

Draft report for distribution and comment.

Prepared by: Jeb Brown, PhD & Ashley Simon, MA; Center for Clinical Informatics

Abstract: Purpose & Findings

The purpose of this report is to address an increased call for accountability and demands that providers demonstrate results. The initial intended audience is mental health professionals presently involved in the ACORN collaboration. As such, the paper is also intended to be accessible and meaningful to practicing clinicians and those with clinician/administrative responsibilities. As the results demonstrate, there are clear implications for patient care. The paper outlines practical and empirically validated pathways to improve patient outcomes.

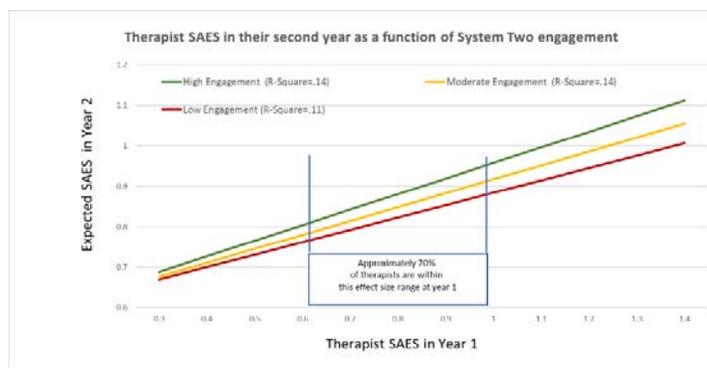
The paper assumes the reader has some familiarity with the ACORN Toolkit and its capabilities.

The report will eventually be prepared for publication in a peer reviewed journal, at which time more background information and references relevant to the ACORN collaboration will be included. However, preparation for publication too often involves altering the readability for the average practicing professional, and may delay dissemination substantially.

Organizations involved with the ACORN collaboration have the capacity to implement quality improvement interventions which could have immediate impact on patient care. Any delay in the dissemination of relevant information simply means that more time passes before the information can have an impact on clinical care.

This report summarizes the evidence that clinicians participating in the ACORN collaboration are, as a group, achieving better outcomes over time. Outcomes for patients treated in the therapist's second year of participation are better than the year before at a high level of statistical confidence.

Furthermore, therapists who logged in an average of once a week or more (High Engagement) to review their results via the Toolkit showed much larger gains in the Severity Adjusted Effect Size (SAES) in the second year compared to those who did not view their results at all (Low Engagement).



Introduction

The past decade has seen increased research interest in benchmarking outcomes, and more recently, searching for evidence that outcomes can in fact be improved over time.

Goodyear, Wampold, Tracey & Lichtenberg (2017) argue for a definition of therapist expertise that includes both evidence of effective treatment and the ability to improve results over time.

To quote from the abstract:

“How the field understands psychotherapy expertise is important. It affects how we practice and how we prepare others for practice. As in our other work, we argue that the most meaningful definition of expertise must involve steady improvement over time to achieve superior performance on some meaningful measure, which typically is client outcome.”

While there is ample evidence that routine measurement and feedback improve outcomes in general, primarily by reducing premature dropout of patients faring poorly in treatment, it is much less clear that individual therapists will improve their results over time. In fact, until quite recently the evidence would suggest just the opposite: therapists do not become more effective over time. From Wampold & Imel (2017)

“Therapists do not get better with time or experience. That is, over the course of their professional careers, on average, it appears that therapists do not improve, if by improvement we mean ‘achieve better outcomes’.”

Two articles found that experienced clinicians working in clinics serving as training sites for graduate interns did not have better outcomes than the trainees at those sites (Okiishi, Lambert Neilson & Ogels, 2003; Minami et al., 2009). Wampold & Brown (2005) investigated outcomes in a large sample of therapists in private practice and found no relationship between the provider’s outcomes and type of graduate training or years of experience.

The mean effect of approximately .8 reported from large samples of therapists is essentially equivalent to effect sizes reported in meta-analyses combining decades of well conducted clinical trial of psychotherapy efficacy. (Minami, Brown, McCulloch, & Bolstrom, 2012; Brown, Simon, Cameron, & Minami, 2015; Minami, Serlin, Wampold, Kircher, & Brown; 2008.)

In other words, the typical therapist achieves results comparable to the so-called gold standard of clinical trials. It is not clear how the average therapist is expected to improve if their outcomes are already quite good. Are they expected to exceed benchmarks from clinical trials?

Three recent studies have attempted to look at whether there is evidence that therapists improve over time if they are practicing routine outcomes measurement combined with continuous feedback. Goldberg et al. (2016a) reported on a sample of 6,191 patients treated by 170 therapists over a period of almost five years. The results were not encouraging. To quote from the abstract:

“Therapists achieved outcomes comparable with benchmarks from clinical trials. However, a very small but statistically significant change in outcome was detected indicating that on the whole, therapists’ patient prepost d tended to diminish as experience (time or cases) increases.”

A second article, involving many of the same authors presented a more encouraging result. Goldberg et al. (2016b) reported on 5,128 patients seen by 153 therapists from a single clinic over a period of 7 years. They found that effect size increased at a rate of .03 per year for the agency. They further determined that the improvement was not due to turnover in therapists or hiring more effective therapist. Rather, it was due to the individual therapists improving over time. While encouraging, the reported effect size annual gains are quite modest.

The largest gains reported in year to year effect size come from the ACORN collaboration (Brown, Simon, Cameron & Minami, 2016). This article reported results for 704 therapists treating 40,204 clients over their first few years of participation in the collaboration. The overall effect size gain was .07.

This study was limited to data from one adult outcome questionnaire and may not accurately reflect results likely to be obtained in a more heterogeneous treatment population using a wide variety of questionnaires. The current study will attempt to address this limitation.

The study will also provide a deeper exploration of the relationship between therapist engagement (time spent on monitoring and improving outcomes) and the probability of improvement for therapists across the continuum of effectiveness.

Thinking fast and slow: System 1 & System 2

A significant body of research exists on what is often called deliberate practice, or practice outside of practice. This research attempts to explain the pattern behind improved performance across a broad area of skilled activities. The main idea is that the accomplished expert does not achieve their results by practice alone, but by thinking about what they are doing, monitoring results, and actively planning how to improve. This concept has been applied to the practice of psychotherapy in Chow et al., 2015; Goldberg et al., 2016.

We suggest that this investigation can be understood as fitting into a larger body of research on heuristics and expert decision making that has revolutionized many industries over the past 25 years. This phenomenon has been chronicled in recent bestselling books.

The *Undoing Project: The Friendship That Changed Our Minds* (Lewis, 2016) tells the story of the remarkable and odd couple collaboration between two Israeli psychologists, Daniel Kahneman and Amos Tversky, that revolutionized our understanding of human cognition and decision making. Kahneman, who was awarded the Nobel Prize for Economics in 2002, summarized this research in his bestselling book *Thinking Fast and Slow* (Kahneman, 2011). Sports fan will of course recognize the impact of data to improve expert decision making as described in *Moneyball: The Art of Winning an Unfair Game* (Lewis, 2003). It only took a few

years for the implications of the use of data and algorithms to augment expert judgement to revolutionize baseball and other sports.

Researchers have settled on the use of the terms “System 1” and “System 2” to make sense of this evidence. System 1 is for thinking fast and making quick decisions, without taxing the brain too much. This is the system we are largely using throughout the day and in social situations requiring flexible and rapid responses. System 2 (thinking slow) requires more mental work in the form of focused attention, problem solving, use of data, reflective thought, rational analysis, etc. For purposes of this paper, we will use the term System 2 for mental activities that are attributed to concept of deliberate practice.

When it comes to making judgements, System 1 is the default system in use, even among experts. System 1 draws on personal experience and relies on various heuristics (rules of thumb, mental shortcuts) to arrive at a decision quickly and efficiently. Not only is this the system we primarily use in social interactions, but it is also the one we use in the practice of psychotherapy. The more experience a therapist has, the more they are likely to rely on System 1 during the therapy session, relying on past experience to guide them.

Research has demonstrated that System 1 is subject to certain kinds of biases and systematic errors. In mental health care, these errors often result in the tendency to overestimate patient improvement. Many mechanisms may account for this effect, including confirmation bias (therapists looking for evidence that their therapy is working).

The use of data in the service of decision support is inherently a System 2 activity. It requires the expert to look actively look at the data and think about what it is saying, aided by algorithms that are empirically derived from analyzing large sets of relevant data. This takes time and effort.

In virtually every inquiry into expertise, it has become clear that the use of data combined with decision support algorithms can inform decision making and minimize the impact of errors due to System 1 heuristics. This fact should no longer be a surprise to anyone who has done even a cursory review of research on expert judgement. Unfortunately, mental health professionals have lagged behind other professions in recognizing this essential fact of life. This is despite the strong evidence emerging from multiple clinical trials that the use of outcome measurement and algorithm driven feedback in psychotherapy results in better treatment outcomes compared to simple use of questionnaires without algorithm driven feedback to the clinician. (Wampold & Imel, 2016; Lambert, 2010)

It is clearly a challenge to measure the extent at which a therapist engages in deliberate practice or System 2 thought. However, the ACORN Toolkit does provide a practical and readily available proxy measure.

One of the early participants and contributors to the ACORN collaboration, Joyce McColluch, then V.P of Health Informatics at Optum Health, suggested the ACORN Toolkit begin to measure how much time therapists spent looking at their data. Beginning in 2011, the Toolkit began to automatically record each time a therapist logged into their account and the amount of

time they spent on the site. It also recorded the number of different page views requested during each session onsite.

The ACORN collaboration has now accumulated over five years of data measuring both the treatment outcomes and the amount of time therapists spent looking at their results, with algorithm driven alerts for those cases which had a high risk for a poor outcome.

Furthermore, each time a therapist logs in they are also provided information on overall aggregated results, which provides real time performance feedback. The user is also able to compare results to benchmarks and to track changes in aggregate results over time. It is a truism that any activity requiring a high level of skill requires meaningful performance measurement and feedback in order to improve performance. The mental health profession would be a very strange profession indeed if this proved not to be the case.

For the purposes of this report, we will refer to the metric of therapists' time online as a measure of "engagement" in the process of trying to improve outcomes. We assume that at least during this time they are engaged in System 2 thinking. While this an admittedly a crude metric for assessing deliberate practice/System 2 thinking, it does have the virtue of being easy to measure and remarkably predictive of therapist improvement in treatment outcomes.

Therapists vary greatly in their engagement in monitoring results, graphically demonstrated by the fact that over 40% of the therapists using the ACORN system never logged into to view their data in during the second year of participation. Looking at the bright side, this group of unengaged therapists makes an excellent naturalistic control group. Sadly, the evidence shows that this silver lining comes at a cost to patient care and outcomes.

Benchmarking Outcomes

To assess a therapist's effectiveness, it is necessary to have a well-developed methodology that addresses multiple measurement challenges such as use of different questionnaires, differences in case mix, and small sample sizes.

The ACORN system benchmarks outcomes using multivariate statistics to adjust for differences in diagnosis and severity of symptoms at the start of treatment, as measured by the questionnaire at the first assessment. The methodology employed by ACORN was developed over several years using data from thousands of therapists and a variety of questionnaires, culminating in a series of articles (e.g., Minami, Wampold, Serlin, Kircher, & Brown, 2007; Minami, Serlin, Wampold, Kircher, & Brown, 2008a; Minami et al., 2008b; Minami et al., 2009; Minami, Brown, McCulloch, & Bolstrom, 2012; Brown, Simon, Cameron, & Minami, 2015). The resulting statistic is referred to as Severity Adjusted Effect Size (SAES), and is calculated for each patient.

It is not sufficient to simply estimate a therapist's mean effect size based on a simple average of the SAES for all their patients. Therapists with small sample sizes are more likely to have results that vary widely from the mean than those therapists with large sample sizes. To address this problem, Hierarchical Linear Modeling (HLM) was used to estimate mean effect size while adjusting for sample size.

Two recent articles describe the distribution of effect size for individual therapists using this methodology (Minami, Brown, McCulloch, & Bolstrom, 2012; Brown, Simon, Cameron, & Minami, 2016). Though the articles used entirely independent samples and different outcome measures, the resulting distributions were remarkably similar, both with a mean of very close to .8 effect size.

The following table illustrates the use of HLM to adjust for sample size. This analysis was based on a sample of 3,529 ACORN participating therapists with a minimum sample size of 5 patients with a SAES. The table presents the distributions of mean SAES based on a simple averaging (raw SAES) compared to the estimate of mean SAES using HLM, broken down by sample size. Smaller sample sizes are associated with a wider distribution of results simply because of sampling error. Comparing a mean SAES of .9 obtained with sample of 5 patients to a mean SAES of .9 from a sample of 100 is problematic in the extreme. They are not equivalent. The table demonstrates how the use of HLM to calculate the mean SAES adjusts for sample size.

The resulting distribution using HLM has a mean of .79 SAES, with a standard deviation of .17 SAES. This result provides a useful benchmark when comparing the distribution of effect sizes during the first and second years of ACORN participation. For a more complete discussion of the use of HLM in benchmarking treatment outcomes, see Brown, Simon, Cameron, & Minami (2015) available on line at <http://societyforpsychotherapy.org/are-you-any-good-as-a-therapist/>.

Distribution of therapist SAES as a function of sample size

<i>N</i> of clients per therapist	<i>N</i> therapist	Mean # assessments per case	<i>M</i> (<i>SD</i>)	Min	25th	Median	75th	Max
raw SAES								
5 to 14	1069	3.5	0.75 (0.48)	-0.92	0.44	0.73	1.06	2.57
15 to 49	1256	3.8	0.79 (0.32)	-0.23	0.59	0.79	.99	1.97
50 to 99	616	3.6	0.82 (0.26)	0.10	0.67	0.80	0.97	1.58
Over 100	588	3.2	0.82 (0.19)	0.17	0.70	0.81	0.93	1.67
All therapists	3529	3.6	0.79 (0.35)	-0.92	0.59	0.79	.99	2.57
HLM-estimated SAES								
5 to 14	1069	3.5	0.76 (0.13)	0.32	0.68	0.76	0.85	1.28
15 to 49	1256	3.8	0.78 (0.18)	0.28	0.67	0.78	1.0	1.62
50 to 99	616	3.6	0.81 (0.18)	0.28	0.69	0.80	0.93	1.39
Over 100	588	3.2	0.81 (0.17)	0.25	0.71	0.80	0.92	1.58
All therapists	3529	3.6	0.79 (0.17)	0.25	0.68	0.78	0.89	1.62

Study Sample

The study sample was drawn from the ACORN Toolkit database. Therapist inclusion, and by extension patient inclusion, was based on the following inclusion criteria.

First, only therapists who became involved with ACORN after 1/1/2011 were included. The reason was that in that year the ACORN Toolkit began to track user time online.

For each therapist in the database, the date of a first submission on a questionnaire was identified. This date became the start date for that therapist. Subsequent patients seen by that therapist were assigned to a year, based on how long after the therapist start date the client completed treatment, so that patients were grouped by year 1, year 2, etc. In this way, it was possible to look at each therapist's results in their first year and subsequent years, regardless of their start date.

The criteria for patient inclusion in the study included an intake score in the clinical range for the questionnaire employed and at least one subsequent questionnaire within a 120 day period. The

criterion for therapist inclusion was to have at least 5 eligible patients in both the first and second years of ACORN involvement.

Using these criteria, 1,056 therapists who treated a total of 107,178 patients in their first two years of ACORN participation were included. The average sample size per therapist was 28 cases in year 1 and 32 in the second year.

These therapists were associated with 90 different organizations. These were grouped broadly by type of organization. Forty six percent (46%) of the patient sampled were treated at for profit clinics. Another 34% were treated at non-profit clinics, while the last 20% were treated in clinics associated with staff model HMOs.

Age, sex and diagnostic data were available for a subset of the patients, approximately 38% of in the sample depending on the variable in question. Following are the frequencies for cases where data is available.

Frequencies by Diagnostic Groups

Diagnostic Group	Percent
ADHD	6.56%
Adjustment disorders	21.27%
Anxiety disorders	12.26%
Autism spectrum disorders	0.91%
Behavior disorders	3.36%
Bipolar disorders	5.43%
Mood/Depressive disorders	23.93%
Eating disorders	0.27%
Other	1.67%
PTSD	10.17%
Personality disorders	0.19%
Psychotic disorders	4.10%
Substance abuse	9.92%

Frequencies for sex and age

Sex	Percent
Male	57%
Female	43%
Age Group	
Adults	58%
Adolescent	19%
Children	23%

A variety of questionnaires are used within the ACORN collaboration. One of the strengths of this study is that it is not based on a single questionnaire, thus the findings are more generalizable to the broader universe of available outcome questionnaires.

It should be noted however that virtually all the questionnaires in this study include a small number of so called alliance items that ask the patient to give feedback on the prior session. In this way, the therapist has the feedback from the prior sessions at the start of each session. There is no need for the therapist to administer a second alliance questionnaire at the end of each session.

Prior research within the ACORN collaboration had demonstrated that patients who complete these items are likely to remain in treatment longer with significantly larger SAESs. It is possible

therefore that the use of outcome questionnaires without the concurrent use of an alliance measure would not produce the same results.

Following are frequencies for the broad groupings of questionnaires included in this sample.

Questionnaire Type

ACORN -Adult	42.9%
ACORN-Youth-Self	24.6%
ACORN-Youth-Parent	11.6%
PHQ/GAD7	19.1%
ACORN-Addictions-Recovery	0.8%
ACORN-Adult-Severely Mentally Ill	0.8%
ACORN Functional Impairment Rating Scale	0.1%
DLA-20	0.1%

Therapists were grouped into three levels of “Engagement of System 2” activity based on the number of times they logged in during their second year of participation.

Forty two percent (42%) were classified as Low Engagement, based on having never logged in during the second year. Another 46% were classified as showing Moderate Engagement, meaning they logged in occasionally, but less than 50 times during their second year of participation.

A select group, representing 15% of the therapists, logged in at least 50 times during their second year, indicating an average frequency of at least once per week. This group of therapists is classified as High Engagement, though they are spending on average only about 45 minutes a week reviewing their results.

The choice to group therapist into three levels of engagement based on frequency of logins provides an opportunity for a clear graphic demonstration of the impact of therapist engagement in System 2 activity on treatment outcomes. The group of highly engaged therapists provides an example of what is possible, particularly when compared to the group of unengaged therapists.

Results

The following table displays the information on the distribution of SAES in year 1 and year 2. The difference column displays the change from year 1 to year 2 (year 2 value – year 1 value). The mean SAES (using HLM) for year 1 is .79 is identical to the result presented earlier in the sample of all ACORN therapists with at least 5 cases with individual SAES scores.

The mean increase in SAES for all therapists from year 1 to year 2 was .05 ($p < .001$; one sided t-test). This gain is between the results reported by Goldberg et al (2016b) and Brown, Simon, Cameron & Minami (2016).

Year 1 and Year 2 SAES: All Therapists (N-1056)

	Year 1	Year 2	Difference
Mean Sample Size	28	32	4
Mean Annual Logins	50	27	-23
Mean SAES	0.79	0.84	0.05

The differences in SAESs from year 1 to year 2 become more striking when broken out by level of therapist engagement. Note that in all comparisons, the increase in SAES remains significant ($p < .001$).

Following are results for each level of engagement. In the interest of readability, the tables have been condensed to display the most relevant information – Mean SAES, Sample Size, and Mean Login Count.

First is the comparison for the group of therapists with the lowest level of System 2 engagement. In this case, there was an SAES gain of .04 in one year, comparable to the estimated annual increase reported by Goldberg et al (2016b). From this result, the regular use of questionnaires may be resulting in increased effectiveness, regardless of whether therapists look at their results. Note that this group rarely logged in during the first year and not at all in the second year.

Low Engagement (N=440)

	Year 1	Year 2	Difference
Mean Sample Size	29	30	1
Mean Annual Logins	9	0	-9
Mean SAES	0.77	0.81	0.04

The Moderate Engagement group (44% of the sample) showed a one year gain of .05 SAES. Note that their average number of logins also fell substantially from year 1 to year 2, from an average of over twice a month to just over once per month.

Moderate Engagement (N=460)

	Year 1	Year 2	Difference
Mean Sample Size	26	28	2
Mean Annual Logins	34	15	-18
Mean SAES	0.8	0.85	0.05

As expected, the High Engagement group (15% of the sample) showed the largest gain, an increase of .09 SAES. Note also that their average number of logins increased slightly from year 1 to year 2. This group has certainly demonstrated a commitment to monitoring their own results.

High Engagement (N=156)

	Year 1	Year 2	Difference
Mean Sample Size	33	51	18
Mean Annual Logins	144	147	3

Mean SAES

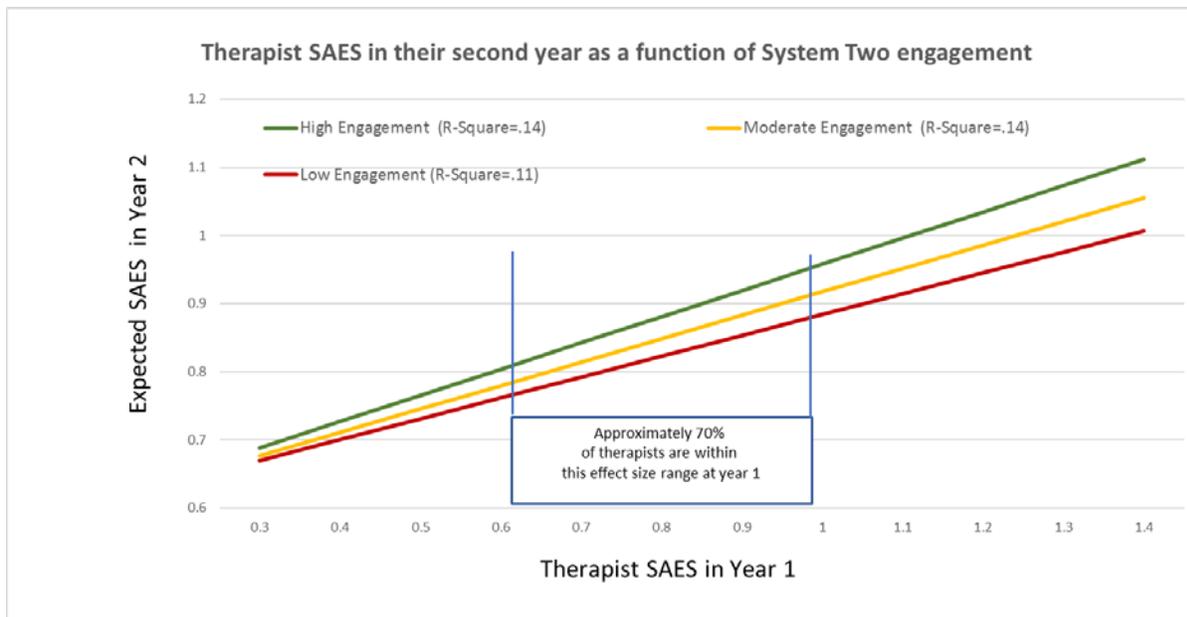
0.79

0.88

0.09

Any attempt to look at changes in performance from one point in time to another must account for regression to the mean. Low performers are most likely to improve, while top performances are likely to see their results diminish somewhat. This is purely a function of sampling and does not reflect a change in the underlying performance. While the use of HLM when calculating mean SAES does help to minimize regression artifacts due to eliminating the effects for small sample sizes, this does not eliminate the phenomena of regression to the mean entirely.

In order control for regression, the relationship between mean SAES at year 1 and the mean SAES at year 2 modeled using a linear model, with year 1 SAES predicting year 2. Using the regression formula for a line that best fits the relationship, it is possible to plot the measured SAES at year 1 against the expected SAES at year 2 for each of the three levels of engagement.



The impact of engagement in System 2 activities is clear. The graph also illustrates that the benefits of System 2 activities are apparent across the range of therapists' mean SAES in year 1. In fact, the larger the SAES was in year 1, the larger the gap in SAES between the high engagement group and the low engagement group. However, while the magnitude of the gap increases, when the magnitude of the gap is calculated as a percent of the SAES at year 1 it increases at a slower rate across continuum of therapist mean SAES, in a range of 6% at the low end of SAES at year 1 to 7.5% at the high end. For those therapists in the middle 70%, the average differences in year 2 for the High Engagement group compared to Low Engagement group was 7.2% of the SAES at year 1.

Discussion and implications for patient care

It appears that no matter how effective the therapist is during their first year, the odds are that their results will be better in the second year if they continue to log in at a high frequency. This suggests that it is in the best interest of patients if therapists guard against complacency and over confidence in their own results.

The improvement in treatment outcomes described in this paper is larger than previously reported by other studies looking at therapist improvement over time. The reason for this is unclear, and it may be that results from the ACORN collaboration will not generalize to other outcomes systems. We hope that others will publish similar findings from other datasets.

It is possible that the routine use of alliance measures concurrent with the outcome measure is contributing to the result of this study. The fact that therapists receive feedback on overall results and not simply case by case results might also contribute to the gains observed.

However, within the community of the ACORN collaboration, the implications for patient care are clear and potentially profound. There appears to be a significant opportunity for improvement if organizations can develop strategies to increase therapist engagement in the process. Any increase in engagement over the next 12 month, particularly among those therapists exhibiting low engagement, is hypothesized to have a measurable impact on patient outcomes in the next 12 months.

To this end, we encourage organizations within ACORN collaboration to consider strategies to increase engagement. This may include disseminating information, conducting in service trainings, and asking clinical supervisors to monitor engagement of their treatment team. Organizations may develop other strategies.

The behind-the-scenes programming of the ACORN clinical information system to produce results for this report is complete. The same set of analyses can be formed quickly for individual clinic locations or for groupings of locations, such as all Medicaid providers in a single state. The output of results directly to the Toolkit for immediate access will require additional programming.

During the coming year, we hope to work with multiple organizations to receive feedback from these reports and to help guide the Toolkit development process. For more information and to give feedback, contact Jeb Brown, PhD. (jebbrown@clinical-informatics.com; 801 541-9720)

References

- 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. doi: 10.1037/pst0000033
- Brown, G. S., Simon, A., & Minami, T. (2015). Are you any good...as a therapist? <http://societyforpsychotherapy.org/are-you-any-good-as-a-therapist/>
- Chow, D. L., Miller, S. D., Seidel, J. A., Kane, R. T., Thornton, J. A., & Andrews, W. P. (2015). The role of deliberate practice in the development of highly effective psychotherapists. *Psychotherapy, 52*, 337–345. doi /10.1037/pst0000015
- Goldberg, S.B., Rousmaniere, T., Miller, S.D., Whipple, J. ⁴, Nielsen, S.L., Hoyt, W.T., & Wampold, B.E. (2016a). Do Psychotherapists Improve With Time and Experience? A Longitudinal Analysis of Outcomes in a Clinical Setting. *Journal of Counseling Psychology, 63* (1), 1-11.
- Goldberg, S.B., Babins-Wagner, R., Rousmaniere, T., Berzins, S., Hoyt, W.T., Whipple, J.L., Miller S.D., & Wampold B.E. (2016b). Creating a climate for therapist improvement: A case study of an agency focused on outcomes and deliberate practice. *Psychotherapy, 53*(3):367-75. doi: 10.1037/pst0000060.
- Goodyear, R.K., Wampold, B.E., Tracey, T.J.G., & Lichtenberg, J.W. (2017). Psychotherapy Expertise Should Mean Superior Outcomes and Demonstrable Improvement Over Time. *The Counseling Psychologist, 45*(1) 54–65
- Kahneman, D. (2011). *Thinking, Fast and Slow*. Thinking, Fast and Slow. Macmillan. ISBN 978-1-4299-6935-2
- Lambert, M.J. (2010). *Prevention of Treatment Failure: The Use of Measuring, Monitoring, and Feedback in Clinical Practice*. Washington, DC:American Psychological Association.
- Lewsi, M. (2003). *Moneyball: The Art of Winning an Unfair Game*. W.W. Norton.
- Lewis, M. (2016) *The Undoing Project: A Friendship That Changed Our Minds*. W.W. Norton.
- Minami, T., Brown, G. S., McCulloch, J., & Bolstrom, B. J. (2012). Benchmarking therapists: Furthering the benchmarking method in its application to clinical practice. *Quality and Quantity, 46*, 1699-1708. doi:10.1007/s11135-011-9548-4
- Minami, T., Davies, D. R., Tierney, S. C., Bettmann, J. E., McAward, S. M., Averill, L. A., Huebner, L. A., Weitzman, L. M., Benbrook, A. R., Serlin, R. C., & Wampold, B. E. (2009). Preliminary evidence on the effectiveness of psychological treatments delivered at a university counseling center. *Journal of Counseling Psychology, 56*, 309-320. doi:10.1037/a0015398

- Minami, T., Serlin, R. C., Wampold, B. E., Kircher, J. C., & Brown, G. S. (2008). Using clinical trials to benchmark effects produced in clinical practice. *Quality and Quantity*, 42, 513-525. doi:10.1007/s11135-006-9057-z
- Minami, T., Wampold, B. E., Serlin, R. C., Kircher, J. C., & Brown, G. S. (2007). Benchmarks for psychotherapy efficacy in adult major depression. *Journal of Consulting and Clinical Psychology*, 75, 232-243. doi:10.1037/0022-006X.75.2.232
- 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. doi:10.1037/0022-006X.76.1.116
- Wampold, B. E., & Brown, G. S. (J.) (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
- Wampold, B. E., Imel, Z. E. (2015). *The great psychotherapy debate: The evidence for what makes psychotherapy work* (2nd ed.). New York, NY: Routledge.
- Wampold BE & Imel Z (2017) <http://societyforpsychotherapy.org/what-do-we-know-about-psychotherapy-and-what-is-there-left-to-debate/>