REVIEW

Risk-Adjusting Outcomes of Mental Health and Substance-Related Care: A Review of the Literature

Richard C. Hermann, MD, MS, Caitlin K. Rollins, AB, and Jeffrey A. Chan, BS

Risk adjustment is increasingly recognized as crucial to refining health care reimbursement and to comparing provider performance in terms of quality and outcomes of care. Risk adjustment for mental and substance use conditions has lagged behind other areas of medicine, but model development specific to these conditions has accelerated in recent years. After describing outcomes of mental health and substance-related care and associated risk factors, we review research studies on risk adjustment meeting the following criteria: (1) publication in a peer-reviewed journal between 1980 and 2002, (2) evaluation of one or more multivariate models used to risk-adjust comparisons of utilization, cost, or clinical outcomes of mental or substance use conditions across providers, and (3) quantitative assessment of the proportion of variance explained by patient characteristics in the model (e.g., R^2 or c-statistic). We identified 36 articles that included 72 models addressing utilization, 74 models of expenditures, and 15 models of clinical outcomes. Models based on diagnostic and sociodemographic information available from administrative data sets explained an average 6.7% of variance, whereas models using more detailed sources of data explained a more robust 22.8%. Results are appraised in the context of the mental health care system's needs for risk adjustment; we assess what has been accomplished, where gaps remain, and directions for future development. (HARV REV PSYCHIATRY 2007;15:52–69.)

Keywords: **mental disorders, quality assessment, quality of care, reimbursement, risk adjustment, substance use disorders**

From the Center for Organization, Leadership and Management Research, Veterans Health Administration, Boston MA (Dr. Hermann), and Center for Quality Assessment and Improvement in Mental Health, Tufts University School of Medicine.

Supported by the Agency for Healthcare Research and Quality (grant no. R01-HS10303), National Institute of Mental Health (grant no. K08-MH001477), and Substance Abuse and Mental Health Services Administration.

Original manuscript received 3 February 2006; revised manuscript received 15 November 2006, accepted for publication 11 December 2006.

Correspondence: Richard C. Hermann, MD, MS, Center for Quality Assessment and Improvement in Mental Health, Tufts–New England Medical Center, 750 Washington St., NEMC #345, Boston, MA 02111. Email: rhermann@cqaimh.org

DOI: 10.1080/10673220701307596

Risk adjustment is a statistical process of controlling for patient characteristics when assessing health outcomes. In this context, "outcomes" refers not only to traditional clinical outcomes such as changes in symptoms, functioning, or health status, but also—departing from Donabedian's lexicon— to such "outcomes" as quality of care and expenditures for health care services.[1,2] The importance of risk adjustment has grown over the past two decades as efforts to assess quality of care, allocate resources, and align incentives have expanded. Many payers and accreditors have developed report cards to compare the performance of plans and providers on the basis of clinical outcomes or quality of care. For comparisons to be fair, one may need to adjust statistically for differences among patient populations, such as the complexity or severity of illness. Similarly, risk adjustment is needed to design reimbursement systems that reward efficiency without penalizing providers who treat patients with greater treatment needs.

Risk adjustment for mental and substance use (M/SU) conditions has lagged behind the medical mainstream.

These conditions receive little attention in the most commonly used commercial risk-adjustment systems. Many prominent initiatives comparing quality of care for mental health and substance-related care lack risk adjustment. In the last few years, however, there has been greater interest and development of models specifically for M/SU conditions.

In addition to the broader needs of the health care system, several features of mental health care contribute to the context for risk adjustment. M/SU conditions are highly prevalent and costly. An estimated 19% to 29% of Americans suffer from a diagnosable mental disorder (of one kind or another) in a given year,[3–5] and $82 billion is spent annually on mental health treatment.[6] Much of the impetus for risk adjustment in mental health care has come from the public sector. Medicaid, Medicare, and state mental health and substance abuse authorities play a major role in financing and providing care, particularly for individuals with severe and persistent mental illness (SPMI). Resource constraints have led policymakers to seek more sophisticated methods for allocating mental health services among those in need. Public and private purchasers and managers are attempting to assess the value of care by applying measurement to assess quality and outcomes of care. Advocacy groups pressing to limit program cuts, expand access, and attain parity in insurance coverage for M/SU conditions are recognizing that measurement, management, and comparative analyses can strengthen their case. In response to growing interest, the U.S. Substance Abuse and Mental Health Administration, National Institute of Mental Health, and the Agency for Healthcare Research and Quality have provided leadership and funding for development of risk-adjustment and quality measurement for M/SU conditions.

Organizational features of mental health services delivery complicate risk adjustment. Care for individuals with SPMI is funded at the local, state, and federal levels. Many employers and health plans contract with organizations that specialize in managing mental health and substance-related care separately from medical and surgical care. Meanwhile, more mental health care is provided in the primary care sector than the mental health specialty sector. These partitions in financing, managing, and delivering care pose challenges to accessing and integrating data needed to evaluate and adjust outcomes of care.

Ettner and Frank[7] and Hendryx[8] have authored valuable introductions to risk-adjustment methodology in mental health care. This report expands on their work by providing a comprehensive review of risk-adjustment models for M/SU conditions. We assess risk adjustment in view of the health care–system needs described above: What has been accomplished? Where are there gaps? What directions emerge for future development? The first section outlines the basic components of risk-adjustment models and the relationships among them. The next sections describe outcomes of M/SU conditions, risk factors for these outcomes, and sources of information about them. The fourth section evaluates risk-adjustment models for mental health and substance-related care in terms of their purpose, components, and predictive ability. In addition to methodologic issues, we discuss practical considerations such as tradeoffs between a model's predictive ability and its data-collection burden.

COMPONENTS OF RISK-ADJUSTMENT MODELS

Basic elements of a risk-adjustment model include an outcome, patient factors hypothesized to predict that outcome, and a statistical equation that quantifies the degree to which significant factors predict the outcome. Patient characteristics are sometimes referred to as "risk factors" because modeling seeks to determine the risk of a given outcome.[2] Risk factors can be specific to an outcome, which is to say, a risk-adjustment model that predicts improvement in depressive symptoms may not predict improvement in patient satisfaction. Statistical equations that relate risk factors and outcomes range from simple stratification to multivariate analysis and more sophisticated methods.

RISK ADJUSTMENT OF MENTAL HEALTH AND SUBSTANCE-RELATED CARE

The outcome of a risk-adjustment model is determined by the model's purpose: what is to be compared or adjusted? In report cards providing feedback to clinicians on the quality of their care, clinical outcomes are typically presented. Examples include symptom reduction in response to treatment and rates of appropriate medication dosages. Financial and utilization outcomes are often of interest to payers and purchasers of care; examples include aggregate annual costs per patient and total number of inpatient days, respectively.

Regardless of its focus, an outcome selected for risk adjustment should ideally meet several criteria. It should be valid, precisely specified, and reliably measured. The outcome may be positive (e.g., improved functioning) or adverse (e.g., relapse). The time period for assessment should be clearly defined (e.g., relapse within 60 days of hospital discharge). Feasibility of measurement is another important consideration: are the data available, affordable, and secure from gaming or manipulation?[9,10]

Clinical Outcomes

Mental health does not have a meaningful, frequent, and easily collected clinical outcome measure such as mortality, a measure that has played an important role in assessing outcomes of myocardial infarction and certain surgical procedures. Objective laboratory tests cannot establish the

severity of mental illness in the way that biophysiological parameters characterize the severity of some medical conditions. Most clinical outcome measures of M/SU conditions assess changes in symptoms, functioning, or quality of life in response to treatment. Individuals with a chronic condition such as schizophrenia face the risk of an exacerbation of symptoms with associated loss of functioning or well-being. Desired outcomes would be a delay or diminution of these occurrences. Individuals with episodic illnesses such as major depressive disorder risk an acute episode, with desired outcomes including remission of symptoms, restoration of functioning and well-being, and prevention of relapse or recurrence.

Many outcomes in mental health and substance-related care mirror those used elsewhere in medicine, including somatic symptoms, health-related quality of life, and patient satisfaction, as well as social, occupational, and other role functioning. Some outcomes are specific to mental health care, such as change in depressive, manic, or psychotic symptoms. In the evaluation of substance-related care, common behavioral outcomes include amount consumed, abstinence, and involvement in the criminal justice system. Functional domains of particular interest in mental health and substance-related care include interpersonal and familial functioning. Social connectedness has been used as a measure of the effectiveness of treatment for substance-use conditions.[11] Other outcomes are specific to demographically based subpopulations. For instance, outcomes of interest in geriatric mental health care include cognitive functioning, self-care, and behavioral disturbances associated with dementia. For children and adolescents, relevant outcomes include adaptation and adjustment, attention, school performance, and conduct-related problems. For individuals with SPMI, additional outcomes relate to social welfare: poverty, exposure to disease, adequacy of food and shelter, and safety from crime.[12]

A long tradition of psychometric study in mental health has produced many instruments in varied domains of symptoms and functioning in mental health and illness. Most were developed for clinical research but are increasingly applied in the evaluation of health care services. Two recent texts review these instruments and their properties in detail;[13,14] illustrative examples are provided below.

Symptom-rating scales assess clinical outcomes by evaluating change over time, or before and after service use. Rating scales can be clinician- or patient-administered and generic or disorder-specific. For example, the Hamilton Rating Scale for Depression is a 24-item, clinician-administered scale of depressive symptoms and severity.[15] In contrast, the Symptom Checklist (SCL-90R) is a 90-item, patient-administered questionnaire covering nine symptom clusters.[16] Both are widely used in clinical and services research, and have been extensively tested. For some conditions, thresholds have been established from severity scores to operationally define remission, relapse, and recurrence.[17]

Instruments assessing mental health functioning and disability vary widely in their sensitivity, specificity, and burden. The Global Assessment of Functioning Scale (GAF) illustrates one end of the spectrum. Clinicians assess a patient's symptoms and social and occupational functioning on a single-item, 1-to-100 scale.[18] Strengths of the GAF include its brevity, wide use in clinical care, and moderately good reliability in formal testing. A weakness is that the scale attempts to assess symptoms and functioning with a single rating, although their severity may differ. In addition, little data assess the reliability of the GAF in routine use, where application is subject to variations in training, assessment, and financial incentives. A more extensive instrument, though correspondingly more burdensome, is the World Health Organization Disability Assessment Schedule (WHODAS II), a 36-item, clinician-administered assessment of six functional domains, compatible with the WHO's International Classification of Functioning, Disability, and Health (ICF).[19]

The Addiction Severity Index (ASI) is widely used for clinical assessment, treatment planning, and outcome evaluation among individuals with substance-use conditions.[20] One hundred sixty-one items lead to composite scores in seven problem areas: medical status, employment and support, drug use, alcohol use, legal status, family/social status, and psychiatric status. Although abbreviated versions of the ASI have been developed, administration of the full instrument by a trained clinician interviewer results in superior reliability and validity.

A number of multidimensional instruments have been developed specifically for assessing outcomes of M/SU conditions in facilities and delivery systems. Researchers at the University of Arkansas have created a series of disorder-specific instruments for schizophrenia, depression, alcohol abuse/dependence, and panic disorder. The Schizophrenia Outcomes Module assesses symptoms and functional status over the course of treatment, along with selected patient characteristics for risk adjustment.[21] Other instruments can be applied across disorders. The Behavior and Symptom Identification Scale (BASIS-32) is a patient self-report instrument designed for heterogeneous populations of psychiatric inpatients. Thirty-two items yield scores in five domains: depression and anxiety, psychosis, impulsive and addictive behavior, relation to self and others, and daily living and role functioning.[22]

Clinical Processes

Quality-of-care measures are a common focus of risk adjustment. Under Donabedian's framework,[1] one would

describe these as process measures, though in the context of risk-adjustment models, they are grouped broadly under outcomes. Domains of process measurement include detection, access, assessment, treatment, continuity, coordination, and safety.[23–25] Some measures assess provider performance in comparison to evidence- or consensus-based guidelines. In areas of clinical practice where standards or benchmarks are not defined, measures are used to identify statistical outliers (e.g., inpatient services with very high or low rates of physical restraints). The National Association of State Mental Health Program Directors Research Institute (NRI) has developed a widely used quality-measurement system for inpatient care in the public sector. The HEDIS measures of the National Committee for Quality Assessment (NCQA) assess the quality of various processes of care in commercial health plans—including, for example, adherence to antidepressant medications, rates of aftercare following hospital discharge, and continuity of care for children treated with medication for attention deficit disorder.

Process measures are often said to be less in need of risk adjustment than measures of outcome. Detailed denominator specifications can sometimes define sufficiently homogeneous populations to allow comparisons of practice across providers. Some clinical processes are fully under the provider's control, making adjustment for patient characteristics unnecessary. Other processes are so important that they should always be performed, making comparisons among providers less important than achieving 100% conformance. Many other process measures are intended to compare performance across providers and require statistical adjustment of results so that these comparisons can be fair. This is particularly true for process measures that depend on both provider and patient performance. An example is the HEDIS measure of the proportion of patients discharged from psychiatric hospitals who attend an outpatient visit within seven days, which can be influenced by provider practice as well as patient adherence. Although it is the responsibility of the inpatient team to ensure that a patient is stabilized prior to discharge, that the patient participates in the development of a treatment plan for postdischarge care, and that an aftercare visit is scheduled proximal to discharge, the inpatient service will receive "credit" on the measure only if the patient attends the appointment. Meanwhile, no-show rates for scheduled outpatient appointments can exceed 50% among individuals with SPMI, and patients with certain characteristics—for example, with a comorbid substance use disorder or greater severity of mental illness—are less likely to attend scheduled follow-up care than others.[26] Risk adjustment might be considered if inpatient facilities are to be compared on this measure.

Cost/Resource Utilization

Spurred by rising health care costs, purchasers and payers are developing reimbursement systems to allocate health care resources more effectively by aligning financial incentives with patients' need for care. Instead of paying for each unit of services, fixed-fee or capitated reimbursement pays providers on a per-patient basis calculated from average costs adjusted for patient characteristics.[27]

Overall health care costs also serve as the basis for reimbursement systems. Case rates cover the costs of service utilization for an acute episode of care once services have been initiated. Capitation rates cover all service-related costs for a population. In capitation models that integrate medical and mental health care, risk adjustment can reduce incentives to avoid individuals with M/SU conditions. In plans that set a separate capitation rate for M/SU conditions, risk adjustment can diminish selection against sicker patients.

Risk adjustment is applied to comparisons of resource use across providers, such as length of inpatient stay or number of outpatient visits per patient per episode of illness. These models attempt to control for patient characteristics, leading to a better understanding of the variability in utilization attributable to providers. Risk adjustment has also been applied to readmission rates after hospital discharge. This measure is a utilization-based outcome that is frequently used as a proxy for relapse. Its use as a quality measure is controversial. Research has studied the relationship between the quality of inpatient care and subsequent readmission rates, but has failed to find an association.[28]

Dimensions of Risk for Mental Illness

Potential risk factors for mental health outcomes include sociodemographic information, health-status data, and other patient characteristics. Data relating to risk should ideally satisfy the same criteria as for outcomes: they should be reliable, valid, and not subject to gaming. In addition, a risk factor's contribution to a model must be weighed against the cost of obtaining the data. In this section we highlight risk factors having particular relevance to mental health and substance-related care.

Model developers identify prospective risk factors based on empirical evidence or clinical experience suggesting an association with the outcome. Another consideration in selecting risk factors for testing is whether they are consistent with the model's purpose. For example, minority racial or ethnic status may be significantly associated with the quality of care that a patient receives;[29] however, adjusting report-card results for the racial composition of clinical populations would obscure differences in provider performance

associated with race—a result inconsistent with the goals of narrowing disparities in quality of care.

Diagnosis

Clinical diagnosis of M/SU conditions is based on criteria defining threshold signs and symptoms, level of impairment, and time course. The *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.) (DSM-IV)[30] is used throughout mental health and substance-related care and is cross-indexed to the *International Classification of Diseases* (9th ed.) (ICD-9) codes used in administrative claims. DSM-IV uses a multiaxial system, with Axis I comprising clinical disorders (such as major depression or schizophrenia); Axis II, personality disorders and mental retardation; and Axis III, medical conditions associated with mental illness. Axis IV describes social and environmental problems, and Axis V consists of the GAF rating of symptoms and functioning.

Most risk-adjustment methods for mental health incorporate diagnosis because the information is clinically meaningful and readily available in administrative databases. The simplest models assign patients, based on their primary diagnosis, to an aggregate category of Axis I disorders (e.g., affective, psychotic, or substance-use conditions). Finer-grained models assign patients to groups consisting of specific Axis I diagnoses or common combinations of Axis I and Axis II diagnoses such as a mental disorder with a co-occurring substance-use or personality disorder. Concurrent medical illness can be represented in a model by a simple binary variable (present/absent) or by more sophisticated measures of the number, complexity, and severity of medical conditions.[31] In general, the expectation is that patients with more-severe conditions and more-complex comorbidities would, on average, experience greater utilization and costs, and worse outcomes.

Adoption of DSM criteria has led to improved diagnostic reliability for many psychiatric disorders. However, accuracy and reliability varies with the clinician's fidelity to the DSM process, the amount of information available at the time of diagnosis, and consistency in the patient's presentation at different times. In addition, diagnoses recorded in administrative claims have demonstrated limited accuracy. An analysis of Medicaid claims from six states found that 25% of individuals with a claim for schizophrenia in 1994 had at least one claim the following year for a nosologically incompatible condition, such as bipolar disorder, psychotic disorder NOS, or schizoaffective disorder.[24] Model designers have developed a variety of strategies to deal with inconsistencies in claims-based diagnoses—for example, using algorithms to assign patients to the primary diagnosis that appears most frequently or that is generally most disabling. Other studies have examined the validity of claims-based diagnoses, with data from a variety of other sources serving as a gold standard. Lurie and colleagues[32] compared diagnoses from administrative claims for schizophrenia to medical record–based assessments by psychiatrists, finding them to have good specificity but lesser sensitivity. Geiger-Brown and colleagues[33] compared diagnoses from Medicaid claims to those from structured clinical interviews and patients' reports, finding agreement rates to be good for schizophrenia, fair for bipolar disorder, and fair to poor for other mental disorders. Despite these limitations, administrative data are often used in risk-adjustment models because of their relative affordability compared to other data sources.

Illness Severity

Rating scales assessing symptoms and functioning can provide detailed information about severity of illness. However, several obstacles limit their use for report cards or reimbursement. Currently, only a small proportion of mental health care systems has invested the resources necessary to incorporate rating scales into routine practice. Among those that have, there is limited agreement on which instruments should be used—which restricts opportunities to compare results across health care systems for benchmarking and quality improvement.[34] Alternatively, medical records contain data relevant to severity—including symptoms, functional impairments, and mental status exam findings—but documentation of this information is inconsistent, and abstracting records is labor intensive. Administrative databases are the least costly sources of severity data. The fifth digit of DSM-IV diagnostic codes provides information on severity of illness. For example, major depressive disorder can be coded as mild, moderate, severe, severe with psychotic features, in remission, or in partial remission. However, these descriptors are often not completed. An analysis of 1994–95 Medicaid claims for major depression from six states found that only 30.2% of outpatient claims and 64.5% of inpatient claims specified a fifth-digit code describing severity level.[24] Codes related to psychiatric disability from Medicaid, Medicare, and VA databases have also been used as severity indicators.

In contrast to medical conditions, which may have objective results from diagnostic testing that establish severity, the severity of M/SU conditions is ascertained through clinical observation and interview. These subjective processes may be more susceptible to error or gaming, particularly when the severity assessments are used to determine reimbursement rates. For instance, research evidence suggests that implementation of Medicare's prospective payment system was accompanied by upcoding of depressive disorders to a more severe subtype—a practice also seen in the coding of treatment outside of mental health care.[35] In addition, mental and substance-use conditions may not be documented in the medical record or on billing claims due to clinician or

patient concerns about stigmatization or the consequences for employment, insurance coverage, or legal status.[36]

Age

Age is one of a number of sociodemographic characteristics that may be included in risk-adjustment models. Medical and surgical outcomes generally worsen in a linear relationship with age, but the relationship between outcomes and age is more complex with mental disorders. Whereas the course of schizophrenia typically worsens following onset in the early to mid-20s, "positive" symptoms, such as psychosis, typically diminish in later years, and functional impairment tends to plateau.[37] Similarly, length of hospital stay for mental disorders does not increase linearly with age; on average, children and geriatric patients have longer hospital stays than adults.[38]

Social Support

Because individuals with SPMI are often unable to care fully for themselves, outcomes can be influenced by the amount of social support received. Risk-adjustment models draw on a variety of data sources to address social support, with different levels of validity and associated burden. Administrative databases typically record marital status. Medical records may describe whether a patient lives alone or with others, and if family members are involved in their care. Structured instruments provide a more consistent assessment of living status and social support, with commonly used categories that include independent residence; residence with family; group home; group home with staff; shelter; and homeless. Other systems have evaluated the degree to which a patient depends on clinical staff for support.[39,40]

Other Risk Factors

In the absence of easily obtained data on patient severity of illness, risk-adjustment models for hospital care have drawn on available data to serve as proxies. For example, involuntary legal status on admission is often used to characterize more severely ill patients. Although useful, such proxies should be used cautiously. Involuntary status may signal high acuity at admission rather than greater severity of illness over time. Furthermore, research studies have found that standards applied to involuntary commitment vary widely, even among facilities regulated by the same statutes.[41,42] Other data elements that have been used as proxies for severity include an unplanned discharge (e.g., eloped or against medical advice) and source of admission (e.g., another inpatient service, emergency room, outpatient office, home).

Prior utilization of mental health services is used in risk-adjustment models as an indicator of the severity and chronicity of mental illness. Several studies have found prior utilization to be significantly associated with outcomes of hospital or community-based care.[40,43,44] Of course, prior utilization reflects patient status only in part; it also incorporates provider practices and the availability of services.

Early risk-adjustment models in mental health included *current* treatment as a risk factor. Electroconvulsive therapy (ECT), physical restraints, and antipsychotic medications—interventions recommended for severely depressed, agitated, and psychotic patients, respectively—have been used as indicators of high severity of illness. Studies found that patients receiving one or more of these treatments had higher costs and utilization than patients who did not.[45–48] However, more-recent studies have heeded concerns that incorporating specific treatments into risk-adjustment models could lead to incentives to overuse them. These concerns may be particularly warranted for procedures that lack consensus among clinicians on indications for their use. Despite practice guidelines and government regulations, wide variations are seen in the use of both ECT and physical restraints.[49,50]

Other risk factors are specific to the outcome of interest. A report-card initiative to rank mental health service agencies in Vermont in terms of client criminal-justice outcomes used prior involvement in the criminal justice system as a risk factor in the risk-adjustment model.[43] A risk-adjustment model for workplace outcomes of depressed employees might include such risk factors as the number of days in a prior period absent from work.

RISK-ADJUSTMENT MODELS FOR MENTAL HEALTH AND SUBSTANCE-RELATED CARE

In general, risk-adjustment models assess the statistical association between the outcome and risk factors in a sample of patients. Stratification is the simplest method, useful when few risk factors are present. By dividing the sample into categories (i.e., strata) consisting of different presentations of a risk factor, stratification allows the outcome to be assessed within more homogeneous groups. For example, patient satisfaction is known to vary by age; hence, results are frequently stratified by adolescents, adults, and elderly survey respondents.

When numerous risk factors are present, multivariate regression analysis is used to assess the influence of patient characteristics (i.e., independent variables) on the outcome (the dependent variable). Regression analysis yields a summary statistic reflecting the model's predictive power. For continuous outcomes this statistic is typically an R^2 value, which ranges from 0 (no predictive power) to 1 (high predictive power). A c-statistic is often used to characterize the performance of models with a dichotomous outcome—for example, mortality.

In the absence of widely disseminated models for mental health and substance-related care, groups wishing to implement risk adjustment either develop a model de novo or search the research literature for related efforts. In this section, we review risk-adjustment models for M/SU conditions. We conducted searches of Index Medicus and PsycLit and examined citations to identify studies meeting the following criteria: (1) publication in a peer-reviewed journal between 1980 and 2002; (2) evaluation of one or more multivariate models used to risk-adjust comparisons of utilization, cost, or clinical outcomes of mental or substance-use conditions across providers; and (3) quantitative assessment of the proportion of variance explained by patient characteristics in the model (e.g., R^2 or c-statistic).

The sample population of each model was described in terms of its demographics (e.g., adult, children, and elderly), diagnostic groups, and health care status (e.g., plan enrollees, service utilizers, and level of care). Models were then classified on the basis of risk factors. Because one aim was to compare the performance of models based on administrative data to models drawing from more extensive data sources, we sorted risk factors into three categories. "Basic demographic" referred only to demographic data typically available in administrative health care databases, such as age, gender, and race. "Clinical status" included data obtained from rating scales, patient interviews, and other sources beyond claims-based diagnostic codes. "Other" included risk factors not classifiable under prior categories (e.g., discharged against medical advice) and sociodemographic data typically unavailable from administrative databases, such as living status, income, and education. Because including current treatment in risk-adjustment models is controversial, we categorized models containing these data separately. Under "% Variance" we provide the range and median R^2 results for models in each subgroup.

Tables 1 through 3 (pages 59–63) report the results. A total of 36 research articles met our review criteria, describing 72 models of utilization (Table 1), 74 models of expenditures (Table 2), and 15 models of clinical outcomes (Table 3). All but one model examined continuous outcomes; R^2 values or their equivalents are reported for these studies. One model examined a dichotomous outcome and assessed model performance in terms of a c-statistic, a metric that is not interpretable as a percentage of explained variance. Comparisons of the explanatory power from individual rows in Tables 1 through 3 should be made cautiously, as study characteristics other than risk factors influence the variance explained by multivariate models. Sample size and composition, setting, and the use of concurrent versus prospective modeling may also have an impact. Examining these studies in aggregate, however, we draw preliminary conclusions and describe illustrative examples below. In addition, we briefly describe risk-adjustment models that have been adopted to address real-world needs but that have not been formally assessed for their predictive capacity.

Risk Adjustment of Utilization

Risk-adjustment models of utilization for M/SU conditions are summarized in Table 1. Despite decades of "deinstitutionalization," the inpatient sector continues to consume a large proportion of the mental health care dollar,[6] and as a result, inpatient length of stay (LOS) has been the subject of the majority of risk-adjustment efforts in mental health care. Diagnostic-related groups (DRGs) were developed for the prospective payment system, a fixed-fee method of paying for inpatient stays that Medicare implemented in 1983. DRGs included 9 diagnostic categories for mental disorders and 5 for substance-use conditions. These 14 groups explained 3.2% to 15% of the variance in LOS, with marked clustering at the lower end of the range. The predictive ability of the M/SU DRGs was similar to the medical diagnostic groups but considerably lower than the surgical DRGs.[80] In mental health care, studies found that hospital type explained a significant proportion of the remaining variance, raising the question of whether facility type (e.g., general hospitals versus psychiatric specialty hospitals) was serving as a proxy for unmeasured differences in patient severity.[81] In part due to these concerns, DRGs were applied to psychiatric hospitalizations in medical wards of general hospitals, but specialty psychiatric hospitals and psychiatric units were permitted to apply for exemption.

In an effort to increase explanatory power, subsequent investigations augmented DRGs. Disease staging is a method that sorts diagnostic codes into subgroups of increasing severity and comorbidity.[51] Stage 1 for depressive disorders consists of dysthymic disorder (a mild, chronic form of depression), followed by dysthymia comorbid with a personality or substance-use disorder in Stage 2, and major depression in Stage 3. Clinically related groups (CRGs) separate diagnoses into hierarchies based on clinical features and age distribution.[51] Psychotic conditions are first divided by diagnosis (e.g., schizophrenia and bipolar disorder) and then by age group (under 18, 18–64, and 65 and older). Other initiatives supplemented diagnosis with data readily available from administrative databases, such as age, sex, marital status, comorbidity, number of previous hospitalizations, and discharge against medical advice.[47,52] While most of these efforts yielded some improvement over DRGs, none explained more than 15% of the variance in LOS.

More elaborate models have drawn on additional data collected from staff, medical records, or patient report, and attained a higher rate of explained variation.[39,45,53,63] The Psychiatric Severity of Illness Index (PSI) grouped inpatients into four categories on the basis of signs and symptoms;

TABLE 1. Risk-Adjustment Models for Mental Health and Substance-Related Utilization

Dependent variable Model[a]	Population	Independent variables: Diagnosis	Basic demographic	Clinical status	Treatment	Other	% Variance (median, range)
Inpatient length of stay, M/SUD							
DRG[48,51–55]	Inpatients w/M/SUD	X					5.8 (0.1–28.4)
DRG[39,53]	Inpatients w/MD						
DRG[53]	Inpatients w/SUD						
Diagnosis;[56,57] Disease Staging;[51] RGN[58]	Inpatients w/M/SUD						
Diagnostic Categories[59]	Inpatients w/M/SUD or neurological disorders						
DRG;+.[47] Gates[60]	Inpatients w/M/SUD	X				X	16.9 (16.6–17.2)
Schumacher et al.;[54] Paths[60]	Inpatients w/M/SUD		X				6.5 (5.3–7.8)
Schumacher et al.;[54] CRG[51]	Inpatients w/M/SUD	X	X				8.5 (5.4–12.3)
DRG+;[52] Cyr & Haley;[61] Paths+Gates;[60] DRG+;[47] ADRG;[47] ADRG+[47]	Inpatients w/M/SUD	X	X			X	21.4 (7.3–30.7)
PSI[39]	Inpatients w/MD			X			26.8 (18.5–35.0)
Strain Ratio[62]	Inpatients w/M/SUD						
DRG+PSI[39]	Inpatients w/MD	X		X			20.9 (13.6–40.0)
Creed et al.[56]	Inpatients w/M/SUD						
CSI[63]	Adult inpatients w/psychotic disorders						
Creed et al.[56]	Inpatients w/M/SUD	X		X		X	34.1 (31.6–36.6)
DRG+[62]	Inpatients w/M/SUD	X	X	X			33.6 (15.6–62.0)
CSI[63]	Adult inpatients w/depressive disorders						
Schumacher et al.;[54] Choca et al.[64]	Inpatients w/M/SUD	X	X	X		X	15.9 (7.8–24.0)
Creed et al.;[56] Essock-Vitale[48]	Inpatients w/M/SUD	X			X	X	37.5 32.9–42.0)
Creed et al.[56]	Inpatients w/M/SUD			X	X	X	36.2 (10.0–71.2)

(Continued on next page)

TABLE 1. Risk-Adjustment Models for Mental Health and Substance-Related Utilization *(Continued)*

Dependent variable Model[a]	Population	Independent variables: Diagnosis	Basic demographic	Clinical status	Treatment	Other	% Variance (median, range)
CPSI;[45] CPSI+[46]	Inpatients w/schizophrenia or affective disorder						
CPSI;[45] CPSI+[46]	Inpatients w/schizophrenia						
CPSI;[45] CPSI+[46]	Inpatients w/affective disorder						
Taube et al.[57]	Inpatients w/M/SUD	X	X		X	X	16.9
CPSI+[46]	Inpatients w/schizophrenia or affective disorder		X	X	X	X	32.6 (14.6–38.9)
Schumacher et al.[54]	Elderly inpatients w/M/SUD						
Schumacher et al.[54]	Adult inpatients w/M/SUD						
PPC[53]	Inpatients w/MD	X	X	X	X	X	18.0 (11.5–31.6)
PPC[53]	Inpatients w/SUD						
PPC[53]	Inpatients w/M/SUD						
Annual inpatient days, M/SUD							
Diagnosis[56]	Inpatients w/M/SUD	X					11.0
Creed et al.[56]	Inpatients w/M/SUD	X				X	18.9
Creed et al.[56]	Inpatients w/M/SUD	X			X	X	21.4
Inpatient nursing hours, M/SUD							
DRG[48]	Inpatients w/M/SUD	X					6.0
Essock-Vitale[48]	Inpatients w/M/SUD	X			X	X	14.0
Inpatient readmission, M/SUD							
Paths[60]	Inpatients w/M/SUD		X				13.9
Gates[60]	Inpatients w/M/SUD	X				X	12.8
Hendryx[8]	Adult inpatients w/M/SUD		X			X	0.66[b]
Paths+Gates[60]	Inpatients w/M/SUD	X	X			X	36.0 (26.7–36.0)
Peterson et al.;[65] Phibbs et al.[66]	Inpatients w/SUD						
Outpatient visits, M/SUD							
Wood & Beardmore[67]	Adult outpatients w/M/SUD	X	X	X		X	Each <1.0
Outpatient visits, all health care							
DCG/HCC[68]	Adult in/outpatients w/SUD	X	X				24.6
DCG/HCC[68]	Adult in/outpatients w/SUD	X	X	X			27.8 (24.6–27.9)

ADRG, Alternative DRG; CRG, Clinically Related Group; CPSI, Computerized PSI; CSI, Computerized Severity Index; DCG/HCC, Diagnostic Cost Group/Hierarchical Coexisting Condition; DRG, Diagnosis-Related Group; MD, mental disorder; M/SUD, mental or substance use disorder; PPC, Psychiatric Patient Class; PSI, Psychiatric Severity Index; RGN, Refined Group Number; SUD, substance use disorder.

[a]When a model has a name, that name is used; when a model has no name, the name of the author is used.

[b]c-statistic.

Table 2. Risk-Adjustment Models for Mental Health and Substance-Related Costs

Dependent variable Model[a]	Population	Independent variables: Diagnosis	Basic demographic	Clinical status	Treatment	Other	% Variance (median, range)
Costs, health care							
Diagnosis[69]	In/outpatients w/M/SUD	X					5.0 (3.0–7.0)
Diagnosis[70]	Adult long-stay inpatients w/MD						
Ettner & Notman[71]	Adults w/health care utilization		X				1.3 (0.4–2.2)
Ettner & Notman[71]	Children w/health care utilization						
ACG[71]	Adults w/health care utilization	X	X				2.4 (0.9–31.5)
ACG;[71] Ettner & Notman[71]	Children w/health care utilization						
DCG/HCC[68]	Adult in/outpatients w/SUD						
Disability[69]	In/outpatients w/M/SUD					X	1.6
GAF[69]	In/outpatients w/M/SUD			X			3.1
GAF+[69]	In/outpatients w/M/SUD			X		X	2.5
DCG/HCC[68]	Adult in/outpatients w/SUD	X	X	X			32.4 (31.6–32.4)
Knapp et al.[70]	Adult long-stay inpatients w/MD		X	X		X	35.0
Chisholm et al.[72]	Adult residential-care patients w/M/SUD	X	X	X		X	19.8 (18.5–21.1)
Costs, M/SUD							
Demographic[73]	Health plan enrollees		X				0.3 (-0.3–0.8)
Demographic[73]	Adult health plan enrollees						
Demographic[73]	Child health plan enrollees						
Demographic[74]	Health care utilizers						
Ettner & Notman[71]	Adults w/health care utilization						
Ettner & Notman[71]	Children w/health care utilization						
Demographic[75]	Adult in/outpatients w/M/SUD						

(Continued on next page)

Table 2. Risk-Adjustment Models for Mental Health and Substance-Related Costs *(Continued)*

Dependent variable Model[a]	Population	Independent variables: Diagnosis	Basic demographic	Clinical status	Treatment	Other	% Variance (median, range)
Kapur[75]	Adult in/outpatients w/M/SUD		X			X	0.9
ACG;[73] ADG;[73] HCC;[73] Comorbidity[73]	Health plan enrollees	X	X				6.9 (1.0–19.3)
ACG;[73] ADG;[73] HCC;[73] Comorbidity[73]	Adult health plan enrollees						
ACG;[73] ADG;[73] HCC;[73] Comorbidity[73]	Child health plan enrollees						
ACG;[71] Ettner & Notman[71]	Adults w/health care utilization						
ACG;[71] Ettner & Notman[71]	Children w/health care utilization						
Disability Presence;[74] Disability Type;[74] ACG+;[74] ADG+;[74] HCC+[74]	Health care utilizers						
DCG/HCC[68]	Adult in/outpatients w/SUD						
DCG/HCC[68]	Adult in/outpatients w/SUD	X	X	X			20.0 (18.1–20.0)
Diagnosis;[75] Kapur[75]	Adult in/outpatients w/M/SUD	X	X	X		X	8.7 (2.7–10.3)
Inpatient costs, M/SUD							
DRG;[51] Disease Staging[51]	Inpatients w/M/SUD	X					4.8 (3.6–10.5)
Diagnosis+[69]	Adult in/outpatients w/M/SUD	X				X	38.6
CRG[51]	Inpatients w/M/SUD	X	X				5.0 (4.4–9.6)
SCIPP[40]	Adult inpatients w/M/SUD	X		X	X	26.3	
GAF+[69]	Adult in/outpatients w/M/SUD			X		X	25.0 (11.4–38.6)
LPPC[76]	Adult long-stay inpatients w/M/SUD						

ACG, Ambulatory Care Group; DCG/HCC, Diagnostic Cost Group/Hierarchical Coexisting Condition; GAF, Global Assessment of Functioning; LPPC, Long-Stay Psychiatric Patient Classification; MD, mental disorder; M/SUD, mental or substance use disorder; SCIPP, System for Classification of In-Patient Psychiatry; SUD, substance use disorder.

[a]When a model has a name, that name is used; when a model has no name, the name of the author is used.

Table 3. Risk-Adjustment Models for Clinical Outcomes of Mental and Substance Use Disorders

Dependent variable Model[a]	Population	Independent variables: Diagnosis	Basic demographic	Clinical status	Treatment	Other	% Variance (median, range)
Functioning							
SF-36 MCS[77]	Outpatients w/depression		X				27.0
GAF+[78]	Adult inpatients w/M/SUD	X				X	14.8
SF-12 MH score+[79]	Adult outpatients w/M/SUD	X	X				3.0
SF-12 MH score+[79]	Adult outpatients w/schizophrenia			X			38.0 (20.0–56.0)
SF-12 MH score+[44]	Adult outpatients w/M/SUD						
SF-12 MH score+[44,79]	Adult outpatients w/M/SUD	X		X			39.5 (23.0–56.0)
SF-36 MCS[77]	Outpatients w/depression		X	X			37.0
Symptom severity							
D-ARK[77]	Outpatients w/depression			X			20.0
D-ARK[77]	Outpatients w/depression	X		X		X	19.0
Satisfaction							
CSQ+[44]	Adult outpatients w/M/SUD			X			19.0
CSQ+[44]	Adult outpatients w/M/SUD	X	X	X			23.0
BHRS[78]	Inpatient adults w/M/SUD		X	X		X	4.8.0
Quality of life							
QOLI+[44]	Adult outpatients w/M/SUD			X			31.0
QOLI+[44]	Adult outpatients w/M/SUD		X	X			34.0

BHRS, Behavioral Healthcare Rating of Satisfaction Scale; CSQ, Client Satisfaction Questionnaire; D-ARK, Depression–Arkansas Scale; GAF, Global Assessment of Functioning; MH, mental health; MCS, mental component summary; M/SUD, mental or substance use disorder; PCS, physical component summary; QOLI, Qualify of Life Index; SF-12[-36], Medical Outcomes Study Short Form–12[–36].

[a]When a model has a name, that name is used; when the model has no name, the name of the author is used.

complications; comorbid conditions; availability of, and need for, support; and treatment response. The four PSI groups explained 34% to 50% of the variance in LOS.[39] Later versions were revised to reduce burden, eliminate circular items, and focus on specific disorders—but at the cost of a decrease in predictive power.[45,63]

Risk Adjustment of Costs

Risk-adjustment models of costs for mental health and substance-related services are summarized in Table 2. Studies examining inpatient costs in addition to inpatient LOS yielded results corresponding closely to those described above, under inpatient utilization. Other studies, described here, specifically examined costs. The System for Classification of Inpatient Psychiatry (SCIPP) was developed by the Ontario Joint Policy and Planning Committee to explain the variance in nonphysician staff costs attributable to differences among psychiatric inpatients. Drawing on data collected from the Resident Assessment Instrument–Mental Health (RAI-MH),[82] the SCIPP used more than 80 sociodemographic and clinical attributes to predict approximately 26% of variance.[40,17] The Ontario Health Ministry is reviewing this system for use in reimbursing psychiatric hospitals for inpatient stays. The Center for Medicaid and Medicare Services has funded U.S. investigators involved in the SCIPP to develop a shorter, less burdensome system for inpatient psychiatry. The resulting instrument, the Case Mix Assessment Tool (CMAT), will soon be ready for field testing.

Researchers have evaluated whether risk-adjustment systems developed for capitation of medical and surgical care can adequately adjust for mental health costs. Adjusted clinical groups (ACG), adjusted diagnostic groups (ADG), diagnostic cost groups (DCG), and hierarchical condition categories (DCG-HCC) are each based on demographic and diagnostic data. Ettner and colleagues[71,73] found that none of the systems explained more than 10% of the variance in total health care costs for populations that included individuals with M/SU conditions. In general, the models underestimated total health care costs for individuals with M/SU conditions and overestimated costs for those without them. Modifying the commercial systems by adding more precise M/SU diagnoses resulted in somewhat improved predictions of M/SU costs.[83] Examining approaches to risk adjustment for VA patients with substance-use conditions, Rosen and colleagues[68] applied DCG-HCCs to predict variance in service days (a proxy for costs), finding them to explain a lower proportion of variance in substance-use conditions than has generally been found in non-M/SU populations. Augmenting the system by creating distinct categories for substance-use conditions and subdividing diagnostic groups based on severity did not lead to significant improvement in the model.

A number of states have risk-adjusted reimbursement rates for mental health services. Maryland has used ACGs to adjust capitation rates for mental health services under Medicaid, and Rhode Island uses age, sex, and eligibility status. Indiana evaluated risk-adjusted models for case-rate reimbursement for children and adults receiving mental health services financed by the state's Department of Mental Health. Using hierarchical regression models, they identified subgroups with similar service costs in the 90-day period after initial assessment. For example, adults with severe mental illness were subdivided into nine groups based on three diagnostic clusters and three functional levels. Applying regression results to historical costs, they established case rates for the nine groups ranging from $1,194 to $7,981.[84] The explained variance of this model has not been evaluated.

Risk Adjustment of Clinical Outcomes

Risk-adjustment models of clinical outcomes of M/SU conditions are summarized in Table 3. Several research studies demonstrate that risk adjustment influences outcome-based ratings of mental health facilities.[44,77,78] Hendryx and colleagues[44] compared and ranked six community mental health agencies on the basis of risk-adjusted outcomes—in particular, patient satisfaction, functioning, and health-related quality of life. Several risk factors predicted one or more of these outcomes: "severe diagnoses" (schizophrenia, bipolar disorder, and major depression), substance abuse, age, and baseline functioning and quality of life. Rankings among agencies varied, however, depending on which outcome measure was used.

Other models have focused on patient populations with specific disorders. Kramer and colleagues[77] used detailed demographic and clinical data from the University of Arkansas Depression Outcome Module and the Medical Outcomes Study Short Form–36 (SF-36) to develop a risk-adjustment model for major depression. Baseline depression severity, income, and medical comorbidity explained 26% of the variance in depression severity at three months. The addition of baseline functional status (SF-36 physical and mental health summary scores) increased the R^2 to 0.38.

More than half of state mental health authorities have implemented routine outcome measurement to assess symptoms or functioning of individuals with SPMI,[85] but few states adjust these data for risk. Massachusetts and Indiana perform stratification of outcome-measure results. The Vermont mental health authority has employed risk adjustment in comparisons of performance of community mental health care programs. Outcomes assessed include consumer satisfaction and pre/post differences in hospitalization and criminal justice involvement following community-based treatment.[43,86–88]

Risk Adjustment of Clinical Processes

Our literature review found no published studies evaluating the adequacy of risk adjustment for clinical process measures in mental health care. In addition, a national survey of 246 process measures developed for quality assessment and improvement in mental health found only 7% to be risk adjusted—13 with multivariate modeling and 5 with stratification.[89] For example, the Washington Circle measure results are stratified by age, while some HEDIS results are stratified by age, sex, and gender. The NRI performance-measurement system, used by more than 240 psychiatric hospitals, has developed multivariate adjustment models for three measures: 30-day readmission rates, percentage of inpatients restrained, and percentage secluded. In addition to diagnostic and sociodemographic data, risk factors include legal status, referral source, living arrangement, clinical focus, and average LOS. The VA's Northeast Program Evaluation Center risk-adjusts the results of hospital report cards by sociodemographic and diagnostic information, receipt of VA compensation, discharge GAF score, and the patient's distance from the nearest VA facility. Measures examine timeliness, continuity, and intensity of treatment. An analysis of the model found explained variance to average 2.7% across measures (0.3–9.7%).[90]

Risk Adjustment in Mental Health Services Research

Mental health services researchers frequently employ statistical adjustment in observational studies examining the impact of organizational and financial change on quality and outcomes.[91,92] Many of these studies did not compare care across providers or did not provide information about the statistical model's predictive power. Services research may nonetheless serve as an incubator of adjustment methods that can be evaluated for applied purposes. Recent studies highlight advanced approaches. Schoenbaum and colleagues[93] employed instrumental-variable analysis to control for unmeasured differences between groups in comparing outcomes between patients receiving appropriate versus inappropriate treatment for depression. Dickey and colleagues[94] used propensity scores to compare guideline adherence between managed and unmanaged cohorts in a study of the quality of care for Medicaid beneficiaries with schizophrenia.

Comparison of Explained Variance by Data Source

We calculated both the mean rate of explained variance for risk-adjustment models based on administrative data and the mean for models that used clinically richer data sources (e.g., rating scales or medical records). Models based on the diagnostic and sociodemographic information found in administrative claims explained, on average, only 6.7% of variance, whereas models using more-detailed data sources explained a more robust 22.8%.

DISCUSSION

Considerable progress has been made on risk adjustment for M/SU conditions since DRGs sparked wide interest in the subject two decades ago. Risk factors have advanced beyond diagnosis to encompass more detailed clinical and sociodemographic information. Rating scales have been adapted to assess severity of mental illness in clinical populations. Multivariate models have been developed to adjust clinical, utilization, and financial outcomes of mental health care. Over time these models have advanced in terms of methodologic rigor and predictive ability. Public and private organizations are beginning to apply risk adjustment to reimbursement systems and to comparative report cards of provider performance.

Risk adjustment for mental health and substance-related care remains at an early stage of development. While many models have been tested, few are in routine use. Risk adjustment for reimbursement of inpatient care—the most extensively studied area—has yet to produce a model that combines adequate explanatory power with an acceptable data-collection burden. Researchers evaluating risk adjustment for capitation have concluded that current models do not adequately diminish incentives to avoid or undertreat individuals with mental illness. Gaining favor instead are mixed payment systems, which provide reimbursement partially on the basis of actual costs and partially on adjusted rates. These systems provide some incentive to contain costs but lessen the motivation to restrict enrollment or treatment.[83,95] Report cards that rank providers on the basis of patient outcomes also require further development, as rankings have proven sensitive to the choice of adjustment method, the diversity of the clinical population, and the outcome selected.

Taken together, the array of outcomes (financial, utilization, clinical, and quality) and the diversity of conditions, treatments, and settings within the mental health care system form a multidimensional matrix of factors requiring risk adjustment. Only a few cells within this matrix have been filled in thus far. For example, disorder-specific models are currently limited to depression (five) and schizophrenia (four). Fewer than 9% of all models focus on children and adolescents. Further development of methods for risk adjustment, along with broader application and testing of models for M/SU conditions, constitutes a substantial agenda for the years to come.

The application of more sophisticated analytic methods, such as instrumental-variable analysis and the use of

propensity scores, may lead to models with greater accuracy. In addition, hierarchical models can account for correlated observations when subjects are nested within groups, such as patients nested within clinician caseloads, which are, in turn, within clinics and delivery systems. Decision-tree or classification and regression-tree (CART) models can take into account interactions that occur among risk factors.

We found that risk-adjustment models relying on administrative data alone explained less than one-third of the variance explained by models incorporating data from medical records or clinical assessment instruments. This finding is fairly consistent across the models in our sample; nonetheless, differences in sample size and in the dependent variables studied may have contributed to the differences in variance observed. The relatively low explanatory power of models based on administrative data suggests that risk adjustment for mental health care will need to tap clinically richer sources of data.

In an era of resurgent increases in health care costs, it is unlikely that providers, purchasers, and payers will individually lead efforts to expand data-collection requirements any time soon. Instead, stakeholders in the mental health system will need to work together to improve the quality of administrative data, to review and refocus resources dedicated to chart review, and to develop consensus on clinical assessment instruments. Use of administrative data for risk adjustment is also complicated by rules protecting patient confidentiality—often more strict for mental health care—and tightened further under Health Insurance Portability and Accountability Act of 1996.

In the long term, electronic health records (EHRs) may provide the basis for enriching administrative data sets by capturing key clinical variables in a consistent format. However, EHRs are unlikely to be in widespread use soon, particularly among providers with limited resources, such as private practice clinicians, community mental health centers, and small hospitals. In the meantime, accelerated efforts are needed to establish a minimal data set of clinical information crucial to risk adjustment, reimbursement, and assessment of quality and outcomes of M/SU conditions. The aim of a minimal data set would be to standardize clinical information that providers obtain, thereby increasing its consistency, accuracy, and usefulness.

Payers, regulators, and managed behavioral health organizations already mandate chart review for numerous incompatible quality-assessment activities.[23] These organizations need to find common ground with other stakeholders in mental health care on a manageable number of quality measures.[10] In order to provide crucial information on illness severity and M/SU outcomes, clinicians and methodologists need to achieve consensus on the selection of clinical instruments for assessing mental health symptoms, functioning, and general health status in clinical practice. Nearly a dozen severity-rating scales are available for depression alone. At a Depression Diagnosis and Severity Measure Consensus Meeting in October 2002, the American College of Physicians, American Psychiatric Association, and American Academy of Family Practice initiated a dialogue that envisioned the eventual adoption of a single instrument. Like many sources of information for risk-adjustment models, the selection of this instrument will inevitably reflect a compromise between optimal measurement properties and a feasible burden of use.

REFERENCES

1. Donabedian A. Explorations in quality assessment and monitoring: the definition of quality and approaches to its assessment. Ann Arbor, MI: Health Administration, 1980.
2. Iezzoni LE. Risk adjustment for measuring healthcare outcomes. 3rd ed. Chicago: Health Administration, 2003.
3. Kessler RC, Chiu WT, Demler O, Merikangas KR, Walters EE. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry 2005;62:617–27.
4. Narrow W, Rae D, Robins L, Regier D. Revised prevalence estimates of mental disorders in the United States. Arch Gen Psychiatry 2002;59:115–23.
5. Substance Abuse and Mental Health Services Administration. Results from the 2005 National Survey on Drug Use and Health: national findings. Rockville, MD: SAMHSA, Office of Applied Studies, 2006.
6. Mark T, McKusick D, King E, Harwood H, Genuardi J. National expenditures for mental health, alcohol and other drug abuse treatment. Rockville, MD: U.S. Department of Health and Human Services, 1996.
7. Ettner S, Frank R. The use of risk adjustment to set capitation rates. In: Mandersheid RW, Henderson MJ, eds. Mental health, United States, 1998. Washington, DC: Government Printing Office, 1998.
8. Hendryx M. Introduction: risk-adjustment issues in mental health services. J Behav Health Serv Res 2001;28:225–34.
9. McGlynn EA. Choosing and evaluating clinical performance measures. Jt Comm J Qual Improv 1998;24:470–9.
10. Hermann RC, Palmer RH. Common ground: a framework for selecting core quality measures. Psychiatr Serv 2002;53:281–7.
11. National outcome measures (NOMs) for co-occurring disorders. 2006. http://www.oas.samhsa.gov/NOMsCoOccur2k6.pdf
12. Rosenblatt A, Attkisson C. Assessing outcomes for sufferers of severe mental disorder: a conceptual framework and review. Eval Program Plann 1993;16:347–63.
13. IsHak W, Burt T, Sederer L, eds. Outcome measurement in psychiatry: a critical review. Washington DC: American Psychiatric Publishing, 2002.
14. Task Force for the Handbook of Psychiatric Measures. Handbook of psychiatric measures. Washington, DC: American Psychiatric Association, 2000.

15. Hamilton M. Development of a rating scale for primary depressive illness. Br J Soc Clin Psychol 1967;6:278–96.
16. Derogatis L. SCL-90-R, Brief Symptom Inventory, and matching clinical rating scales. In: Maruish M, ed. Psychological testing, treatment planning, and outcome assessment. New York: Erlbaum, 1994.
17. Frank E, Prien R, Jarrett J, et al. Conceptualization and rationale for consensus definitions of terms in major depressive disorder: response, remission, recovery, relapse, and recurrence. Arch Gen Psychiatry 1991;48:851–5.
18. Endicott J, Spitzer RL, Fleiss JL, Cohen J. The Global Assessment Scale: a procedure for measuring overall severity of psychiatry disturbance. Arch Gen Psychiatry 1976;33:766–71.
19. Epping-Jordan A, Ustun T. The WHODAS II: leveling the playing field for all disorders. WHO Ment Health Bull 2000;6: 4–5.
20. McLellan AT, Kushner H, Metzger D, et al. The fifth edition of the Addiction Severity Index. J Subst Abuse Treat 1992;9:199–213.
21. Cuffel B, Fischer E, Owen Jr. R, Smith G Jr. An instrument for measurement of outcomes of care for schizophrenia: issues in development and implementation. Eval Health Prof 1997;20:96–108.
22. Eisen SV. Behavior and Symptom Identification Scale (BASIS-32). Belmont, MA: McLean Hospital, 1986.
23. Hermann R, Leff H, Palmer R, et al. Quality measures for mental health care: results from a national inventory. Med Care Res Rev 2000;57 suppl 2:136–54.
24. Hermann RC. Improving mental healthcare: a guide to measurement-based quality improvement. Washington, DC: American Psychiatric Press, 2005.
25. Selecting indicators for the quality of mental health care at the health systems level in OECD countries. Organization for Economic Co-operation and Development, 2004. http://www.oas.samhsa.gov/NOMsCoOccur2k6.pdf
26. Smoller J, McLean R, Otto M, Pollack M. How do clinicians respond to patients who miss appointments? J Clin Psychiatry 1998;59:330–8.
27. Frank RG, McGuire TG, Newhouse JP. Risk contracts in managed mental health care. Health Aff (Millwood) 1995;14:50–64.
28. Lyons J, O'Mahoney M, Miller S, Neme J, Kabat J, Miller F. Predicting readmission to the psychiatric hospital in a managed care environment: implications for quality indicators. Am J Psychiatry 1997;154:337–40.
29. Institute of Medicine. Unequal treatment: confronting racial and ethnic disparities in health care. Washington, DC: National Academy Press, 2002.
30. American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 4th ed. Washington, DC: APA, 1994.
31. Schneeweiss S, Maclure M. Use of comorbidity scores for control of confounding in studies using adminstrative databases. Int J Epidemiol 2000;29:891–8.
32. Lurie N, Popkin M, Dysken M, I, Finch M. Accuracy of diagnoses of schizophrenia in Medicaid claims. Hosp Community Psychiatry 1992;43:69–71.
33. Geiger-Brown J, Steinwachs D, Fahey M, Skinner E, Lehman A. The concordance of Medicaid claims, survey and research interview diagnoses for patients with serious and persistent mental illnesses. On file with author.
34. Hermann RC, Provost SE. Interpreting measurement data for quality improvement: means, norms, benchmarks, and standards. Psychiatr Serv 2003;54:655–7.
35. Kiesler C, Simpkins C. Changes in diagnostic case mix in psychiatric care in general hospitals, 1980–1985. Gen Hosp Psychiatry 1992;14:156–61.
36. Rushton JL, Felt BT, Roberts MW. Coding of pediatric behavioral and mental disorders. Pediatrics 2002;110(1 pt 1):e8.
37. Carpenter W, Buchanan R. Schizophrenia. In: Kaplan H, Sadock B, eds. Comprehensive textbook of psychiatry/VI. Baltimore, MD: Williams & Wilkins, 1995:889–902.
38. Average length of stay for discharges from short-stay hospitals, by age and first-listed diagnosis: United States, 1998. National Center for Health Statistics, 1998. http://www.cdc.gov/nchs/data/hdasd/sr13_148t9.pdf
39. Horn S, Chambers A, Sharkey P, Horn R. Psychiatric severity of illness: a case mix study. Med Care 1989;27:69–84.
40. Hirdes J, Fries B, Botz C, Ensley C, Marhaba M, Perez E. The System for Classification of In-Patient Psychiatry (SCIPP): a new case-mix methodology based on the RAI–Mental Health (RAI-MH). On file with author.
41. Rubin W, Snapp M, Panzano P, Taynor J. Variation in civil commitment processes across jurisdictions: an approach for monitoring and managing change in mental health systems. J Ment Health Adm 1996;23:375–88.
42. Faulkner L, McFarland B, Bloom J. An empirical study of emergency commitment. Am J Psychiatry 1989;146:182–6.
43. Banks S, Pandiani J, Bramley J. Approaches to risk-adjusting outcome measures applied to criminal justice involvement after community service. J Behav Health Serv Res 2001;28:235–46.
44. Hendryx M, Dyck D, Srebnik D. Risk-adjusted outcome models for public mental health outpatient programs. Health Serv Res 1999;34:171–95.
45. Stoskopf C, Horn S. The Computerized Psychiatric Severity Index as a predictor of inpatient length of stay for psychoses. Med Care 1991;29:179–95.
46. Stoskopf C, Horn S. Predicting length of stay for patients with psychoses. Health Serv Res 1992;26:743–66.
47. Taube C, Lee E, Forthofer R. Diagnosis-Related Groups for mental disorders, alcoholism, and drug abuse: evaluation and alternatives. Hosp Community Psychiatry 1984;35:452–5.
48. Essock-Vitale S. Patient characteristics predictive of treatment costs on inpatient psychiatric wards. Hosp Community Psychiatry 1987;38:263–9.
49. Hermann R, Dorwart R, Hoover C, Brody J. Variation in ECT use in the United States. Am J Psychiatry 1995;152:869–75.
50. Okin R. Variation among state hospitals in use of seclusion and restraint. Hosp Community Psychiatry 1985;36:648–52.
51. Mitchell J, Dickey B, Liptzin B, Sederer L. Bringing psychiatric patients into the Medicare prospective payment system: alternatives to DRGs. Am J Psychiatry 1987;144:610–5.
52. English J, Sharfstein S, Scherl D, Astrachan I. Diagnosis-Related Groups and general hospital psychiatry: the APA study. Am J Psychiatry 1986;143:131–9.

53. Ashcraft M, Fries B, Nerenz D, et al. A psychiatric patient classification system: an alternative to diagnosis-related groups. Med Care 1989;27:543–57.
54. Schumacher D, Namerow M, Parker B, Fox P, Kofie V. Prospective payments for psychiatry—feasibility and impact. N Engl J Med 1986;315:1331–6.
55. Taube C, Lee E, Forthofer R. DRGs in psychiatry: an empirical evaluation. Med Care 1984;22:597–610.
56. Creed F, Tomenson B, Anthony P, Tramner M. Predicting length of stay in psychiatry. Psychol Med 1997;27:961–6.
57. Taube C, Goldman H, Lee E. Use of specialty psychiatric settings in constructing DRGs. Arch Gen Psychiatry 1988;45:1037–40.
58. Wellock C. Is a diagnosis-based classification system appropriate for funding psychiatric care in Alberta? Can J Psychiatry 1995;40:507–13.
59. McCrone P. Allocating resources in psychiatric hospitals according to casemix. Psychiatr Bull 1994;18:212–3.
60. Gruber J. Paths and gates: the sources of recidivism and length of stay on a psychiatric ward. Med Care 1982;20:1197–208.
61. Cyr J, Haley G. Use of demographic and clinical characteristics in predicting length of psychiatric hospital stay: a final evaluation. J Consult Clin Psychol 1983;51:637–40.
62. Gordon R, Vijay J, Sloate S, Burkey R, Gordon K. Aggravating stress and functional level as predictors of length of psychiatric hospitalization. Hosp Community Psychiatry 1985;36:773–4.
63. Durbin J, Goering P, Pink G, Murray M. Classifying psychiatric inpatients: seeking better measures. Med Care 1999;37:415–23.
64. Choca J, Peterson A, Shanley L, Richards H, Mangoubi E. Problems in using statistical models to predict psychiatric length of stay: an illustration. Hosp Community Psychiatry 1988;39:195–7.
65. Peterson K, Swindle R, Phibbs C, Recine B, Moos R. Determinants of readmission following inpatient substance abuse treatment: a national study of VA programs. Med Care 1994;32:535–50.
66. Phibbs C, Swindle R, Recine B. Does case-mix matter for substance abuse treatment? A comparison of observed and case mix-adjusted readmission rates for inpatient substance abuse treatment in the Department of Veterans Affairs. Health Serv Res 1997;31:775–1.
67. Wood W, Beardmore D. Prospective payment for outpatient mental health services: evaluation of diagnosis-related groups. Community Ment Health J 1986;22:286–91.
68. Rosen A, Loveland S, Anderson J, Hankin C, Breckenridge J, Berlowitz D. Diagnostic cost groups (DCGs) and concurrent utilization among patients with substance abuse disorders. Health Serv Res 2002;37:1079–103.
69. Leslie D, Rosenheck R, White W. Capitated payments for mental health patients: a comparison of potential approaches in a public sector population. J Ment Health Policy Econ 2000;3:35–44.
70. Knapp M, Beecham J, Fenyo A, Hallam A. Community mental health care for former hospital inpatients: predicting costs from needs and diagnoses. Br J Psychiatry 1995;166(27 suppl):10–8.
71. Ettner S, Notman E. How well do ambulatory care groups predict expenditures in mental health and substance abuse patients? Adm Policy Ment Health 1997;24:339–57.
72. Chisholm D, Knapp M, Asin J, Lelliott P, Audini B. The mental health residential care study: predicting costs from resident characteristics. Br J Psychiatry 1997;170:37–42.
73. Ettner S, Frank R, McGuire T, Newhouse J, Notman E. Risk adjustment of mental health and substance abuse payments. Inquiry 1998;35:223–40.
74. Ettner S, Frank R, Mark R, Smith M. Risk adjustment of capitation payments to behavioral health care carve-outs: how well do existing methodologies account for psychiatric disability. Health Care Manag Sci 2000;3:159–69.
75. Kapur KYAMD. Risk adjustment for high utilizers of public mental health care. J Ment Health Policy Econ 2000;3:129–37.
76. Fries B, Nerenz D, Falcon S, Ashcraft M, Lee C. A classification system for long-staying psychiatric patients. Med Care 1990;28:311–23.
77. Kramer T, Evans R, Landes R, Mancino M, Booth B, Smith R. Comparing outcomes of routine care for depression: the dilemma of case-mix adjustment. J Behav Health Serv Res 2001;28:287–300.
78. Dow M, Boaz T, Thornton D. Risk adjustment of Florida mental health outcomes data: concepts, methods, and results. J Behav Health Serv Res 2001;28:258–72.
79. Hendryx M, Teague G. Comparing alternative risk-adjustment models. J Behav Health Serv Res 2001;28:247–57.
80. Frank R, Lave J. The psychiatric DRGs: are they different? Med Care 1985;23:1148.
81. Horgan C, Jencks S. Research on psychiatric classification and payment systems. Med Care 1987;25(9 suppl):S22–36.
82. Hirdes J, Marhaba M, Smith T, et al. Development of the Resident Assessment Instrument–Mental Health (RAI-MH). Hosp Q 2000;4:44–51.
83. Ettner SL, Frank RG, McGuire TG, Hermann RC. Risk adjustment alternatives in paying for behavioral health care under Medicaid. Health Serv Res 2001;36:793–811.
84. DeLiberty R, Newman F, Ward E. Risk adjustment in the Hoosier Assurance Plan: impact on providers. J Behav Health Serv Res 2001;28:301–18.
85. NASMHPD Research Institute. Implementation of the NASMHPD framework of mental health performance measures by states to measure community performance: 2001. Alexandria, VA: NASMHPD Research Institute, 2002.
86. Banks S, Pandiani J, Schacht L, Gauvin L. A risk-adjusted measure of hospitalization rates for evaluating community mental health program performance. Adm Policy Ment Health 1999;26:269–79.
87. Pandiani J, Banks S, Schacht L. Using incarceration rates to measure mental health program performance. J Behav Health Serv Res 1998;25:300–11.
88. Bramley J, Maynard A, Pandiani J. Evaluating and researching mental health programs from different points of view. Presented at the University of South Florida 14th Annual Research Conference—A System of Care for Children's Mental Health: Expanding the Research Base, Tampa, FL, February 2001. http://healthvermont.gov/mh/docs/pips/2001/pip030901.pdf

89. Hermann RC, Leff HS, Palmer RH, et al. Quality measures for mental health care: results from a national inventory. Med Care Res Rev 2000;57(suppl 2):135–54.
90. Department of Veterans Affairs Office of Quality and Performance. Department of Veterans Affairs national mental health program performance monitoring system: fiscal year 1999 report. West Haven, CT: Department of Veterans Affairs, 1999.
91. Wells K, Rogers W, Davis L, et al. Quality of care for hospitalized depressed elderly patients before and after the implementation of Medicare prospective payment system. Am J Psychiatry 1993;150:1799–805.
92. Wells KB, Burnam MA, Rogers W, Hays R, Camp P. The course of depression in adult outpatients: results from the medical outcomes study. Arch Gen Psychiatry 1992;49:788–94.
93. Schoenbaum M, Unutzer J, McCaffrey D, Duan N, Sherbourne C, Wells K. The effects of primary care depression treatment on patients' clinical status and employment. Health Serv Res 2002;37:1145–58.
94. Dickey BD, Normand S, Hermann RC, et al. Guideline recommendations for treatment of schizophrenia: the impact of managed care. Arch Gen Psychiatry 2003;60:340–8.
95. Newhouse JP. Patients at risk: health reform and risk adjustment. Health Aff (Millwood) 1994;13:132–45.