



Faculty Publications

2015-9

Removing very low-performing therapists: A simulation of performance-based retention in psychotherapy

Scott A. Baldwin
Brigham Young University - Provo

Zac E. Imel

Elisa Shang

David C. Atkins

Follow this and additional works at: <https://scholarsarchive.byu.edu/facpub>



Part of the [Psychology Commons](#)

BYU ScholarsArchive Citation

Baldwin, Scott A.; Imel, Zac E.; Shang, Elisa; and Atkins, David C., "Removing very low-performing therapists: A simulation of performance-based retention in psychotherapy" (2015). *Faculty Publications*. 6069.

<https://scholarsarchive.byu.edu/facpub/6069>

This Peer-Reviewed Article is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in Faculty Publications by an authorized administrator of BYU ScholarsArchive. For more information, please contact ellen_amatangelo@byu.edu.



Published in final edited form as:

Psychotherapy (Chic). 2015 September ; 52(3): 329–336. doi:10.1037/pst0000023.

Removing very low-performing therapists: A simulation of performance-based retention in psychotherapy

Zac E. Imel^a, Elisa Sheng^b, Scott A. Baldwin^c, and David C. Atkins^b

^aUniversity of Utah

^bUniversity of Washington

^cBrigham Young University

Abstract

Therapists can impact the likelihood a given patient will benefit from psychotherapy. However, therapists are rarely held accountable for their patients' outcomes. As a result, low performing providers likely continue to practice alongside providers with high response rates. In the current study, we conducted a Monte Carlo simulation to illustrate a thought experiment—what happens to patient outcomes if therapists with the worst outcomes were removed from practice? We drew initial samples of 50 therapists from three simulated populations of 1,000 therapists with a mean patient response rate of 50% and different effect sizes for therapist variability in outcomes. We simulated 30 patient outcomes for each therapist, with outcome defined as response to treatment versus no response. We removed therapists with response rates in the bottom 5% and replaced them with a random sample of therapists from the population. Over 10 years, the difference in responses between the lowest and highest performing therapists was substantial (between 697 and 997 additional responses to treatment). After repeatedly removing the lowest performing providers 40 times (simulating a 10 year time span), response rates increased substantially. The cumulative number of patient responses (i.e., summing the total number of responses across 10 years) increased by 4266, 6404, and 9307 when therapists accounted for 5%, 10%, or 20% of the patient outcome variance, respectively. These findings indicate that performance-based retention of therapists could improve the quality of psychotherapy in health systems by improving the average response rate and decreasing the probability that a patient will be treated by a therapist who has little chance of helping.

Keywords

Therapist Effects; Simulation; Quality Improvement; Outcomes Monitoring

“Nothing in the science of prediction and selection beats observing actual performance in an equivalent role”

(Capelli, 2009, p. 184).

Psychotherapy is an effective treatment for many patients; however, anywhere from 40% to 70% of patients do not respond depending on the problem being treated and how success is calculated (e.g., completers vs. 'intent to treat'; Westen & Bradley, 2005; Westen & Morrison, 2001). Additionally, therapists vary in their effectiveness in both clinical trials with intensive training and supervision as well as in less controlled community-based settings (Baldwin & Imel, 2013). Thus, although the average response rate for major depressive disorder is approximately 50% (Westen & Morrison, 2001), a given patient's chance of responding to treatment can vary widely depending on the particular therapist they see (i.e., some therapists will surpass the average rate of 50%, but some may do much worse). Patient outcomes must be assessed to estimate provider-specific response rates, but this is not common in routine care. Consequently, key stakeholders (e.g., patients and mental health system administrators) cannot make informed choices about treatment providers based on the providers' performance, and relatively ineffective providers may continue to practice, with no objective method to identify them relative to those who are much more effective. Furthermore, an ineffective provider is not likely to be aware he or she has poor outcomes (Walfish, McAlister, O'Donnell, & Lambert, 2012). What if patient outcomes and therapist effectiveness were systematically tracked? What if health systems based decisions about therapist retention on this information? Would this course of action enhance patient outcomes? These are the questions that motivated the current simulation study.

Treatments Involve Providers and Providers Differ

The evaluation and dissemination of specific treatments has been a central focus of the Evidence Based Practice (EBP) movement in psychology (American Psychological Association, 2006; Westen, Novotny, & Thompson-Brenner, 2004). The premise of this approach is that using more effective treatments – identified by research – will lead to better patient outcomes. Accordingly, dissemination in this tradition has focused on ensuring specific treatments are available to patients and has largely ignored how therapist variability influences all quality improvement efforts in psychotherapy.

Despite the lack of research focus on differences in therapist outcomes, therapist effectiveness varies even in well-controlled clinical trials that provide ongoing symptom monitoring, training, supervision and feedback (Baldwin et al., 2011). Evidence from several dissemination studies suggests that providers often do not maintain adherence to new treatments (Walters, Matson, Baer, & Ziedonis, 2005), and variability in patient outcomes across providers persists even when being supervised in the use of evidence-based treatments (Laska, Smith, Wislocki, Minami, & Wampold, 2013). A recent meta-analysis indicated that therapists accounted for 7% of the variance in clinical outcomes in naturalistic settings where supervision and training are less available, and 3% in clinical trials wherein supervision and training are typically provided as a part of the study design. Notably, the estimates of therapist differences in patient outcomes vary widely across studies (0 to 55% of the variance in outcomes; Baldwin & Imel, 2013). This evidence suggests that even with supervision and training there are likely to be outcome differences between therapists, and these differences are magnified in naturalistic settings where consistent feedback, supervision, and training can be absent. Thus, to protect patients from harm and improve quality of care, it remains important for therapists and health care administrators to

understand how using outcome data to make personnel decisions can impact patient outcomes.

Differential Therapist Effectiveness has a Real Impact on Patients

As noted above, the importance of provider effects in psychotherapy is typically estimated as the percentage of variance in clinical outcomes that is attributable to providers. In light of effects suggesting that providers account for 3-7% of the variance in outcomes in research studies, it would be tempting to conclude that therapist selection is not a major determinant of patient outcomes. Indeed, what about the other 93-97% of the variance in outcomes?¹ If the outcomes of two therapists within the middle two quartiles of effectiveness are compared, the difference in outcomes would be minor.

Although the absolute amount of patient outcome variance attributable to therapists is small, such effects can have a large impact. For example, in a classic paper, Abelson (1985) demonstrated that counter to the belief of most baseball fans, the individual hitter only explains about 1/3 of 1% (i.e., .003 proportion) of the variance in getting a hit in a given at bat. However, when viewed cumulatively (over say 1,000 at bats), the difference in number of hits between a below average and above average hitter can become sizeable (hits are almost 50% more frequent in the above average hitter). Abelson's paradox of small explained variance and large cumulative impact has parallels in the evaluation of therapist outcomes. For example, a large study of therapist differences in patient outcomes in a managed care system found that therapists accounted for 5% of the variance in outcomes. However, the average effect size of patients who saw therapists in the top quartile of outcomes was more than twice as large as the therapists in the bottom quartile (Wampold & Brown, 2005; see also Brown, Lambert, Jones, & Minami, 2005; Okiishi, Lambert, Nielsen, & Ogles, 2003). Thus, small differences across therapists (e.g., 5%) can be important, and the presence of large differences (e.g., 20%) can have a dramatic effect on the likelihood that a given patient will benefit from treatment.

How are Therapists Currently Evaluated?

Despite evidence that therapist differences are important, even the best providers are likely to have periodic bad outcomes in their caseloads (Baldwin & Imel, 2013). As the number of patients per therapist in primary research studies is typically low, estimates of a given therapist's effectiveness may be particularly error prone. Appropriate judgments about providers will require ongoing measurement and large samples of patients for each provider. However, at present, evaluation and retention of therapists is typically based on limited, direct observation of their work by a supervising therapist prior to licensure. After licensure, evaluation may be restricted to utilization metrics (e.g., do the providers' patients come back, does the provider fill a sufficient number of clinical slots), an absence of client complaints or ethical violations, and informal professional reputation (Tracey, Wampold, Lichtenberg, & Goodyear, 2014). Quantitative measures of patient satisfaction are sometimes used to

¹Note, however, that receiving treatment (vs. a wait list) only accounts for 20% of the variance in outcomes. Thus maybe 3-7% is not so bad (see Baldwin & Imel, 2013, p. 277).

evaluate therapists and these may even include a brief measure of patient outcome (e.g., one question asking the patient if they felt improved). However, satisfaction measures are typically only given to a subset of patients and are notoriously skewed (e.g., most patients report very high satisfaction) and not uniformly related to treatment outcomes (Imel, Hubbard, Rutter, & Simon, 2012; Simon, 2009; Simon, Imel, & Ludman, 2012). As an exception to the rule, Lambert and colleagues have designed a feedback system where outcomes are regularly tracked and each patients progress is compared to a predicted treatment response curve. Providers are notified when a patient is not improving as predicted. This feedback produces better patient outcomes relative to no feedback and reduces patient deterioration (Lambert, 2010).

Current Study

While current health care reform efforts emphasize that providers and organizations should be accountable for patient outcomes (Fisher & Shortell, 2010; Shortell, Casalino, & Fisher, 2010), we are not aware of any attempts to retain or remove providers from a specific clinical role (e.g., being part of an HMO panel) based on the prior clinical outcomes of their patients. This is understandable, as the data necessary for evaluating therapist performance with their clients is often unavailable, and even when available, within-therapist variability in patient outcomes is high (i.e., both the best and worst therapists have good and bad outcomes). Also, differences in the complexity of patients seen by different therapists are not likely to be random. Thus, the best therapists may have attenuated response rates due to working with very difficult patients. Understandably, many may question whether making therapist retention decision based on poor patient response would ever be feasible given the above concerns and accompanying political challenges that would be encountered.

Nevertheless, our goal in this study was to illustrate a thought experiment—what might happen if a health system implemented a quality improvement strategy that held therapists accountable for consistently poor clinical outcomes with their patients? To do so, we used a simulation study to examine the impact of replacing therapists with poor patient outcomes (the lowest performing 5%) with new therapists randomly selected from a pool of potential therapists. Simulations explored three different effect sizes for therapist influence on patient outcomes (i.e., small=5%, medium=10%, large=20%). First we examined the magnitude of therapist differences in outcomes over a 10 year span. Second, we examined the impact of removing poor performing therapists on average patient response rate and the cumulative number of patient responses. We also considered how removing poor performing therapists affected the overall distribution of therapist effectiveness within our simulated healthcare system.

Method

The conceptual framework for the current simulation study is as follows. First, collect patient outcomes for each therapist in a simulated healthcare system for a set period of time. Second, at regular intervals, evaluate therapist effectiveness by examining the proportion of patients who responded to treatment for each therapist. Third, the lowest performing 5% of

therapists are then removed (i.e., fire, drop from payer panel, etc.) and are replaced (randomly) from a pool of potential therapists.

We used Monte Carlo methods to simulate patient response rates for a set of 50 therapists. Our goal in designing the simulation was to approximate the theoretical caseload of a full-time therapist. We assumed each therapist works 2,000 hours per year, of which 1,000 hours would be direct service. With 2 weeks of vacation, we calculated approximately 20 direct clinical hours (sessions) per week (i.e., 1,000 hours/50 weeks). Assuming the average length of a treatment episode is 8 sessions, a therapist would complete 2.5 cases in a week, approximately 30 cases in a quarter, and 125 cases in a year. Based on these calculations, we generated 30 cases for each therapist before making a retention decision (i.e., removing the lowest performing therapists), essentially once a quarter. As a consequence, four retention iterations would require 1 year, 20 in 5 years, and 40 iterations in 10 years.

The effect size for variability in therapist effectiveness is the intra-class correlation coefficient (ICC), which is a measure of how correlated patient outcomes are within a therapist. An ICC of zero indicates no association in the outcomes of patients within a therapist (essentially, the therapist a patient was assigned to had no impact on their likelihood of response), and an ICC of one would mean that all patient outcomes are identical within a therapist. As ICC increases, a patient's outcome becomes more associated with the particular therapist (Raudenbush & Bryk, 2002). To initiate the simulation, we generated a population of 1,000 therapists for three different estimates of therapist variability in outcomes (ICC = .05, .10, .20). These ICC's broadly correspond to small, medium, and large effects of therapist on patient outcomes. The average response rate across all patients was assumed to be 50% (% clinically significant improvement; see Westen & Morrison, 2001). Each ICC value corresponds to a different distribution of "true" therapist response rates (i.e., the average patient outcome for each therapist over infinite patients). Taking 1,000 draws from this distribution represents the population of all possible therapists that could ever be selected by a given clinic. Although the average response rate is consistent, the number of extremely high and low response rates increases as the ICC goes up (i.e., when the ICC is lower, comparatively more providers have patient response rates close to the average response rate). However, even when the effect is small (e.g., ICC = .05) some therapists have response rates substantially below the mean of 50% (e.g., a low performing provider would likely have a response rate of less than 20%).

The overall simulation design is shown in Figure 1. For each ICC value, we generated our simulated clinic by randomly selecting 50 therapists from the therapist population (which reduced the total population of therapists by 50). For each iteration (roughly ¼ of a year), we generated 30 patient outcomes for each therapist based on their population or true response rate described above. For example, if a therapist had a true underlying response rate of 60%, we made 30 random draws for that therapist from a binomial distribution (Response vs. Non-Response) with a 60% chance of response. Due to sampling error, the therapist's observed response rate will have some deviation from the therapist's true response rate. We then calculated the ICC, number of responses, and overall response rate based on the patient outcomes (in total, each iteration involved generating outcomes for 1,500 patients). Next, we dropped the therapists with response rates in the lowest 5% of therapist response rates—

approximately four therapists on average across iterations. These therapists were replaced with a random draw of therapists from the original population, which further reduced the size of the population. The retention decision process was repeated 40 times (or about 10 years), each decision represented a single iteration in the simulation. Finally, we performed 10,000 Monte Carlo simulations of the above process.

All calculations were done in the statistical software R (R Core Team, Team, 2011), and multilevel models were fit using lme4 package (Bates, Maechler, Bolker, & Walker, 2014) to estimate ICCs. See the online appendix for additional technical detail on the simulation, and the R code used to generate the results.

Results

As an initial illustration of the simulated differences in patient outcomes between high and low performing providers, we calculated the average number of accumulated patient responses the top and bottom two providers for each amount of therapist differences over 1, 5, and 10 years (4, 20, and 40 iterations) without removing low performing providers. As highlighted in Figure 2, the difference in the number of patient responses between the lowest and highest ranked provider becomes quite substantial over time. Notably, differences were large no matter the true amount of therapist differences. Even with the most conservative estimate of therapist influence on patient outcome (5%), after 1 year, the best provider has an average of 94 responses (out of 120 possible) and the worst has 25. After 10 years, the best provider had almost 700 more responses. When therapist differences were large (i.e., ICC = .20), after 10 years, the number of patient responses obtained from the best therapist was almost 10 times larger than the lowest performing therapist (101 vs. 1098). Thus, although the average response rate in our simulated data is 50%, if a patient happens to get a low performing provider, they have a dramatically reduced chance of benefiting from treatment.

After using the selective retention strategy, the number of responses and response rate increased substantially. Over 10 years, the cumulative number of patient responses (i.e., summing the total number of responses across all 40 iterations) was substantially larger when low performing therapists were removed with approximately 4266, 6404, and 9307 more responses when ICC = .05, .10, .20 respectively. This increase in the number of responses corresponded to cumulative response rates of 57, 61, and 66% across therapist difference values (as compared to 50% when no therapists were removed; see Table 1). All differences between the first and last iteration were significant; $p < .001$ from a chi-squared test).

There were also substantial effects of the retention strategy on the consistency of therapist outcomes. After 10 years, therapist variability reduced by approximately half. For an initial ICC of .20, the final ICC was .10. For an initial ICC of .10, the final value was .10, and for an initial value of .05, the final value was .03. However, much of the reduction in therapist variability happened in the first 10 iterations (2.5 years) and stopped reducing entirely by the 14th, 19th, and 30th iteration of selective retention for each ICC value (ICC = .05, .10, .20).

Discussion

Therapists vary in their effectiveness. While therapist differences in clinical trials where the number of patients and therapists are low may appear to primarily be a statistical nuisance that complicates the analysis of treatment effects, these differences can accumulate over time and begin to have relevance for public health. In particular, the current simulation study showed that even when therapist influence on patient outcomes was small (i.e., ICC = .05), a patient's response to treatment can deviate dramatically from the average response rate across therapists. Differences between therapists may not appear substantial as a percentage of variance in outcome, but over time, the best therapists helped hundreds more patients than the worst.

We have also provided initial evidence from a simulation study that removing the therapists with the lowest patient response rates can have an important impact on the likelihood of a patient responding to treatment. Although the impact of this quality improvement intervention depends partially on the pre-existing amount of variability between therapists, there was a notable impact on patient response rate across all initial ICC values.

Removing poorly performing therapists is perhaps best viewed as a long-term quality improvement strategy where the benefit to patients becomes clearer over time. Over 40 iterations (approximately 10 years and 60,000 patients), response rates improved from 50 to 57, 61, and 66% respectively depending on the initial ICC. At least 4,200 (and up to 9,300) more patients benefited from treatment. As with the baseball example cited in the introduction, small differences may have large impacts when observed in a large healthcare system.

Limitations

The findings of this simulation study are limited both by the constraints of the chosen parameter values and practical difficulties related to implementation. In this initial simulation, we fixed all parameter values except the initial ICC (e.g., number of iterations, average response rate, % of therapists to replace, number of patients per therapist). However, future work might explore the impact of varying any combination of specific values. For example, it is possible that average response rates vary along with therapist differences in outcomes such that a particular clinic might have a lower average response rate relative to the population of therapists. In this case, bringing in new therapists might have a more dramatic impact than when both the clinic and therapist population perform similarly on average. Also, we fixed the number of patients per therapist to 30 meaning that we evaluated therapists based on their last 30 patients, ignoring how they may have done with prior patients. An alternative procedure would be to make decisions about retention based on cumulative response rate across all patients, rather than a specific cut point. However, it is possible that 30 patients is not a sufficient number to make a reliable judgment about a provider. Thus, it will also be important to determine the specific number of patients required to prevent the accidental removal of a high performing therapist who may have had a short-term run of especially challenging patients.

Many might question whether the frequency of removing therapists (once a quarter) would be feasible in the real world. Our caseload estimates are based on a number of assumptions that may or may not hold for specific clinical situations. In addition, it is not clear that replacing therapists four times a year would be administratively feasible even if our assumptions hold. However, all the numbers in this simulation are subject to modification. For example, one could examine the impact of removing therapists after 100 cases (rather than 30). This simulation represents an initial attempt to quantify several of the factors that should be considered and tested before a strategy like this one is implemented.

The notion that therapists should be held accountable for treatment outcomes has important overlap with recent work on the impact of “pay for performance” models of health care. In these models, providers, systems, hospitals may be incentivized for achieving specific benchmarks (e.g., bonuses for keeping patients blood pressure within certain limits; Serumaga et al. 2015), using certain evidence based practices, or punished for other outcomes (e.g., a pay or refusing re-imbursement for hospital acquired conditions; Lee et al. 2012). However, the effects of these quality improvement interventions are far from clear (Bardach et al. 2013; Flodgren et al., 2011; Jha et al., 2012; Scott et al., 2011), and there are broader concerns about the impact of these programs on the practice of medicine (see Carroll, 2014 and Hartzband & Groopman, 2014 for recent opinion pieces in the popular press). For example, there could be unintended consequences such as external incentives decreasing provider intrinsic motivation to provide quality care (e.g., Deci, Koestner, & Ryan, 1999), or providers simply beginning to treat towards the financial bonus rather than the unique needs of a given patient. The most direct correlate of many of these pay for performance models in psychotherapy might be incentivizing therapists to only offer specific forms of evidenced based treatments (presumably some form of CBT), or failing to reimburse therapists who choose not to use specific treatments. Such an approach would indeed be concerning given current evidence that many providers currently achieve outcomes that meet or exceed established benchmarks from clinical trials (Minami et al. 2008). The simulated strategy evaluated in the current study represents an alternative to this approach, offering providers flexibility in how they achieve outcomes, but holding them to some minimal standard.

Nevertheless, there are clear potential limitations of therapist outcome accountability. One unintended consequence of an outcome based retention policy might be limiting access to psychotherapy for the most severe or complicated patients (e.g., personality disorders, etc.). That is, therapists might not want to work with patients who will harm their overall response rate. Here it will be important to conduct adequate statistical adjustment of therapist level metrics based on different patient characteristics – both those that might be captured directly in symptom checklists and diagnostic interviews as well as other factors that may influence the likelihood of improvement (e.g., prior treatment history, socio-economic status). To project against this possibility, systems might offer “bonuses” to providers who elect to work with difficult patients or more strongly weight getting a treatment response with a more difficult patient in the calculation of a provider metric (e.g., the rapid improvement of a patient with an adjustment disorder might be weighted so as to have little influence on a providers response rate, while a patient with a history of borderline personality disorder and bipolar disorder who has stayed out of the hospital for two years might be weighted so as to have a strong impact on that providers evaluation). Finally, there are potential broad

implications for provider morale. For example, a system that is perceived as punitive or shaming may decrease the overall attractiveness of the profession (e.g., see the publication of teach effectiveness ratings in the Los Angeles Times; <http://projects.latimes.com/value-added/>), increase burnout, and/or have an overall negative impact on the quality and availability of care. Despite attempts to protect against negative outcomes, any attempt at altering the incentives in a system of care will have unintended consequences that cannot be fully anticipated. Perhaps the only solution is ongoing research, testing, and experimentation first in small and then in larger scale settings.

As noted directly above, providers can be evaluated on more than just improvement on symptom measures in one domain. Patients may value different areas of functioning more than others (e.g., acute vs. chronic symptoms, interpersonal, social, occupational, personality) and systems might value particular outcomes (e.g., prevention of hospitalization, decrease in smoking). In addition, there is evidence that therapist performance can vary across different diagnostic domains, raising the question if it is reasonable to model therapist performance on the basis of one outcome in isolation (Kraus et al., 2011). This suggests that there might not simply be a “best” or “worst” therapist, but a range of high and low performing therapists across a range of outcomes. As a result, it might only make sense to remove therapists who are consistently poor across a majority of primary outcomes.

More broadly, we imagine that our basic premise might be troublesome for some therapists and healthcare professionals. Preventing poorly performing therapists from seeing patients might seem draconian. Indeed, implementing this strategy in the real world would introduce numerous and important concerns about the quality of outcome data upon which decisions are based, provider morale, and the ethics of basing retention decisions on patient outcomes. In addition, in some clinics it may be that the lowest 5% of therapists are performing quite well with response rates at or approaching 50%. In such a clinic it may be important to tie retention decisions not only to a therapist's ranking on patient outcomes but also the attainment of specific outcome benchmarks. In addition, removing therapists from practice might be a last and final option after other potential interventions are considered. These may include simply providing feedback to therapists on the outcome of their patients directly (a strategy known to improve outcomes; e.g., Lambert, 2010), providing remedial training, or inspecting case-mix differences in the therapist's caseload that may be affecting their outcomes. However, after therapists have been offered opportunities for remediation and alternative explanations for outcome differences have been explored, it is ethically questionable to continue to expose patients to therapists who provide them little opportunity for improvement.

A thoughtful implementation of this policy would be complex and would necessarily take place within a broader context of quality monitoring and improvement. Specifically, actual evaluation decisions may not be able to be done quarterly. For example, such regular decisions might damage moral. The practicality of these decisions is likely to be far more complex than our simulation described (probation period, appeals process, etc.). In addition, we randomly selected new therapists, which will likely not be possible or even advisable in a real system that evaluates providers based on experience, recommendations, prior training,

etc. In the same way that student test scores are not the sole evaluation metric for teachers, patient outcomes should not be the sole evaluation metric for therapists. However, we also have an ethical obligation to patients, and if there is clear and consistent evidence of poor performance by a therapist, then at some point they should be removed from practice.

Conclusions

Despite important advances in psychotherapy science and the popularization of evidence-based practice in mental health, for patients, choosing a therapist remains a gamble. Although there is data from clinical trials and naturalistic settings on the mean response rate for specific disorders, a given patient may receive treatment from a provider with a response rate 3 to 4 times above or below the mean. At present, there is no way to determine which of these therapists a patient has encountered. Therapists are not hiding this information—the data simply do not exist. When the data necessary to compute provider-level response rates is collected in select clinics, systems, or payer networks, this information is almost never shared with patients or used as an evaluation metric for providers such that their retention, promotion, or pay is linked to patient recovery.

Given the state of the science in psychotherapy on therapist variability in patient response, modern medical records technology, and ongoing pressures for outcome based accountability in healthcare, this state of affairs is unacceptable and unsustainable. The integrity of psychotherapy as a professional activity, and the well being of patients who trust us with their care requires that we begin the difficult work of determining how to hold therapists accountable for their performance with patients. This must be done in a way that protects therapists from wrongful action, but that also protects patients from harm (or the illusory expectation of benefit). Some of this work will involve additional and more complex simulations studies such that we can explore the impact of quality improvement strategies when all other things are held constant. It will also include messy, controversial implementation efforts. Despite these difficulties, we have shown that the potential benefit to patients is worth the effort.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Author Note. Funding for the preparation of this manuscript was provided by National Institute of Drug Abuse (NIDA) of the National Institutes of Health under award number R34/DA034860 and the National Institute on Alcohol Abuse and Alcoholism (NIAAA) under award number R01/AA018673. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

- Abelson RP. A variance explanation paradox: When a little is a lot. *Psychological Bulletin*. 1985; 97(1):129.
- American Psychological Association. Evidence-based practice in psychology. *American Psychologist*. 2006; 61(4):271–285. [PubMed: 16719673]

- Baldwin, SA.; Imel, ZE. Therapist Effects: Findings and Methods. In: Lambert, MJ., editor. *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change*. 5th. New Jersey: Wiley; 2013. p. 258-297.
- Baldwin SA, Murray DM, Shadish WR, Pals SL, Holland JM, Abramowitz JS, et al. Intraclass Correlation Associated with Therapists: Estimates and Applications in Planning Psychotherapy Research. *Cognitive Behaviour Therapy*. 2011; 40(1):15–33. [PubMed: 21337212]
- Bardach NS, Wang JJ, De Leon SF, Shih SC, Boscardin J, Goldman LE, Dudley RA. Effect of Pay-for-Performance Incentives on Quality of Care in Small Practices With Electronic Health Records: A Randomized Trial. *Journal of the American Medical Association*. 2013; 310:1051–1059. [PubMed: 24026600]
- Bates, D.; Maechler, M.; Bolker, B.; Walker, S. lme4: Linear mixed-effects models using Eigen and S4. R package version 1.0-6. 2014. <http://CRAN.R-project.org/package=lme4>
- Brown GS, Lambert MJ, Jones ER, Minami T. Identifying highly effective psychotherapists in a managed care environment. *American Journal of Managed Care*. 2005; 11(8):513–520. [PubMed: 16095437]
- Capelli, P. What's Old is New Again: Managerial “Talent” in an Historical Context. In: Martocchio, J.; Liao, H., editors. *Research in Personnel and Human Resources Management*. Vol. 28. 2009. p. 179-218.
- Caroll, AE. (2014, July 28). The Problem With ‘Pay for Performance’ in Medicine. *The New York Times*. 2014. Retrieved from <http://www.nytimes.com>
- Deci EL, Koestner R, Ryan RM. A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*. 1999; 125:627–668. [PubMed: 10589297]
- Fisher ES, Shortell SM. Accountable Care Organizations: Accountable for What, to Whom, and How. *Journal of the American Medical Association*. 2010; 304(15):1715–1716. DOI: 10.1001/jama.2010.1513 [PubMed: 20959584]
- Flodgren G, Eccles MP, Shepperd S, Scott A, Parmelli E, Beyer FR. An overview of reviews evaluating the effectiveness of financial incentives in changing healthcare professional behaviours and patient outcomes. *Cochrane Database of Systematic Reviews* 2011. 2011; (7) Art. No.: CD009255. doi: 10.1002/14651858.CD009255
- Hartzband, P.; Groopman, J. How Medical Care Is Being Corrupted. *The New York Times*. 2014 Nov 18. Retrieved from <http://www.nytimes.com>
- Imel ZE, Hubbard R, Rutter C, Simon GE. Patient rated alliance as a measure of therapist performance in two clinical settings. *Journal of Consulting and Clinical Psychology*. 2012; 81(1):154–165. [PubMed: 23231576]
- Jha AK, Joynt KE, Orav J, Epstein AM. The Long-Term Effect of Premier Pay for Performance on Patient Outcomes. *The New England Journal of Medicine*. 2012; 366:1606–15. [PubMed: 22455751]
- Kraus DR, Castonguay L, Boswell JF, Nordberg SS, Hayes JA. Therapist effectiveness: implications for accountability and patient care. *Psychotherapy Research*. 2011; 21(3):267–76. DOI: 10.1080/10503307.2011.563249 [PubMed: 21623550]
- Lambert, MJ. Prevention of treatment failure: The use of measuring, monitoring, and feedback in clinical practice. *American Psychological Association*; 2010.
- Laska KM, Smith TL, Wislocki AP, Minami T, Wampold BE. Uniformity of evidence-based treatments in practice? Therapist effects in the delivery of cognitive processing therapy for PTSD. *Journal of Counseling Psychology*. 2013; 60(1):31–41. [PubMed: 23356465]
- Lee GM, et al. Effect of Nonpayment for Preventable Infections in U.S. Hospitals. *The New England Journal of Medicine*. 2012; 367:1428–37. [PubMed: 23050526]
- Minami T, Wampold BE, Serlin RC, Hamilton EG, Brown GS, Kircher JC. Benchmarking the Effectiveness of Psychotherapy Treatment for Adult Depression in a Managed Care Environment: A Preliminary Study. *Journal of Consulting and Clinical Psychology*. 2008; 76:116–124. [PubMed: 18229989]
- Okiishi J, Lambert MJ, Nielsen SL, Ogles BM. Waiting for supershrink: an empirical analysis of therapist effects. *Clinical Psychology & Psychotherapy*. 2003; 10(6):361–373.

- R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2011.
- Raudenbush, SW.; Bryk, AS. Hierarchical linear models: Applications and data analysis methods. Newbury Park, CA: Sage; 2002.
- Scott A, Sivey P, Ait Ouakrim D, Willenberg L, Naccarella L, Furler J, Young D. The effect of financial incentives on the quality of health care provided by primary care physicians. The Cochrane Collaboration. 2011; (9) Art.No.: CD008451. doi: 10.1002/14651858.CD008451.pub2
- Serumaga B, Ross-Degnan D, Avery AJ, Elliott RA, Majumdar SR, Zhang F, Soumerai SB. Effect of pay for performance on the management and outcomes of hypertension in the United Kingdom: interrupted time series study. *British Medical Journal*. 2015; 342:d108. [PubMed: 21266440]
- Shortell SM, Casalino LP, Fisher ES. How The Center For Medicare And Medicaid Innovation Should Test Accountable Care Organizations. *Health Affairs*. 2010; 29:71293–1298. DOI: 10.1377/hlthaff.2010.0453
- Simon G, Rutter C, Crosier M, Scott J, Operskalski B, Ludman E. Are comparisons of consumer satisfaction with providers biased by nonresponse or case-mix differences? *Psychiatric Services*. 2009; 60(1):67–73. [PubMed: 19114573]
- Simon G, Imel ZE, Steinfield BJ. Is dropout after a first psychotherapy visit always a bad outcome? *Psychiatric Services*. 2012; 63:705–7. [PubMed: 22752034]
- Tracey TJG, Wampold BE, Lichtenberg JW, Goodyear RK. Expertise in Psychotherapy: An Elusive Goal? *American Psychologist*. 2014; doi: 10.1037/a0035099
- Walfish S, McAlister B, O'Donnell P, Lambert MJ. An investigation of self-assessment bias in mental health providers. *Psychological Reports*. 2012; 110(2):639–644. [PubMed: 22662416]
- Walters ST, Matson SA, Baer JS, Ziedonis DM. Effectiveness of workshop training for psychosocial addiction treatments: A systematic review. *Journal of Substance Abuse Treatment*. 2005; 29(4): 283–293. [PubMed: 16311181]
- Wampold BE, Brown GSJ. Estimating variability in outcomes attributable to therapists: A naturalistic study of outcomes in managed care. *Journal of Consulting and Clinical Psychology*. 2005; 73(5): 914–923. [PubMed: 16287391]
- Westen D, Bradley R. Empirically Supported Complexity Rethinking Evidence-Based Practice in Psychotherapy. *Current Directions in Psychological Science*. 2005; 14(5):266–271.
- Westen D, Morrison K. A multidimensional meta-analysis of treatments for depression, panic, and generalized anxiety disorder: An empirical examination of the status of empirically supported therapies. *Journal of Consulting and Clinical Psychology*. 2001; 69(6):875–899. [PubMed: 11777114]
- Westen D, Novotny CM, Thompson-Brenner H. The Empirical Status of Empirically Supported Psychotherapies: Assumptions, Findings, and Reporting in Controlled Clinical Trials. *Psychological Bulletin*. 2004; 130(4):631–663. [PubMed: 15250817]

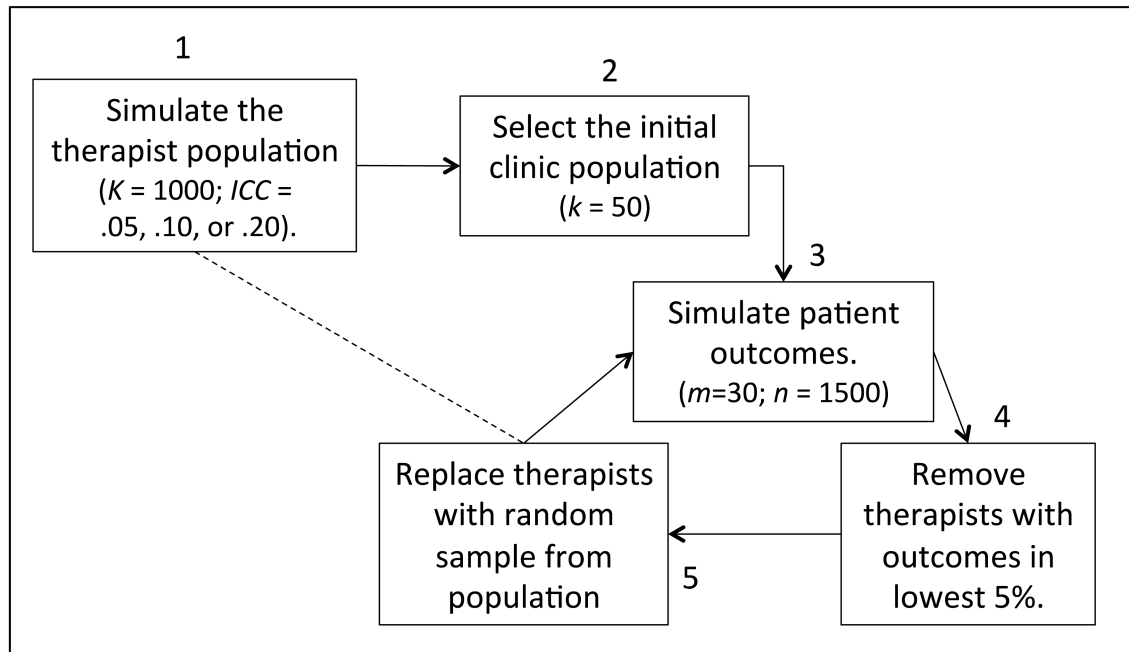


Figure 1.

Flow chart illustrating the simulation algorithm. K = number of therapists in population; ICC = intra-class correlation; k = number of therapists selected for clinic population; m = number of clients per therapist, n = overall client sample in clinic.

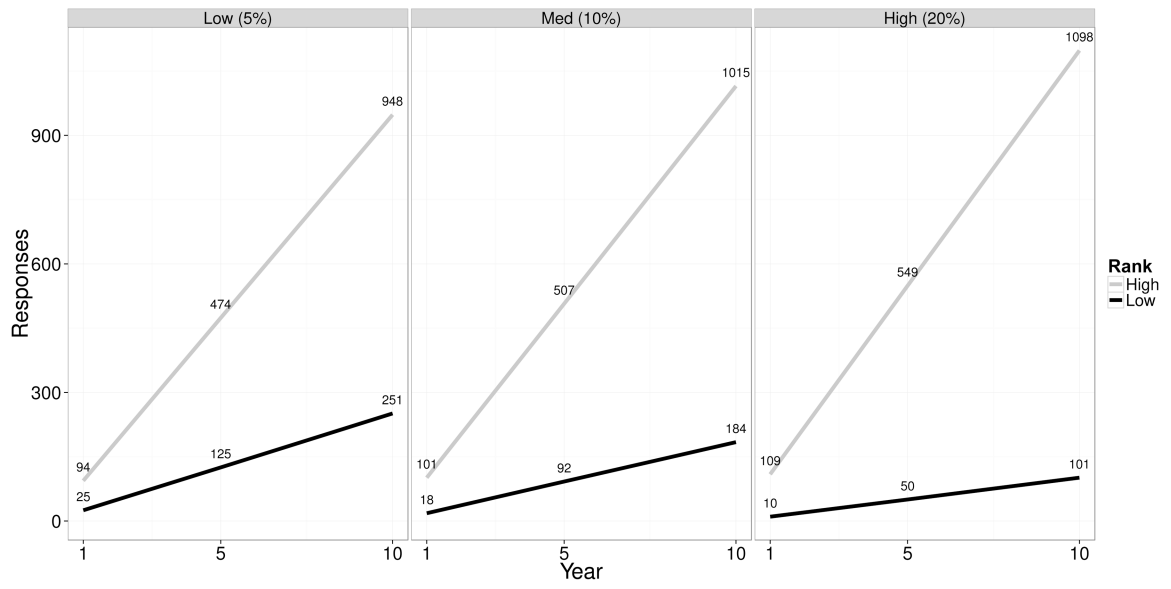


Figure 2. Bar graph depicting average accumulated patient responses from the highest and lowest rank providers over 1,5, and 10 years. Differences between high and low ranked providers are illustrated for each ICC value, small; .05, medium; .10, and large; .20.

Table 1

Change in total responses after removing lowest performing therapists.

	Total Responses (%)[*]
No Removal	30000 (50)
ICC = .05 (Small)	34266 (57)
ICC = .10 (Medium)	36404 (61)
ICC = .20 (Large)	39307 (66)

* Note. Total number of potential responses over 40 iterations (10 years) is 60000.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript