Psychiatric Epidemiology: Progress and Prospects in the Year 2000

William W. Eaton¹ and Kathleen R. Merikangas²

PROGRESS

In the last half century, the field of psychiatric epidemiology has been born and is rapidly coming of age. The emergence of chronic disease epidemiology in the 1940s and 1950s was conceptually joined to the revival of interest in psychiatric disorders that came about after World War II, in a publication by the Milbank Memorial Fund, in 1950, which used the term psychiatric epidemiology (1). For 25 years following, sometimes called the “second generation” of psychiatric epidemiology (2), the field was dominated by sociologists and social psychiatrists, who applied newly developed household survey techniques to the problem of estimation of the prevalence of mental problems and who studied causal factors such as socioeconomic stratification (3, 4), social integration (5), and social stress (6). There were relatively few epidemiologic studies of specific mental disorders, and those that existed focused efforts on measuring the extent of disorders in the population, that is, descriptive epidemiology (7). Research directed toward understanding the etiology of disorders—analytic epidemiology—used measures of psychological distress or global impairment, precluding integration with laboratory and clinical research and hindering construction of credible etiologic hypotheses.

During this second generation of psychiatric epidemiology, there were developments in related fields that were to overtake and change the field drastically. One important advance was the development of the empirical basis for psychiatric diagnosis. In the latter half of the decade of the 1960s and into the 1970s, several studies (8, 9) showed that the reliability of psychiatric diagnosis was so weak as to allow the charge that diagnoses were “myths” or “ideology” (10, 11). Empirical work on diagnostic methods, such as carefully structured examinations with rigorous training of diagnosticians (12, 13), was led by epidemiologically oriented clinicians but driven by the need for epidemiologic comparisons, such as in the US/UK study of schizophrenia (14). This research set the stage for the development of operational criteria for the range of psychiatric disorders (15, 16), which in turn generated a wide range of assessment instruments directed at specific diagnostic targets and a nearly fanatical concern for the reliability of diagnosis (17–19).

A second important advance was demonstration of the utility of psychiatric diagnoses, by developments in genetic research and by new psychotropic medications. Diagnostic utility means the degree to which the diagnosis parsimoniously represents important etiologic pathways, thus facilitating prevention of the disorder; and the degree to which it predicts the natural history, facilitating control of the disorder and its consequences. Although there had been family and twin studies of disorders since the early part of the 20th century, the findings had always been subject to the interpretation that transmission within the family was cultural, not genetic. Even twin studies were open to this interpretation, since the argument could be made that monozygotic twins are treated in a more similar fashion than are dizygotic twins. In the late 1960s and 1970s, several adoption studies presented supporting data that were nearly immune to this interpretation (20). The effect was to provide credibility for the entire corpus of genetic research and, thereby, for the validity of the diagnostic process. Likewise, during these decades, there was increasing recognition of the specificity of the targets of new psychotropic medications, again supporting the validity of the diagnostic process.

The utility of specific diagnoses was further enhanced by reviews of research around the world, which suggested that many, or even most, disorders found in modern Western settings could also be found throughout the world (4). There remained many so-called culture bound syndromes, but even these are sometimes understood as reflecting basic physiologic and cultural processes existing in humans generally (21, 22). The prototypical study is the International Pilot Study of Schizophrenia (IPSS) and its descendants (23).
The IPSS applied new standardized diagnostic procedures, developed out of the classic US/UK study of diagnosis (14), in a variety of settings around the world. In each setting there was a syndrome that could be reliably diagnosed as schizophrenia. In later related studies, the range of variation in rates of incidence was within one order of magnitude, suggesting that methodological factors that had threatened prior cross-national comparisons could be addressed successfully (24).

A result of improvements in the reliability and credibility of specific diagnoses was a push to link outcome variables in psychiatric epidemiology to those in clinics (facilitating services research) and laboratories (facilitating building etiologic theories), introducing the so-called "third generation" of psychiatric epidemiology (2). Field survey instruments were developed in conjunction with new operational definitions of diagnosis, providing the potential for such an advantageous link. These survey instruments were in a highly structured, verbatim format, which could be administered by interviewers without medical or clinical training (25, 26). The Epidemiologic Catchment Area study in the United States (27-29) was followed by the National Comorbidity Survey there (30) and by similar surveys in England (31) and in a variety of cultures around the world (32, 33). The verbatim instruments were less successful where insight is a principal feature of the disorder and, thus, did not contribute much to the understanding of schizophrenia or bipolar disorder, even in estimating their respective prevalences credibly (34). The diagnostic field surveys remained in the descriptive tradition, for the most part, not contributing much to the understanding of etiology. These field surveys did contribute to our knowledge regarding the range of expression of psychiatric disorders in the general community and the extent to which the major mental disorders remain untreated. Despite the high prevalence of psychiatric disorders in community residents, only a minority had treatment, only a minority had treatment, and an even smaller proportion has had contact with a clinic (facilitating services research) and laboratories (facilitating building etiologic theories), introducing the so-called "third generation" of psychiatric epidemiology (2). Field survey instruments were developed in conjunction with new operational definitions of diagnosis, providing the potential for such an advantageous link. These survey instruments were in a highly structured, verbatim format, which could be administered by interviewers without medical or clinical training (25, 26). The Epidemiologic Catchment Area study in the United States (27-29) was followed by the National Comorbidity Survey there (30) and by similar surveys in England (31) and in a variety of cultures around the world (32, 33). The verbatim instruments were less successful where insight is a principal feature of the disorder and, thus, did not contribute much to the understanding of schizophrenia or bipolar disorder, even in estimating their respective prevalences credibly (34). The diagnostic field surveys remained in the descriptive tradition, for the most part, not contributing much to the understanding of etiology. These field surveys did contribute to our knowledge regarding the range of expression of psychiatric disorders in the general community and the extent to which the major mental disorders remain untreated. Despite the high prevalence of psychiatric disorders in community residents, only a minority had treatment, and an even smaller proportion has had contact with a mental health specialist. This finding stimulated interest in increasing awareness of psychiatric problems among primary care practitioners (35). Increasing evidence shows that the common psychiatric disorders have major social, economic, and personal impact; for example, of all human disorders, depression is the seventh leading source of disability (36).

Community-based studies, particularly those with a longitudinal design, have also contributed to the development of the diagnostic nomenclature by highlighting the importance of comorbidity between and within the major psychiatric syndromes. The comorbidity arises in part from the lack of clear boundaries and thresholds for diagnosis (37, 38), from varying patterns of symptom expression and longitudinal stability and course, and from consequences of psychopathology for the onset of a range of physical conditions (39, 40).

Improvements in diagnosis could well be linked to future developments in genetics, but this is not guaranteed. With advances in characterization of the human genome, the specific genetic basis of numerous human diseases has been identified, and increasing research has been devoted to identify genetic factors underlying psychiatric disorders as well. Although the history of genetic studies in psychiatry is more abundant than that of many other chronic diseases, specific etiologic genes for the major psychiatric disorders have, thus far, not been identified (41). Studies of familial aggregation suggest that the most compelling model involves susceptibility genes that convey increased disease risk but only in the presence of additional environmental, genetic, or biologic factors. The most notable example is schizophrenia for which there is emerging evidence for gene-environment interaction (42) and oligogenic inheritance (43). It now seems likely that many, or most, mental problems involve a complex mixture of multiple genetic and environmental influences, interacting in nonlinear and nonadditive fashion. This complex web of interacting factors will be understood only with careful choice of genetically informative samples, as well as carefully selected measures of the environment. Environmental influences will include so-called biologic factors such as obstetric complications, infections, and exposure to toxins, as well as the social and psychological environment.

**PROSPECTS**

The enhanced biologic credibility of the specific diagnoses, as well as the survey tools for diagnostic assessment which were developed, facilitated the transition from descriptive epidemiology to analytic epidemiology, which is still underway. Currently the field is in a state of ferment, reflecting the tensions of transition to a new era, as shown by a flurry of conceptual reviews (44-49). Paradoxically, developments in diagnosis and genetics, which have brought about the major advances so far, now produce contradictions that set the stage for new modes of thinking.

The Human Genome Project will spawn new understandings of human physiology and new chemical therapies for diseases generally, and there undoubtedly will be important spinoff results for psychiatry (50). Advances in genetics and neuroscience will gradually inform classification and assist in identifying sources of heterogeneity at both the genotypic and phenotypic levels. Designing research to elucidate these multiple influences will require changes in the (mostly genetic) paradigm that has generated the question itself. New
research will incorporate many causal factors, from many disciplines, into longitudinal designs that allow for relatively long latencies of causal effect. The new research will take advantage of developments of DNA and bioassays that facilitate use under field conditions (51), as well as independent advances in statistical techniques for analysis of genetic linkage and association (52), gene-environment interactions (53–55), and longitudinal data (56). In addition, the identification of susceptibility genes solely on the basis of high density families will require population studies that examine the extent to which these genetic markers are associated with disease risk in the general population. Likewise, population data will be critical for determining the health policy implications of newly discovered human gene variants.

Interest in social factors as causes of mental disorders waned as the biologic revolution got underway, culminating recently in the Human Genome Project. Paradoxically, however, new understandings of animal and human physiology and new capabilities in measurement may reassert the importance of social factors. Recent research in animal populations has shown that changes in the basic dimensions of social structure, such as social stratification (57) or social integration (58) (the dual thrusts of the second generation of psychiatric epidemiology), produce strong changes in physiology, which are arguably quite close to the changes occurring in humans during episodes of psychopathology. The capability for understanding the effects of social interaction is greatly potentiated by the development of statistical approaches that model group-level effects and individual risk for disorder (59).

Advances in diagnosis have likewise produced a paradoxical need for revolution in thinking about what diagnosis means. Earlier near-fanatical concern with the reliability of diagnosis has been replaced by near-fanatical concern with the precision of estimates of the prevalence of specific disorders. The prevalence of mental disorders is a topic garnering attention that is inconceivable for its analog, the prevalence of physical disorders (60). Too much concern with differences in prevalence from one study to another reflects unwarranted defensiveness about the scientific value of research in psychiatric epidemiology (61). The focus on sample weights, response bias, and interview techniques is not wrong, unless it distracts from attempts to understand the etiology of mental disorders, as is apparently the situation in psychiatric epidemiology. Evidence for the distraction is the underutilization, until the last decade or so, of time-honored epidemiologic methods, such as the case-control design, in understanding etiology.

The increasing specificity of mental disorders, as displayed in the increase in the numbers of diagnoses in the Diagnostic and Statistical Manual of Mental Disorders (DSM) from 159 in DSM-II, to 227 in DSM-III, to 357 in DSM-IV, may have resulted more from the social, political, and financial nature of the process of producing the manual than from scientific evidence (62). Given the increasing specificity of diagnosis, the amount of comorbidity (63) should not be surprising. Moving from an interest in the description of comorbidity to an understanding of the nature of the psychopathology itself requires a shift in conceptual framework. The shift will involve empirical study of multiple patterns of symptoms, not necessarily in the form of one or another specific diagnosis (64). As the genetic basis of the components of psychiatric disorders unfolds, diagnostic categories are likely to be dramatically changed to reflect both genetic heterogeneity (different genes leading to similar phenotypes) and phenotypic heterogeneity (single genes resulting in multiple phenotypic outcomes). The shift in emphasis will take advantage of recent developments in statistical techniques for the analysis of high-dimensional data (65, 66).

Progress in the research of major risk factors for mental disorders has been cumulative over the last 25 years, and there are beginning to be explicit integrations across disciplines, consistent with the eclectic framework provided by the epidemiologic approach. For any given major disorder, it is possible to tick off the major risk factors, often on the fingers of one hand, and, as with many chronic physical conditions, usually beginning with a family history of the disorder. For example, there exists a large body of work on the effects of disruptions in the immediate social environment of the individual on that person’s risk for onset of depressive disorder (6, 67). Genetic vulnerability enhances the effect of loss (68). These epidemiologic findings can be integrated into models of depression that included biochemical, ethological, social, and clinical research (69). Other consistent and credible risk factors include female gender (70, 71) and childhood loss, abuse, and trauma (72). Another example is the series of case-control studies directed at obstetric complications as a risk factor for schizophrenia (73). These results could be integrated into neurodevelopmental models of schizophrenia (74) that include studies of the family history of the disorder and of brain structure (75). Other consistent risk factors for schizophrenia include birth in an urban area (76) and winter birth (77). What remains is the large challenge of understanding how risk factors from formerly competing paradigms (often genetic vs. environment) interact with each other. So far, only a few studies have included genetic

Epidemiol Rev Vol. 22, No. 1, 2000
and environmental influences in the same research (68, 78). Psychiatric epidemiology will increasingly adopt the methods of genetic epidemiology to examine how genetic and environmental factors combine in producing disease etiology. Application of designs that either hold constant the environment while allowing the variation in genetic factors (e.g., monozygotic vs. dizygotic twins or step-sibs reared together) or those that hold constant the genetic background while allowing environmental variation (e.g., adoptees or sibs reared apart) are especially powerful as counterparts to the large population-based studies that include both familial history and environmental risk factors.

Increasing evidence reveals that the majority of adulthood psychiatric disorders begin in adolescence (79, 80). The power of psychiatric epidemiology to enhance our understanding of the early manifestations of mental disorders has become evident from its growing application to childhood mental disorders (81). Similar to the contributions made in recent years relative to adult psychiatric disorders, data from community-based samples using standardized diagnostic methods, particularly those with longitudinal designs, are beginning to emerge. The epidemiologic approach can advance knowledge through the collection of information on early signs and risk factors gleaned from high-risk studies and from studies on protective factors and processes related to the development of psychopathology in childhood. Beginning in the 1960s with the Isle of Wight study, the importance of often subtle nosological definitions in determining the rates of childhood disorders became evident, as did the importance of complex issues of diagnostic comorbidity and the relation between the age of onset of psychiatric syndromes and their environmental and developmental correlates (82). To fully understand these processes, it will be critical to institute population-based interventions during childhood or adolescence. With the discouraging results of many universal prevention trials, efforts are likely to shift to use the selective rather than universal interventions or a combination of strategies (83). Both descriptive and analytic epidemiology will provide a valuable empirical basis for identifying targets and outcomes of such efforts.

In summary, the task for the next half century in psychiatric epidemiology is to understand how multiple risk factors interact over time in producing multiple outcomes. This understanding will be achieved only if the study of risk factors is not narrowed by disciplinary orientation and only if the study of outcomes is not narrowed by strict dichotomous approaches to diagnosis. Breaking loose from the confines of discipline and diagnosis will lead to progress in comprehending and influencing the web of causation for the complex combination of phenomena we call mental disorders.

ACKNOWLEDGMENTS

This work was supported by National Institute of Mental Health grants MH47447 and MH53188.

REFERENCES

60. Spitzer RL. Diagnosis and need for treatment are not the same. (Comment). Arch Gen Psychiatry 1998;55:120.