Criteria for the Evaluation of Quality Improvement Programs and the Use of Quality Improvement Data

American Psychological Association

This document provides a set of criteria to be used by psychologists in evaluating quality improvement programs (QIPs) that have been promulgated by health care organizations, government agencies, professional associations, or other entities. These criteria also address the privacy and confidentiality issues evoked by the intended use of patient data gathered by such QIPs. Although developed for psychologists, these criteria may be useful across health service areas and professions.

The health care marketplace has witnessed an increased interest on the part of third-party payers, both public/governmental and private, in the development of pay-for-performance and other "quality improvement" programs for the purpose of improving the health care outcomes of patients.[1] These programs vary in design, program implementation, and quality measures. Psychologists who are health care providers will soon find themselves, if they have not already, confronted with having to make decisions about participating in one or more of these programs.

Recognizing that psychologists have considerable expertise in program development and evaluation, the American Psychological Association (APA) can make a useful contribution to the evaluation of such programs. To date, the available QIPs have been of varying quality and relevance to the practice of psychology. Psychologists are supportive of programs that genuinely improve the benefits of health care services to the public and improve the quality of services provided. However, well-designed QIPs achieve these goals while also protecting the rights of patients (e.g., confidentiality) and respecting the professional responsibilities and clinical judgment of psychologists.

Quality Improvement Programs in the Public and Private Sectors

Quality is defined by the Institute of Medicine (2001) as the degree to which services and treatment increase the likelihood of desired outcomes and are consistent with current professional knowledge. QIPs include all programs that systematically collect information from providers or patients with the intention of drawing conclusions about the quality of care provided and improving provider performance, treatment outcome, or efficiency. QIP activities are both prospective and retrospective, including ongoing assessment of change models and continuous reevaluation of process and outcome targets that QIPs aspire to change. Quality assessment mechanisms include the following:

- structural measures that examine professional and technical resources or infrastructure;
- process measures that reflect treatment protocols or procedures;
- performance measures that assess the level of care provided, measure patient outcomes, and/or identify areas in need of improvement.

The federal government has various programs within the U.S. Department of Health and Human Services designed to improve quality in health care (e.g., Agency for Healthcare Research and Quality, 2001; Centers for Medicare & Medicaid Services, n.d.-a). The Agency for Healthcare Research and Quality supports research to improve the quality of health care and to assist consumers and policymakers in making more informed health care decisions. The Centers for Medicare & Medicaid Services (n.d.-b) utilizes a national network of Quality Improvement Organizations to promote delivery of "the right care for every

These criteria were approved as American Psychological Association (APA) policy by the APA Council of Representatives on August 13, 2008. The criteria were developed by the following members of APA's Performance Improvement Advisory Group: Katherine C. Nordal (Chair), Ann M. Doucette, Marvin R. Goldfried, Walter E. Penk, and Bruce E. Wampold. The Performance Improvement Advisory Group acknowledges the helpful comments received from APA boards, committees, and divisions as well as individual members in response to drafts of the criteria and is especially grateful for the insightful contributions of Lisa R. Grossman, Kristin A. Hancock, and Suzanne Bennett Johnson. The Advisory Group also wishes to acknowledge Russ Newman for his foresight regarding the need for these criteria and for initiating this project and thanks the following APA staff members for their consultation and assistance: Lynn Bufka, Alan Nessman, Diane Pedulla, and Elizabeth Winkelman.

This document is scheduled to expire as APA policy after 10 years (2018). After this date, users are encouraged to contact the Practice Directorate, American Psychological Association, to confirm that this document remains in effect.

Correspondence concerning this article should be addressed to the Practice Directorate, American Psychological Association, 750 First Street, NE, Washington, DC 20002-4242.

[1] To be consistent with discussions of quality improvement in other areas of health care, the term *patient* is used here to refer to the recipient of psychological services. However, the APA Performance Improvement Advisory Group recognizes that in many situations there are important and valid reasons for using such terms as *client, consumer,* or *person* in place of *patient* to describe the recipient of services.

© 2009 American Psychological Association 0003-066X/09/$12.00
Vol. 64, No. 6, 551–557 DOI: 10.1037/a0016744

person every time." It contracts with these organizations in each state to ensure that Medicare services are reasonable and necessary and that the care provided to Medicare beneficiaries meets professionally recognized standards.

The Centers for Medicare & Medicaid Services has instituted quality improvement initiatives for hospitals, home health agencies, and nursing facilities. In addition, federal regulations (Health and Human Services Quality Assessment and Performance Improvement Program, 2004) require managed care organizations contracting with state Medicaid plans to have ongoing quality assessment and performance improvement programs.

More pertinent to individual providers is that Medicare instituted a pay-for-reporting program in July 2007. Known as the Physician Quality Reporting Initiative (PQRI), the program awards eligible professionals a bonus payment for successful reporting on a designated set of quality measures. It is likely that the PQRI will evolve into a pay-for-performance (PFP) program with a larger set of quality measures.

In the private sector, more than half of commercial health maintenance organizations have already begun using PFP programs (Rosenthal & Frank, 2006; Rosenthal, Landon, Normand, Frank, & Epstein, 2006). A survey conducted by Med-Vantage, Inc. found that as of 2007, there were 148 different PFP programs affecting more than 57 million Americans (Baker & Delbanco, 2007). PFP programs operate on the principle that providing financial rewards will promote improvement and excellence in the delivery of health care (de Brantes, 2006). These programs vary in design but typically involve some type of incentive payment for psychologists or other health care professionals who meet specified objectives. Many of the initial PFP programs focused on hospitals, but over time the trend has shifted to include individual health care professionals (Leapfrog Group & Bridges to Excellence, 2004).

Need for Criteria

A number of important issues are raised by QIPs, relating to the types of data collected, the ways in which data are used, program design, and program implementation. These issues are briefly reviewed in this section.

Concerns have been raised regarding the appropriateness of the types of measures that have been used in QIPs. The variables that are the focus of measurement, which may include clinician behavior and targeted outcomes, may not be linked to improvement or excellence in the delivery of health care (e.g., Kessler, 2007). Poor or irrelevant measures or targets of change may have little bearing on treatment outcomes or delivery of health care. Inappropriate measures may inadvertently incent behavior that is not appropriate for all clinical situations, increasing the use of certain clinician behaviors (e.g., administering a particular treatment or an assessment instrument to every patient with a given diagnosis) that may not be in the best interest of every patient treated.

Moreover, satisfactory models for ensuring privacy and confidentiality in the collection and use of data have not always been used in these programs. Personal health information is protected under federal law (Health Insurance Portability and Accountability Act of 1996 [HIPAA], 1996), and information related to psychotherapy notes is generally afforded extra protections (HIPAA Privacy Rule, 2003, Section 164.508 (a)(2) in particular and Part 164 generally). It is unclear whether information collected in a QIP would also be afforded any extra protection and, if not, who would have access to what kind of information.

PFP programs are a type of QIP that seeks to link health care costs with quality of service through the use of financial incentives. These incentives are intended to improve the performance of health care providers, with the goal of achieving more favorable outcomes. While this principle provides a rationale for many PFP programs and QIPs, the literature has been equivocal regarding whether or not financial incentives will promote improvement in health care (Epstein, Lee, & Hamel, 2004; Rattray, Andrianos, & Stam, 2004). Small financial incentives may be insufficient for motivating any change in delivery of services, and larger incentives may result in compliance with protocols unrelated to patient outcomes. Furthermore, many are not convinced that individual clinician behavior is the appropriate level of focus in attempting to effect improvement in the health care delivery system.

Many other concerns have been raised about potential negative consequences that may result from the implementation of PFP programs and other similar QIPs. These additional concerns relate to issues such as the effectiveness, fairness, and accuracy of such programs; the relevance of certain types of measures to the practice of psychology; possible conflicts of interest or other interference with the psychologist–patient relationship as a result of financial incentives; potential bias against patients with complex or chronic conditions (which could discourage providers from treating these patients and potentially reduce access to services); possible negative impact on health disparities (Casalino & Elster, 2007); and ensuring that PFP programs are voluntary and do not penalize those who choose not to participate. In addition, concerns have been raised regarding how data gathered by QIPs about individual providers are analyzed and presented. Several recent lawsuits by physicians have challenged the rationale and fairness of "quality" ratings that were made available to the public (Kessler, 2007).

Given the importance of the concerns and issues described above, it would be useful to provide a set of criteria for psychologists to use when evaluating or considering participation in a QIP.

Purpose of Criteria

The purpose of this document is to provide criteria to assist psychologists in the determination of the strengths and weaknesses of QIPs. QIPs are neither inherently beneficial nor inherently detrimental, and this document is not intended either to encourage or discourage their development. The burden of proof of the utility and usefulness of such systems rests on those implementing them. Clear

demonstration that a QIP is likely to benefit patient care is needed in order to justify its implementation.

QIPs have the potential to influence the provision of care to many patients, and therefore their processes and development need to be open to public scrutiny. Moreover, failure to disclose information related to the rationale and development of the QIP and the intended uses of the collected information is likely to lead to low participation rates and inferior data quality. Disclosure of this type of information to both patients and providers increases the likelihood that programs will achieve their aims. It will also enable psychologists to evaluate programs according to the criteria described below.

The criteria listed below provide a framework for psychologists to use in evaluating programs and determining whether their participation in them is warranted. Additionally, this document provides policy guidance for advocacy efforts at the federal and state levels regarding the design of QIPs that may impact psychologists. This document is not intended to promote or discourage psychologists' participation in QIPs, nor is it intended to imply that psychologists must review each of these criteria prior to participating in any QIP.

QIPs may be evaluated along four dimensions: (1) program design, (2) program implementation, (3) indicators used to measure quality, and (4) privacy and confidentiality. The criteria listed below describe important issues that program developers are encouraged to address in the best possible manner and are not intended as mandatory standards.

1. Program Design

Criterion 1.0. The primary goal of a QIP is to improve quality of care. QIPs are designed to ensure and promote quality of care. Cost containment is never the sole purpose of a well-designed QIP.

Criterion 1.1. Representatives from affected stakeholder groups, including practicing psychologists and recipients of psychological services, are included in the process of program design. Involvement of these groups ensures that various perspectives are represented, ideally resulting in a program that is relevant and acceptable to all stakeholders.

Criterion 1.2. Programs include an articulated model for improving quality, based on the best available research evidence. Well-designed QIPs have a clearly written rationale readily available to psychologists and patients. The rationale is based on sound psychological principles and research evidence. The definition of quality balances patient and clinical perspectives. How data are to be used to improve quality is clearly stated in the rationale.

Criterion 1.3. Program design conforms to the principles of evidence-based practice in psychology (EBPP). According to the APA (2005), EBPP "is the integration of the best available research with clinical expertise in the context of patient characteristics, culture, and preferences" (p. 1). Accordingly, QIPs are encouraged to balance the three elements of EBPP such that clinical expertise and patient preference inform the interpretation of data collected. Well-designed QIPs allow for the role of professional judgment in determining treatment interventions for individual patients. For example, waivers or exclusions for particular treatment protocols may be appropriate in some cases if a valid rationale is provided.

Criterion 1.4. Program design ensures that reporting systems protect the integrity of data collected so that they are as accurate and complete as possible. Well-designed QIPs specify the methods for verifying data accuracy in advance of data collection. Protections to prevent "gaming" of the system, such as selectively reporting data only on patients who are progressing well in treatment or who report high levels of distress in order to obtain additional services, are considered in the design of reporting systems.

Criterion 1.5. Data analyses and presentation of results are appropriately designed and statistically sound. Well-designed QIPs conduct data analyses using appropriate methods for the questions being studied. For example, sample sizes are sufficient to produce stable estimates. Decisions about quality improvement account for sampling and measurement error. Confidence intervals are provided for quality estimates. Decisions about individual psychologist quality are derived from data for an adequate number of representative patients.

Criterion 1.6. Data take into account patient characteristics and context. Well-designed QIPs take into account the diversity of patients and the contexts in which they live. Patients will have differential outcomes based on the health and environmental challenges they face. Therefore, adjustments are needed to ensure comparability of data across psychologists, patients, and settings by taking into account patient characteristics and context and adjusting estimates accordingly. These adjustments include variables such as severity at intake, history of hospitalization, environmental stressors, complicating physical illnesses, socioeconomic status, race, ethnicity, culture, age, sex/gender, disability status, benefit plan, co-pay, and diagnosis.

Criterion 1.7. Programs are designed to reduce health disparities. Well-designed QIPs not only improve overall quality of service but reduce any preexisting disparity in services provided to particular patient populations (e.g., traditionally underserved populations; Hasnain-Wynia et al., 2007). Unless carefully designed, QIPs may have the unintended negative consequence of increasing disparities (Casalino & Elster, 2007). Therefore, QIPs include appropriate methods of risk adjustment and address their potential impact on health disparities. (See Criterion 3.1 for additional information on risk adjustment.)

Criterion 1.8. Programs that make determinations about the quality of care provided by individual psychologists or that provide ratings or rankings of psychologists do so in a way that is accurate, fair, and designed primarily to improve quality of care. The amount of relevant data and the sample size (e.g., all or a subset of an individual psychologist's caseload) may not be sufficient to accurately calculate individual quality ratings (see Crite-

rion 1.5). If a psychologist works in a community that suffers from health disparities or works with patients diagnosed with particularly complex or chronic disorders, these data are risk-adjusted, per Criterion 3.1 below, in order to provide a fair representation of the quality of care provided. Any disclosure of information about individual psychologists to the public or to other third parties is specified in advance and has a valid rationale that supports quality improvement.

Criterion 1.9. Programs provide a clearly articulated procedure to allow individual psychologists to comment on or appeal any quality ratings. This mechanism ensures that psychologists have the ability to challenge any rating that they believe to be misleading, inaccurate, or unfair.

Criterion 1.10. The cost of implementing a QIP is reasonable in the context of the treatment setting. For example, costs associated with the technology needed for gathering or reporting QIP data do not place a disproportionate burden on the practitioner.

Criterion 1.11. PFP programs provide financial incentives in addition to payments that psychologists are otherwise entitled to receive as usual and customary fees. Well-designed PFP programs do not reduce or delay payments that psychologists are otherwise entitled to receive and do not subject psychologists to financial penalties if they choose not to participate, are unable to participate, or treat patients who decline to participate.

Criterion 1.12. Program developers consider effective nonmonetary incentives for quality improvement. Alternative approaches to quality improvement may be more cost-effective than PFP (Rosenthal & Frank, 2006). Depending on the treatment setting, education programs, infrastructure subsidies, performance feedback, and recognition may be equally effective approaches to stimulating quality of care.

2. Program Implementation

Criterion 2.0. Effectiveness of QIPs is evaluated in an ongoing manner, and programs are modified accordingly. Well-designed QIPs demonstrate effectiveness in improving quality outcomes as measured by the chosen indicators in order to justify continued implementation.

Criterion 2.1. QIPs focus on attainment of benchmark indicators or on demonstrable progress toward meeting benchmarks (Eagar, Burgess, & Buckingham, 2003; Hermann, Chan, Provost, & Chiu, 2006; Hermann, Mattke, et al., 2006; Hermann & Provost, 2003; Sluyter & Barnette, 1995). Benchmarks are typically thought of as measurement references, a goal against which improvement or progress is measured. In certain settings, such as those specializing in the treatment of underserved populations or the treatment of patients with severe or complex health problems, demonstrating improvement may be a more appropriate goal than meeting prespecified targets (Casalino & Elster, 2007).

Criterion 2.2. Incentives, such as PFP bonus payments, are structured to reward the maintenance of care meeting benchmark indicators as well as to encourage continued improvement. Equity issues may also arise when incentives are only used to reward improvement. Well-designed PFP programs include rewards for consistently meeting or exceeding quality indicators.

Criterion 2.3. Benchmarks are based on empirical evidence and are psychometrically sound, clinically informed, reasonable, and achievable in the context in which the services are delivered. Well-designed PFP programs use benchmarks that reflect the complexity of the problems being treated and are appropriate for the patient population receiving services. (More detailed requirements for statistically sound benchmarks and other quality indicators are described in Criterion 3.0 below.)

Criterion 2.4. Incentives for meeting benchmarks or making progress toward benchmarks appropriately account for potential sources of error. Sampling error arises when the sample selected for analysis does not adequately reflect the population from which it was drawn. Other potential sources of error include patient refusal to complete the surveys on which benchmarked improvement is based, differences in how the benchmarking information is collected (e.g., telephone, face to face, self-report), differences in patient screening protocols, and geographical differences.

Criterion 2.5. Program design includes timely and ongoing feedback to psychologists about their performance. Research has shown that feedback improves effectiveness and efficiency of care (Howard, Moras, Brill, Martinovich, & Lutz, 1996; Lambert, 2005; Lambert, Hansen, & Finch, 2001; Lambert, Whipple, et al., 2001).

3. Indicators Used to Measure Quality

Criterion 3.0. Indicators used to measure quality are based on empirical evidence and are psychometrically sound, relevant, actionable, auditable, and feasible.

1. Psychometric properties. Measures used to assess quality are reliable, producing the same results when repeated in the same populations and settings. While there are accepted estimates of reliability, there is no single estimate of validity. Assumptions of validity rely on the evidence that the instrument is appropriate for its intended use and for the population being studied. For example, valid quality indicators correlate well with other measures of the same aspects of care and are linked to desired outcomes.

2. Sensitivity to change. The quality of the data has marked implications for investigating change over time as a result of a particular intervention or treatment. For example, many QIP measures use Likert-type response options (e.g., *strongly agree, agree, strongly disagree*). Patient responses are typically interpreted as being equally spaced (interval data), but seldom is this the case. Because score changes using raw scores are not always intervalized, a change of 10 points may not have equivalent meaning independent of where the change occurs on the measured construct (Bond & Fox, 2001). As a result, raw score

change is not always a reliable or valid indicator of change. More sophisticated measurement models (e.g., item response theory) can be used to transform Likert-type response categorical data into interval data.

3. Indicator relevance. Quality indicators are meaningful to practicing psychologists and to patients for making treatment choices (Newman & Tejeda, 1996). Quality indicators yield data that target aspects of care that can be changed and/or provide information about how to strategically improve service delivery. Indicators and benchmarks may also be used to stimulate patients' internal improvement efforts and encourage activities that maximize patient well-being.

4. Indicator auditability. Quality indicators are not susceptible to manipulation or "gaming" that would be undetectable in an audit.

5. Indicator feasibility. Well-designed QIPs specify the data sources of the program's indicators and benchmarks and the methods for data collection and reporting. The collection of data does not violate any accepted standards of patient confidentiality (see, e.g., HIPAA Privacy Rule, 2003) and is feasible in the treatment context.

Criterion 3.1. Indicators are appropriately risk adjusted. Risk adjustment is essentially the process of adjusting the outcome probabilities for unlike groups so that comparisons can be made. Treatment outcomes are adjusted to produce greater accuracy in interpreting outcomes when external influences on treatment such as age, gender, socioeconomic status, race, ethnicity, chronicity, acuity, and comorbidity are nonrandomly distributed across the groups to be compared. Methods for achieving this goal may include adjusting for case mix (types of patients seen) or service mix (types of services provided).

Typically, risk adjustment focuses on two distinct categories: (a) predicting service utilization and cost and (b) comparing treatment outcomes. Utilization and cost estimates are adjusted with the goal of yielding more precision in setting capitation, case and premium rates, and the like. Adjusting treatment outcomes is a complex endeavor. Multiple factors, including demographic characteristics, clinical and functional attributes, diagnosis, presence of comorbid conditions, and quality of the services received by an individual, are likely influences on both treatment and service utilization/cost outcomes. While individual attributes are significant in articulating risk adjustment strategies, other considerations are also important if risk adjustment techniques are to be meaningful. The unit of analysis (e.g., individual clients, client groups, psychologists) and the interval of time observed are critical considerations. The ability to evaluate outcomes more precisely with risk-adjusted probabilities is constrained by the availability of data and by the methodological designs employed to address outcome questions of interest.

Criterion 3.2. Indicators used to determine payment levels or quality ratings for psychologists are comparable across practice settings and measure variables under the psychologists' control. Quality measures used to determine payment levels or quality ratings for providers are based on provider behavior, accurately measure what is actually happening in treatment, and are not affected by any variables that are beyond the practicing psychologist's or practice network's control. Risk stratification or a validated model for calculating an adjusted result can be used to ensure comparability across practicing psychologists and psychologist networks (see also Criteria 1.6 and 3.1).

Criterion 3.3. Indicators used to measure quality are related to patient health or well-being. Measures may include both psychologist and patient measures. Patient measures might include, for example, indicators of patient functioning, well-being, and symptom severity. Psychologist measures might include the delivery of important intervention components (e.g., appropriate screening for suicidality) or reflect important therapy principles (e.g., formation of a therapeutic alliance).

Criterion 3.4. Representatives from affected stakeholder groups, including practicing psychologists and recipients of psychological services, are involved in the selection of relevant indicators. Involvement of these groups ensures that various perspectives are represented, ideally resulting in the development of measures that are relevant and acceptable to both patients and providers.

4. Privacy and Confidentiality Issues

Criterion 4.0. QIPs provide informed consent forms that are clear, thorough, linguistically appropriate, and easily understood by patients. Well-designed QIPs inform psychologists as to how the QIP will safeguard confidentiality and provide patients with an informed consent form clearly describing any potential privacy risks (e.g., see Criteria 4.3 and 4.6 below).

Criterion 4.1. Patient and psychologist participation in any PFP program is voluntary. Well-designed PFP programs clearly inform patients and psychologists that participation is voluntary and do not pressure or penalize patients or psychologists if a patient chooses not to provide self-report data.

Criterion 4.2. QIPs provide appropriate safeguards to protect the confidentiality of data. Patient data collected directly from the patient are typically less protected than data collected via the traditional method of the company asking the psychologist about the patient. Data that are identified with a particular patient have some protection under the HIPAA regulations. While identified patient data are covered under protected health information, these data may not be protected under the HIPAA psychotherapy notes provision and may not be privileged. Instruments sometimes record data in ways that psychologists would not in their own records. Well-designed QIPs report performance or quality data to third parties only with the patient's express written consent, by court order, or as otherwise required by law. This does not prohibit the use of HIPAA-compliant, anonymous aggregate data for research or quality improvement purposes, where appropriate safeguards are used to protect psychologist and patient confidentiality (Kraus, Wolf, & Castonguay, 2006).

Criterion 4.3. QIPs provide a clear rationale, empirically documented utility, and appropriate confidentiality safeguards for the collection of particularly sensitive patient information (e.g., illegal activities, drug use). Additional safeguards may be advisable for particularly sensitive data.

Criterion 4.4. QIPs specify and fully disclose, in advance, the ways in which individually identified patient data will be collected and used. A well-designed QIP discloses all data sharing (e.g., with health care providers, health insurers, disability insurers) in advance and gives the patient the opportunity to opt out. Decisions about individual psychologists or patients are limited to the uses specified in the written QIP.

Criterion 4.5. If QIP survey data use codes identified with a particular patient, codes are assigned such that it is not possible to decode the identity of the patient by using collateral data. This is a particularly important concern with small sample sizes. For example, if data are being collected for only a few patients being treated by a particular psychologist, very little additional information may be necessary to decode a patient's identity.

Criterion 4.6. QIPs notify patients that there are other situations under which sensitive information collected as data for quality improvement could be revealed. For example, this information may subpoenaed in custody or personal injury litigation or may be requested on employment applications for the military, government positions, or jobs requiring a security clearance.

Criterion 4.7. QIPs clearly and fully inform psychologists about any data that are routinely collected directly from the patient relating to treatment by that psychologist. The more the psychologist is involved in the process (e.g., discusses survey content with the patient, reviews survey), the more the data are likely to be subject to privilege.

Conclusion

This document presents a description of pay for performance and other quality improvement programs and outlines criteria to be used by psychologists in evaluating these programs and/or when considering participating in them. Although these criteria are written for psychologists, many of the concepts are equally relevant to other health care providers and their patients. Psychologists support continuous quality improvement and professional development to ensure that their patients receive the best possible care. Careful evaluation of quality improvement strategies helps ensure improved quality of care while avoiding unintended negative consequences to the patient and/or the therapeutic relationship.

REFERENCES

Agency for Healthcare Research and Quality. (2001). *AHRQ profile: Advancing excellence in health care.* Retrieved August 31, 2007, from U.S. Department of Health and Human Services website: http://www.ahrq.gov/about/profile.htm

American Psychological Association. (2005). *Policy statement on evidence-based practice in psychology.* Washington, DC: American Psychological Association. (Available at http://www2.apa.org/practice/ebpstatement.pdf)

Baker, G., & Delbanco, S. (2007). *Pay for performance: National perspective. 2006 longitudinal survey results with 2007 market updates.* Retrieved March 2008, from the Med-Vantage website: http://www.medvantage.com/Pdf/2006NationalP4PStudy.pdf

Bond, T., & Fox, C. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences.* Mahwah, NJ: Erlbaum.

Casalino, L. P., & Elster, A. (2007). Will pay-for-performance and quality reporting affect health care disparities? *Health Affairs, 26*(3), w405–w414.

Centers for Medicare & Medicaid Services. (n.d.-a). *Physician Quality Reporting Initiative.* Retrieved August 31, 2007, from U.S. Department of Health and Human Services website: http://www.cms.hhs.gov/PQRI/

Centers for Medicare & Medicaid Services. (n.d.-b). *Quality improvement organizations.* Retrieved August 31, 2007 from U.S. Department of Health and Human Services website: http://www.cms.hhs.gov/QualityImprovementOrgs/

de Brantes, F. (2006). Pay for performance and beyond: A recipe for improving healthcare. In *The quality conundrum: Practical approaches for enhancing patient care* (pp. 110–114). PricewaterhouseCoopers Health Research Institute. Retrieved February 14, 2008, from http://www.bridgestoexcellence.org/Documents/P4PandBeyondarticlebyFSB.pdf

Eagar, K., Burgess, P., & Buckingham, B. (2003). *Towards national benchmarks for Australian mental health services* (ISC Discussion Paper No. 4). Retrieved June 22, 2009, from http://www.health.gov.au/internet/main/publishing.nsf/Content/mental-pubs-t-bench

Epstein, A. M., Lee, T. H., & Hamel, M. B. (2004). Paying physicians for high-quality care. *New England Journal of Medicine, 350*(4), 406–410.

Hasnain-Wynia, R., Pierce, D., Haque, A., Hedges Greising, C., Prince, V., & Reiter, J. (2007). *Health Research and Educational Trust disparities toolkit.* Retrieved February 25, 2008, from http://www.hretdisparities.org

Health and Human Services Quality Assessment and Performance Improvement Program, 42 C.F.R. § 438.240 (2004).

Health Insurance Portability and Accountability Act of 1996, Pub. L. No. 104–191, 110 Stat. 1936 (1996).

Hermann, R. C., Chan, J. A., Provost, S. E., & Chiu, W. T. (2006). Statistical benchmarks for process measures of quality of care for mental and substance use disorders. *Psychiatric Services, 57*(10), 1461–1467.

Hermann, R. C., Mattke, S., Somekh, D., Silfverhielm, H., Goldner, E., Glover, G., et al. (2006). Quality indicators for international benchmarking of mental health care. *International Journal for Quality in Health Care, 18* (Suppl. 1), 31–38.

Hermann, R. C., & Provost, S. (2003). Interpreting measurement data for quality improvement: Means, norms, benchmarks, and standards. *Psychiatric Services, 54*(5), 655–657.

HIPAA Privacy Rule, 45 C.F.R Parts 160 and 164 (2003).

Howard, K. I., Moras, K., Brill, P. L., Martinovich, Z., & Lutz, W. (1996). Efficacy, effectiveness, and patient progress. *American Psychologist, 51*, 1059–1064.

Institute of Medicine. (2001). *Crossing the quality chasm: A new health system for the 21st century.* Washington, DC: National Academy Press.

Kessler, M. (2007). Connecticut physicians file lawsuit challenging "elite" physician designation. *BNA's Health Law Reporter, 16*(34), 1029.

Kraus, D., Wolf, A., & Castonguay, L. G. (2006). The outcomes assistant: A kinder philosophy to the management of outcomes. *Psychotherapy Bulletin, 41*(3), 23–31.

Lambert, M. J. (2005). Emerging methods for providing clinicians with timely feedback on treatment effectiveness: An introduction. *Journal of Clinical Psychology, 61*, 141–144.

Lambert, M. J., Hansen, N. B., & Finch, A. E. (2001). Patient-focused research using patient outcome data to enhance treatment effects. *Journal of Consulting and Clinical Psychology, 69*, 159–172.

Lambert, M. J., Whipple, J. L., Smart, D. W., Vermeersch, D. A., Nielsen, S. L., & Hawkins, E. J. (2001). The effects of providing therapists with feedback on patient progress during psychotherapy: Are outcomes enhanced? *Psychotherapy Research, 11*(1), 49–68.

Leapfrog Group & Bridges to Excellence. (2004). *Measuring provider efficiency version 1.0.* Retrieved June 22, 2009, from http://www

.bridgestoexcellence.org/Documents/Measuring_Provider_Efficiency_Version1_12-31-20041.pdf
Newman, F. L., & Tejeda, M. J. (1996). The need for research that is designed to support decisions in the delivery of mental health services. *American Psychologist, 51,* 1040–1049.
Rattray, M. C., Andrianos, J., & Stam, D. T. (2004). *Quality implications of efficiency-based clinician profiling.* Retrieved June 22, 2009, from http://carevariance.com/images/quality_Implications_Efficiency-Based_Clinician_Profiling.pdf
Rosenthal, M. B., & Frank, R. G. (2006). What is the empirical basis for paying for quality in health care? *Medical Care Research and Review, 63,* 135–157.
Rosenthal, M. B., Landon, B. E., Normand, S. L., Frank, R. G., & Epstein, A. M. (2006). Pay for performance in commercial HMOs. *New England Journal of Medicine, 355,* 1895–1902.
Sluyter, G. V., & Barnette, J. E. (1995). Application of total quality management to mental health: A benchmark case study. *Journal of Mental Health Administration, 22*(3), 278–285.