# Introduction: Risk-Adjustment Issues in Mental Health Services

**Michael S. Hendryx, PhD**
**Astrid Beigel, PhD**
**Ann Doucette, PhD**

## Abstract

*State mental health authorities and other public and private entities are developing outcome measures and comparing results across providers, programs, and systems. To make comparisons equitable, outcomes must be risk adjusted. This article provides an introduction to mental health risk adjustment and outlines issues involved in the selection of outcome and risk variables, data collection protocols, and analytic methods. It stresses the importance of proper identification of risk-adjustment variables and models. The article concludes with the next steps necessary to develop a valid approach to the risk-adjustment methodology.*

## Introduction and Definition of Risk Adjustment

The demand for outcome or performance measurement in the public mental health sector is expanding rapidly. Performance measurement serves a number of important functions including setting outcome expectations, guiding quality improvement, monitoring the effects of specific interventions, assisting purchasers in choosing providers, and comparing the performance of groups of providers.[1] Performance measurement is mandated in many state public mental health systems and managed care organizations. These efforts are the focus of large financial investments as the development and implementation of outcome monitoring systems become more complex in their attempt to address the issues of treatment effectiveness; quality of care; and improvement at the client, program, and system levels.

For example, psychiatric hospitals accredited by the Joint Commission on Accreditation of Healthcare Organizations are mandated to participate in a measurement system; many choose the ORYX system to monitor quality of care. ORYX performance indicators include hospital rates of readmissions, seclusion and restraint use, and medication errors, among others. Another example is found in the Substance Abuse and Mental Health Services Administration/Center for Mental Health Services–funded Mental Health Statistics Improvement Program (MHSIP). MHSIP focuses on the

Address correspondence to Michael Hendryx, PhD, Associate Professor, Washington Institute for Mental Illness Research and Training, Washington State University, 601 W. First Avenue, Spokane, WA 99201; e-mail: Hendryx@wsu.edu.

Astrid Beigel, PhD, is the Chief, Quality and Outcome Bureau, Los Angeles County Department of Mental Health, Los Angeles.

Ann Doucette, PhD, is a Research Associate, Center for Mental Health Policy, Vanderbilt University.

*Journal of Behavioral Health Services & Research*, 2001, 28(3), 225–234. 

public mental health sector and offers a framework of recommended indicators that monitor access, appropriateness, satisfaction, and outcome of care.

Another example of the proliferation of outcome monitoring systems includes published comparative report cards indicating service system and treatment improvements. States and local communities have developed such report cards for both external and internal use. Report card approaches allow agencies and organizations to track changes within their own systems as well as across systems within particular boundaries. These reports may be published or unpublished, and they may include a wide variety of agency-, community-, or state-specific performance indicators. Although these reports recognize the need for systematic outcome monitoring, they often fail to acknowledge the issues associated in comparisons made using different population characteristics, assessment instruments, and data collection methodologies.

The instrumentation and data collection approaches are not the only concerns in cross-agency or -state comparisons. The populations served by various behavioral health care providers also can be vastly different. Providers serving individuals with severe and comorbid impairment cannot be equitably compared with providers serving individuals with less challenging mental health concerns. When outcomes or other performance indicators are monitored at multiple agencies, the question of risk adjustment or case-mix adjustment invariably arises, necessarily so if we are to understand differential performance outcomes across service providers. The outcomes that agencies strive for, and for which they are held accountable, are only partly under their control; many person and environmental variables affect outcomes independently of care. We need to ask ourselves "How can indicators be compared across different agencies if the agencies treat a different mix of clients or otherwise vary in important ways beyond agency control?" This suggests the importance of risk adjustment if meaningful comparisons are to be made.

However, risk adjustment is poorly articulated in the mental health services field. At the time of this writing, risk-adjustment strategies were being introduced in the ORYX and MHSIP initiatives. Some states employ simple risk-adjustment strategies (eg, stratification of results by diagnostic or other risk groups), but in many states risk adjustment has not begun.

This document contributes to a better understanding of risk adjustment by discussing what risk adjustment is, what it can and cannot do, and the prerequisites to risk adjustment. The purpose of this article is to provide an introduction to mental health risk adjustment for persons who are largely unfamiliar with the concept and its applications. Issues involved in the selection of outcome variables, risk variables, and data collection and analysis are outlined. The article concludes with recommended steps to undertake to develop a mental health risk-adjustment methodology.

## What Is Risk Adjustment?

Risk adjustment may be defined as a means of statistically controlling for group differences when comparing nonequivalent groups on outcomes of interest. The groups are often providers or treatment agencies, but they also can be individuals defined by characteristics such as diagnosis or severity of the mental health problem. They are nonequivalent in the sense that the persons in each group are assumed unequal in their opportunity for a good outcome for reasons beyond the control of the provider. In other words, risk variables are those that influence outcomes but are not a part of treatment.

There is no guarantee that a given set of risk variables will account for the majority of the outcome variance. It may be that outcome variance is primarily due to idiopathic patient differences and to subtle site-specific differences in treatment delivery. Moreover, it is not appropriate to include the provider or agency as a risk predictor or to otherwise include as predictors those variables that providers should be accountable for, such as treatment modality or quality. As noted by Iezzoni,[2] risk adjustment is a partial fix that cannot create perfectly equivalent groups or duplicate the rigor of experimental assignment to groups.

Outside of mental health there is a longer history of the development of risk adjustment. For example, the corporate sector has used risk-adjustment algorithms in considering various scenarios to increase the potential for profit. Risk adjustment has been developed in the assessment of physical health care. The APACHE III (Acute Physiology and Chronic Health Evaluation) system is such an illustration.[3,4] APACHE III is used in intensive care units (ICUs) to estimate the probability of patient in-hospital mortality following admission to the ICU. It has undergone years of development and is currently in its third iteration. The APACHE III is a valuable system because it (1) includes the systematic assessment of objective physiologic and health indicators routinely collected on all admissions in the first 24 hours; (2) focuses on a single, important, objective outcome; and (3) is used not only to estimate group mortality probabilities but also to assist in individual patient treatment planning.

The mental health field is far from identifying or agreeing on a single objective outcome indicator, or even a set of indicators. Mental health clients present with emotional, behavioral, and attitudinal manifestations of mental health disorders, making it difficult to identify a discrete set of objective outcomes. Behavioral health care is far from the more objective markers presented in physical health care settings (this is not to negate the difficulties encountered in assessing physical health care symptomatology). Although there is not a defined set of objective physiologic risk variables, it nevertheless is possible to aspire to develop risk-adjustment models that use reliable and valid indicators that reflect the outcome of services and treatments. The commitment to develop outcome monitoring systems in the public mental health care sector is evident. Given this commitment, failure to advance the precision of these outcome systems with risk-adjustment approaches is tantamount to comparing outcomes across nonequivalent, noncomparable groups. As outcome monitoring systems are defined, it is critical to develop the corresponding risk-adjustment models.

## Outcomes Measures

Performance indicators or outcomes may be grouped into two major types: (1) utilization and cost indicators and (2) treatment indicators including clinical and functional status, service satisfaction, and other results of treatment. Regarding indicators such as utilization and costs, the risk-adjustment objective is to predict use levels or costs. Without careful scrutiny of client case- and service-mix characteristics and risk-adjusted cost estimates, providers may choose to cream and skim or dump by underselecting high-cost/-service patients. Risk-adjustment approaches modify capitation estimates based on the characteristics of the treatment population, and they discourage creaming and skimming practices. Agencies that treat a higher proportion of patients at risk of heavy utilization receive a higher capitation rate payment. Readers who are interested in more detail regarding cost- or utilization-based risk adjustment may consult another source.[5]

The second type of outcome—patient treatment results—is the major focus of this article. This type of outcome is an attempt to estimate whether or not, or to what degree, patients benefited from treatment. Risk adjustment is necessary to consider because many person-level and environmental-level variables outside of provider control can influence outcome measures. As previously noted, the nature of mental illness is such that there is little or no consensus on which of the multiple outcomes is the most important. Mental health indicator outcomes range from reducing symptoms, improving social functioning, avoiding hospitalization or incarceration, gaining employment, reducing danger to self or others, improving sense of well-being or happiness, delaying the onset of subsequent service episodes, and so forth. The National Association of State Mental Health Program Directors (NASMHPD) presidential task force developed a list of 28 initial outcome indicators falling into four domains: access, quality, outcomes, and system performance. The consumer survey used in the MHSIP includes four consumer perception domains: access, appropriateness, outcome, and satisfaction.[6] Rosenblatt and Attkisson[7] provide a framework that groups mental health outcomes into four types: clinical, functional, safety and welfare, and life satisfaction. Most of these variable types have yet to be defined operationally and studied empirically.

However, before embarking on a risk-adjustment initiative, it is necessary to think carefully about the outcomes that are of interest in terms of their utility and feasibility. Caution must be exercised before holding providers accountable for outcomes with intervening variables between the mental health intervention and the final result. Rather than holding mental health providers accountable for school performance, for example, it may be more appropriate to focus on whether or not the child is given an opportunity to go to school. Only if a mental health treatment system was directed to the goal of improved school performance should it be held accountable for this outcome. The question to ask in identifying outcomes is "What are the important results of care that providers have the responsibility and the opportunity to improve?"

## Criteria for Selecting Outcomes and Risk Variables

It is likely that risk variables are specific to the groups under study and the outcome variables of interest.[8,9] The variables that are important to include as adjusters for one outcome may be irrelevant for another outcome. For example, severity of clinical symptoms at admission may be a critical predictor of clinical symptoms at follow-up, but it may not be useful as a predictor of client satisfaction with care. Another example is suicidal behavior. Suicide attempts within a year or so of treatment is an indicator of much greater risk for potentially poor outcomes than attempts made many years ago. Although indicators such as diagnosis, symptom severity, functional level, and prior service history may be generally relevant to examine, which indicator variables to include must be determined for each outcome and purpose. If important risk predictors of an outcome are not included, incorrect conclusions about provider performance and patient outcomes are likely to occur. Moreover, the inclusion of inappropriate indicator variables will result in inaccurate risk-adjustment models. It is likely that valid models will be specific to diagnostic and other important subgroups as well.

Consider schizophrenia and major depression as examples. Poorer outcomes and prognosis in general for schizophrenia are associated with family attitudes, lack of family support, family history of mental illness, substance abuse, poor medication adherence history, history of more severe medication side effects, low social support, early age onset, and poor premorbid adjustment.[10] For major depression, poorer prognosis is associated with severe symptoms, co-occurring substance abuse or dysthymia, co-occurring physical illness, family history of depression or alcohol abuse, greater number of previous depressive episodes, greater number of previous psychiatric hospitalizations, young age at onset, and lack of social support.[11] Although schizophrenia and major depression have some similar markers, the differences in the disorders may result in distinct risk-adjustment models.

Table 1 lists suggested criteria that potential risk variables and potential outcome variables should satisfy. In developing risk-adjustment models, one might use these criteria as checklists. If the potential risk variable or the potential outcome fails to satisfy any criteria, then its value is reduced.

Criteria for potential risk variables include the following:

- *Variables must be measures of the consumer or the consumer's extra-treatment environment.* Consumer variables include, for example, diagnosis, demographic characteristics, and clinical and functional status at admission. The extra-treatment environment includes family and social support and features of the community beyond provider control such as rural/urban location and availability of treatment alternatives. The individual provider, the treatment modality, and the treatment agency are clearly not risk variables, because the purpose of risk adjustment is to identify and control variables that influence outcome other than treatment.
- *Variables should be theoretical and empirical predictors of outcome.* For example, only variables that have been shown to predict increased risk for re-hospitalization, or that are theoretically related to this outcome, should be included in risk-adjustment models of re-hospitalization. The same holds true for other outcomes. If a suspected risk variable is uncorrelated or unrelated to outcomes, then it is not a risk variable. It is possible that a risk variable will not be correlated with an outcome at a bivariate level but still be important in a multivariate context.

**Table 1**

Suggestive criteria for potential risk variables and potential outcomes

| Criteria for potential risk variables | Criteria for potential outcome variables |
|---|---|
| Be non-treatment measures of the consumer or the consumer's extra-treatment environment | Outside the control of providers to game |
| Be significantly related to outcome, conceptually and empirically | Within the control of providers to impact through treatment, without intervening variables |
| Be outside control of providers to game | Possible to measure reliably and validly |
| Be possible to measure reliably and validly | Substantial (ie, be of importance to consumers, providers, payers, and the public) |
| Discriminate among providers | |
| Make a difference in final interpretation of performance relative to unadjusted results | |

- *Variables should be measured in a way that is not under the control of providers or other risk-bearers.* This ability to control is referred to as "gaming." Although this is related to measurement validity (see below), it deserves special mention. An example is a case manager rating of patient severity. If case managers know that they look good if patients have high severity at intake and low severity at follow-up, and case managers control the rating of severity, there is an obvious perverse incentive to manipulate scores. If the risk-bearers are conducting the data collection, data should be subject to audits for accuracy, as is done in the ORYX system and in the reporting system used in the Indiana public mental health system.[12]
- *Variables should discriminate among plans, providers, or other groups being compared.* If age, for example, correlates with an outcome measure and satisfies other criteria, but the age distribution is the same across groups, then it is less useful as a risk indicator because all groups face the same risk.
- *Variables must be measured reliably and validly.* This is simply a basic measurement requirement. When survey or other data collection instruments are used, they must be psychometrically sound and established. Data quality must be high. Clients included must be representative of populations.
- *Variables must make a difference.* After the risk-adjustment models are developed and implemented, the conclusions that one draws about comparative provider performance must be meaningfully altered. In other words, risk-adjusted performance is different from unadjusted performance at the group level. This can be determined by identifying which groups score statistically higher or lower than the overall mean using unadjusted and adjusted scores; if any of the groups identified are not the same, then risk adjustment makes a difference.[13]

Outcome variables also must satisfy a number of criteria in order to be useful.

- *Outcome indicator assessment must remain outside the control of providers to manipulate via measurement.* This is the same as a criterion mentioned above for risk variables. Assessment approaches must be objective in that measures known to be sensitive for selected groups are not used to game better outcome measurement for those groups.
- *Outcomes must be within the control of providers to impact through treatment, without intervening variables.* That is, what outcomes can providers reasonably impact? These outcomes should be discriminated from others that might be desirable but not realistically under the control of the provider.

- *Outcome indicator measurement must be reliable and valid.* It must satisfy accepted psychometric standards, particularly in terms of item and scale biases (racial/ethnic, gender, age, and so forth).
- *Outcomes must be substantive.* That is, they must be of importance to consumers, providers, payers, and the public.

Existing efforts to develop risk-adjustment models have not paid sufficient attention to these criteria for inclusion. Efforts to develop risk-adjustment models in mental health have used conveniently available databases and conventional risk adjusters like demographic variables. Empirical examination of the utility and sensitivity of commonly accepted measurement instruments in assessing change or improvement over time is needed. In addition, there is a need for empirical grounding in the history and course of many mental health disorders that could potentially illuminate and distinguish the risk variables that should be examined and distinguish whether these variables remain unchanged during the course of the disorder.

In addition to a clear choice of predictor and outcome variables, it is necessary to have a clearly stated goal or purpose of the risk adjustment. Risk-adjusted outcomes may be used for multiple purposes—rate setting, contract decisions, quality monitoring, public reporting, or financial behavior modification. For example, a state mental health authority (SMHA) may choose several performance indicators. These may include consumer-reported satisfaction, 30-day re-hospitalization rates following inpatient discharge, or case manager ratings of consumer community functioning ability. The SMHA collects data on these indicators from providers across the state. For each indicator, a valid risk-adjustment model passing the criteria in Table 1 is developed. The SMHA enters the risk-adjustment effort with a clear understanding of the quality and limits of the data, a clear plan of analysis and presentation of findings, and clear objectives in mind. In this hypothetical example the objectives might be to use risk-adjusted findings to (1) create internal reports for providers to use for their own quality improvement studies and (2) set benchmarks that will be used in later years that all providers must meet in order to contract with the state.

## Data Questions in Doing Risk Adjustment

Doing the actual risk-adjustment data collection, analysis, and reporting requires attention to questions of the quality and rigor of the data and the design. Readers are referred to other sources that discuss general research measurement, design, and analysis.[14–16] Obviously, the data must be reliable, valid, and representative of the population to which inferences will be made. For example, questionnaires with poor reliability, studies with large attrition leading to unrepresentative samples, or outcome indicators that providers can easily manipulate to their advantage can lead to results that are at best useless and at worst counterproductive.

In order to conduct at least some risk-adjustment studies, it is necessary to follow clients over time, collect risk variables at one point, such as entry into a treatment episode, and collect outcomes at a later point; thus, longitudinal designs are sometimes preferred over cross-sectional ones, and these designs require time and resources to conduct successfully. When developing and implementing longitudinal risk-adjustment models, it also is necessary to decide what the time interval for follow-up will be and how often the models will be recalibrated. Risk-adjustment models must be updated as significant changes are made in treatment modalities, environmental risks, or patient populations.

## Methods of Analysis

The following analytic methods are most often used in development of risk-adjusted results: baseline-post-test; direct weighting; regression models; and classification and regression trees, growth curves, and other advanced models.

### Baseline-post-test

The baseline-post-test approach is the simplest model. It measures an outcome indicator at baseline and measures the same variable later in treatment; improvement is defined as the difference between the two measurements. Its advantage lies in the fact that the best predictor of an outcome is often the same variable earlier in time (eg, the best predictor of functioning at discharge may be functioning at admission). However, this model is severely limited in terms of interpretation. Critical risk variables may be ignored in a baseline-post-test design. In addition, this model cannot be used in situations where the outcome (eg, re-hospitalization) does not have a corresponding baseline level. Other problems are that the value of the variable at baseline may, in some cases, be the responsibility of the provider (based on earlier treatment) and so cannot be used as a risk-adjuster, and that a model based on one predictor may be easier to game. Measuring change also creates a number of psychometric and interpretation difficulties.[17] Although this method may be useful in some circumstances, in general it is likely to be unsatisfactory.

### Direct weighting

This approach weighs the outcome indicator by risk variables so that the contribution of a risk variable for a given group matches the distribution of that variable over the entire multigroup population. For example, let us assume that age is a risk variable because age is correlated with an outcome of interest and the age distribution is different at different agencies. Then, outcome scores will be calculated separately for different age groups, weighted by the overall population percentage that falls into that age group, and summed. This approach, illustrated in Table 2, may be simpler than regression models, but it also has difficulties. It is cumbersome when more than a few risk variables are important, and it creates arbitrary breaks in the risk-variable distribution. In addition, selection of the appropriate risk variables requires correlating risk to outcome in order to identify which risk variables are important. Since this correlation is what regression models do directly, a regression-based approach may be preferable to direct weighting.

There is a variation of direct stratification called propensity weighting[18,19] that is capable of combining multiple risk variables into a single propensity score. By dividing the distribution of the propensity score into quintiles, about 90% of the bias due to the risk variables can be removed. Outcome scores then can be examined within each quintile for each pair of treatment groups. This works well when there are two groups (eg, two treatment agencies) being compared, but the number

**Table 2**

Illustration of direct weighting risk-adjustment approach*

| Age distribution | Agency A sample size ($n = 100$) | Agency A outcome score (A)† | Age distribution in entire multi-agency population (B)‡ | Contribution to outcome score (A×B)§ |
|---|---|---|---|---|
| 18–29 | 15 | 43 | .25 | 10.8 |
| 30–45 | 30 | 48 | .35 | 16.8 |
| 46–64 | 30 | 51 | .30 | 15.3 |
| 65+ | 25 | 61 | .10 | 6.1 |

*A hypothetical outcome measure (mean = 50) at one agency (Agency A) is age adjusted for the multi-agency population age distribution.

†Unadjusted Agency A mean outcome = 51.4

‡Sum = 1.00

§Sum of this column is 49.0, which is the age-adjusted Agency A mean outcome.

of comparisons increases rapidly as the number of treatment groups increases, and the approach may not be practical when there are 10 to 20 or more treatment groups. Perhaps it would be possible to combine the quintile scores into a single weight for each observed outcome score and compare the weighted outcome score of each treatment group against all other groups combined.

### Regression models

This approach is the most widely used in the physical health sciences[16] and is the most likely candidate for application to mental health. Other forms of linear models such as analysis of variance (ANOVA) and analysis of covariance (ANCOVA) fall into this category as well. The dependent variable is a dichotomous or interval variable leading to logistic or linear regression, respectively. Independent variables are the risk variables. Regression and other linear models can easily incorporate both main effects and interactions including interactions among risk variables and between risk variables and providers (see the article in this issue titled "Risk Adjustment of Florida Mental Health Outcomes Data: Concepts, Methods, and Results"). Regression models may be tested against a number of statistical criteria and validated in separate samples.[9] The capacity to include multiple main effects and interactions and to conduct statistical validity tests is an advantage of regression models over baseline or direct weighting models. Disadvantages of regression models can arise when model assumptions are not met (eg, normal error distributions, linear relationships), which may result in misinterpretation of results.

Once a valid regression model is developed for a population, expected scores may be calculated for each group within that population and compared with observed scores to determine whether risk-adjusted outcomes are better or worse than expected. The populationwide regression equation is used to create expected scores. For example, let us suppose that in the development of our regression model, we find that (1) baseline functioning, (2) diagnosis, and (3) marital status predict better functioning at a follow-up measurement. The individual's values for baseline functioning, diagnosis (ie, schizophrenia—yes or no), and marital status (married or not) are entered into the regression equation and an expected follow-up functioning score for that individual is obtained. An average or mean expected score can be calculated overall of the individuals in a given agency. The mean *difference* between the observed follow-up functioning score and the expected score is a measure of agency performance adjusted for baseline functioning, diagnosis, and marital status. The mean difference score for an agency will be positive if agency performance is better than expected and negative if performance is worse than expected. Analyses such as $t$ tests or $F$ tests can be conducted to determine whether these differences are statistically significant (ie, a $t$ test to see whether the agency difference scores are significantly higher or lower than 0; an $F$ test to see if difference scores vary significantly across agencies).

Now the critical importance of specifying the correct set of variables to include in the model may be more apparent. If data measuring baseline functioning had not been available, and only diagnosis, marital status, and demographic variables like age and sex were used, would the resulting expected scores be different? A different set of agencies may be identified as performing better or worse than expected based on the particular set of risk variables used.[13] For each outcome of interest, the correct set of risk variables must be identified from theory and research and then collected and analyzed along with the outcome measures of interest.

### Classification and regression trees, growth curves, and other advanced models

More complex models derived from regression or other general linear models also may be used. The classification and regression tree (CART) approach builds regression models iteratively, focusing on interactions among possible risk variables. The choice of a CART or a traditional regression model may depend on whether or not the risk prediction may be captured adequately through single main effect variables. CART models may potentially offer advantages to regression models in identifying sets of risk variables, but they are more computationally advanced. Because they are constructed iteratively they may be sensitive to initial starting values and so will require careful validation if they are used. Other approaches such as growth curve models, hierarchical models, or neural networks also are possible but are more demanding analytically and require a greater time and resource investment in data collection.

## Implications for Behavioral Health Services

This article introduced readers to the issues surrounding the valid development and implementation of mental health risk-adjustment models. In efforts to develop performance indicators and to introduce greater accountability into public mental health services, the question of how to report fair risk-adjusted outcomes is an appropriate one to ask. However, risk adjustment is not a simple uniform method that applies to every study. The choice of risk variables, outcome variables, data collection design, analysis method, and intent each must be carefully considered.

In this development of a mental health risk-adjustment methodology, there are five critical activities to undertake in sequential order. The first activity is the careful identification of key outcomes that mental health provider organizations can and should impact through treatment. The second activity is the identification of the correct set of risk variables for each outcome. This identification should be based on literature review and empirical test. The correct set may be specific to population groups defined by diagnosis, age groups, and other categories. The third activity is the specification of data collection protocols that ensure that outcomes and risk variables are collected reliably and validly, in a common way across providers, and representative of treatment populations. The fourth activity is the identification of the analytic methods best suited to the task. If a simple baseline difference method is adequate for some outcome measures, then there is no need to employ more advanced techniques. Other outcomes may require analysis using regression models, CART models, or other approaches; still other outcomes may not require risk adjustment at all. The fifth activity is to translate risk-adjusted results into information that can be used at multiple levels (SMHAs, agencies, providers, consumers, and consumer advocates) for accountability and quality improvement. To the extent that rigorous risk-adjustment models are not developed, caution in comparative analysis of providers must be exercised, and consideration must be given to refraining from conducting such comparisons.

## Introduction to the Special Section

The articles in the Special Section illustrate the issues introduced in this article. First, they demonstrate that there is no standard methodology for conducting mental health outcomes risk adjustment. Banks, Pandiani, and Bramley (see the article in this issue titled "Approaches to Risk-Adjusting Outcome Measures Applied to Criminal Justice Involvement after Community Service") employ a combination of weighting and baseline/follow-up methods to construct risk-adjusted scores for criminal justice involvement. This article illustrates creative use of existing administrative databases in Vermont to derive risk-adjusted scores in the absence of person-level-identifiable indicators.

Hendryx and Teague (see the article titled "Comparing Alternative Risk-Adjustment Models") and Dow, Boaz, and Thornton (see the article titled "Risk Adjustment of Florida Mental Health Outcomes Data: Concepts, Methods, and Results") illustrate the linear model approach. The Hendryx and Teague article is a cross-sectional analysis that compares different model specifications—models limited to administrative variables and models that add consumer survey variables, as well as general versus diagnosis-specific models. The form of the model leads to different conclusions about agency performance and illustrates that proper specification of the risk-adjustment model remains an important goal. The Dow, Boaz, and Thornton article makes important contributions by adding interaction effects into model specifications and by examining mental health outcomes in a baseline-follow-up linear model design. It also is a good read because it thoughtfully critiques and refines some of the ideas regarding analytic methods and risk-adjustment criteria that are suggested here.

Lambert, Doucette, and Bickman (see the article titled "Measuring Mental Health Outcomes with Pre-Post Designs") distinguish individual from group prediction and show that pre-post designs are inadequate for individual prediction. This article also shows that difference scores versus residuals will lead to different conclusions and that multiwave designs (3+ observations) offer improved statistical power and improved individual and group prediction. Kramer and colleagues (see the article titled "Comparing Outcomes of Routine Care for Depression: The Dilemma of Case-Mix Adjustment") offer risk-adjustment models specifically for depression and illustrate the strength of clinical risk predictors that most risk-adjustment research to date has not incorporated. Even in a homogeneous group defined by a common diagnosis, the form of the model depended on the specific outcome in question. This article also employs a baseline–follow-up design and regression analysis.

Finally, DeLiberty, Newman, and Ward (see the article titled "Risk Adjustment in the Hoosier Assurance Plan: Impact on Providers") describe a risk-adjustment approach that is actually in place and used in a state public mental health system. It is the only article in this issue that attempts to develop mental health risk-adjustment models for children. Taken collectively, these articles demonstrate that work remains to be done to identify the best risk-adjustment models for particular purposes; however, they do elucidate key principles to guide this effort.

A second general lesson to gather from these articles is that risk adjustment makes a difference. Invariably, when group performance is evaluated using unadjusted scores, the results are different than when using adjusted scores. The bottom-line conclusion made about group performance depends on conducting risk adjustment. These results confirm that performance measurement systems for mental health that examine group performance, whether for report cards, contract monitoring, quality assurance, or other purposes, must incorporate risk adjustment or refrain from making comparisons.

## References


1. Eddy DM. Performance measurement: problems and solutions. *Health Affairs.* 1998;17(4):7–25.
2. Iezzoni LI. The risks of risk-adjustment. *Journal of the American Medical Association.* 1997;278:1600–1607.
3. Marsh HM, Krishan I, Naessens JM, et al. Assessment of prediction of mortality by using the APACHE II scoring system in intensive-care units. *Mayo Clinic Proceedings.* 1990;65:1549–1557.
4. Zimmerman JE, Wagner DP, Draper EA, et al. Evaluation of acute physiology and chronic health evaluation III predictions of hospital mortality in an independent database. *Critical Care Medicine.* 1998;26:1317–1326.
5. Ettner SL, Frank RG. 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. DHHS publication (SMA) 99–3285.
6. Wackwitz JH, Hall J. *Structure of the MHSIP Consumer Survey.* Denver, CO: Colorado State Mental Health Services; 1997.
7. Rosenblatt A, Attkisson CC. Assessing outcomes for sufferers of severe mental disorder: a conceptual framework and review. *Evaluation and Program Planning.* 1993;16:347–363.
8. Hendryx M. Risk-adjusting mental health outcomes: access and quality applications under managed care. In: Coughlin K, ed. *1999 Behavioral Outcomes and Guidelines Sourcebook.* New York: Faulkner and Gray; 1999:364–371.
9. Hendryx M, Dyck D, Srebnik D. Risk-adjusted outcome models for public mental health outpatient programs. *Health Services Research.* 1999;34:171–195.
10. Fischer EP, Cuffel BJ, Owen RR. *Schizophrenia Outcomes Module User's Manual.* Little Rock, AR: Center for Outcomes Research and Effectiveness; 1997.
11. Smith GR, Burnam A, Burns B, et al. *Major Depression Outcomes Module User's Manual.* Little Rock, AR: Center for Outcomes Research and Effectiveness; 1997.
12. Newman FL, DeLiberty RN, Ward EO. Hoosier Assurance Plan IV: consumer progress/outcomes by risk-adjustment group for those enrolled for more than 3 years. Presented at the Sixth Annual Florida Conference on Behavioral Healthcare Evaluation; December 10, 1999; Orlando, FL.
13. Hendryx M, Moore R, Leeper T, Reynolds M, Davis S. Psychiatric re-hospitalization rates across treatment agencies: a risk-adjustment analysis. Presented at the Tenth Annual Conference on State Agency Mental Health Services Research, Program Evaluation and Policy; February 2000; Washington, DC.
14. Creswell JW. *Research Design: Qualitative and Quantitative Approaches.* Thousand Oaks, CA: Sage; 1994.
15. Fink A. *Evaluation Fundamentals.* Newbury Park, CA: Sage; 1993.
16. Iezzoni LI, ed. *Risk Adjustment for Measuring Health Care Outcomes.* Ann Arbor, MI: Health Administration Press; 1994.
17. Collins LM, Horn JL, eds. *Best Methods for the Analysis of Change.* Washington, DC: American Psychological Association; 1991.
18. Rosenbaum PR, Rubin DB. Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association.* 1984;79:516–524.
19. Rubin DB. Estimating causal effects from large data sets using propensity scores. *Annals of Internal Medicine.* 1997;127:757–763.