COMMENTARY

Routine Outcome Monitoring: Coming of Age—With the Usual Developmental Challenges

Bruce E. Wampold
University of Wisconsin and Modum Bad Psychiatric Center, Vikersund, Norway

The articles in this series present a variety of systems that involve routine outcome monitoring and the provision of feedback to therapists and/or patients with the goal of improving the quality of mental health care. Sufficient evidence exists for the adopting 1 of these systems (or 1 that was not discussed in this series); nevertheless, a number of challenges exist. Issues related to identifying the efficacious components, implementation, utilization, scientific inquiry, and the next generation are discussed.

Keywords: routine outcome monitoring, patient progress, feedback, practice-based evidence

"Where should I go?"

—Alice.

"That depends on where you want to end up."

—The Cheshire Cat.

The monitoring of patient progress, often called routine outcome monitoring (ROM) among other terms (e.g., practice-based evidence), evidently has come of age. There now exist multiple measures as well as systems for using and interpreting the outcome measures completed by patients over the course of therapy. The articles in this series present the case for the utility of many of these, including the OQ-System (Lambert, 2015), PCOMS (Duncan & Reese, 2015), TOP (Boswell, Kraus, Castonguay, & Youn, 2015), CORE System (Barkham, Mellor-Clark, & Stiles, 2015), CCAPS (Youn et al., 2015), BHM (Kopta, Owen, & Budge, 2015), and ACORN (Brown, Simon, Cameron, & Minami, 2015).[1] There are others, as well (e.g., the STIC, appropriate for systemic assessments; Pinsof et al., 2009). Some of these systems have found their way to SAMHSA's National Registry of Evidence-based Programs and Practices (viz., the OQ and the PCOMS) and others have been adopted by various systems of care including insurance companies, the U.S. military, state agencies, colleges and universities, and even entire countries. Coming of age, here used as a metaphor for the acceptance as a fully engaged participant in the mental health delivery system, brings with it challenges related to identity, legitimacy, life trajectory, generational changes in culture, potential progeny, as well as host of others. Miller, Hubble, Chow, and Seidel (2015), in their contribution to this series, discuss several of the challenges as ROM comes of age. In this article, I mention what I think are some others, with an emphasis on what ROM has to offer, currently and in the future, as well as some cautions. Entering society comes with risks of many types but most come of age without much introspection, which I hope to encourage. In my view, the use of ROM is the most noteworthy advance in psychotherapy in the last 25 years, yet the potential of this method will be realized only if several important challenges are met.

Coming of Age Challenges

What is Coming of Age, Anyway?

I have used the acronym ROM generically, but what exactly is it we are discussing here? It seems to be well established that collecting information about patient progress and providing that information to therapists, often called *feedback*, improves outcomes (Lambert & Shimokawa, 2011; Shimokawa, Lambert, & Smart, 2010). Yet, each of the systems are composed of various elements, including a scale that is administered to patients on a regular basis (typically every session) and presentation of the information collected to the therapist, the patient, or both the patient and the therapist. The information provided to the recipient has various components, which may include simply the scores on the scale or various subscales (see below), a graphical presentation of the scores over time (illustrated in several of the articles in this series), comparison of trajectory with normative data for patient progress (usually presented graphically as well), various "signals" that indicate whether the patient is making adequate progress or not or whether the patient is in the severe range, computer generated messages, as well as ancillary scales that might provide

[1] The measures and associated systems typically are so well known that their abbreviations are initialisms if not acronyms and thus I use the abbreviations initially. Also note that reference to the measure and reference to systems using the measure often use the same abbreviation, creating confusion if care in not taken.

information about why the patient is or is not making normal progress (e.g., scales measuring alliance, readiness for change, and social support). When it is said generically, "feedback improves outcomes," what is meant by feedback may often be ambiguous, and this is not a trivial issue, as I shall argue.

Evidence, Science, and the Black Box

The evidence for two of the systems (viz., the OQ-system and PCOMS) appears to be strong, particularly for preventing deterioration of cases at risk for failure (Lambert & Shimokawa, 2011; Shimokawa et al., 2010). However, before accepting the efficacy of "feedback" as a means to improve the quality of care, some issues need to be considered. First, as Miller et al. (2015) stated, the effect sizes for feedback versus treatment as usual are modest and there is a danger that they are overestimates of what might be expected in the future, given how most effects decline over years in the sciences, and in psychology in particular (Yong, 2012), including the effect of CBT as a treatment for depression (Johnsen & Friborg, 2015), may not represent "truth" (Ioannidis, 2014), and may not be applicable in practice settings. Of course, such concerns are not unique to ROM, but nevertheless there is some cause for concern. Most of the research cited in meta-analyses has been conducted by advocates of a particular system, which reduces some of the implementation problems discussed below. Well-conducted large trials have not produced effects (de Jong, van Sluis, Nugter, Heiser, & Spinhoven, 2012; see also Rise, Eriksen, Grimstad, & Steinsbekk, 2015) and several recent failed trials, with which I am familiar, have not been published. With some notable exceptions (e.g., Anker, Duncan, & Sparks, 2009), the instrument used to assess outcomes was the same instrument used for ROM. That is, therapists are given particular information about the patient, which is then discussed or used in subsequent sessions; it might well be that the focus on the particular instrument increases the likelihood of change on *that* instrument but not on another measures of outcome (symptom measures, quality of life, well-being, interpersonal functioning). Moreover, having clients complete the same measure each session more than likely creates some testing effects. The effects of feedback would be more robust if the success of therapy was measured by an instrument *not* used for ROM. By themselves, these issues may not be surprising for an emerging technology, but only serve as cautionary notes, to consider vis-à-vis the very promising results that have been reported.

However, in terms of sustaining ROM as a viable method to improve the quality of mental health services, several scientific questions need to be resolved. As discussed above, there are many components of ROM and, for the most part, the efficacious ingredients of the package have not been identified (sound familiar?). In some studies, it appears that a warning signal (e.g., the patient is off track!) is critical (Probst et al., 2013) where in other studies it is not (Amble, Gude, Ulvenes, Stubdal, & Wampold, 2015). It is also unclear whether integrating the information about patient progress in therapy (e.g., by discussing it with the patient) is important (see Amble et al., 2015). Discussing patient progress with the patient each session is endemic to PCOMS (Duncan & Reese, 2015), a system has produced encouraging results. Therapists report that the graph and discussion with the patient are the most important aspects of ROM (Amble et al., 2015). As well, it seems that the effects of feedback are limited to therapists who are

enthusiastic about ROM and use the system (de Jong et al., 2012; Lutz et al., 2015). And here is something incredible: After all the years that we have studied ROM, we do not know how therapists use ROM to improve the quality of service—and maybe even more incredible, we have not investigated this important question. What about ROM creates effects? The authors in this series provided excellent examples of how ideally to use their systems and now research is needed to identify the important ingredients. Although many RCTs of ROM have been conducted, there has been a lack of attention to a theory of how it works. For the most part the pursuit of improvement has been pragmatic—if it works, great! Pragmatism has a prominent place in progress but it is my view that "there is nothing as practical as a good theory" (Van de Ven, 1989; see also Miller, 1994; Popper, 1963). There is an immense literature about the relationship between feedback vis-à-vis learning and performance; Len Bickman has used this literature to understand how feedback to therapist might be optimally structured (e.g., Sapyta, Riemer, & Bickman, 2005), but reference to this literature or its application was not emphasized in the articles in this series, although this was not the focus of the articles.[2] I will return to this issue when I discuss expertise below.

Implementation

Anyone who has been involved with the implementation of ROM in practice settings, either in the context of a randomized clinical trial or as part of standard practice, knows that implementation is difficult. Indeed, developers of ROM have discussed the challenges involved (Boswell, Kraus, Miller, & Lambert, 2015) and this issue was discussed in several of the articles in this series (see Barkham et al., 2015; Boswell et al., 2015; Brown et al., 2015). However, the challenges are formidable and should not be minimized (see Wolpert, 2014 as well). Similar to the situation with ROM, there is a science (theory and empirical results) relative to implementation, which was discussed in some detail by Miller et al. (2015) in this issue. The challenges to implementation will not be rehashed here, but some related issues that need attention will be mentioned (and hopefully are not a rehash).

Managers of care need to consider the financial cost of ROM, which include costs of system, but also opportunity costs associated with training and time to use the system, technology infrastructure, and so forth. Some of the systems belong to for profit companies and some belong to nonprofits. Personally, I am agnostic in this regard as the profit motive can create innovation and progress, although profit motives can create conflicts of interest.

Implementation of ROM raises other challenges for managers of care. With good intentions and based on empirical evidence, a manager implements ROM in a system of care, using the principles of implementation science. There is a modest increase in effect sizes, as expected from the literature. However, there is a nagging fact that depresses the manager's ebullient mood—some therapists consistently are not achieving expected outcomes, a fact that is predicted by what is known about therapist effects in naturalistic settings (Baldwin & Imel, 2013)—indeed, it is clear that the bottom 5% to 10% of therapists are a consistent drag on effects of a system of care (Saxon & Barkham, 2012; Wampold & Brown, 2005). Moreover, it appears that in general use of ROM is not

---

[2] Kopta et al. (2015) did discuss the relevance of phase theory for ROM.

improving the outcome of therapists over time (see next section). What should the manager do if despite efforts some therapists continue to underperform? Imel, Sheng, Baldwin, and Atkins (2015) have shown in a simulation that removing the bottom 5% of therapists periodically and replacing them at random has a remarkable effect on overall performance. However, economists have demonstrated that various strategies of using performance data for employment decision, including various pay for performance incentives, can have perverse effects, so I am not making a recommendation in this regard. However, once ROM is employed, systems of care are presented with some dilemmas (e.g., what do with chronically underperforming therapists) given their responsibilities to patients and payers. Boswell et al. (2015) suggested that patients be assigned to therapists based on therapist's expertise in particular areas.

Development of Expertise

Tracey, Wampold, Lichtenberg, and Goodyear (2014) argued that it is difficult to develop expertise in psychotherapy. That is to say, under normal conditions therapists do not improve over time. How might ROM help therapists develop expertise. Mere exposure to ROM (i.e., being a therapist in an agency that uses ROM) does not seem to lead to improvement—indeed, it appears that therapists actually get poorer outcomes over time in such an agency, although the decrease was very small (Goldberg et al., 2015). This phenomenon is not unexpected, given the conditions that are needed to develop expertise (see Tracey, Wampold, Lichtenberg, & Goodyear, 2014). Miller et al. (2015) discussed how ROM, as a component of deliberate practice, can lead to continued improvement.

Having to Choose One

Suppose that you are a manager of clinic or a system of care and have decided that ROM should be used, notwithstanding the issues discussed. Which of the measures/systems would you select? They all come highly recommended and seem to have more than adequate credentials (e.g., adequate reliability and validity). Yet, they are not the same, but is what differentiates them important?[3] Here I discuss a few important distinctions. The length and complexity of the measures varies. The Outcome Rating Scales (ORS), used in the PCOMS system, consists of four items and no subscales, whereas the adult TOP clinical scales contain 58 items assessing 12 symptom and functional domains. Clearly, there are trade-offs to be made here: information provided versus time and effort to collect it. The more complex instruments contain subscales, but it must be determined how useful they are in practice and how well validated they are. Although it might appears that longer instruments will provide greater information, this might not be the case. The factor structures of instruments appear not to be particularly robust, as the structure varies from sample to sample, cannot be replicated, and is often complex and hierarchical (e.g., see Bludworth, Tracey, & Glidden-Tracey, 2010, for a bilevel factor analysis of the OQ). Moreover, there is clearly a general factor in such instruments, which is not surprising because there appears to be a strong general psychopathology factor (Caspi et al., 2014). The correlations among these instruments, although largely untested, will be quite high. However, clinicians find the subscales and even

individual items, especially related to risk, to be informative and useful in discussions with patients. However, adopters of ROM will have to weigh costs, benefits, incremental validity, and acceptability of the longer scales vis-à-vis shorter ones.

Education, Training, and Supervision

Several of the articles in this series mention the role ROM can play in education, training, and supervision. Psychology has placed emphasis on competencies (Kaslow, 2004; Kaslow et al., 2004), but from my perspective I would much rather know that graduates of professional training program demonstrated that they consistently help patients, which of course is demonstrable when ROM are adopted in such programs. However, the utility of ROM extends beyond documentation of effectiveness. If ROM is used in conjunction with an assessment of skills in various areas, then training can be targeted toward particular areas that need growth (e.g., empathic responding or delivery of a coherent and cogent treatment protocol). In a survey of trainees, it was found that about 50% of respondents had used ROM with at least one client, although over 30% were not aware that ROM exists (Overington, Fitzpatrick, Hunsley, & Drapeau, 2015)! However, of those who used ROM in training, most found that it was useful and were twice as likely to state that they would use ROM in postdegree practice than was the case for nonusers. From my own experience, it is clear that a major impediment to using ROM in university training clinics is the cost and I recommend that the owners of the various systems provide them at no or low-cost to training programs.

Research

One of the most exciting positive impacts of the coming of age of ROM from my perspective is that it has spawned a plethora of research on routine mental health service. Assessing outcome as well as various process measures at every session allows for sophisticated longitundinal analyses (see, e.g., Falkenström, Granström, & Holmqvist, 2013, 2014), and benchmarking (Barkham et al., 2015; Minami et al., 2008). There will be much learned from practice-based networks in the coming years. From my perspective, albeit with a small dose of bias, we are learning much more about psychotherapy from naturalistic data sets with frequent measurements than we have learned from randomized clinical trials comparing different forms of therapy (Laska, Gurman, & Wampold, 2014).[4]

Spawning the Next Generation

Yes, ROM has come of age, and with it exciting times. But inklings of the next generation are becoming apparent already. The use of paper and pencil instruments at the beginning or end of each session will seem like propeller driven airliners when smart phones can assess mental states, such as depression, from the data that

---

[3] Relative costs, discussed above, are a factor, of course.

[4] It is interesting to note that RCTs are now including frequent measurement and performing longitundinal analyses as well, although RCTs are typically relatively small and unpowered, particularly when therapists are considered.

exists on patient's phone without deliberate user input, using data such circadian movement, normalized entropy location variability, and call duration and frequency (see e.g., Saeb et al., 2015). Ratings of therapy process by humans (either observers or participants) will be replaced by learning machines that detect process measures, such as empathy, without involving human respondents (participants or observers; see, e.g., Imel et al., 2014; Imel, Steyvers, & Atkins, 2015; Lord, Sheng, Imel, Baer, & Atkins, 2015). Our basic methods, which involved paper and pencil instruments, completed by the patient (albeit without the paper and pencil, but essentially the same) and the use of coders, have remained essentially unchanged for three-quarters of a century and have greatly limited our research efforts. Moreover, these data have the potential to improve the quality of mental health care. For example, the clinician will have access the mental health status of patients in real time, rather than status only prior to the session reported retrospectively. Moreover, advances in our knowledge of psychotherapy process will inform the therapist more accurately about what is and is not working in therapy.

Conclusions

Although ROM is now coming of age, Paul W. Clement assessed the progress of every one of his patients since he began practice as a licensed psychologist in July 1966 (Clement, 1994). It was his opinion that failure to answer empirically the question "Are you any good?" (p. 173) was irresponsible and unethical. Yes, there is resistance to ROM by therapists, implementation is difficult, challenges to using the data to improve the quality of care are present, and a number of issues remain. But the pervasiveness of ROM attests to its robustness—at the very least we can answer Clement's question, "Are we any good?" However, the potential to improve the quality of mental health services, as well as be accountable for the services we provide, is the exciting (and necessary) aspect of this movement.

The purpose of this comment was to present the challenges that exist so that ROM can reach its potential. I look forward to watching ROM mature.

References

Amble, I., Gude, T., Ulvenes, P., Stubdal, S., & Wampold, B. E. (2015). How and when feedback works in psychotherapy: Is it the signal? *Psychotherapy Research,* 1–11. http://dx.doi.org/10.1080/10503307.2015.1053552

Anker, M. G., Duncan, B. L., & Sparks, J. A. (2009). Using client feedback to improve couple therapy outcomes: A randomized clinical trial in a naturalistic setting. *Journal of Consulting and Clinical Psychology, 77,* 693–704. http://dx.doi.org/10.1037/a0016062

Baldwin, S. A., & Imel, Z. E. (2013). Therapist effects: Finding and methods. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (6th ed., pp. 258–297). New York, NY: Wiley.

Barkham, M., Mellor-Clark, J., & Stiles, W. B. (2015). A CORE approach to progress monitoring and feedback: Enhancing evidence and improving practice. *Psychotherapy, 52,* 402–411. http://dx.doi.org/10.1037/pst0000030

Bludworth, J. L., Tracey, T. J. G., & Glidden-Tracey, C. (2010). The bilevel structure of the Outcome Questionnaire-45. *Psychological Assessment, 22,* 350–355. http://dx.doi.org/10.1037/a0019187

Boswell, J. F., Kraus, D. R., Castonguay, L. G., & Youn, S. J. (2015). Treatment outcome package: Measuring and facilitating multidimen-

sional change. *Psychotherapy, 52,* 422–431. http://dx.doi.org/10.1037/pst0000028

Boswell, J. F., Kraus, D. R., Miller, S. D., & Lambert, M. J. (2015). Implementing routine outcome monitoring in clinical practice: Benefits, challenges, and solutions. *Psychotherapy Research, 25,* 6–19. http://dx.doi.org/10.1080/10503307.2013.817696

Brown, G. S., Simon, A., Cameron, J., & Minami, T. (2015). A collaborative outcome resource network (ACORN): Tools for increasing the value of psychotherapy. *Psychotherapy, 52,* 412–421. http://dx.doi.org/10.1037/pst0000033

Caspi, A., Houts, R. M., Belsky, D. W., Goldman-Mellor, S. J., Harrington, H., Israel, S., . . . Moffitt, T. E. (2014). The p factor: One general psychopathology factor in the structure of psychiatric disorders? *Clinical Psychological Science, 2,* 119–137. http://dx.doi.org/10.1177/2167702613497473

Clement, P. W. (1994). Quantitative evaluation of 26 years of private practice. *Professional Psychology: Research and Practice, 25,* 173–176. http://dx.doi.org/10.1037/0735-7028.25.2.173

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. http://dx.doi.org/10.1080/10503307.2012.673023

Duncan, B. L., & Reese, R. J. (2015). The Partners for Change Outcome Management System (PCOMS): Revisiting the client's frame of reference. *Psychotherapy, 52,* 391–401. http://dx.doi.org/10.1037/pst0000026

Falkenström, F., Granström, F., & Holmqvist, R. (2013). Therapeutic alliance predicts symptomatic improvement session by session. *Journal of Counseling Psychology, 60,* 317–328. http://dx.doi.org/10.1037/a0032258

Falkenström, F., Granström, F., & Holmqvist, R. (2014). Working alliance predicts psychotherapy outcome even while controlling for prior symptom improvement. *Psychotherapy Research, 24,* 146–159. http://dx.doi.org/10.1080/10503307.2013.847985

Goldberg, S. B., Rousmaniere, T., Whipple, J. L., Nielsen, S. L., Hoyt, W. T., & Wampold, B. (2015). *Do psychotherapists improve with time and experience? A longitudinal analysis of real world outcome data.* (Manuscript submitted for publication.)

Imel, Z. E., Barco, J. S., Brown, H. J., Baucom, B. R., Baer, J. S., Kircher, J. C., & Atkins, D. C. (2014). The association of therapist empathy and synchrony in vocally encoded arousal. *Journal of Counseling Psychology, 61,* 146–153. http://dx.doi.org/10.1037/a0034943

Imel, Z. E., Sheng, E., Baldwin, S. A., & Atkins, D. C. (2015). Removing very low-performing therapists: A simulation of performance-based retention in psychotherapy. *Psychotherapy, 52,* 329–336.

Imel, Z. E., Steyvers, M., & Atkins, D. C. (2015). Computational psychotherapy research: Scaling up the evaluation of patient-provider interactions. *Psychotherapy, 52,* 19–30. http://dx.doi.org/10.1037/a0036841

Ioannidis, J. P. A. (2014). How to make more published research true. *PLoS Medicine, 11,* e1001747. http://dx.doi.org/10.1371/journal.pmed.1001747

Johnsen, T. J., & Friborg, O. (2015). The effects of cognitive behavioral therapy as an anti-depressive treatment is falling: A meta-analysis. *Psychological Bulletin, 141,* 747–768. http://dx.doi.org/10.1037/bul0000015

Kaslow, N. J. (2004). Competencies in professional psychology. *American Psychologist, 59,* 774–781. http://dx.doi.org/10.1037/0003-066X.59.8.774

Kaslow, N. J., Borden, K. A., Collins, F. L., Jr., Forrest, L., Illfelder-Kaye, J., Nelson, P. D., . . . Willmuth, M. E. (2004). Competencies conference: Future directions in education and credentialing in professional psychology. *Journal of Clinical Psychology, 60,* 699–712. http://dx.doi.org/10.1002/jclp.20016

Kopta, M., Owen, J., & Budge, S. L. (2015). Measuring psychotherapy with the Behavioral Health Measure-20: Efficient and comprehensive. *Psychotherapy, 52,* 442–448. http://dx.doi.org/10.1037/pst0000035

Lambert, M. J. (2015). Progress feedback and the OQ-System: The past and the future. *Psychotherapy, 52,* 381–390. http://dx.doi.org/10.1037/pst0000027

Lambert, M. J., & Shimokawa, K. (2011). Collecting client feedback. In J. C. Norcross (Ed.), *Psychotherapy relationships that work: Evidence-based responsiveness* (2nd ed., pp. 203–223). New York, NY: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780199737208.003.0010

Laska, K. M., Gurman, A. S., & Wampold, B. E. (2014). Expanding the lens of evidence-based practice in psychotherapy: A common factors perspective. *Psychotherapy, 51,* 467–481. http://dx.doi.org/10.1037/a0034332

Lord, S. P., Sheng, E., Imel, Z. E., Baer, J., & Atkins, D. C. (2015). More than reflections: Empathy in motivational interviewing includes language style synchrony between therapist and client. *Behavior Therapy, 46,* 296–303. http://dx.doi.org/10.1016/j.beth.2014.11.002

Lutz, W., Rubel, J., Schiefele, A. K., Zimmermann, D., Böhnke, J. R., & Wittmann, W. W. (2015). Feedback and therapist effects in the context of treatment outcome and treatment length. *Psychotherapy Research.* Advance online publication. http://dx.doi.org/10.1080/10503307.2015.1053553

Miller, D. (1994). *Critical rationalism: A restatement and defense.* Chicago, IL: Open Court.

Miller, S. D., Hubble, M. A., Chow, D., & Seidel, J. (2015). Beyond measures and monitoring: Realizing the potential of feedback-informed treatment. *Psychotherapy, 52,* 449–457. http://dx.doi.org/10.1037/pst0000031

Minami, T., Wampold, B. E., Serlin, R. C., Hamilton, E. G., Brown, G. S., & Kircher, J. C. (2008). Benchmarking the effectiveness of psychotherapy treatment for adult depression in a managed care environment: A preliminary study. *Journal of Consulting and Clinical Psychology, 76,* 116–124. http://dx.doi.org/10.1037/0022-006X.76.1.116

Overington, L., Fitzpatrick, M., Hunsley, J., & Drapeau, M. (2015). Trainees' experiences using progress monitoring measures. *Training and Education in Professional Psychology, 9,* 202–209. http://dx.doi.org/10.1037/tep0000088

Pinsof, W. M., Zinbarg, R. E., Lebow, J. L., Knobloch-Fedders, L. M., Durbin, E., Chambers, A., . . . Friedman, G. (2009). Laying the foundation for progress research in family, couple, and individual therapy: The development and psychometric features of the initial systemic therapy inventory of change. *Psychotherapy Research, 19,* 143–156. http://dx.doi.org/10.1080/10503300802669973

Popper, K. (1963). *Conjectures and refutations.* London, UK: Routledge.

Probst, T., Lambert, M. J., Loew, T. H., Dahlbender, R. W., Göllner, R., & Tritt, K. (2013). Feedback on patient progress and clinical support tools for therapists: Improved outcome for patients at risk of treatment failure in psychosomatic in-patient therapy under the conditions of routine practice. *Journal of Psychosomatic Research, 75,* 255–261.

Rise, M. B., Eriksen, L., Grimstad, H., & Steinsbekk, A. (2015). The long-term effect on mental health symptoms and patient activation of using patient feedback scales in mental health out-patient treatment. A randomised controlled trial. *Patient Education and Counseling.* Advance online publication. http://dx.doi.org/10.1016/j.pec.2015.07.016

Saeb, S., Zhang, M., Karr, C. J., Schueller, S. M., Corden, M. E., Kording, K. P., & Mohr, D. C. (2015). Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: An exploratory study. *Journal of Medical Internet Research, 17,* e175. http://dx.doi.org/10.2196/jmir.4273

Sapyta, J., Riemer, M., & Bickman, L. (2005). Feedback to clinicians: Theory, research, and practice. *Journal of Clinical Psychology, 61,* 145–153. http://dx.doi.org/10.1002/jclp.20107

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. http://dx.doi.org/10.1037/a0028898

Shimokawa, K., Lambert, M. J., & Smart, D. W. (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. http://dx.doi.org/10.1037/a0019247

Tracey, T. J. G., Wampold, B. E., Lichtenberg, J. W., & Goodyear, R. K. (2014). Expertise in psychotherapy: An elusive goal? *American Psychologist, 69,* 218–229. http://dx.doi.org/10.1037/a0035099

Van de Ven, A. H. (1989). Nothing is quite so practical as a good theory. *Academy of Management Review, 14,* 486–489. http://dx.doi.org/10.5465/AMR.1989.4308370

Wampold, B. E., & Brown, G. S. (2005). Estimating therapist variability: A naturalistic study of outcomes in managed care. *Journal of Consulting and Clinical Psychology, 73,* 914–923. http://dx.doi.org/10.1037/0022-006X.73.5.914

Wolpert, M. (2014). Uses and abuses of patient reported outcome measures (PROMs): Potential iatrogenic impact of PROMs implementation and how it can be mitigated. *Administration and Policy in Mental Health and Mental Health Services Research, 41,* 141–145. http://dx.doi.org/10.1007/s10488-013-0509-1

Yong, E. (2012). Replication studies: Bad copy. *Nature, 485,* 298–300. http://dx.doi.org/10.1038/485298a

Youn, S. J., Castonguay, L. G., Ziao, H., Janis, R., McAleavey, A. A., Lockard, A. J., . . . Hayes, J. A. (2015). The Counseling Center Assessment of Psychological Symptoms (CCAPS): Merging clinical practice, training, and research. *Psychotherapy, 52,* 432–441. http://dx.doi.org/10.1037/pst0000029