

Positive responses to the stem question for panic disorder (ECA, 6.5%; and NCS, 16.4%) show more than 2-fold differences in screening positive rates with 12-month rates of subsequent full criteria for panic disorders in 1.1% and 2.3% of the respective populations. The proportions of those who screened positive and were subsequently found to meet 12-month full criteria for the diagnosis were 17% for the ECA and 14% for the NCS. From these screening data, one could hypothesize that once someone is screened as having met the A criteria for major depression or panic disorder in either of these instruments, about the same proportion will meet the additional symptom, duration, and severity criteria to reach diagnostic levels. The obvious question is whether each of the final groups contains subjects with valid clinical diagnoses or if either or both have a high proportion of false-positive responses. Unfortunately, validity data are not available to compare studies by their relative rates of false-positive responses. Recent attempts to validate some diagnoses in the NCS have reduced the estimated lifetime prevalence rates of clinically significant schizophrenia.[19]

## Demographic, Disorder, and Diagnostic Criteria Comparison Results

Because these preliminary calculations showed that age-range differences were unable to explain much of the variance, other sources of variance were considered. Common age and demographic group characteristics were next selected for control purposes to minimize the variance due to these factors. The group for which there was adequate sample size across the 2 studies included the following: noninstitutionalized persons aged 18 to 54 years who identified themselves as white, black, or Hispanic living in urban or suburban areas. This was to control for more substantial urban-rural differences noted in the NCS than in the ECA.

To control for variance in the *DSM-III* and *DSM-III-R* diagnostic criteria, more detailed analyses were done of individual symptom and question differences between the ECA and NCS. Symptom questions were first classified from both instruments according to similarity in wording. Questions were classified as having "virtually identical wording" or "same content, different wording"; "similar idea, different coverage"; "different number of questions"; or "no comparable question." (A list of these comparisons for panic disorder, simple phobia, social phobia, agoraphobia, major depression, and alcohol and drug abuse or dependence is available on request from the first author.) The most comparable possible subset of questions, *DSM-III* disorders, and *DSM-III-R* disorders were examined. Furthermore, to ensure comparable age and gender distributions within the common demographic groups considered, we computed prevalence rates standardized to 1990 US census figures by age (4 groups) and gender by the method of direct adjustment. In this context, it is possible to see, in **Table 4**, an expected pattern of closer approximations of single-wave *DSM-III* prevalence estimates (1-wave ECA, 22.2%; and NCS, 22.8%), a slightly higher ECA rate for 2 waves of data collection (25.9%), and a still higher rate of comparable *DSM-III-R*–defined disorders (28.6%) in the NCS.

An interpretation of the higher *DSM-III-R* rates is that the additional criteria-based questions required by the *DSM-III-R* present more opportunities than in the *DSM-III* to qualify for some disorders such as social phobia. In this disorder, there is no increase in the severity threshold to counteract an expected increase in rate proportional to the more inclusive set of criteria. A comparison of

**Table 4. One-Year Prevalence Rates in the ECA Studies and the NCS, Controlling for Common Demographic Variables and *DSM-III* Diagnostic Criteria Questions and Standardized by Age and Sex Groups to the 1990 US Census***

| Mental and Addictive Disorders | ECA | | NCS | |
|---|---|---|---|---|
| | 1-Wave *DSM-III* (n=8553) | 2-Wave *DSM-III* (n=6798) | 1-Wave *DSM-III* (n=5737) | 1-Wave *DSM-III-R* (n=5737) |
| Any disorder† | 22.2 | 25.9 | 22.8 | 28.6 |
| Any substance dependence | 6.2 | 6.9 | 10.0 | 9.7 |
| Alcohol dependence | 4.6 | 5.0 | 8.3 | 8.1 |
| Drug dependence | 2.1 | 2.4 | 3.3 | 3.2 |
| Schizophrenia/schizophreniform‡ | 1.5 | 1.5 | 2.8 | 2.8 |
| Any affective disorder | 7.9 | 10.9 | 11.2 | 11.9 |
| Manic episode | 0.9 | 1.0 | 1.1 | 1.1 |
| Major depressive episode | 4.7 | 6.8 | 10.6 | 10.7 |
| Dysthymia | 4.1§ | 6.2§ | 1.3 | 2.7 |
| Any anxiety disorder | 12.1 | 13.9 | 7.2 | 15.6 |
| Panic | 1.3 | 1.7 | 1.4 | 2.1 |
| Any phobia | 11.5 | 13.1 | 6.6 | 14.9 |
| Agoraphobia | 5.0 | 5.7 | 1.9 | 3.1 |
| Social phobia | 2.4 | 2.5 | 3.2 | 8.1 |
| Simple phobia | 9.5 | 9.9 | 3.4 | 8.5 |

**ECA indicates Epidemiologic Catchment Area; NCS, National Comorbidity Survey.*

*†Excludes substance abuse without dependence, obsessive-compulsive disorder, antisocial personality disorder, cognitive impairment, and generalized anxiety disorder.*

*‡Equivalent to nonaffective psychosis in the NCS.*

*§In the ECA, dysthymia diagnoses were made on a lifetime basis only because of difficulty in determining the onset and offset of a disorder requiring a 2-year duration of symptoms.*

other specific disorders shows higher NCS *DSM-III* rates for alcohol and drug dependence, a screening assessment for schizophrenia, and major depressive episode (Table 4). A striking finding in the analysis of the affective mood disorders is the relatively comparable rates of any affective disorder for the 2-wave ECA and the NCS *DSM-III* diagnoses (2-wave ECA, 10.9%; and NCS, 11.2%). There is a pronounced shift of prevalence rates out of the less symptomatic dysthymia diagnosis in the ECA and into the major depressive episode category in the NCS for *DSM-III* and *DSM-III-R* criteria. This discrepancy can conceivably be accounted for by differences in the 2 instruments in eliciting symptoms (eg, the early placement of stem questions to avoid interview fatigue or experience-based attenuation of response in the UM-CIDI). Those respondents who were identified by the DIS in the ECA instrument to be on the symptom threshold of at least 3 symptoms required for a diagnosis of dysthymia (plus a 2-year duration) are able to recall 4 or 5 such symptoms in the NCS and thus meet criteria for the more symptomatic major depression diagnosis.

A similar pattern of identifying more symptomatic subjects in the NCS is noted in the substance use diagnoses, in which overall rates of substance abuse and dependence (Table 2) are not significantly different for the 2-wave ECA and the NCS. Rates of the more severe dependence criteria are higher in the NCS, however (a finding that explains the lower overall 1- and 2-wave ECA rates in Table 4 wherein substance abuse only is not included), because of the lack of comparability in the questions for substance abuse only.

Table 4 also shows that the prevalence rates for *DSM-III* anxiety disorders as a whole tend to be higher than found with the comparable NCS *DSM-III* questions in both the single-wave and 2-wave ECA, driven mainly by the much higher rates of phobias. This analysis shows higher rates of agoraphobia and simple phobia for the ECA and slightly higher rates of social phobia for the NCS. When *DSM-III-R* criteria and questions are added, allowing either avoidance or distress on exposure, rather than the avoidance-only requirement of the *DSM-III*, rates of simple phobia and social phobia in the NCS more than double, resulting in a higher overall rate of anxiety disorder than in the ECA. This is an excellent example of the differences in prevalence estimates that can occur when the diagnostic criteria are changed and when questions addressing criteria only in the older classification (*DSM-III*) are dropped in revised instruments (UM-CIDI) to make room for the new criteria-based questions (*DSM-III-R*).

## Policy Implications

**High Prevalence Rates and Implied Treatment Need.** The high estimates of lifetime disorders that have recently emerged from the NCS and from the 2-wave analysis of the ECA raise questions about the clinical significance of all these disorders in such a large proportion of the population. This is not unlike what happened at the close of the previous generation of epidemiological studies in which rates had ranged from as low as 10.9% in Baltimore, Md,[20] to highs of 55% in the Stirling County study (a pseudonym to protect the identity of a small rural county in Nova Scotia),[21] and 81.5% in midtown Manhattan, NY.[22] The high rates in the last studies had led to concerns about the clinical significance of some of these conditions and about the comparability of diagnostic assessments in these different studies.

In the current US climate of determining the medical necessity for care in managed health care plans, it is doubtful that 28% or 29% of the population would be judged to need mental health treatment in a year. Hence, additional impairment and other criteria should be developed for future epidemiological surveys to identify those most in need of such treatment. If these population-defined prevalence rates are useful for defining high-risk groups for future prevention purposes, evidence of such clinical course information should also be obtained.

**Diagnostic Criteria.** Even though we now have estimates of individual syndromes, as defined more rigorously by criteria from all *DSM* versions and the *ICD-10*, it is not clear that these disorders in community populations are equivalent to those identified by the same criteria in clinical settings. On the positive side, disorders identified in the ECA and NCS community samples have been noted to have the same risk-factor distributions as are present in clinical samples, including age, sex, marital status, other sociodemographic variables, and some indicators of disability.[13] We do not yet know, however, if these disorders have the same significance for the clinical course, the exclusion of other disorders, family or genetic clustering, laboratory studies (eg, modern functional imaging studies), and response to treatment that were suggested as indicators of validity by Robins and Guze.[1] Comparisons in the longitudinal course of matched clinical and community population samples would be a helpful starting point to address these issues.

Although it is possible that all of these community-based disorders are simply milder cases of essentially the same disorders seen in clinical settings, there are other possibilities as well. Based on the high prevalence rates identified in both the ECA and the NCS, it is reasonable to hypothesize that some syndromes in the community represent transient homeostatic responses to internal or external stimuli that do not represent true psychopathologic disorders. The human organism has a limited repertoire of response patterns to various physical, biological, and emotional stresses. Transient changes in blood pressure, pulse rate, body temperature, anxiety, or mood are not always indicators of pathology but of appropriate adaptive responses. It is possible that many people with currently defined mental syndromes (in particular among the affective and anxiety disorders) not brought to clinical attention may be having appropriate homeostatic responses that are neither pathologic nor in need of treatment—eg, other equivalents of grief reactions that meet clinical criteria but are not considered pathologic if they are time-limited.

Regardless of whether the high rates of syndromes defined by various versions of the *DSM* and the *ICD-10* contain mild forms of true psychopathologic disorders or nonpathologic homeostatic responses, it appears that additional severity, impairment, comorbidity, and duration criteria beyond those in the *ICD-10*, and the *DSM-IV* will be required to define a need for treatment when the criteria are applied to community populations. The first steps in conceptualizing a definition of treatment need have been made by the National Advisory Mental Health Council[23] and by the Substance Abuse and Mental Health Services Administration,[24] with continuing work along these lines now being done by WHO.[25]

**Standardizing Instruments.** Although diagnostic criteria are the framework for any clinical or epidemiological assessment, no assessment of clinical status is independent of the reliability and validity of the methods used to determine the presence of a diagnosis—be it by an unstructured clinical interview, a structured clinical assessment, or a highly structured instrument administered by lay interviewers. We might have thought that the wide variation in methods that characterized the previous generation of epidemiological studies was a thing of the past. We have seen, however, that "drift" or mutation in the structure of even the various versions of the CIDI has now become substantial. One purpose of this analysis was to help evaluate the extent of this phenomenon, which can contribute to both observation and criteria variance of methods.[26]

If we are to have a cumulative scientific base in this field, we need to address the problem by standardizing our assessment methods beyond what has been accomplished to date. A number of method effects, including research sponsorship, study context, and changes in the sequence or range of disorders covered, can affect response patterns and may be beyond the ability of instrument developers to control. It should be feasible, however, to reduce vaguely defined criteria and experiences that do not have high psychological salience[27-29] and to focus calibration efforts on the sorts of serious and severe psychiatric disorders that are of most significance for public health and public policy. In the absence of such efforts, it could rapidly be impossible to determine if any differences in prevalence estimates across several studies are due to differences in nature or differences in criteria or methods. Such differences have already prevented us from determining if there has been a real change in the prevalence of depressive disorders in age cohorts of subjects in the 10 years between the ECA and the NCS. The deficits in upward compatibility or crosswalks between the DIS, the CIDI, and variants of the CIDI now being used in various parts of the world will need to be made explicit and reduced if we are to have a cumulative science in which successive versions in instrument development produce increasing reliability and validity in estimating prevalence rates.

**Public Presentation of Epidemiologic Findings.** In addition to having common scientific criteria and technical methods, we also are in need of clear information on the prevalence rates of psychiatric disorders for public policy and constituent groups who support and ultimately benefit from our research activities. Major differences in reported prevalence rates of common disorders are extremely confusing. This problem can be addressed adequately by careful attention to assuring comparability of methods before initiating new stud-

ies. Rigorous scholarship in the later analysis and publication process should aid comparison with previous research findings to contribute to the development of a cumulative research field.

## CONCLUSIONS

Although the previous analyses have shown that standardization of demographic variables, diagnostic criteria, and comparable questions can narrow discrepancies between studies, these 2 major epidemiological studies do not adequately differentiate between diagnosis and treatment need. More in-depth studies of the validity of diagnoses of the *DSM-IV* and the *ICD-10* in community populations should use additional symptom threshold, impairment or disability, and duration criteria in existing assessment techniques. If some greater degree of standardizing assessment methods can then be accomplished, psychiatric epidemiological and services research will be positioned for another leap forward in both improving the understanding of the cause(s) of mental disorder and helping to focus limited resources in the most cost-effective manner.

*Accepted for publication April 4, 1997.*

*Reprints: Darrel A. Regier, MD, MPH, National Institute of Mental Health, Suite 4A52, 31 Center Dr, Bethesda, MD 20892-2475.*

## REFERENCES


1. Robins E, Guze SB. Establishment of diagnostic validity in psychiatric illness: its application to schizophrenia. *Am J Psychiatry*. 1970;126:983-987.
2. Feighner JP, Robins E, Guze SB, Woodruff RA, Winokur G, Munoz R. Diagnostic criteria for use in psychiatric research. *Arch Gen Psychiatry*. 1972; 26:57-63.
3. Spitzer RL, Endicott J, Robins E. Research diagnostic criteria: rationale and reliability. *Arch Gen Psychiatry*. 1978;35:773-782.
4. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, Third Edition*. Washington, DC: American Psychiatric Association; 1980.
5. Myers JK, Weissman MM. Psychiatric disorders in a US urban community: the New Haven Study. In: Weissman MM, Myers JK, Ross CE, eds. *Community Surveys of Psychiatric Disorders*. New Brunswick, NJ: Rutgers University Press; 1986: 155-175.
6. Robins LN, Helzer JE, Croughan J, Ratcliff K. National Institute of Mental Health Diagnostic Interview Schedule: its history, characteristics, and validity. *Arch Gen Psychiatry*. 1981;38:381-389.
7. Robins LN, Wing J, Wittchen HU, Helzer JE, Babor TF, Burke JD, Farmer A, Jablenski A, Pickens R, Regier DA, Sartorius N, Towle LH. The Composite International Diagnostic Interview: an epidemiologic instrument suitable for use in conjunction with different diagnostic systems and in different cultures. *Arch Gen Psychiatry*. 1988;45: 1069-1077.
8. Wittchen HU, Robins LN, Cottler LB, Sartorius N, Burke JD, Regier DA, and Participants in the Multicentre WHO/ADAMHA Field Trials. Cross-cultural feasibility, reliability, and sources of variance of the Composite International Diagnostic Interview (CIDI). *Br J Psychiatry*. 1991;159:645-653.
9. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, Third Edition, Revised*. Washington, DC: American Psychiatric Association; 1987.
10. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition*. Washington, DC: American Psychiatric Association; 1994.
11. World Health Organization. *International Statistical Classification of Diseases, 10th Revision (ICD-10)*. Geneva, Switzerland: World Health Organization; 1992.
12. Regier DA, Boyd JH, Burke JD, et al. One-month prevalence of mental disorders in the United States: based on five Epidemiologic Catchment Area sites. *Arch Gen Psychiatry*. 1988;45:977-986.
13. Robins LN, Locke BZ, Regier DA. An overview of psychiatric disorders in America. In: Robins LN, Regier DA, eds. *Psychiatric Disorders in America*. New York, NY: Free Press; 1991.
14. Regier DA, Narrow WE, Rae DS, Manderscheid RW, Locke BZ, Goodwin FK. The de facto US mental and addictive disorders service system: Epidemiologic Catchment Area prospective 1-year prevalence rates of disorders and services. *Arch Gen Psychiatry*. 1993;50:85-94.
15. Regier DA, Rae DS, Narrow WE. Stability of diagnoses in the Epidemiologic Catchment Area Study. Presented at the annual meeting of the American Psychiatric Association; May 14, 1991; New Orleans, La.
16. Regier DA, Bromet E, Helzer J, Narrow WE, Weissman M. Course of mental disorders: an epidemiological and clinical comparison. Presented at the annual meeting of the Society of Biological Psychiatry; May 1, 1992; Washington, DC.
17. Kessler RC, McGonagle KA, Zhao S, Nelson CB, Hughes M, Eshleman S, Wittchen HU, Kendler KS. Lifetime and 12-month prevalence of *DSM-III-R* psychiatric disorders in the United States: results from the National Comorbidity Survey. *Arch Gen Psychiatry*. 1994;51:8-19.
18. Blazer DG, Kessler RC, McGonagle KA, Swartz MS. The prevalence and distribution of major depression in a national community sample: the National Comorbidity Survey. *Am J Psychiatry*. 1994; 151:979-986.
19. Kendler KS, Gallagher TJ, Abelson JM, Kessler RC. Lifetime prevalence, demographic risk factors, and diagnostic validity of nonaffective psychosis as assessed in a US community sample: the National Comorbidity Survey. *Arch Gen Psychiatry*. 1996;53:1022-1031.
20. Pasamanick B, Roberts DW, Lemkau PV. A survey of mental disease in an urban population. *Am J Public Health*. 1956;47:923-929.
21. Leighton DC, Harding JS, Macklin DB, MacMillan AM, Leighton AH. *The Character of Danger: Psychiatric Symptoms in Selected Communities*. New York, NY: Basic Books; 1963.
22. Srole L, Langner TS, Michael ST, Opler M, Rennie T. *Mental Health in the Metropolis: The Midtown Manhattan Study*. New York, NY: McGraw-Hill Book Co; 1962.
23. National Advisory Mental Health Council. Health care reform for Americans with severe mental illnesses: report of the National Advisory Mental Health Council. *Am J Psychiatry*. 1993;150:1447-1465.
24. Kessler RC, Berglund PA, Zhao S, Leaf PJ, Kouzis AC, Bruce ML, Friedman RM, Grosser RC, Kennedy C, Narrow WE, Kuehnel TG, Laska EM, Manderscheid RW, Rosenheck RA, Santoni TW, Schneier M. *The 12-Month Prevalence and Correlates of Serious Mental Illness (SMI) in Center for Mental Health Services (CMHS) Mental Health, United States, 1996*. Rockville, Md: US Dept of Health and Human Services; 1996: 59-70.
25. Üstün TB, Cooper JE, van Duuren-Kristen S, Kennedy C, Hendershot G, Sartorius N, in collaboration with WHO/MNH Disability Working Group. Revision of the ICIDH: mental health aspects. *Disabil Rehabil*. 1995;17:202-209.
26. Regier DA, Burke JD. Quantitative and experimental methods in psychiatry: epidemiology. In: Kaplan HI, Sadock BJ, eds. *Comprehensive Textbook of Psychiatry*. Baltimore, Md: Williams & Wilkins; 1995;6:377-397.
27. Clark HH, Schober MF. Asking questions and influencing answers. In: Tanur JM, ed. *Questions About Questions: Inquiries Into the Cognitive Bases of Surveys*. New York, NY: Russell Sage Foundation; 1992:15-48.
28. Schuman H, Pressor S. *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context*. New York, NY: Academic Press; 1981.
29. Sudman S, Bradburn NM, Schwarz N. *Thinking About Answers: The Application of Cognitive Processes to Survey Methodology*. San Francisco, Calif: Jossey-Bass Publishers Inc; 1995.

COMMENTARY

# Problems in Defining Clinical Significance in Epidemiological Studies

REGIER ET al[1] summarize discrepancies in prevalence rates between the Epidemiological Catchment Area (ECA) study and the National Comorbidity Survey (NCS) and provide a useful overview of plausible explanations. I will extend their comments by emphasizing 3 inherent limitations to defining clinical cases in epidemiological studies: (1) the definition of mental disorder in *DSM-IV*[2] fails to provide a clear boundary between psychopathology and normality; (2) the concepts "clinical significance" and "medical necessity" are difficult to operationalize and to assess reliably; and (3) lay interviewers do not have the experience necessary to judge clinical significance.

*See also page 109*

In *DSM-IV*, the mental disorders are defined as clinically significant behavioral or psychological syndromes causing distress or disability (ie, impairment in functioning). The syndrome must not be an expected response to a particular event (eg, the death of a loved one).

Although this definition assists in decisions on what should be included in the manual, it is necessarily vague in setting the boundary between clinically significant psychopathology and the aches and pains of "normal" life. Unfortunately, there is no accepted way to define this boundary and decisions about it vary across clinicians, times, and cultures. The definition of clinical caseness is most divergent across the ECA and NCS studies for the milder depressive and phobic disorders precisely because these are the presentations that most frequently appear without clinical significance in "normal" people.

Recognizing this problem, *DSM-IV* places great emphasis on the judgment of "clinical significance" in defining caseness. It includes the following item as a requirement in the criteria sets of dozens of disorders: "The disturbance causes clinically significant distress or impairment in social, academic (occupational), or other important areas of functioning." This appeal to clinical judgment is a reminder to evaluate not only the presence of the symptoms in the criteria set, but also whether they are severe enough to constitute mental disorder. Unfortunately, this method of defining caseness contains the seeds of tautology; mental disorder is present only when there is "clinically significant" impairment, but this determination is based on the clinician's judgment. We have not been more precise in defining clinical significance because it is an inherently difficult concept to operationalize across disorders and settings. The managed care concept of medical necessity will be subject to the same problems in establishing reliability and generalizability.

Epidemiological studies use lay interviewers to avoid the high cost of clinician time. Lay interviewers do not have clinical experience and cannot be expected to make judgments about clinical significance. This problem is further compounded in epidemiological studies because clinical significance is inherently more difficult to evaluate in the community than in clinical samples, especially for symptoms like mild phobia or depression that frequently occur normally.

Given all these problems, it is no surprise that epidemiologic studies using different sampling methods, assessments, diagnostic criteria, and interviewers arrive at different prevalence rates of the milder psychiatric disorders in the community. The methods used in existing studies probably all bias to the overdiagnosis of the milder disorders.

Regier et al suggest how to improve future epidemiologic studies. We should study fewer disorders—only the more severe ones that are most likely to affect public health policy. Definitions of caseness should go beyond symptom evaluation to require measurable functional impairment. The severity and duration criteria for depression and phobia may need stiffening. Finally, it will be important to validate lay diagnoses with clinical interviewers, particularly for those disorders most likely to render false positives.

Psychiatric epidemiology is beset by both conceptual problems (what is the boundary of normality) and practical problems (which instruments, which criteria, and which interviewers). This should not discourage us. Determining rates of disorders and changes over time is worth the effort and is an increasingly necessary guide to policy. The ECA and NCS studies have been invaluable in refining our questions and methods and providing reasonably consistent rates for the more severe and easily defined disorders. This is an excellent foundation on which to build the future epidemiological studies that will need to be repeated every decade to provide us with an updated census of mental disorders.

*Allen Frances, MD*
*Department of Psychiatry and Behavioral Sciences*
*Duke University Medical Center, Box 3950*
*Durham, NC 27710*

## REFERENCES

1. Regier DA, Kaelber CT, Rae DS, Farmer ME, Knauper B, Kessler RC, Norquist GS. Limitations of diagnostic criteria and assessment instruments for mental disorders: implications for research and policy. *Arch Gen Psychiatry.* 1998;55:109-115.
2. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition.* Washington, DC: American Psychiatric Association; 1994.

COMMENTARY

# Diagnosis and Need for Treatment Are Not the Same

IF, LIKE the senior authors of the article by Regier et al,[1] you have the responsibility of providing accurate data about the prevalence of mental disorders in the community as a basis for justifying the allocation of scarce treatment resources, you are haunted by 2 findings: major discrepancies in the prevalence of mental disorders in 2 large community studies and lifetime and 1-year prevalence rates for major disorders that are higher than, to many, seem reasonable. Unfortunately, there is no sharp boundary between mental disorder and psychological health or between the various mental disorders. Readers, therefore, should not be surprised that prevalence rates can vary markedly with even minor changes in diagnostic criteria and thresholds for defining mental disorders and changes in assessment questions used in community surveys. As the authors note, however, the risk factors for the *DSM-III* and *DSM-III-R* disorders[2,3] remained relatively constant even when the prevalence rates varied greatly. This suggests that, despite these often puzzling differences in prevalence, both surveys measured something valid.

*See also page 109*

What is a reasonable lifetime or 1-year prevalence for mental disorders? Maybe it is time to present the argument that mental disorders, like physical disorders, vary in their severity and associated functional impairment. Dermatology has skin cancer and warts. So, too, some mental disorders are devastating in their associated impairment (eg, schizophrenia), whereas others (eg, some animal phobias) are distressing but rarely cause serious impairment. No one is interested in the lifetime or 1-year prevalence of any physical disorder, so why the interest in prevalence rates for any mental disorder?

The thrust of the article is to suggest that our current diagnostic criteria are "limited" as guides to the need for treatment. To confuse making a mental disorder diagnosis with demonstrating treatment need, however, would be a serious mistake. Consider examples from physical medicine. Many physical disorders are often transient and mild and may not require treatment (eg, acute viral infections or low back syndrome). It would be absurd to recognize such conditions only when treatment was indicated.

The authors question whether mental disorders identified in the community with current diagnostic criteria "have the same significance for the clinical course, the exclusion of other disorders, family or genetic clustering . . . and response to treatment that were suggested as indicators of validity by Robins and Guze." They suggest that "comparisons in the longitudinal course of matched clinical and community population samples would be a helpful starting point to address these issues." Not true. It is hardly news that treated cases of mental (and physical) disorders tend to be more severe than untreated cases. A follow-up study of community cases of untreated major depression (or low back syndrome or acute rheumatoid arthritis) will almost certainly show a different clinical course (less chronicity and impairment) but that will not indicate that the community cases are invalid. This all goes to show that in the absence of a biological gold standard for determining true "caseness," diagnostic criteria for mental disorders will have to continue to rely on the expert clinical judgment of the *DSM* subcommittee members.

Should future surveys include data on impairment, disability, and duration to better assess the need for treatment? Of course, but let us not revise diagnostic criteria that help us make clinically valid standard diagnoses in order to make community prevalence data easier to justify to a skeptical public.

*Robert L. Spitzer, MD*
*Biometrics Research Department*
*New York State Psychiatric Institute*
*722 W 168th St*
*New York, NY 10032*

## REFERENCES

1. Regier DA, Kaelber CT, Rae DS, Farmer ME, Knauper B, Kessler RC, Norquist GS. Limitations of diagnostic criteria and assessment instruments for mental disorders: implications for research and policy. *Arch Gen Psychiatry*. 1998;55:109-115.
2. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, Third Edition*. Washington, DC: American Psychiatric Association; 1980.
3. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, Third Edition, Revised*. Washington, DC: American Psychiatric Association; 1987.