

# The Expanding Relevance of Routinely Collected Outcome Data for Mental Health Care Decision Making

James F. Boswell<sup>1</sup> · Michael J. Constantino<sup>2</sup> · David R. Kraus<sup>3</sup> · Matteo Bugatti<sup>4</sup> · Jennifer M. Oswald<sup>4</sup>

Published online: 11 April 2015

© Springer Science+Business Media New York 2015

**Abstract** Evidence shows that routine outcome monitoring (ROM) and feedback using standardized measurement tools enhances the outcomes of individual patients. When outcome data from a large number of patients and clinicians are collected, patterns can be tracked and comparisons can be made at multiple levels. Variability in skills and outcomes among clinicians and service settings has been documented, and the relevance of ROM for decision making is rapidly expanding alongside the transforming health care landscape. In this article, we highlight several developing core implications of ROM for mental health care, and frame points of future work and discussion.

**Keywords** Routine outcome monitoring · Clinician effects · Psychotherapy · Mental health care services

## Introduction

In their seminal paper, Howard et al. (1996) suggested using standardized session-to-session measures of patient progress to evaluate and improve treatment outcome by using data-driven feedback. In doing so, they launched a new area of research labeled *patient-focused research*, which asks: Is this treatment, however constructed, delivered by this particular clinician, helpful to this patient at this point in time? Patient-focused research symbolized a paradigm shift from the more common nomothetic methods that emphasized aggregating data among patients in controlled trials conducted in formal research settings.

The significance of this paradigm shift cannot be understated. For instance, the prevailing model of communication and dissemination between researchers and practitioners had heretofore been largely unidirectional (Castonguay et al. 2013). The implicit role of the practitioner would be to read an article and promptly begin to apply group-level findings to their individual patients. Although still employed today, due to a variety of factors, this unidirectional dissemination model has been largely ineffective (Boswell et al. 2015; Brownson et al. 2012; Safran et al. 2011). Consistent with Stricker and Trierweiler's (1995) concept of the "local clinical scientist," patient-focused research moves the "lab" to routine practice settings, and emphasizes the integration of routine assessment and data-driven feedback to inform the treatment of this patient, as well as groups of patients in a clinician's practice. This paradigm shift can empower individual clinicians and function as a bridge between the local clinical scientist and traditional academia.

---

✉ James F. Boswell  
jboswell@albany.edu

Michael J. Constantino  
mconstantino@psych.umass.edu

David R. Kraus  
dkraus@outcomereferrals.com

Matteo Bugatti  
mbugatti@albany.edu

Jennifer M. Oswald  
joswald@albany.edu

<sup>1</sup> Department of Psychology, Social Science 399, University at Albany SUNY, 1400 Washington Avenue, Albany, NY 12222, USA

<sup>2</sup> Department of Psychological and Brain Sciences, University of Massachusetts Amherst, 612 Tobin Hall - 135 Hicks Way, Amherst, MA, USA

<sup>3</sup> Outcome Referrals, Inc., 1 Speen St., Framingham, MA 01701, USA

<sup>4</sup> Department of Psychology, University at Albany SUNY, Albany, NY 12222, USA

Several well-powered meta-analyses (e.g., Shimokawa et al. 2010) have provided strong empirical support for the integration of routine outcome monitoring (ROM) and feedback in psychological treatment. Owing to such evidence, ROM and feedback systems have ascended to a prominent role in mental health care policies, practice settings, and research agendas (Bickman et al. 2006; IOM 2007). Work in this area is also part and parcel of health care's increasing emphasis on "quality measurement" (Hoagwood 2013). Patient-focused research methods and the data collected through these methods are being integrated into broader systems of care. Howard et al. (1996) presciently noted the potential utility of ROM to supervisors, case managers, and systems of care, in addition to individual patients and clinicians. Castonguay et al. (2013) underscored the usefulness of the umbrella term *practice-oriented research* to capture not only patient-focused research, but also similar data-driven methods for enhancing practice-research integration and practitioner-researcher collaboration, as well as mental health care decision making. In line with Bickman et al. (2006), Constantino et al. (2013) argued that ROM could potentially promote clinical responsiveness at multiple levels, spanning individuals and health care systems.

It stands to reason that ROM and feedback can be used to enhance the patient-centeredness and responsiveness of mental health care delivery and decision making across multiple levels. Precisely how this is accomplished, however, raises many important questions. For example, what types of measures yield the most valid and useful information? What are the potential consequences of not measuring progress and outcomes? How do outcomes (broadly defined) fit into health care's operationalization of "quality indicators"? What outcomes do patients value? How can we develop new treatments or employ existing ones in a way that leverages information about patients and providers that informs "best matches" and "precise decision making" for different people? Although these issues are extraordinarily complex, the aims of this article are (a) to highlight our perceptions of several developing core implications of ROM for mental health care delivery and decision making, and (b) to frame fundamental points of future work and discussion.

### Has ROM Become a Necessity?

Although the potential benefits of ROM are wide-ranging (Lambert 2010; Youn et al. 2012), arguably the most important is the ability to identify whether a current course of treatment for a given patient is at risk of being ineffective or harmful. Based on tens-of-thousands of individual cases, reliable estimates of deterioration and non-response rates in

psychotherapy have been established (Lambert 2010). The estimated rate of deterioration is approximately 10 % for adult patients (Hansen et al. 2002; Lambert 2010); estimated rates of non-response (including in controlled trials) range between 30 and 50 % (Hansen et al. 2002). In a study of over 4000 child and adolescent patients, the estimated average deterioration rate across two naturalistic samples was 19 % (Warren et al. 2010).

Given these sobering statistics, it appears that ROM has become a necessary practice. Simply stated, ROM is an effective tool to help identify whether a given patient is at risk for a negative outcome or on track to experience benefit. Research to date has demonstrated that ROM and feedback significantly reduces deterioration and dropout rates in routine psychological treatment (Lambert 2013). Predictive analytics indicating that a given patient is at risk for deterioration or non-response can promote clinical responsiveness. An individual clinician may respond to this feedback by (a) offering more sessions or increasing the frequency of sessions, (b) ramping up pointed assessment (e.g., of suicidality, motivation to change, social support), (c) altering the micro-level or macro-level treatment plan, (d) referring for a medication consultation, (e) calling a family meeting, and/or (f) seeking additional consultation and supervision. This is by no means an exhaustive list of potential responses, but it represents a starting point. In the absence of feedback and timely responsiveness, however, treatment may continue to follow the same problematic course (Harmon et al. 2007).

### What if Clinicians did not Monitor Outcomes?

If clinicians did not monitor outcomes, they would be unaware of many patients who fail to experience benefit from a given course of treatment. A lack of relevant, accurate information about a patient's current status and/or prognosis impedes appropriate responsiveness (Constantino et al. 2013; Tracey et al. 2014). Furthermore, psychology is not only interested in understanding behavior, but also in its prediction. Unfortunately, predictions that are based solely on human judgment tend to be inaccurate, even when such predictions are made by experienced professionals (Garb 2005). For example, Hannan et al. (2005) asked psychotherapists at the end of each session if that particular patient would experience deterioration during treatment. This method of prediction was compared to a statistical algorithm derived from routine outcome data. Participating therapists were initially informed of the base-rate of deterioration in routine treatment. Yet, therapists predicted that less than 1 % of all patients (3 out of 550) would deteriorate; in actuality, more than 7 % (40 out of 550) evidenced deterioration. The statistical algorithm

correctly identified 77 % of the deteriorating cases. Additional research (e.g., Ellsworth et al. 2006) has suggested that such algorithms can identify 85 to 100 % of eventual deteriorators prior to leaving treatment, which consistently exceeds clinical judgment alone.

Although the data on statistical prediction can stand alone, its importance is enhanced when considered alongside research on clinicians' beliefs about their own effectiveness. A recent survey found that clinicians estimated that about 85 % of their own patients improve or recover (Walfish et al. 2012). In addition, clinicians had the common impression that they are unusually successful, with 90 % rating themselves in the upper quartile and none seeing themselves as below average in relation to their peers. Thus, while clinicians tend to believe that they are extremely effective with most of their patients, this overvaluation is inconsistent with the extant research.

To put it bluntly, failing to monitor outcomes is a choice to ignore a practice that can reliably enhance clinical responsiveness to patients and significantly reduce rates of deterioration, non-response, and dropout (Boswell et al. 2015). Of course, it is important to emphasize that ROM and ROM-based feedback is an evidence-based tool that practitioners can integrate into routine practice. Just as a physician must interpret and respond to the results of a blood glucose test when monitoring a patient with diabetes, the psychotherapist is still ultimately responsible for seeking out, interpreting, and deciding how to respond to ROM-based feedback as treatment unfolds. It is also important to note that the effect sizes of ROM and feedback interventions compared to treatment as usual for cases at risk of deterioration hover around  $r = .25-.30$ , representing a medium effect (Lambert and Shimokawa 2011). Although modest, the scalability of integrating ROM broadens its potential impact on patient outcomes.

### Is the Relevance of ROM Expanding?

Imagine that the blood glucose levels of a diabetic patient are recorded over an extended period and plotted to obtain a graphical display of his or her response to a particular treatment regimen. Similar recordings and plots are obtained for each patient being treated by this particular physician, and this physician's colleagues in the diabetes clinic have obtained similar recordings from their own patients' over time. A diabetes clinic at a partnering hospital has also been collecting such recordings on individual patients over an extended period. What begins with monitoring the progress of an individual patient rapidly becomes a significant amount of ecologically valid comparative treatment information. This is already happening

in medicine (Farley et al. 2002; IOM 2007), and is inevitable in behavioral health.

As data accumulate, the implications for understanding treatment response and improving treatment outcomes are considerably broadened. For example, one could ask whether or not significant differences in diabetes outcomes emerge among different types of patients within a physician's caseload (e.g., those with early vs. late onset), among physicians in a particular clinic (e.g., those with large vs. small caseloads), or among clinics within a system of health care delivery (e.g., those in economically advantaged vs. disadvantaged neighborhoods). Based on such information, interventions can be tailored to the individual patient (e.g., administer an alternative treatment to those patients with an early onset), or resources can be allocated in a manner that is more likely to maximize impact. In their comprehensive review of the therapist effects literature, Baldwin and Imel (2013) described a review of outcomes data from an archival data registry in which 100 accredited Cystic fibrosis treatment centers participate and contribute data (Gawande 2004). Even with consistent and stringent accreditation criteria requiring the use of detailed treatment guidelines across all centers, significant differences in life expectancy and lung functioning have emerged between the centers. ROM not only allows such analysis, but also spotlights people and places within healthcare systems that might require professional guidance/consultation, additional training, or even recognition/acceptance of certain practice strengths and limitations.

### Addressing the Clinician Uniformity Myth

Kieler (1966) famously wrote about the participant uniformity myth in psychotherapy research; specifically, that it is problematic to consider patients within any particular group (diagnostic, cultural) as homogeneous, as well as to assume that all therapists function as the same social stimulus for all patients. In fact, treatment outcome researchers have historically viewed the therapist as a "nuisance" variable that needs to be controlled through standardized training, supervision, and fidelity monitoring. Historically, health plans analyzing claims data have similarly treated clinicians as commodities or interchangeable parts, each of whom is assumed to achieve the same outcome. However, in more recent years, health plans have started to offer "tiered networks" that are beginning to integrate differences in provider and hospital performance (Scanlon et al. 2008).

Spanning multiple decades, research has consistently identified significant variability in skill and outcomes among therapists, in both naturalistic and controlled settings (e.g., Baldwin and Imel 2013; Crits-Christoph and Mintz 1991;

Lutz et al. 2007; Okiishi et al. 2003). Furthermore, differences among practitioners frequently account for a greater portion of treatment outcome variance than the specific interventions delivered in controlled trials (Krause et al. 2007; Wampold 2001). In a reanalysis of the National Institute of Mental Health (NIMH) Treatment of Depression Collaborative Research Program's (TDCRP) pharmacotherapy outcomes, differences between psychiatrists accounted for more of the outcome variance than the anti-depressant medications they prescribed (McKay et al. 2006).

It is important to note that estimates of outcome variance attributable to therapists vary considerably between studies, with some reporting estimates that are not significantly different from zero, and others reporting up to 50 % of the outcome variance being attributable to therapists (Baldwin and Imel 2013). Some of the variability across studies is certainly due to sampling error and/or differences in statistical model specification (e.g., treating therapist as a fixed vs. random effect; Wampold and Bolt 2006). It also stands to reason that greater variability in therapist behavior and outcomes will be observed in routine community practice compared to controlled research settings, given the homogeneous clinician selection, training, and supervision typically found in efficacy studies. Furthermore, although significant variability among therapists remains when controlling for within-therapist (i.e., between-patient) variability, it is likely that moderators play a role in accounting for variability in therapist outcomes, especially interactions with patient factors (e.g., Hayes et al. 2014).

For example, research has demonstrated that therapist differences in fidelity to evidence-based treatment protocols can be a function of patient severity (Imel et al. 2011) and trait interpersonal aggression (Boswell et al. 2013). In a study of nearly 700 therapists' naturalistic treatment outcomes over multiple problem domains (e.g., depression, anxiety, substance use, sleep), involving a sample of 6960 patients, the majority of therapists demonstrated a differential pattern of effectiveness depending on the problem domain (Kraus et al. 2011). For example, some therapists demonstrated substantial effectiveness in depression reduction, while others evidenced particular effectiveness in the substance abuse domain. Many therapists demonstrated effectiveness over multiple problem domains, yet no therapists demonstrated reliable effectiveness across all domains. Furthermore, a small, but notable 4 % of the therapists failed to demonstrate positive outcomes on any domain. With stable estimates of a therapist's relative performance with different groups of patients, stakeholders can make a priori predictions regarding a patient's likelihood of experiencing benefit depending on the clinician to whom she or he is referred or assigned (Wampold and Brown 2005).

In summary, when ROM data from a large number of individual patients are aggregated, patterns in outcomes

can be tracked and comparisons can be made at multiple levels—within a therapist's caseload, between different therapists, within a treatment setting [e.g., community mental health center (CMHC)], and between different treatment settings. Treatment outcome data from both naturalistic and controlled settings, spanning multiple areas of health care, demonstrate significant differences between providers and clinics (even when similar training and evidence-based guidelines are employed), and there is evidence to suggest that these results may be moderated by patient factors (e.g., interpersonal variables, cultural variables, problem domain, and problem severity). These results highlight the additional complexity that emerges once the use of ROM data extends beyond the individual patient. Case mix variables such as age, gender, health status, severity, and previous treatment history have long been collected by insurers and health care providers to manage and inform service delivery (Hirdes et al. 1996). The utility and precision of analytics based on ROM data will depend on the collection of case mix. As cogently stated by Saxon and Barkham (2012) “By including in the model measures of a therapist's case mix that are predictive of outcome, not only are they controlled for but their relative impact on outcome can be estimated” (p. 536). However, the identification of stable and reliable case mix algorithms remains a work in progress.

### Use of Clinician and Systems Level Outcome Data

The above research findings and recent trends in mental health care are pressing patients, clinicians, health care systems, and researchers to grapple with complicated questions. For example, as a health care consumer in need of care for diabetes or Cystic fibrosis, would you want access to a list of top performing clinicians and centers in your area? The field of medicine has already begun to explore this issue through pay-for-performance initiatives (Dudley 2005; IOM 2007) and public reporting of provider and hospital outcomes (Akbari et al. 2008; Henderson and Henderson 2010; Scanlon et al. 2008).

In their Cochrane Review, Henderson and Henderson (2010), attempted to examine research on the effects of providing surgeon performance data to people considering elective surgery. The rationale for disseminating these data is that patients should be informed of a surgeon's past performance before making the decision to allow a particular physician to perform surgery. Although a number of studies were available for initial review, no studies were deemed to be of sufficient quality to meet inclusion criteria [i.e., randomized controlled trial (RCT), quasi-randomized controlled trial, or controlled pre-post design]. In a subsequent review of studies involving the public release of

medical provider performance data (Ketelaar et al. 2011), only four studies were deemed methodologically suitable (i.e., RCT, quasi-RCT, interrupted time series, or controlled pre-post design). The release of provider performance data was linked to small improvements in acute myocardial infarction mortality rates and increased quality improvement activity within care organizations. None of the identified studies (or studies cited but excluded from systematic review due to inadequate methods) involved mental health.

In addition to the absence of mental health treatment, there has been a glaring lack of stakeholder involvement in such outcomes data initiatives and research to-date. Much remains to be learned about using outcomes data to inform mental health care decisions (e.g., Do patients value this information? How should this information be disseminated? Does access to this information result in better care?). Theoretically, access to routine psychotherapy outcomes data would encourage patients to compare individual clinicians and preferentially choose the best performing clinician in a particular area of need or geographic location. Nevertheless, shared decision-making would ideally involve diverse stakeholder input on the value of different types of performance data, methods of data presentation, and how data are used. Furthermore, a significant number of patients are initially referred to local CMHCs, group practices, and hospital-based outpatient clinics where assignment to a particular clinician is largely an administrative decision based on a number of purely practical considerations, such as insurance coverage, openings on a caseload, and scheduling considerations.

The obvious, and still quite open, question is whether or not patients, referral sources, and mental health care administrators might make better informed decisions about potential providers if information regarding a clinician's outcome track record with similar patients were made available. Furthermore, rather than continue to assume that we are better than 90 % of our peers across the board, could the same data assist clinicians in identifying relative strengths and weaknesses with particular types of patients (which could, in turn, inform referral acceptances, choice of subsequent continuing education experiences, or case consultations)? Could CMHC administrators use these same data to inform case assignments and resource allocation (e.g., additional training/supervision for clinicians or clinics that are struggling in a particular area)?

In a pilot survey of mental health stakeholders' attitudes toward expanding the use of ROM data and feedback, patient, clinician, and CMHC administrator groups all endorsed seeing value in ROM and feedback (Boswell et al. 2014). This is consistent with previous work by Bickman et al. (2000) and Hatfield and Ogles (2004), who found that a large percentage of therapists held interest in receiving

regular reports of patient progress and access to reliable outcome information. However, other research indicates that there may be a discrepancy between endorsed values and beliefs regarding ROM and its actual utilization (De Jong et al. 2012; Hatfield and Ogles 2007). Informed by existing theory (see Riemer et al. 2005), future research should attempt to understand this discrepancy because, in many ways, it parallels the longstanding science-practice chasm; that is, clinicians often acknowledge the importance of science, but remain skeptical adopters (or even absent consumers) of research-based practices (Safran et al. 2011).

Results from this survey also indicated that a large majority of patients experience perceived difficulty finding/selecting a mental health care provider. Of course, the selection of a clinician may be somewhat out of the patient's control. For example, a clinician may not accept Medicare or Medicaid. However, the health care climate is shifting toward increased personalization and choice, and empowering patients with information to aid in their own treatment decision making (Edwards and Elwyn 2009). As noted above, there have been recent attempts to disseminate provider and clinic-level outcomes with the explicit goal of enhancing stakeholder decision making (Fung et al. 2008; Ketelaar et al. 2011; Scanlon et al. 2008). Presently, patients generally lack access to valid mental health outcomes data.

There is a growing need to contrast clinician outcome data derived from repeated administration of standardized assessments on a large number of patients with unsystematic consumer satisfaction ratings. In a recent article, Chamberlin (2014) discussed the issue of reviews (in particular, negative reviews) of clinicians on consumer websites, such as HealthGrades.com and ZocDoc.com. At a fundamental level, the reliability and validity of such ratings are questionable. One implication, therefore, is that a clinician may protect oneself by collecting routine outcome and satisfaction data, using well-established, standardized measures. In fact, it may be a disservice to an individual who is seeking a clinician to dismiss contacting a particular provider who may be likely to help him or her based on a single anecdotal rating and testimonial from a former patient. Reliably collected and validly analyzed ROM data, however, are a different story; not only can these data inform patient decision-making, but they can also more accurately represent clinicians' skills based on their personal track record. Although consumer satisfaction is important information (i.e., we should not rely on symptom change alone to define the "quality" of services provided), we cannot simply trust judgment in either direction; clinicians generally over-estimate their own broad-based effectiveness, and dissatisfied patients may lose credible objectivity. Data from well-established outcome measures can paint an important, and reliable, picture that provides context to health care decisions.

Furthermore, although mental health stakeholders appear to value the idea of using ROM data to aid decision making, there is less clarity on how outcomes data should be disseminated. To our knowledge, this is a crucial issue that has not been directly addressed. Some evidence suggests that patients are uncertain about their ability to interpret and use outcome data reports to make health care decisions (Boswell et al. 2014; Hibbard et al. 2001). In addition to its importance for arriving at a sufficiently broad operationalization of quality, straightforward consumer satisfaction ratings and testimonials may be more appealing to the average patient. Regardless, patients will need to be able to understand and use direct-to-consumer information that is disseminated, and this will only be accomplished through patients' direct involvement in the research and development process. Marketing research techniques, such as deliberative focus groups, may be useful in this regard.

### Current Developments and Recommendations

Within the broad context of health care reform, a number of federal and non-federal “quality improvement” initiatives that place ROM and reporting at the center are already underway (IOM 2007; Zima et al. 2013). For example, the Medicare physician quality reporting system (PQRS) now penalizes providers with reduced reimbursement rates if data are not reported on designated service measures. State agencies and managed care companies have begun to apply alternative payment models, including pay-for-performance initiatives (Bremer et al. 2008). In addition, a number of the United States' largest health systems have recently formed a “Health Care Transformation Task Force,” which aims to develop new payment models based on quality performance, patient experience, and cost containment (Herman and Evans 2015). These efforts have potential merit and seemingly some support among stakeholders. However, whether or not they ultimately improve the quality of care and reduce costs remains to be seen. What is clear, however, is that significant changes to payment and service delivery models are already underway, and that such changes will rely on rigorous, ongoing assessment of process (e.g., guideline fidelity) and outcome indicators. As such, the relevant question is no longer “if” the collection of behavioral health outcomes data will be standard practice and expanded to inform a variety of health care decisions, but rather “when” and “how” the use of such information will be expanded. In order to help frame the relevance of these issues for mental health, we offer the following list of key issues and preliminary recommendations.

### Identifying, Defining, and Measuring the Relevant Outcomes

We have focused on the use of outcomes data, which typically involves patient self-reports of general distress or severity in various symptom and functioning domains. It is important to state that quality of care cannot be reduced to symptom improvement alone, and there is an active debate regarding what processes and outcomes should be measured. For some patients, remaining out of the hospital over a 6 month period represents a substantial treatment gain, yet they might continue to present with a high level of symptom severity on self-report measures. Furthermore, “quality of care or services” implies some level of process assessment – the nature and frequency of, and the manner in which, an intervention is delivered. It may be similarly problematic, however, to define quality by the use of pre-determined intervention guidelines alone (Hoagwood 2013; Kravitz et al. 2004). In both controlled and naturalistic settings, even empirically-supported treatments can be delivered in a manner that results in a negative impact (see Castonguay et al. 2010).

The IOM (2001) outlined several quality criteria: (a) safe, (b) effective (i.e., using evidence-based interventions), (c) patient-centered (tailoring treatment based on patient preferences, needs, and values), (d) timely, (e) efficient, and (f) equitable. Two things are striking about these criteria. First, patient outcomes are not addressed explicitly, perhaps with the exception of documenting lack of harm. Second, routine and simultaneous assessment of these criteria will present significant challenges. In order to maximize utility, quality indicators must be measurable, easily assessed on a repeated basis, and demonstrate predictive validity. In terms of efficiency, feasibility, and importance to public health, we believe that outcome domains (e.g., symptom change, clinical threshold, quality of life, days out of the hospital) should ultimately remain at the center of discussions regarding what should contribute to calculations of a “quality quotient.”

### The Use and Integration of Technology

Technology will play a vital role in both the growing integration of ROM in practice settings (Boswell et al. 2015), and the regular assessment of diverse process and outcome indicators being considered in initiatives, such as the Health Care Transformation Task Force. For example, integration with electronic medical records provides a built in case mix resource and promises a more seamless integration of mental health care with other areas of medicine. Technology also potentially reduces administrative burden. A patient can receive an automated reminder to complete a self-report measure on his or her smartphone prior to an

appointment, which can be scored and disseminated practically instantaneously to the provider(s). The capabilities of ROM and feedback can be enhanced by reduced burden and increased efficiency because more data from more individuals results in better predictive models.

### Addressing the Implementation Elephant

As noted above, there appears to be a discrepancy between clinician reported valuing of ROM and its consistent utilization. It is difficult to change provider behavior, as the recent growth in health service implementation process and outcome research can attest (Brownson et al. 2012). In the absence of reliable and consistently collected data, the potential utility of ROM data for decision making will not be realized. Additional research is sorely needed to better understand the barriers and facilitators to provider ROM adoption and compliance.

### The Necessity of Risk Adjustment

Evidence from both controlled and naturalistic outcome research points to the importance of accounting for patient level variability and moderators when examining outcome differences among clinicians or clinics. Any direct comparison between practitioners, clinics, hospitals, or systems of care should be based on risk adjusted and benchmarked data (Tremblay et al. 2012; Weissman et al. 1999). Risk-adjustment algorithms are important for increasing the comparability of collected data because they adjust for patient characteristics (e.g., case-mix) that could account for differences among clinicians or mental health centers. Importantly, the development of risk adjustment models is a continuous process. Enhanced integration of other (and more) health care data would theoretically result in better risk adjustment models. Given the implications of not applying risk adjustments when interpreting outcomes data (e.g., falsely concluding that a particular provider or hospital is underperforming relative to peers or other hospitals), this represents a crucial area for future research. However, future research does not guarantee that a stable, reliable algorithm will be achieved. Furthermore, given the iterative nature of such model development, the question of whether or not a particular model has reached a point of “sufficient” reliability for use in applied settings is a critical one. It is our hope that decision makers will proceed cautiously.

### The Inevitability of New Payment Models

Recent developments at the federal, state, and private level are a consequence of the wide recognition of a need for new payment models (IOM 2007). For example, the

Centers for Medicare and Medicaid Services (CMS) have spent billions of dollars in recent years funding research that places the development and testing of new payment models at the center. To a large degree, suggested models prioritize demonstration of performance, quality, and efficiency (Berwick et al. 2008). Although financial incentives for demonstration of good outcomes makes logical sense, the field must proceed cautiously because (a) as noted, a number of factors interact to determine a particular outcome, let alone overall quality, and (b) the relationships between motivation and performance are extraordinarily complex. For example, a recent meta-analysis found that intrinsic motivation and incentives must be considered in combination when predicting performance (Cerasoli et al. 2014). When incentives were directly tied to performance (e.g., a financial bonus for achieving a certain outcome), intrinsic motivation was less predictive of performance. In addition, intrinsic motivation was more strongly related to performance *quality*, and incentives were more strongly related to performance *quantity*. Certainly much more research is needed to understand how these relationships operate within health care, specifically. Nevertheless, it would be unfortunate if the core benefits of ROM (e.g., aiding decision making and thereby enhancing responsiveness) are lost when stakeholders begin focusing more on “carrots and sticks.”

### The Importance of a Scientific Attitude

Similar to any service-level intervention, initiatives involving the dissemination and implementation of ROM should be rigorously, empirically scrutinized. As noted, research on the dissemination of provider performance data has been sparse and methodologically flawed (Henderson and Henderson 2010). Furthermore, to-date there is limited available knowledge regarding stable and replicable risk adjustment algorithms. Any program or system should be evidence-based prior to full scale implementation. Fortunately, funding agencies such as the Patient Centered Outcomes Research Institute (PCORI) and the National Institutes of Health (NIH), with its recent prioritizing of “services research,” are offering more support for these types of projects. In the absence of this work, initiatives in this area are likely to evidence minimal impact.

### Diverse Stakeholder Collaboration

The complexity of these issues will require a continuous dialogue among stakeholder groups, including patients (and their families), clinicians, administrators, policy makers, private industry, and researchers. We find this issue to be particularly important. In order for any data-driven initiative to be successful, relevant stakeholders

must be adequately represented in the design, implementation, and testing of such mental health care initiatives. Because each group is potentially motivated by different priorities, this represents a significant challenge. The issue of information privacy should be a principal focus of discussion among stakeholders.

## Conclusion

The uses of ROM and its implications for mental health care are rapidly expanding. Whether or not this movement benefits patients and providers will be dependent on active and open collaborations among stakeholder groups, including researchers who are interested in putting developing methods and programs to the empirical test. Beyond documenting differences between clinicians and health care systems (and the development and investigation of reliable methods for doing so), we agree with Baldwin and Imel's (2013) recommendation to learn more about what accounts for such differences, and to begin learning directly from clinicians and systems of care that demonstrate consistently superior outcomes.

## References

- Akbari, A., Mayhew, A., Al-Alawi, M. A., Grimshaw, J., Winkens, R., Glidewell, E., et al. (2008). Interventions to improve outpatient referrals from primary care to secondary care. *Cochrane Database of Systematic Reviews*, 4, CD005471. doi:10.1002/14651858.CD005471.pub2.
- Baldwin, S. A., & Imel, Z. E. (2013). Therapist effects: Findings and methods. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (pp. 258–297). Hoboken, NJ: Wiley.
- Berwick, D. M., Nolan, T. W., & Whittington, J. (2008). The triple aim: Care, health, and cost. *Health Affairs*, 27, 759–769.
- Bickman, L., Riemer, M., Breda, C., & Kelley, S. D. (2006). CFIT: A system to provide a continuous quality improvement infrastructure through organizational responsiveness, measurement, training, and feedback. *Report on Emotional and Behavioral Disorders in Youth*, 6(86–87), 93–94.
- Bickman, L., Rosof-Williams, J., Salzer, M. S., Summerfelt, W. T., Noser, K., Wilson, S. J., et al. (2000). What information do clinicians value for monitoring adolescent client progress and outcomes? *Professional Psychology: Research and Practice*, 31, 70–74.
- Boswell, J. F., Constantino, M. J., & Kraus, D. R. (2014). Promoting context-responsiveness with outcomes monitoring. *Paper presented at the society for the exploration of psychotherapy integration conference*. Montreal.
- Boswell, J. F., Gallagher, M. W., Sauer-Zavala, S. E., Bullis, J., Gorman, J. M., Shear, M. K., et al. (2013). Patient characteristics and variability in adherence and competence in cognitive-behavioral therapy for panic disorder. *Journal of Consulting and Clinical Psychology*, 81, 443–454.
- Boswell, J. F., Kraus, D. R., Miller, S., & Lambert, M. J. (2015). Implementing routine outcome assessment in clinical practice: Benefits, challenges, and solutions. *Psychotherapy Research*, 25, 6–19. doi:10.1080/10503307.2013.817696.
- Bremer, R. W., Scholle, S. H., Keyser, D., Knox Houtsinger, J. V., & Pincus, H. A. (2008). Pay for performance in behavioral health. *Psychiatric Services*, 59, 1419–1429.
- Brownson, R. C., Colditz, G. A., & Proctor, E. K. (2012). *Dissemination and implementation research in health*. New York: Oxford University Press.
- Castonguay, L. G., Barkham, M., Lutz, W., & McAleavey, A. (2013). Practice-oriented research. In M. J. Lambert's (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (Vol. 6, pp. 85–133). Hoboken, NJ: Wiley.
- Castonguay, L. G., Boswell, J. F., Constantino, M. J., Hill, C. E., & Goldfried, M. R. (2010). Training implications of harmful effects of psychological treatments. *American Psychologist*, 65, 34–49.
- Cerasoli, C. P., Nicklin, J. M., & Ford, M. T. (2014). Intrinsic motivation and extrinsic incentives jointly predict performance: A 40-year meta-analysis. *Psychological Bulletin*, 140, 980–1008.
- Chafe, R., Neville, D., Rathwell, T., & Deber, R. (2008). A framework for involving the public in health care coverage and resource allocation decisions. *Health Manage Forum*, 21, 6–13.
- Chamberlin, J. (2014). One-star therapy? *Monitor on Psychology*, 45, 52–55.
- Constantino, M. J., Boswell, J. F., Bernecker, S. L., & Castonguay, L. G. (2013). Context-responsive integration as a framework for unified psychotherapy and clinical science: Conceptual and empirical considerations. *Journal of Unified Psychotherapy and Clinical Science*, 2, 1–20.
- Crits-Christoph, P., & Mintz, J. (1991). Implications of therapist effects for the design and analysis of comparative studies of psychotherapies. *Journal of Consulting and Clinical Psychology*, 59, 20–26.
- De Jong, K., van Sluis, P., Nugter, M. A., Heiser, W. J., & Spinhoven, P. (2012). Understanding the differential impact of outcome monitoring: Therapist variables that moderate feedback effects in a randomized clinical trial. *Psychotherapy Research*, 22, 464–474. doi:10.1080/10503307.2012.673023.
- Dudley, R. A. (2005). Pay-for-performance research: How to learn what clinicians and policy makers need to know. *JAMA*, 294, 1821–1823.
- Edwards, A., & Elwyn, G. (2009). *Shared decision-making in health care: Achieving evidence-based patient choice* (2nd ed.). New York, NY: Oxford University Press.
- Ellsworth, J. R., Lambert, M. J., & Johnson, J. (2006). A comparison of the outcome questionnaire-45 and outcome questionnaire-30 in classification and prediction of treatment outcome. *Clinical Psychology and Psychotherapy*, 13, 380–391. doi:10.1002/cpp.503.
- Farley, D. O., Elliott, M. N., Short, P. F., Damiano, P., Kanouse, D. E., & Hays, R. D. (2002). Effect of CAPHS performance information on health plan choices by Iowa Medicaid. *Medical Care Research and Review*, 59, 319–336.
- Fotaki, M., Roland, M., Boyd, A., McDonald, R., Scheaff, R., & Smith, L. (2008). What benefits will choice bring to patients? Literature review and assessment of implications. *Journal of Health Service Research Policy*, 13, 178–184.
- Fung, C., Yee-Wei, L., Soeren, M., Damberg, C., & Shekelle, P. (2008). Systematic review: The evidence that publishing patient care performance data improves quality of care. *Annals of Internal Medicine*, 148, 111–123.
- Garb, H. N. (2005). Clinical judgment and decision making. *Annual Review of Clinical Psychology*, 1, 67–89. doi:10.1146/annurev.clinpsy.1.102803.143810.
- Gawande, A. (2004). The bell curve. *The New Yorker*. Retrieved from <http://www.newyorker.com/magazine/2004/12/06/the-bell-curve>. Accessed 1 Aug 2014.



- Hannan, C., Lambert, M. J., Harmon, C., Nielsen, S. L., Smart, D. W., Shimokawa, K., et al. (2005). A lab test and algorithms for identifying clients at risk for treatment failure. *Journal of Clinical Psychology, 61*, 155–163. doi:10.1002/jclp.20108.
- Hansen, N. B., Lambert, M. J., & Forman, E. V. (2002). The psychotherapy dose-response effect and its implications for treatment delivery services. *Clinical Psychology: Science and Practice, 9*, 329–343.
- Harmon, S. C., Lambert, M. J., Smart, D. W., Hawkins, E. J., Nielsen, S. L., Slade, K., et al. (2007). Enhancing outcome for potential treatment failures: Therapist/client feedback and clinical support tools. *Psychotherapy Research, 17*, 379–392. doi:10.1080/10503300600702331.
- Hatfield, D. R., & Ogles, B. M. (2004). The current climate of outcome measures use in clinical practice. *Professional Psychology: Research and Practice, 35*, 485–491. doi:10.1037/0735-7028.35.5.485.
- Hatfield, D. R., & Ogles, B. M. (2007). Why some clinicians use outcome measures and others do not. *Administrative Policy Mental Health and Mental Health Services Research, 34*, 283–291. doi:10.1007/s10488-006-0110-y.
- Hayes, J. A., Owen, J., & Bieschke, K. J. (2014). Therapist differences in symptom change with racial and ethnic minority clients. *Psychotherapy*. doi:10.1037/a0037957.
- Henderson, A., & Henderson, S. (2010). Provision of a surgeon's performance data for people considering elective surgery. *Cochrane Database of Systematic Reviews, 11*, CD006327. doi:10.1002/14651858.CD006327.pub2.
- Herman, B., & Evans, M. (2015). Transformers' push to modify pay must overcome legacy system. Modern Healthcare. <http://www.modernhealthcare.com/article/20150131/MAGAZINE/301319987>. Accessed 27 Feb 2015.
- Hibbard, J. H., Peters, E., Slovic, P., Finucane, M. L., & Tusler, M. (2001). Making health care quality reports easier to use. *Joint Commission Journal on Quality Improvement, 27*, 591–604.
- Hirdes, J. P., Botz, C. A., Kozak, J., & Lepp, V. (1996). Identifying an appropriate case-mix measure for chronic care: Evidence from an Ontario Pilot Study. *Healthcare Management Forum, 9*, 40–46.
- Howard, K. I., Moras, K., Brill, P. L., Martinovich, Z., & Lutz, W. (1996). Evaluation of psychotherapy. *American Psychologist, 51*, 1059–1064. doi:10.1037/0003-066X.51.10.1059.
- Hoagwood, K. E. (2013). Don't mourn: Organize. Reviving mental health services research for healthcare quality improvement. *Clinical Psychology: Science and Practice, 20*, 120–126.
- Imel, Z. E., Baer, J. S., Martino, S., Ball, S. A., & Carroll, K. M. (2011). Mutual influence in therapist competence and adherence to motivational enhancement therapy. *Drug and Alcohol Dependence, 115*, 229–236. doi:10.1016/j.drugalcdep.2010.11.010.
- Institute of Medicine. (2001). *Crossing the quality chasm: A new health system for the 21st century*. Washington, D.C.: National Academy Press.
- Institute of Medicine Committee on Redesigning Health Insurance Performance Measures, Payment, and Performance Improvement Programs. (2007). *Rewarding provider performance: Aligning incentives in Medicare: Pathways to quality health care series*. Washington, D.C.: National Academy Press.
- Ketelaar, N. A. B. M., Faber, M. J., Flottorp, S., Rygh, L. H., Deane, K. H. O., & Eccles, M. P. (2011). Public release of performance data in changing the behaviour of healthcare consumers, professionals or organisations. *Cochrane Database of Systematic Reviews, 11*, CD004538.
- Kielser, D. J. (1966). Some myths of psychotherapy research and the search for a paradigm. *Psychological Bulletin, 65*, 110–136.
- Kraus, D. R., Castonguay, L. G., Boswell, J. F., Nordberg, S. S., & Hayes, J. A. (2011). Therapist effectiveness: Implications for accountability and patient care. *Psychotherapy Research, 21*, 267–276.
- Krause, M. S., Lutz, W., & Saunders, S. M. (2007). Empirically certified treatments or therapists: The issue of separability. *Psychotherapy, 44*, 347–353.
- Kravitz, R. L., Duan, N., & Braslow, J. (2004). Evidence based medicine, heterogeneity of treatment effects, and the trouble with averages. *The Milbank Quarterly, 82*, 661–687. doi:10.1111/j.0887-378X.2004.00327.x.
- Lambert, M. J. (2010). *Prevention of treatment failure: The use of measuring, monitoring, and feedback in clinical practice*. Washington, D.C.: American Psychological Association Press.
- Lambert, M. J. (2013). The efficacy and effectiveness of psychotherapy. In M. J. Lambert (Ed.), *Bergin & Garfield's handbook of psychotherapy and behavior change* (6th ed., pp. 169–218). New York, NY: Wiley.
- Lambert, M. J., & Shimokawa, K. (2011). Collecting client feedback. *Psychotherapy, 48*, 72–79. doi:10.1037/a0022238.
- Lutz, W., Leon, S. C., Martinovich, Z., Lyons, J. S., & Stiles, W. B. (2007). Therapist effects in outpatient psychotherapy: A three-level growth curve approach. *Journal of Counseling Psychology, 54*, 32–39. doi:10.1037/0022-0167.54.1.32.
- McKay, K. M., Imel, Z. E., & Wampold, B. E. (2006). Psychiatrist effects in the psychopharmacological treatment of depression. *Journal of Affective Disorders, 92*, 287–290. doi:10.1016/j.jad.2006.01.020.
- Okiishi, J., Lambert, M. J., Nielsen, S. L., & Ogles, B. M. (2003). Waiting for Supershrink: An empirical analysis of therapist effects. *Clinical Psychology and Psychotherapy, 10*, 361–373.
- Riemer, M., Rosof-Williams, J., & Bickman, L. (2005). Theories related to changing clinician practice. *Child and Adolescent Psychiatric Clinics of North America, 14*, 241–254.
- Rogers, E. M. (2003). *Diffusion of innovations* (5th ed.). New York: Free Press.
- Safran, J. D., Abreu, I., Ogilvie, J., & DeMaria, A. (2011). Does psychotherapy research influence the clinical practice of researcher-clinicians? *Clinical Psychology: Science and Practice, 18*, 357–371.
- Saxon, D., & Barkham, M. (2012). Patterns of therapist variability: Therapist effects and the contribution of patient severity and risk. *Journal of Consulting and Clinical Psychology, 80*, 535–546. doi:10.1037/a0028898.
- Scanlon, D. P., Lindrooth, R. C., & Christianson, J. B. (2008). Steering patients to safer hospitals? The effect of a tiered hospital network on hospital admissions. *Health Services Research, 43*, 1849–1868.
- Shimokawa, K., Lambert, M. J., & Smart, D. (2010). Enhancing treatment outcome of patients at risk of treatment failure: Meta-analytic and mega-analytic review of a psychotherapy quality assurance system. *Journal of Consulting and Clinical Psychology, 78*, 298–311. doi:10.1037/a0019247.
- Stricker, G., & Trierweiler, S. J. (1995). The local clinical scientist: A bridge between science and practice. *American Psychologist, 50*, 995–1002.
- Tracey, T. J. G., Wampold, B. E., Lichtenburg, J. W., & Goodyear, R. K. (2014). Expertise in psychotherapy: An elusive goal? *American Psychologist, 69*, 218–229. doi:10.1037/a0035099.
- Tremblay, M. C., Hevner, A. R., & Berndt, D. J. (2012). Design of an information volatility measure for health care decision making. *Decision Support Systems, 52*, 331–341. doi:10.1016/j.dss.2011.08.009.
- Walfish, S., McAlister, B., O'Donnell, P., & Lambert, M. J. (2012). An investigation of self-assessment bias in mental health providers. *Psychological Reports, 110*, 639–644. doi:10.2466/02.07.17.

- Wampold, B. E. (2001). *The great psychotherapy debate: Models, methods, and findings*. Mahwah, NJ: Erlbaum.
- Wampold, B. E., & Bolt, D. M. (2006). Therapist effects: Clever ways to make them (and everything else) disappear. *Psychotherapy Research, 16*, 184–187. doi:[10.1080/10503300500265181](https://doi.org/10.1080/10503300500265181).
- Wampold, B. E., & Brown, G. S. (2005). Estimating variability in outcomes attributable to therapists: A naturalistic study of outcomes in managed care. *Journal of Consulting and Clinical Psychology, 73*, 914–923. doi:[10.1037/0022-006X.73.5.914](https://doi.org/10.1037/0022-006X.73.5.914).
- Warren, J. S., Nelson, P. L., Mondragon, S. A., Baldwin, S. A., & Burlingame, G. M. (2010). Youth psychotherapy change trajectories & outcome in usual care: Community mental health versus managed care. *Journal of Clinical and Consulting Psychology, 78*, 144–155.
- Weissman, N. W., Allison, J. J., Kiefe, C. I., Farmer, R., Weaver, M. T., Williams, O. D., et al. (1999). Achievable benchmarks of care: The ABCs of benchmarking. *Journal of Evaluation in Clinical Practice, 5*, 269–281. doi:[10.1046/j.1365-2753.1999.00203.x](https://doi.org/10.1046/j.1365-2753.1999.00203.x).
- Youn, S. J., Kraus, D. R., & Castonguay, L. G. (2012). The Treatment Outcome Package: Facilitating practice and clinically relevant research. *Psychotherapy, 49*, 115–122. doi:[10.1037/a002](https://doi.org/10.1037/a002).
- Zima, B. T., Murphy, J. M., Scholle, S. H., Hoagwood, K. E., Sachdeva, R. C., Mangione-Smith, R., et al. (2013). National quality measures for child mental health care: Background, progress, and next steps. *Pediatrics, 131*(1), S38–S49.