

It Is Time to Flip the Script and Leverage the Point of Care to Discover Ways of Improving Treatment Outcomes for Posttraumatic Stress Disorder: Commentary on “A Sobering Look at Treatment Effectiveness of Military-Related Posttraumatic Stress Disorder” (Levi et al., 2021)

Clinical Psychological Science
2023, Vol. 11(2) 381–387
© The Author(s) 2022
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/21677026221101068
www.psychologicalscience.org/CPS


Brett T. Litz^{1,2,3} 

¹VA Boston Healthcare System, Boston, MA, ²Department of Psychological and Brain Sciences, Boston University, and ³Department of Psychiatry, Boston University

Abstract

Levi et al. (2021) reported posttreatment outcomes of 709 Israeli veterans seeking treatment for service-related distress and concluded that their results reveal the “limited overall effectiveness” of treatments for posttraumatic stress disorder (PTSD). In this commentary, I critically appraise the categorical/diagnostic, model-based methods Levi et al. employed to examine effectiveness and use their observational study as a point of departure to describe an approach that uses benchmarked, repeated assessments of individual and aggregated outcomes (measurement-based care) to leverage these types of observational results to improve outcomes at the point of care as a way of advancing knowledge about how to best treat PTSD.

Keywords

effectiveness, measurement-based care, PTSD, treatment outcome

Received 1/27/22; Revision accepted 4/28/22

Levi et al. (2021) reported the outcomes of 709 Israeli veterans, who, on the basis of a case conference, were assigned and subsequently completed either a year of psychodynamic psychotherapy, a face-valid mixture of 20 sessions of exposure and cognitive therapy (the core change agents in cognitive-behavioral treatments [CBTs] for posttraumatic stress disorder [PTSD]), or typical regimens of pharmacological treatments, each occurring in a well-staffed, multidisciplinary, specialty-care setting. The authors aimed to provide a glimpse into the real-world impact of the treatments they provide. Levi et al. found that regardless of the length and type of treatment, 39.4% of veterans “remitted.” Levi et al. also found that all treatments were least likely to lead to “remittance” of prototypic reexperiencing symptoms.

I agree with Levi et al. (2021) that there is little evidence of the real-world effectiveness of evidence-based practices derived from efficacy trials and that it is valuable to discern outcomes using large observational cohorts. The authors concluded that their results reveal the “limited overall effectiveness” of PTSD treatments “in real-world settings.” However, with respect to effectiveness, the authors may not be aware that, at best, clinicians should expect a mean effect-size change in PTSD symptom severity if their patient matches the mean characteristics of patients treated in a given trial

Corresponding Author:

Brett T. Litz, VA Boston Healthcare System, Boston, MA
Email: litzb@bu.edu

(including trauma type and context) and the treatment is provided in the same manner (e.g., Kraemer, 2003). Consequently, this should lead to the prediction that real-world findings *will* typically diverge from efficacy trials.

The Levi et al. (2021) findings are inherently valid with respect to the care provided *in their setting*, and this commentary is devoted to this aspect. Although the last 30 years of clinical trials of nomothetic approaches to treatment have accomplished a great deal, like others (e.g., Davison, 2019; Hofmann & Hayes, 2019), I believe that the predominant categorical/diagnostic (taxon) model of PTSD (the way it is conceptualized, assessed, and treated) has limited validity and that the exclusive reliance on randomized experiments to demonstrate efficacy has constrained the discovery of approaches to improve outcomes otherwise based on real-world observational and quasi-experimental methods (Diener et al., 2022). Efficacy trials are not scalable with respect to comparative effectiveness questions, have thus far failed to address which approach best fits a particular patient and context, and have poor external validity. Consequently, researchers need to flip the script and generate new knowledge from local or clinical-system-based large-scale observational studies and, when feasible, test hypotheses using pragmatic quasi- and cluster-randomized experiments at the point of care (POC). This is especially needed to improve outcomes in clinics that treat service members and veterans because of the multifarious nature of military trauma and the complexity of cases (e.g., Litz et al., 2019). By contrast, instead of generating evidence to answer questions about “How are *we* doing?” and “What do *we* need to do to improve outcomes, and how can *we* evaluate *our* efforts?” Levi et al. offered their results to the scientific community.

Leveraging Clinical-Care Contexts to Generate New Knowledge

To address the disconnect between science and care, researchers need to leverage the POC to discover and test ways of improving outcomes in those contexts (e.g., Borckardt et al., 2008). This is the aspirational goal of a *learning health-care system* (LHS) that creates local knowledge that identifies its own needs, uses its own infrastructure, and is capable of directly implementing research results into practice (e.g., Dzau & Ginsburg, 2016). In an LHS, “science, informatics, incentives, and culture are aligned for continuous improvement and innovation, with best practices seamlessly embedded in the delivery process and new knowledge captured as an integral by-product of the delivery experience” (Smith et al., 2013, p. 2). Robust findings in a clinical context

may be generalizable to similar settings, but this is an empirical question.

In health care, there are considerable ethical, logistical (principally the need for little, if any, perturbations in workflow and seamless data capture), and professional challenges (e.g., role and motivational conflicts pertaining to the boundaries of clinical care) to leveraging a clinical-care context to generate hypotheses on the basis of patient-reported outcomes, let alone testing ideas about ways of improving outcomes. In medicine, POC research requires equipoise between two therapies or procedures that target a specific health problem (in theory, patients could be ethically assigned randomly). The problems with generating local evidence to improve quality and outcomes in PTSD clinics are amplified because the challenges are to generate and test *new*, flexible, ideographic, treatment strategies (or *newly* sequenced and matched empirically supported change agents) that can flexibly respond to a lack of change over the course of treatment. These challenges are addressable if researchers follow Hofmann and Hayes (2019) and others’ recommendation to obviate the categorical/diagnostic-model frame and use sound theory and empirically established principles of change (or flexibly dismantle existing change agents in nomothetic packages). A return to the Boulder ethos, preparing local scientists and agents for questioning and testing hypotheses and contributing to practice-based research to improve care, will help flip the script about how this new kind of knowledge is generated.

Accordingly, I argue that the categorical/diagnostic-model lens Levi et al. (2021) used to guide and conceptualize care (which is universal) greatly limits the probative value of their real-world observational evidence. In this article, I describe some solutions that the authors may wish to use to reanalyze their data and others to consider programmatically to generate practice-based evidence to improve outcomes. I also recommend a clinical-care, intellectual, and data-capture infrastructure to support the generation and testing of ways to improve outcomes in the real world (see Fig. 1). The main issues I describe are problems with using the *Diagnostic and Statistical Manual of Mental Disorders (DSM)* to determine outcome, measurement-based care (MBC) and the promise of MBC, and methods of generating individual-level indices of clinically significant change (CSC) and benchmarks for CSC to support MBC.

A Summary of Problems With the Categorical/Diagnostic-Model-Based Outcome Assessment

Levi et al. (2021) primarily operationalized effectiveness as counts of the number of cases that no longer met

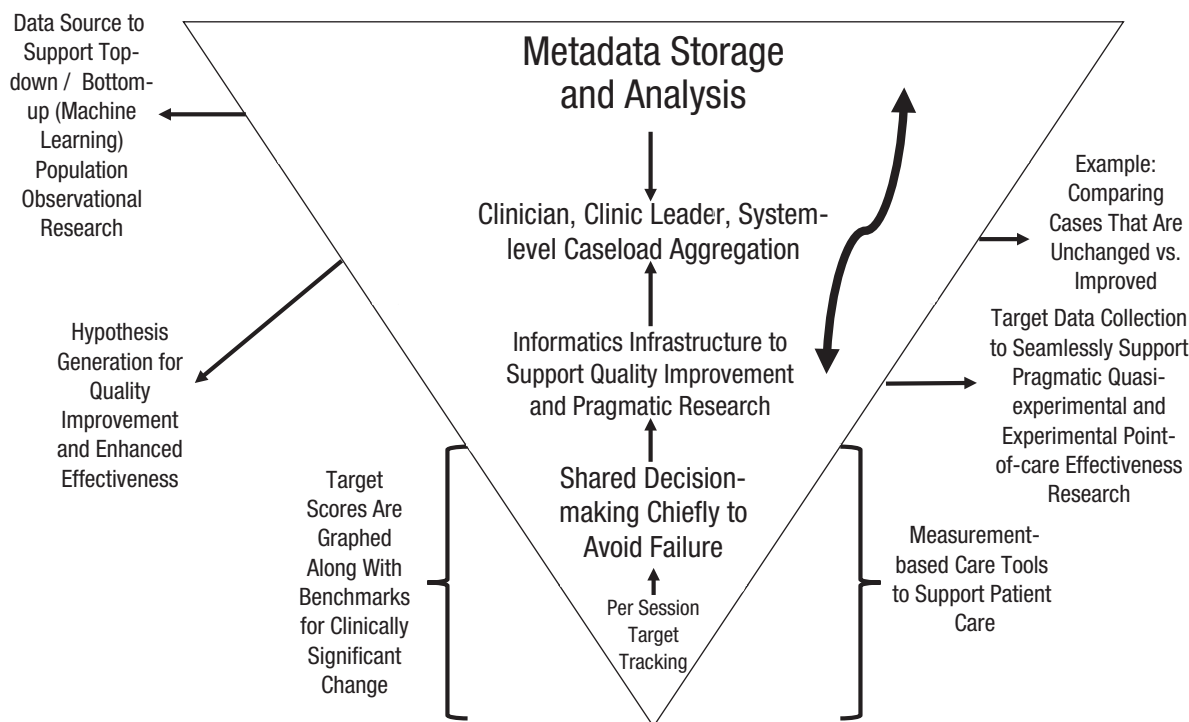


Fig. 1. The form, structure, and function of measurement-based care.

the diagnostic criteria for PTSD, the diagnostic criteria for PTSD subclusters, or an interview-based threshold rule for the frequency and intensity of individual symptoms, all indexed to the posttreatment endpoint cross-section. Second, the authors examined mean change from baseline to the posttreatment interval in PTSD symptom-severity scores and subcluster scores. The validity of the authors' primary indices of putative effectiveness depends on the validity of PTSD and PTSD subclusters as taxons, which has not been confirmed empirically (Ruscio et al., 2002) and is conceptually and empirically problematic given that there are hundreds of thousands of possible distinct PTSD symptom combinations (Galatzer-Levy & Bryant, 2013), the decision rule for PTSD caseness in the *DSM* is arbitrary, cases that just miss meeting diagnostic criteria (so-called subsyndromal cases) have similar PTSD symptom severity as cases with similar associations with suicidal thoughts and functional impairments (e.g., Jakupcak et al., 2007), and at the end of treatment, a person that is no longer diagnoseable can have a higher symptom burden than someone who still has "PTSD" and vice versa. In addition, indices of effectiveness that rely on mean effect sizes, although invaluable for meta-analyses, are problematic in specific studies because there is no standard for determining how large an effect size should be to inform practice.

Measurement-Based Care

Levi et al. (2021) administered a PTSD scale at baseline and at the end of treatment as a means of determining effectiveness. This scheme fails to inform care decisions about whether the treatment is working and to shift approach accordingly. The questions I would like to address are how clinics like Levi et al.'s can improve their measurement scheme to improve patient care and how the local metadata collected over time can be leveraged to generate and test ideas to improve the quality of care and outcomes locally. There can be no practice-based advances in knowledge without a system of care that embraces the value of outcome assessment and the need to track outcomes to facilitate a collaborative discussion with patients about the treatment plan and treatment progress (i.e., shared decision-making [SDM]) to maximize the potential for good clinical outcomes. This requires repeated outcome assessments to track change to avoid treatment failure and dropout because of non-CSC or deterioration, otherwise known as MBC.

To advance knowledge, the examination and evaluation of observational outcome collected in systems (including top-down and machine-learning analyses; see Cristea et al., 2021) requires complete, high-quality data; an informatics solution to make data entry efficient and

functionally useful; and a warehousing system to store and retrieve metadata. It is impossible to achieve a high degree of compliance with repeated assessments of an outcome without clinician and patient buy-in, and this can be accomplished only if assessments are functional for both. Yet in practice, the frequency of repeated assessments is very low. I argue that without benchmarks, test scores may not be actionable and clinicians may not be trained to use evidence to guide treatment planning. Lambert and others have shown that the functional value of repeated assessments of a target are diminished without benchmarks for clinically significant change that promote vigilance about change (Lambert et al., 2008; Youngstrom et al., 2017). Low rates of MBC are unfortunate because this greatly diminishes the validity of metadata and because repeated measurement has been shown to change behavior (Poston & Hanson, 2010), shorten treatment, reduce dropouts, and most importantly, avoid failures (e.g., Lambert et al., 2008). Consequently, the ever-widening shift toward MBC coupled with schemes and procedures for leveraging aggregated data for quality improvement and outcome improvement in clinics and systems hold the promise of providing the data that can be used to improve systems of care and advance clinical science across the board (e.g., Connors et al., 2021; Lewis et al., 2019). In the ideal case, if clinicians are provided informatics tools to examine their cases with respect to benchmarked outcomes, they can be empowered to generate ideas about ways of improving care. Once hypotheses are generated, in theory, the data-collection and storage system would be in place to test those ideas.

Figure 1 depicts the form, structure, and function of MBC to support improvements in patient care, quality improvement, and research. The process starts with a data-collection process that first meets clinical objectives. Clinicians and patients need to be covigilant partners, using session-by-session outcome tracking to determine whether an approach is working. This can be achieved by conceptualizing SDM about treatment as evidence-based (i.e., MBC). In terms of clinical care, MBC is a collaborative process that involves initially deciding on the best treatment as a starting place and revising the treatment plan on the basis of patient experiences and test results over time. Paraphrasing Harding et al. (2011), clinically, MBC is defined as enhanced precision and consistency in assessment, tracking, and treatment to achieve optimal outcomes. The MBC process entails (a) collecting (administering psychometrically sound, clinically appropriate measures at regular intervals), (b) SDM, and (c) acting (using data to develop treatment plans, assess progress, and inform SDM about changes to the treatment plan over time; Resnick & Hoff, 2020). The potential value of MBC to

the provider and individual patient is unequivocal; it empowers patients to be consumers of care and forces clinicians to find ways of modifying the approach when it is not working, and repeated assessments of treatment targets can improve outcomes (Fortney et al., 2018; Janse et al., 2020). Assess–treat–reassess–refine treatment is modal in clinical practice, and MBC is necessary to answer the most important questions to providers and patients: Is the therapy working, and if not, what symptoms or problems need to be targeted differently?

The Joint Commission (2018) recently developed standards for MBC that would be useful for any clinical setting, internationally. These are that organizations (a) use sound assessments for *ongoing* progress monitoring, (b) gather and analyze the data to inform treatment goals, and (c) evaluate outcomes of care by analyzing aggregated outcome and process data. This latter feature of MBC facilitates innovations in practices to improve outcomes and the infrastructure to assess the outcome of these innovations (depicted in Fig. 1). This bridges the functionality of MBC to generate and test ideas about how best to treat individual cases and to leverage cohorts of metadata derived from MBC data to generate ideas about ways to improve outcomes and the quality of care. The Joint Commission suggested that one way to promote MBC is to create an informatics dashboard to allow providers and clinic leaders to examine aggregated cases in their settings and to explore patterns and observational predictors of change. I argue that consistent and clinic-wide MBC coupled with an informatics method of aggregating outcomes and a data-storage and -retrieval process (as shown in Fig. 1) is the vehicle to advance knowledge about how to improve outcomes for service members and veterans with posttraumatic dysfunction.

Indexing Individual-Level CSC

To index CSC, Levi et al. (2021) tallied the percentage of cases that lost their diagnosis in their cohort and described the mean effect sizes for severity scores, each of which obscure the heterogeneity of treatment effects; some patients do poorly, some do not change to a clinically significant degree, and some do very well (Kraemer, 2003; Kravitz et al., 2004), and even substantial mean effect sizes may not be meaningful clinically or functionally. By contrast, individual-level indices of CSC have inherently greater treatment-planning value and, when aggregated, provide a more meaningful way to index of effectiveness.

There has been a lively discourse about the relative validity of various ways of indexing CSC for many decades (e.g., Ogles et al., 2001). In an ideal world,

there would be brief psychometrically sound methods of *independently* assessing the degree to which patients believe they have made substantive, functionally meaningful changes in specific problem areas or consensus best methods for categorizing sufficient change in any problem area that indicates that a patient is no longer dysfunctional. In the absence of these arguably “gold-standard” methods of determining CSC, Jacobson and Truax (1991) recommended a statistical approach that has become the standard way of benchmarking outcomes in MBC because (a) the procedure and the calculations are parsimonious and readily calculated *on any test*, (b) there is a great deal of empirical precedent, (c) the indices allow peers to be compared with peers in a given clinical setting and time period (vs. indices that are derived from efficacy trials, which are not generalizable), (d) the indices are updatable on the basis of new influxes of cases over time, (e) the indices can be aggregated (e.g., the percentage of a clinician’s caseload that made CSC), (f) the indices have been shown to have criterion validity (Eisen et al., 2007; Lunnen & Ogles, 1998; Marx et al., 2022), and (g) the algorithms can be used to define treatment nonresponse and clinically significant deterioration.

In the Jacobson and Truax (1991) scheme, an algorithm defines the clinical significance of an endpoint (or any cross-sectional) score independent of a patient’s baseline score, and an algorithm defines the clinical significance of an individual patient’s change (from baseline) score. With respect to the former task, absent a statistically derived cut score that can be used to define no longer being dysfunctional, Jacobson and Truax reasoned that a posttreatment score that is 2 *SD* below the baseline mean of a valid reference group (in the Levi et al. 2021 study, this would be the baseline values in their clinic), independent of the individual patient’s baseline score, would best represent symptoms outside the dysfunctional range for the group. If a postbaseline score is 2 *SD* below the mean of clinic patients (the dysfunctional reference group), then that putative endpoint is lower than 95% of the baseline scores. The second Jacobson and Truax criterion determines the clinical significance of an individual’s postbaseline score relative to their baseline score, which involves the calculation of a reliable change index (RCI) to ensure that change reliably exceeds measurement error. The RCI is computed according to a formula that is the ratio of an individual change score to the standard error of measurement, which is the standardized error of difference between two test scores, taking into account test reliability (the local internal consistency is modally used).

Jacobson and Truax (1991) recommended applying the two algorithms to classify patients as “recovered” if

their postbaseline test score passes the 2 *SD* criterion and exceeds the RCI criterion, “improved” if the test score passes the RCI criterion but not the 2 *SD* criterion, “unchanged” if the test score fails to pass the RCI criterion (nonresponse), and “deteriorated” if scores pass the RCI criterion in the wrong direction. Clinical tracking software can be generated to plot total scores with horizontal lines that indicate benchmarked RCI thresholds for improvement and deterioration. In addition, each postbaseline cross-sectional score can be categorized as recovered, improved, unchanged, and deteriorated. This may help clinicians and patients determine the end-state CCS status if therapy were stopped at any given point in time. In addition, the Jacobson and Truax benchmark categories can be aggregated to examine effectiveness and to generate hypotheses about the provider, treatment plan, and patient characteristics associated with CSC or the lack thereof. For clinics just starting to collect MBC data, norm-referenced values for the Jacobson and Truax parameters can be used (Marx et al., 2022).

Summary and Recommendations

In many respects, the nomothetic approach to generating and testing PTSD treatments and disseminating and mandating these approaches in practice has revolutionized care in that an unprecedentedly large percentage of clinicians are trained to provide evidence-based CBT to patients that have the resources to devote to multi-session, contiguous, and demanding treatment or agree to initiate a course (see Rosen et al., 2019). The clinical decision-making logic has been simple: If a given patient has “PTSD,” following an evidence-based manual will lead to a better outcome than an ad hoc ideographic approach regardless of the complexity and exigencies of the case or the mismatch between the mean characteristics of the sample of convenience in the trial and the patient. One of the most important research agendas is to use clinics and systems of care to critically interrogate the decision-making behind the selection of non-evidence-based ad hoc approaches by codifying the face-valid strategies that address different case-based exigencies and/or that target problems ideographically, assessing the application of the approaches, and examining outcome. If MBC processes and structures are in place, over time, a clinic or system can generate observational evidence about the differential effectiveness of the set of cases that get an ideographic treatment plan versus those that initiate existing evidence-based CBT. It will also be critical to generate local evidence about the patient characteristics that are associated with the ability to commit to (and choose) existing evidence-based CBTs as is, the patient

characteristics that are associated with completion and CSC from existing treatments, and the typical dose-response (session) curve for patients who make clinically significant gains from these treatments? The latter would identify the session number after which no further change would be expected for cases that are unchanged, which would guide MBC decisions (see Litz et al., 2019). This research would also reveal hypotheses about alternative strategies for patients that do not fit the existing approaches well or who are likely to fail treatment.

The challenge I have tried to address in this commentary pertains to the need to flip the script and leverage local clinical-outcome and -process data to improve clinical outcomes and generate new knowledge about ways of improving outcomes in a clinic. If the findings are robust in a setting, the broader scientific challenge would be to disseminate the idiographic approaches and test these in other similar settings using cluster-randomized pragmatic trials. I have argued that the clinicians and clinics should generate a data-collection methodology that allows and facilitates MBC and SDM and be the drivers of new testable knowledge first and foremost about ways of improving the quality of care and care outcomes in their setting. Arguably, if clinics generate evidence about ways of moving from a nomothetic model to an empirically supported or theory-consistent change-agent model, clinicians could be armed with intervention strategies that can be flexibly chosen to target individual operationalizable, emotional, and behavioral repertoires, some of which are trauma-related and some of which may not be but are pressing problems that, if addressed, may lead to positive collateral change in trauma-related problems.

Transparency

Action Editor: Pim Cuijpers

Editor: Jennifer L. Tackett

Author Contributions

B. T. Litz is the sole author.

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

ORCID iD

Brett T. Litz  <https://orcid.org/0000-0002-0479-8887>

References

- Borckardt, J. J., Nash, M. R., Murphy, M. D., Moore, M., Shaw, D., & O'Neil, P. (2008). Clinical practice as natural laboratory for psychotherapy research: A guide to case-based time-series analysis. *American Psychologist*, *63*(2), 77–95. <https://doi.org/10.1037/0003-066X.63.2.77>
- Connors, E. H., Douglas, S., Jensen-Doss, A., Landes, S. J., Lewis, C. C., McLeod, B. D., Stanick, C., & Lyon, A. R. (2021). What gets measured gets done: How mental health agencies can leverage measurement-based care for better patient care, clinician supports, and organizational goals. *Administration and Policy in Mental Health and Mental Health Services Research*, *48*, 250–265. <https://doi.org/10.1007/s10488-020-01063-w>
- Cristea, I. A., Vecchi, T., & Cuijpers, P. (2021). Top-down and bottom-up pathways to developing psychological interventions. *JAMA Psychiatry*, *78*(6), 593–594.
- Davison, G. C. (2019). A return to functional analysis, the search for mechanisms of change, and the nomothetic-idiographic issue in psychosocial interventions. *Clinical Psychological Science*, *7*(1), 51–53. <https://doi.org/10.1177/2167702618794924>
- Diener, E., Northcott, R., Zyphur, M. J., & West, S. G. (2022). Beyond experiments. *Perspectives on Psychological Science*, *17*(4). <https://doi.org/10.1177/17456916211037670>
- Dzau, V. J., & Ginsburg, G. S. (2016). Realizing the full potential of precision medicine in health and health care. *JAMA*, *316*(16), 1659–1660. <https://doi.org/10.1001/jama.2016.14117>
- Eisen, S. V., Ranganathan, G., Seal, P., & Spiro, A. (2007). Measuring clinically meaningful change following mental health treatment. *The Journal of Behavioral Health Services & Research*, *34*(3), 272–289.
- Fortney, J. C., Unützer, J., Wrenn, G., Pyne, J. M., Smith, G. R., Schoenbaum, M., & Harbin, H. T. (2018). A tipping point for measurement-based care. *Focus*, *16*(3), 341–350. <https://doi.org/10.1176/appi.ps.201500439>
- Galatzer-Levy, I. R., & Bryant, R. A. (2013). 636,120 ways to have posttraumatic stress disorder. *Perspectives on Psychological Science*, *8*(6), 651–662. <https://doi.org/10.1177/1745691613504115>
- Harding, K. J. K., Rush, A. J., Arbuckle, M., Trivedi, M. H., & Pincus, H. A. (2011). Measurement-based care in psychiatric practice: A policy framework for implementation. *The Journal of Clinical Psychiatry*, *72*(8), 1136–1143. <https://doi.org/10.4088/JCP.10r06282>
- Hofmann, S. G., & Hayes, S. C. (2019). The future of intervention science: Process-based therapy. *Clinical Psychological Science*, *7*(1), 37–50. <https://doi.org/10.1177/2167702618772296>
- Jacobson, N. S., & Truax, P. (1992). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. In A. E. Kazdin (Ed.), *Methodological issues & strategies in clinical research* (pp. 631–648). American Psychological Association. <https://doi.org/10.1037/10109-042>
- Jakupcak, M., Conybeare, D., Phelps, L., Hunt, S., Holmes, H. A., Felker, B., Klevens, M., & McFall, M. E. (2007). Anger, hostility, and aggression among Iraq and Afghanistan War veterans reporting PTSD and subthreshold PTSD. *Journal of Traumatic Stress*, *20*(6), 945–954. <https://doi.org/10.1002/jts.20258>
- Janse, P. D., de Jong, K., Veerkamp, C., van Dijk, M. K., Hutschemaekers, G. J. M., & Verbraak, M. J. P. M. (2020). The effect of feedback-informed cognitive behavioral therapy on treatment outcome: A randomized controlled

- trial. *Journal of Consulting and Clinical Psychology*, 88(9), 818–828. <https://doi.org/10.1037/ccp0000549>
- The Joint Commission. (2018, January). *Complying with standard CTS.03.01.09 behavioral health care accreditation program*. <https://www.jointcommission.org/-/media/tjc/documents/accred-and-cert/bhc/rationale-cts030109-revised.pdf>
- Kraemer, H. C. (2003). “Rules” of evidence in assessing the efficacy and effectiveness of treatments. *Developmental Neuropsychology*, 24(2–3), 705–718. <https://doi.org/10.1080/87565641.2003.9651916>
- Kravitz, R. L., Duan, N., & Braslow, J. (2004). Evidence-based medicine, heterogeneity of treatment effects, and the trouble with averages. *The Milbank Quarterly*, 82(4), 661–687. <https://doi.org/10.1111/j.0887-378X.2004.00327>
- Lambert, M. J., Hansen, N. B., & Bauer, S. (2008). Assessing the clinical significance of outcome results. In A. M. Nezu & C. M. Nezu (Eds.), *Evidence-based outcome research: A practical guide to conducting randomized controlled trials for psychosocial interventions* (pp. 359–378). Oxford University Press.
- Levi, O., Ben Yehuda, A., Pine, D. S., & Bar-Haim, Y. (2021). A sobering look at treatment effectiveness of military-related posttraumatic stress disorder. *Clinical Psychological Science*, 10(4). <https://doi.org/10.1177/21677026211051314>
- Lewis, C. C., Boyd, M., Puspitasari, A., Navarro, E., Howard, J., Kassab, H., Hoffman, M., Scott, K., Lyon, A., Douglas, S., Simon, G., & Kroenke, K. (2019). Implementing measurement-based care in behavioral health: A review. *JAMA Psychiatry*, 76(3), 324–335. <https://doi.org/10.1001/jamapsychiatry.2018.3329>
- Litz, B. T., Berke, D. S., Kline, N. K., Grimm, K., Rusowicz-Orazem, L., Resick, P. A., Foa, E. B., Wachen, J. S., McLean, C. P., Dondanville, K. A., Borah, A. M., Roache, J. D., Young-McCaughan, S., Yarvis, J. S., Mintz, J., & Peterson, A. L. (2019). Patterns and predictors of change in trauma-focused treatments for war-related posttraumatic stress disorder. *Journal of Consulting and Clinical Psychology*, 87(11), 1019–1029. <https://doi.org/10.1037/ccp0000426>
- Lunnen, K. M., & Ogles, B. M. (1998). A multiperspective, multivariable evaluation of reliable change. *Journal of Consulting and Clinical Psychology*, 66(2), 400–410.
- Marx, B. P., Lee, D. J., Norman, S. B., Bovin, M. J., Sloan, D. M., Weathers, F. W., Keane, T. M., & Schnurr, P. P. (2022). Reliable and clinically significant change in the clinician-administered PTSD Scale for DSM-5 and PTSD Checklist for DSM-5 among male veterans. *Psychological Assessment*, 34(2), 197–203. <https://doi.org/10.1037/pas0001098>
- Ogles, B. M., Lunnen, K. M., & Bonesteel, K. (2001). Clinical significance: History, application, and current practice. *Clinical Psychology Review*, 21(3), 421–446.
- Poston, J. M., & Hanson, W. E. (2010). Meta-analysis of psychological assessment as a therapeutic intervention. *Psychological Assessment*, 22, 203–212.
- Resnick, S. G., & Hoff, R. A. (2020). Observations from the national implementation of Measurement Based Care in Mental Health in the Department of Veterans Affairs. *Psychological Services*, 17(3), 238–246. <https://doi.org/10.1037/ser0000351>
- Rosen, C. S., Bernardy, N. C., Chard, K. M., Clothier, B., Cook, J. M., Crowley, J., Eftekhari, A., Kehle-Forbes, S. M., Mohr, D. C., Noorbaloochi, S., Orazem, R. J., Ruzek, J. I., Schnurr, P. P., Smith, B. N., & Sayer, N. A. (2019). Which patients initiate cognitive processing therapy and prolonged exposure in department of veterans affairs PTSD clinics? *Journal of Anxiety Disorders*, 62, 53–60.
- Ruscio, A. M., Ruscio, J., & Keane, T. M. (2002). The latent structure of posttraumatic stress disorder: A taxometric investigation of reactions to extreme stress. *Journal of Abnormal Psychology*, 111(2), 290–301.
- Smith, M., Saunders, R., Stuckhardt, L., & McGinnis, J. M. (Eds.). (2013). *Best care at lower cost: The path to continuously learning health care in America*. National Academies Press.
- Youngstrom, E. A., Van Meter, A., Frazier, T. W., Hunsley, J., Prinstein, M. J., Ong, M. L., & Youngstrom, J. K. (2017). Evidence-based assessment as an integrative model for applying psychological science to guide the voyage of treatment. *Clinical Psychology: Science and Practice*, 24(4), 331–363.