Journal of Consulting and Clinical Psychology

Leveraging Available Metadata in VA PTSD Clinics and Generating Benchmarks for Clinically Significant Change

Natasha Benfer, Luke Rusowicz-Orazem, Elliot M. Fielstein, Benjamin C. Darnell, Sheila B. Frankfurt, Joseph Mignogna, and Brett T. Litz

Online First Publication, December 15, 2022. https://dx.doi.org/10.1037/ccp0000779

AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Consulting and Clinical Psychology

ISSN: 0022-006X

https://doi.org/10.1037/ccp0000779

Leveraging Available Metadata in VA PTSD Clinics and Generating Benchmarks for Clinically Significant Change

Natasha Benfer[1], Luke Rusowicz-Orazem[1, 2], Elliot M. Fielstein[3, 4], Benjamin C. Darnell[1], Sheila B. Frankfurt[5, 6], Joseph Mignogna[7], and Brett T. Litz[1, 8]

[1] Massachusetts Veterans Epidemiological Research and Information Center, VA Boston Healthcare System, Boston, Massachusetts, United States

[2] Department of Biostatistics, Boston University School of Public Health

[3] US Department Veterans Affairs, Washington, DC, United States

[4] Department of Biomedical Informatics, Vanderbilt University Medical Center

[5] VISN 17 Center of Excellence for Research on Returning War Veterans, Waco, Texas, United States

[6] Central Texas VA Healthcare System, Temple, Texas, United States

[7] Rocky Mountain MIRECC, Rocky Mountain VA, Aurora, Colorado, United States

[8] Department of Psychiatry, Boston University School of Medicine

Objective: Measurement-based care is designed to track symptom levels during treatment and leverage clinically significant change benchmarks to improve quality and outcomes. Though the Veterans Health Administration promotes monitoring progress within posttraumatic stress disorder (PTSD) clinical teams, actionability of data is diminished by a lack of population-based benchmarks for clinically significant change. We reported the state of repeated measurement within PTSD clinical teams, generated benchmarks, and examined outcomes based on these benchmarks. ***Method:*** PTSD Checklist for *the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* data were culled from the Corporate Data Warehouse from the pre-COVID-19 year for Veterans who received at least eight sessions in 14 weeks (episode of care [EOC] cohort) and those who received sporadic care (modal cohort). We used the Jacobson and Truax (1991) approach to generate clinically significant change benchmarks at clinic, regional, and national levels and calculated the frequency of cases that deteriorated, were unchanged, improved, or probably recovered, using our generated benchmarks and benchmarks from a recent study, for both cohorts. ***Results:*** Both the number of repeated measurements and the cases who had multisession care in the Corporate Data Warehouse were very low. Clinically significant change benchmarks were similar across locality levels. The modal cohort had worse outcomes than the EOC cohort. ***Conclusions:*** National benchmarks for clinically significant change could improve the actionability of assessment data for measurement-based care. Benchmarks created using data from Veterans who received multisession care had better outcomes than those receiving sporadic care. Measurement-based care in PTSD clinical teams is hampered by low rates of repeated assessments of outcome.

What is the public health significance of this article?
We generated benchmarks that indicate clinically significant change in PTSD symptoms based on nationwide data from Veterans seeking PTSD care. However, the available data represented a small fraction of total Veterans seeking PTSD care due to a lack of repeated measurements. More repeated measurements are needed to ensure the promises of measurement-based care.

Keywords: measurement-based care, PTSD, clinically significant change, assessment benchmarks, PCL-5

Supplemental materials: https://doi.org/10.1037/ccp0000779.supp

Natasha Benfer https://orcid.org/0000-0002-8179-5356

The authors have no conflicts of interest to declare. Data are available if approved by the Boston institutional review board. Analytic methods are available upon request. This data set has not been used in any published or in press publications to date.

Natasha Benfer played lead role in writing of original draft, supporting role in methodology and project administration and equal role in writing of review and editing. Luke Rusowicz-Orazem played lead role in formal analysis and supporting role in data curation, methodology and writing of original draft. Elliot M. Fielstein played lead role in data curation and resources and supporting role in conceptualization, methodology and writing of review and editing. Benjamin C. Darnell played supporting role in conceptualization and writing of original draft. Sheila B. Frankfurt played supporting role in conceptualization and writing of review and editing. Joseph Mignogna played supporting role in conceptualization and writing of review and editing. Brett T. Litz played lead role in supervision and writing of review and editing, supporting role in resources and writing of original draft and equal role in methodology and project administration.

This project was determined to not require oversight by the Research and Development Committee at the VA Boston Healthcare System. This study was not preregistered.

Correspondence concerning this article should be addressed to Natasha Benfer, Massachusetts Veterans Epidemiological Research and Information Center, VA Boston Healthcare System, 2 Avenue de Lafayette, Boston, MA 02111, United States. Email: natasha.benfer@va.gov

The primary goal of measurement-based care is to monitor change in patient reported outcomes, or the lack thereof, and use shared decision-making to adjust treatment plans accordingly (Joint Commission, 2018). Measurement-based care also entails archiving clinic and system-wide outcome data to improve the quality of care and outcomes in clinics, hospitals, and systems of care. The latter is the foundation of a learning health care system, which dynamically leverages ecologically valid outcome data for practice-based improvements in quality and outcome (Institute of Medicine et al., 2007). The Veterans Health Administration (VHA) is well-positioned to ensure measurement-based care and to leverage large-scale, high-volume patient, and treatment data to improve care because it promotes measurement-based care in their best practice guidelines (Joint Commission, 2018), promotes consistent care, and uses a standardized electronic medical records system to capture and archive patient reported outcome data.

With respect to the mental health care, advancements in patient care and generating knowledge about ways of improving care and care outcomes requires a high degree of compliance with administering, using, and archiving assessments, which has been challenging in VHA. In 2017, VHA launched a national initiative focused on implementing measurement-based care across its mental health services. At baseline, a survey of 230 providers across 47 VHA medical centers found that just 58% of mental health providers collected at least one measure for at least 50% of patients (Oslin et al., 2019). Further, a chart review study of 28,376 Veterans who received any mental health care at a VA hospital found that only 25% of patients had received at least one outcome measure (King et al., 2018). We can infer from these findings that measurement-based care, which requires, at a minimum, *repeated measurement to track patient progress*, is infrequent in VHA mental health care.

Within the context of posttraumatic stress disorder (PTSD) clinical teams, significant resources have been dedicated to increasing measurement-based care in pursuit of evidence-based care. VHA requires all PTSD clinical teams offer VA/Department of Defense guideline consistent care, and the first-line evidence-based psychotherapies (EBPs), namely prolonged exposure (Foa et al., 2007) and cognitive processing therapy (Resick & Schnicke, 1992), recommend outcome tracking. However, these treatment manuals fail to provide information about how to systematically use progress tracking to inform shared decision-making to avoid failure, perhaps leading to less compliance of repeated measurements, even in EBP practice. Nevertheless, given the recommendation for at least repeated assessments of outcome, the expectation is that outcome tracking would be prevalent in PTSD clinical teams (e.g., Sripada et al., 2018). Although this might be the case if administration of EBPs was common practice, EBPs are by far not the modal form of care in PTSD clinical teams. EBPs are provided for only approximately 3.0%–6.3% of patients in PTSD clinical teams in VHA (Sripada et al., 2018; Watts et al., 2014), with many PTSD clinical teams providing higher rates of supportive therapy or providing non-EBP treatment before and/or after a course of EBP (e.g., Finley et al., 2015; Rosen et al., 2016). The state of repeated administration of patient reported outcome in PTSD clinical teams irrespective of EBP initiation (a necessary but not sufficient condition for measurement-based care) is unclear but one study found that <1% of patients with PTSD diagnoses in the VHA have at least four measurements documented in the VHA Corporate Data Warehouse (Shiner et al., 2019). When VHA medical records are interrogated to glean test administration from treatment notes using natural language processing, it appears that there are many more cases of repeated administration of tests (e.g., up to 43% of patients receiving at least one PTSD Checklist [PCL]; Maguen et al., 2021; Shiner et al., 2019). However, mentioning test administration and test scores in chart notes is not equivalent to conducting measurement-based care, which requires data that are presented and used in an actionable manner (i.e., with benchmarks to interpret when action should be taken), and data that are captured, stored, accessed, and able to be aggregated.

Provider-reported barriers to repeated assessments of outcomes outside of VHA include a perceived lack of utility, lack of sufficient training and refresher training, and insufficient support for interpretation and taking follow-up action (e.g., Callaly et al., 2006; Tauscher et al., 2021); the underpinnings of *actionability*. To the last point, when tests are administered solely for putative diagnostic or administrative purposes, and there is no strategy for using the results to guide shared decision-making to avoid failure, patients have little motivation to comply; these practices are antithetical to measurement-based care. Perhaps most important for improving outcome tracking and data capture efforts is addressing the lack of guidance for interpreting collected data in real time to generate clinically useful information regarding patient progress. This problem is a practical one; even for providers or clinics that have a high rate of repeated outcome monitoring, there are no clear guidelines on how to interpret or use the data, and few studies or organizations (e.g., VHA) report how assessment data can be used to produce actionable information about progress (Peterson et al., 2019), which would in theory improve patient and provider satisfaction and outcomes (e.g., lower dropouts, reduced symptom severity that arise after shifting treatment approach and targets). Without these guidelines, measurement-based care risks acquiring a mostly bureaucratic quality, which may explain why 90% of providers in a public mental health system that *required* outcome assessment (not VHA) reported little impact of measurement on their decisions about clinical care (Garland et al., 2003).

The measurement-based care initiative in VHA mental health care emphasizes collaboratively deciding when and how to take action given the results of valid outcome measures as a core feature of shared decision-making (see Resnick & Hoff, 2020). However, without reliable and actionable benchmarks that signal clinically significant, and ultimately, graphical representations of raw scores and benchmarked thresholds, patients and providers are left to either interpret the data subjectively or use heuristics derived from efficacy trials, which likely do not generalize to practice (Fortney et al., 2017; Litz, in press).

Jacobson and Truax's (1991) approach to statistically indexing clinically significant change is generally regarded as the best approach to standardizing clinical benchmarks (Lambert & Ogles, 2009; Speer & Greenbaum, 1995). The J&T approach is appealing because it provides an index of the clinically significant change of change scores at any cross-section after baseline and the magnitude of any endpoint postbaseline score independent of baseline values and can be easily calculated on any clinical sample, and with any measure. The first step is to calculate the reliable change index (RCI), which provides a threshold of statistically reliable change. The second step is to determine a threshold that is clinically meaningful for an endpoint score, which is either two standard deviations (SDs) below the baseline of a patient sample, or,

if nonpatient norms are available, two SDs within the nonpatient population mean (or a midpoint between the patient and nonpatient population means). J&T recommended applying these two clinically significant change algorithms to classify patients as *recovered* (we prefer the term "probable recovery") if their postbaseline score passes the end-state and the RCI criteria; *improved* if the test score passes the RCI criterion; *unchanged* if a test score fails to pass the RCI criterion; and *deteriorated* if worsened scores pass the RCI criterion (statistical and clinically significant worsening).

Despite the potential utility of the J&T approach, it has been applied mostly in variable ways in efficacy trials in the PTSD literature. The most common benchmark used until recently was a 10-point change from baseline on the PTSD Checklist for *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* [*DSM-5*] (PCL-5; Weathers et al., 2013), which was generated from an unspecified calculation of J&T parameters using the PTSD Checklist for *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition* data from an efficacy trial (PCL; Monson et al., 2008). The validity of this threshold at best, is questionable for the PCL-5, as benchmarks from efficacy trials are may not be generalizable to real-world care in all settings.

Marx et al. (2022) calculated clinically significant change criteria and RCI-based thresholds for the Clinician Administered PTSD Scale for *DSM-5* (CAPS-5; Weathers et al., 2018) and the PCL-5, using samples of Veterans and service members from two efficacy trials (randomized controlled trials; group cognitive behavioral therapy, Sloan et al., 2018; and integrated prolonged exposure Norman et al., 2019), and a single site study in primary care patients without PTSD as a normative group (from Bovin et al., 2021). They found a reliable change threshold value of 12–13 or greater points on the CAPS-5 and 15–18 or greater on the PCL-5 and thresholds of total scores of ≤8 and 28, respectively, for probable recovery (Marx et al., 2022). The RCI threshold for clinically significant change varied across trials, suggesting that sample characteristics may limit generalizability of benchmarks across contexts. Indeed, Reponen et al. (2021) describes how clinically useful benchmarks are influenced by variability in intraorganizational, regional, national, and international contexts. For example, the further two clinics or clinical trials differ from one another geographically (e.g., for Marx et al., 2022: the group-based cognitive behavioral therapy randomized controlled trial was conducted at VA hospitals in Boston, MA and Providence, RI; the integrated prolonged exposure randomized controlled trial was conducted at a VA hospital in San Diego, CA), the greater likelihood of sample differences that lead to heterogenous recommendations of benchmarks (Reponen et al., 2021). Consequently, any recommendation for a universal benchmark of clinically significant change for a test are likely not universally valid. Further, aggregation of locally benchmarked outcomes at the clinic or even provider level enhances the validity of data to generate and eventually test ideas in service of quality and outcome improvement (Fortney et al., 2017).

The goals of this study were to generate a status report of the current state of outcome measurement in PTSD clinical teams nationally (indexed by the PCL-5), the extent to which there are usable cases of repeated administrations of the PCL-5 within an episode of care (EOC), and to generate benchmarks for clinically significant change from the available data. In specific terms, we used (a) national data of PCL-5 tests administered in PTSD clinical teams in the 13 months prior to the COVID-19 pandemic in cases that received some form of individual psychotherapy; (b) generated a cohort of cases that had sustained contiguous individual psychotherapy (an EOC), as well as, cases who received a more typical course of care (modal cohort); (c) calculated clinic-level benchmarks for clinically significant change, using various methodological approaches to calculating J&T; (d) examined variation in benchmarks at the clinic, region (represented by Veteran Integrated Service Network [VISN]), and national levels, as well as variation in benchmarks using Marx et al. (2022) parameters across both EOC and modal cohorts; (e) generated the frequency of J&T categories of clinically significant change as indices of effectiveness of treatment within EOCs nationally; and (f) examined variation in these outcomes at various levels of locality and care and methodological approaches.

Method

Participants and Procedure

The sample was drawn from Veterans seen within PTSD clinical teams in VHA nationwide in the year prior to the start of COVID-19 (January 1st, 2019 to January 30th, 2020), to capture patients seen before the shift to telehealth, which challenged the nature of routine psychotherapy and outcome measurement opportunities. We wanted to capture a year's worth of the most recent data before COVID to allow the best chance for repeated measurements to be observed, if they were to be observed at all. Because the discourse about measurement-based care is still new and the efforts to promote it in VHA are nascent, we presumed that reaching further back than 2019 would capture less repeated measurements. Thirteen months of data versus 12 months were captured because we decided to be inclusive of all January dates. Patient data, comprised of the variables outlined below, were extracted from the VHA Corporate Data, which stores patient and encounter (e.g., sessions in a given clinic) metadata. The variables of interest were pulled and analyzed within the secure Veterans Informatics and Computing Infrastructure environment. This project was determined to not require oversight by the Research and Development Committee at the VA Boston Healthcare System. This study was not preregistered.

The full cohort was comprised of patients seen in a PTSD clinical team who received individual therapy, group therapy, medication management, crisis management, or assessment services in any PTSD clinical team across the nation. We then interrogated the full cohort to identify a cohort of patients who engaged in an EOC for individual psychotherapy. We first included any case who had current procedural terminology codes of 30, 45, or 60 min of individual psychotherapy or evaluation and management (E/M). We then selected cases beginning a new EOC in 2019, defined as not having any other sessions in the PTSD clinical team in the preceding 10 weeks. To identify cases that would be most likely to have repeated measurements of outcome, we then selected cases with at least eight sessions within 14 weeks (consistent with the conservative definitions of an EOC in PTSD clinical teams used in a Corporate Data Warehouse-based study by Shiner et al., 2020). Because of our unique aims, we further narrowed the cohort to those cases that had at least one PCL-5 during their EOC. To generate benchmarks for the PCL-5, at least two PCL-5s are needed: at least, a baseline and another administration at or near the end of an EOC. Consequently, we retained cases with at least two PCL-5s, wherein the first PCL-5 was administered within 1 week of the first

individual therapy session of the EOC and the last PCL-5 was administered either within 1 week of the last individual therapy session of the EOC or by at least the fourth individual therapy session of the EOC, since PTSD symptoms might be expected to decrease midway through treatment (i.e., by the sixth session in a 12-session protocol; Resick et al., 2017). We selected cases also with at least four individual psychotherapy encounters before the first and last PCL-5 administration, representing pretreatment to posttreatment change in outcome (half the number of sessions required to meet definition of EOC). We also examined the breakdown in receipt of an EBP (cognitive processing therapy or prolonged exposure) during the EOC. Receipt of EBP was identified using "health factors" associated with each encounter progress note, which indicates that an EBP progress note template was used.

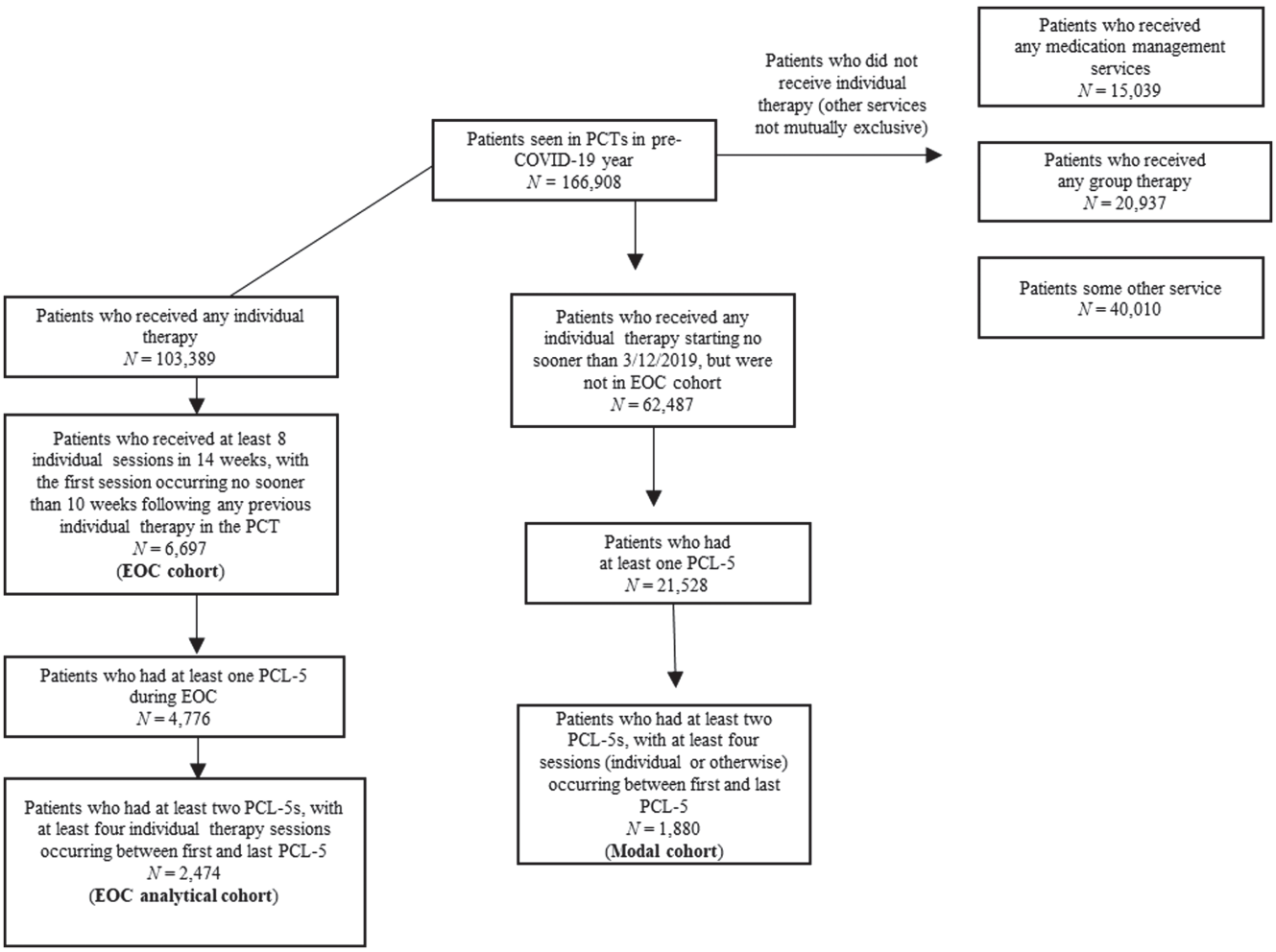
As seen in Figure 1, the EOC cohort we created represented a fraction of the full extracted cohort. To capture the putatively modal form of treatment provided in PTSD clinical teams, we identified a comparison cohort, which we refer to as the *modal cohort*. We first identified non-EOC cohort cases who started individual psychotherapy in a PTSD clinical team on or after March 13, 2019 (i.e., as mentioned above, this represents a period of 10 weeks without any PTSD clinical team sessions), with the last session no later than January 31st, 2020. Next, we selected cases that had at least two PCL-5s, with the first occurring no more than a week from the first individual therapy session and the last occurring no more than a week from the final individual therapy session or the fourth individual therapy session (analogous to the EOC cohort requirement). We finally selected cases who had at least four individual or group psychotherapy encounters between first and last PCL-5 administrations.

Measures and Variables

PTSD Symptoms

The PTSD Checklist for *DSM-5* (PCL-5; Weathers et al., 2013) is a 20-item measure of PTSD symptoms as defined by the *DSM-5*

Figure 1
Cohort Extraction



Note. PCT = PTSD Clinical Team; PCL-5 = PTSD Checklist for *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*; EOC = episode of care; PTSD = posttraumatic stress disorder.

(American Psychiatric Association Division of Research, 2013). Respondents indicate how much they have been bothered by each symptom on a scale of 0 (*Not at all*) to 4 (*Extremely*). The standard version of the PCL-5 is anchored to the past month, but a weekly version is often used clinically to track session-by-session changes. However, due to the inconsistency and confusion among providers about which version to use, both the PCL-5 weekly and PCL-5 monthly were extracted and treated as weekly measures for this study. The PCL-5 has excellent validity and reliability (Bovin et al., 2016). *Service utilization characteristics*. We extracted the types of encounters in the PTSD clinical team in the given timeframe (e.g., medication management, individual psychotherapy), number of individual psychotherapy sessions attended as well as the dates of those sessions, number of PCL-5s completed, whether a session was an EBP was delivered in the span of an identified EOC, and the clinic and VISN in which the sessions were delivered. Clinic was defined as the institution in which the care occurred (e.g., Baltimore PTSD clinical team), rather than individual providers' named clinics.

Data Analysis

All data were extracted using Structured Query Language and analyzed using Statistical Analysis Software (Version 9.4). To generate benchmarks that index clinically significant change, we followed the recommendations of Jacobson and Truax (1991). We first calculated the RCI—a sample-specific change in score magnitude that is statistically reliable. This is calculated by dividing the individual patient change score (i.e., posttreatment minus pretreatment scores) by the standardized difference of the measure. RCI values above or below ±1.96 are reflective of statistically reliable change. For that standardized difference to be sample-specific, local internal consistencies or test–retest coefficients are used. However, there is some debate as to whether internal consistency or test–retest coefficient is preferred (Martinovich et al., 1996; Tingey et al., 1996). Thus, we used both internal consistencies, extracted from all available baseline PCL-5s, and a widely used published test–retest coefficient ($r = .84$; Bovin et al., 2016) in our calculation of RCI, and compared the resulting thresholds for clinically significant change. To calculate the threshold for the clinical significance of endpoint scores, we used the Criterion A method, in which the threshold is defined as endpoint scores that are two standard deviations below the sample baseline mean. Using results from both steps, we categorized cases as: *deteriorated* (reliably changed, but in the negative direction; i.e., RCI ≤ -1.96), *unchanged* (did not reliably change; i.e., RCI between ±1.96), *improved* (reliably changed in the positive direction [RCI ≥ 1.96], but posttreatment score did not fall below the Criterion A threshold), and *probable recovery* (reliably changed in the positive direction, and posttreatment score fell below Criterion A threshold).

Prevalence of rates in each clinically significant change category from each clinic, VISN and the national aggregation were calculated and compared to the previously published thresholds from Marx et al. (2022) and the combined approach, using McNemar's tests for matched pairs, as well as across levels of locality within analytic method (e.g., internal consistency-derived clinic prevalence compared to internal consistency-derived VISN prevalence). To simplify pairwise comparisons, we only examined differences in rates of improved and probable recovery across methods. For our purposes, in using the Marx et al. (2022) thresholds, we used the RCI calculated from the sample that received individual therapy (RCI = 15) since treatment in PTSD clinical teams tends to be focused on PTSD alone (vs. the co-morbid PTSD/alcohol use disorder sample in the other reported study), henceforth referred to as the trial-referenced RCI. The PCL-5 cutoff score of 28 from Marx et al. (2022) is a norm-referenced score calculated from the midpoint between the baseline mean of a sample of primary care patients without PTSD and the patient sample with PTSD. This is known as the Criterion C threshold from Jacobson and Truax (1991), and it is the preferred method to using two standard deviations below the mean of the patient sample (i.e., Criterion A). Finally, we created a benchmark, henceforth referred to as the combined benchmark, that utilized an RCI based on national Corporate Data Warehouse data, using internal consistencies, and the trial-referenced cutoff from Marx et al. (2022). We then compared clinically significant change outcomes using the combined method to clinic, VISN, national, and trial and norm-referenced (i.e., from Marx et al., 2022) clinically significant change outcomes using a series of McNemar's tests.

Data are available if approved by the Boston institutional review board. Analytic methods are available upon request.

Results

Description of Care Received

The sample sizes at each level of cohort extraction are depicted in Figure 1. In the subset of those who received at least one PCL-5 ($N =$ 4,776), an average of 1.20 PCL-5s were completed (mode = 0; median ratio of PCL-5 administrations to individual therapy sessions attended = 0). Of the cases that met criteria for an EOC, 10.43 sessions occurred on average, with a mean of 8.22 days between sessions. These EOC cases had an average of 3.82 PCL-5 administrations (median ratio of PCL-5 administrations to sessions = 0.25). In the EOC cohort, 68.90% received at least one session of an EBP; this latter subset had an average of 5.45 PCL-5 administrations, with mean ration of PCL-5 administrations to sessions of 0.57 (median ratio = 0.56). Of those who received an EBP in the EOC, 68.33% of sessions were categorized as an EBP session. In the modal cohort, 4.77 sessions occurred on average, with a mean of 24.30 days between sessions, and an average of 0.95 PCL-5 administrations (median ratio of PCL-5 administrations to sessions = 0). PCL-5 administration behavior for those who received services other than individual therapy are listed in Supplemental Table 1. There were 2,474 (36.90%) cases who had an EOC who also had sufficient pre- to posttreatment data to calculate clinically significant change (i.e., two PCL-5s, with four sessions between administrations). We labeled this latter subcohort as the EOC analytical cohort. There were 1,880 cases that were analyzable in the modal cohort. They attended an average of 6.41 sessions and completed an average of 4.92 PCL-5s during that time.

PCL-5 Benchmarks

Baseline and final PCL-5 scores were used to assess clinically significant change in the EOC analytical and modal cohorts. At each level of locality—per-clinic, per-VISN, and national aggregate—the internal consistency (Cronbach's α) coefficient was determined for

Table 1
EOC Analytical Cohort Clinic Versus VISN Versus National Versus Marx Versus Combined Outcomes

Category	Clinic—α	VISN—α	National—α	Clinic—Test–retest	VISN—Test–retest	National—Test–retest	Norm-referenced	Combined
RCI threshold	12.34	12.47	12.69	16.05	16.65	16.80	15	16.80
Criterion cutoff score	23	22	22	23	22	22	28	28
Deteriorated (%)	4.1%	4.2%	4.1%,	2.2%	1.9%	1.9%	2.3%	1.90%
Unchanged (%)	51.9%	53.1%	52.8%	62.4%	63.6%	63.1%	61.1%	63.1%
Improved (%)	26.1%[a,d,n]	26.7%[b,e,o]	27.8%[c,f]	18.7%[g,j,p,q]	19.6%[h,k,r]	21%[i,l]	17.4%[a,b,c,g,h,i]	16.3%[d,e,f,j,k,l]
Probable recovery (%)	17.8%[a,m,n]	16.0%[b,e,o]	15.3%[c,f]	16.7%[g,j,q]	15.0%[h,k,r]	14.2%[i,l]	19.2%[a,b,c,g,h,i]	18.8%[e,f,j,k,l]

Note. EOC = episode of care; VISN = Veteran Integrated Service Network; RCI = reliable change index.
[a,b,c] indicate statistically significant prevalence difference (at the 0.05 level of significance) between local internal consistency generated outcomes at the clinic, VISN, and national aggregate level compared to the norm-referenced benchmark. [d,e,f] indicate statistically significant prevalence difference (at the 0.05 level of significance) between local internal consistency generated outcomes at the clinic, VISN, and national aggregate level compared to the combined benchmark. [g,h,i] indicate statistically significant prevalence difference (at the 0.05 level of significance) between published test–retest generated outcomes at the clinic, VISN, and national aggregate level compared to the norm-referenced benchmark. [j,k,l] indicate statistically significant prevalence difference (at the 0.05 level of significance) using published test–retest generated outcomes at the clinic, VISN, and national aggregate level compared to the combined benchmark. [m] indicates statistically significant prevalence difference (at the 0.05 level of significance) between local internal consistency generated outcomes between clinic and VISN level. [n] indicates statistically significant prevalence difference between local internal consistency generated outcomes between clinic and national-level parameters. [o] indicates statistically significant prevalence difference between local internal consistency generated outcomes between VISN and national-level parameters. [p] indicates statistically significant prevalence difference (at the 0.05 level of significance) between published test–retest generated outcomes between clinic and VISN level. [q] indicates statistically significant prevalence difference between published test–retest generated outcomes between clinic and national-level parameters. [r] indicates statistically significant prevalence difference between published test–retest generated outcomes between VISN and national-level parameters.

baseline PCL-5. Uniformly, each value indicated acceptable internal consistency ($\alpha \geq .70$). The distributions of internal consistencies by locality are depicted in Supplemental Figure 1a–b.

The mean baseline PCL-5 score was 53.54 (95% CI [53.00, 54.10]) in the EOC analytical group (n = 2,474), while the modal cohort's mean baseline PCL-5 score was 50.74 (n = 1880; 95% CI [50.50, 51.42]). The mean final score was 41.74 (95% CI [40.99, 42.49]) in the EOC analytical group, while the mean final score in the modal cohort was 45.25 (95% CI [44.43, 46.07]). The difference in baseline-final PCL-5 change score between the EOC analytical group and the modal cohort was −6.31 ($p < .001$, 95% CI [−7.21, −5.41], Cohen's d = 0.43), such that those in the EOC had greater reductions in PCL-5 scores, compared to the modal cohort.

The RCI and criterion cutoff values, by level of locality and method, are displayed in Table 1 (for EOC) and Table 2 (for modal cohort), along with the frequencies of clinically significant change categories. The ranges of RCI and criterion A values are displayed in Figures 2–7. Overall, most Veterans were categorized as unchanged, regardless of method or level of locality. A breakdown of clinically significant change frequencies by VISN can be found in Supplemental Tables 2–5.[1]

Comparison of Outcomes

Tables 1 and 2 display the significant differences in good outcomes (improved and probable recovery) across methods and levels of locality, per the paired McNemar's tests. Significant differences are also described below.

A Comparison of Methods for the EOC Analytical Cohort

Across pairwise tests, thresholds from Marx et al. (2022) led to fewer cases categorized as improved, and more as probable recovery, compared to the locally derived methods. The combined method thresholds led to more cases categorized as probable recovery across all methods and levels of locality, except the clinic-level using internal consistency. The combined method thresholds led to fewer cases categorized as improved, compared all other methods. More cases were classified as probable recovery using the internal consistency-derived clinic benchmarks compared to internal consistency-derived VISN and national benchmarks and more classified as probable recovery, but fewer were deemed improved, when using the internal consistency-derived clinic benchmarks compared to internal consistency-derived national benchmarks. More cases were classified as probable recovery, but fewer as improved, when using internal consistency-derived VISN benchmarks compared to internal consistency-derived national benchmarks. Across levels of locality using test–retest coefficients, fewer cases were

[1] The per-VISN outcomes are limited by the nature of the original Jacobson and Truax (1991) clinically significant change analyses—namely, that individual patients are classified by their observed PCL-5 change score, without explicitly accounting for sources of variation which might influence outcome trajectory. Methods of producing a "case-mix" measurement of clinically significant change account for the possible inherent variation due to factors such as age, sex, ethnicity, socioeconomic differentiators (e.g., income), and geographic discrepancies. We explored analytical models accounting for the random effects of between-VISN variation as a method of case-mix adjusting outcomes but did not find sufficient evidence of variance, potentially due to small cell sizes. Subsequent explorations of clinically significant change outcomes should emphasize appropriate case-mix methods.

Table 2
Modal Clinic Versus VISN Versus National Versus Marx Versus Combined

Category	Clinic—α	VISN—α	National—α	Clinic—Test–retest	VISN—Test–retest	National—Test–retest	Norm-referenced	Combined
RCI threshold	12.39	12.47	12.69	16.16	16.65	16.80	15.00	16.80
Criterion cutoff score	23	22	22	23	22	22	28	28
Deteriorated (%)	**6.7%**	**6.0%**	**6.0%**	**3.9%**	3.7%	**3.7%**	**4.2%**	**3.8%**
Unchanged (%)	**69.2%**	**69.8%**	**69.5%**	**78.2%**	**78.9%**	**78.8%**	**76.7%**	**78.8%**
Improved (%)	**16.3%** [a,d]	**16.3%** [b,e,n]	**17.0%** [c,f]	**10.7%** [j]	**10.8%** [h,k]	**11.2%** [i,l]	**9.8%** [a,b,c,h,i]	**8.5%** [d,e,f,j,k,l]
Probable recovery (%)	**8.3%** [a,m]	**7.8%** [b,e]	**7.6%** [c,f]	**7.2%** [g,i,o,p]	**6.6%** [h,k]	**6.4%** [i,l]	**9.4%** [a,b,c,g,h,i]	**9.1%** [e,f,j,k,l]

Note. EOC = episode of care; VISN = Veteran Integrated Service Network; RCI = reliable change index. Bolded indicates statistically significant differences in prevalence between the EOC analytical cohort versus modal cohort.
[a,b,c] indicate statistically significant prevalence difference (at the 0.05 level of significance) between local internal consistency generated outcomes at the clinic, VISN, and national aggregate level compared to the norm-referenced benchmark. [d,e,f] indicate statistically significant prevalence difference (at the 0.05 level of significance) between local internal consistency generated outcomes at the VISN, and national aggregate level compared to the combined benchmark. [g,h,i] indicate statistically significant prevalence difference (at the 0.05 level of significance) between published test–retest generated outcomes at the clinic, VISN, and national aggregate level compared to the norm-referenced benchmark. [j,k,l] indicate statistically significant prevalence difference (at the 0.05 level of significance) between local published test–retest generated outcomes at the clinic, VISN, and national aggregate level compared to the combined benchmark. [m] indicates statistically significant prevalence difference between local internal consistency generated outcomes between clinic and national-level parameters. [n] indicates statistically significant prevalence difference between local internal consistency generated outcomes between VISN and national-level parameters. [o] indicates statistically significant prevalence difference (at the0.05 level of significance) between published test–retest generated outcomes between clinic and VISN level. [p] indicates statistically significant prevalence difference between published test–retest generated outcomes between clinic and national-level parameters.

classified as improved using test–retest-derived clinic benchmarks compared to test–retest-derived VISN benchmarks. Finally, more Veterans were classified as improved, and fewer as probable recovery, using the test–retest-derived national benchmark compared to both the test–retest-derived clinic and VISN benchmarks.

A Comparison of Methods and Locality for the Modal Cohort

We also compared the frequency of improved and probable recovery cases across methods and levels of locality within the modal cohort. For both the internal consistency and published test–retest values, the prevalence rates of each outcome at all levels of locality (clinic, VISN, and national aggregate) were different from the trial-based threshold for change scores and the norm-referenced end-state benchmark (i.e., combined method), such that more cases were categorized as improved, but fewer as probable recovery, using locally derived benchmarks (except rates of probable recovery were not significantly different from the combined method at the clinic-level using internal consistency). Similarly, the thresholds from Marx et al. (2022) led to fewer cases categorized as improved, and more cases categorized as probable recovery when compared to the locally derived methods. Furthermore, more cases were categorized as probable recovery, compared to internal consistency-derived national benchmarks, by locality, using internal consistency-derived clinic benchmarks. More cases were also categorized as probable recovery using the test–retest-derived clinic benchmarks compared to the test–retest-derived VISN benchmarks. Finally, more cases were categorized as probable recovery using test–retest-derived clinic benchmarks compared to test–retest-derived national benchmarks.

A Comparison of EOC and Modal Cohorts

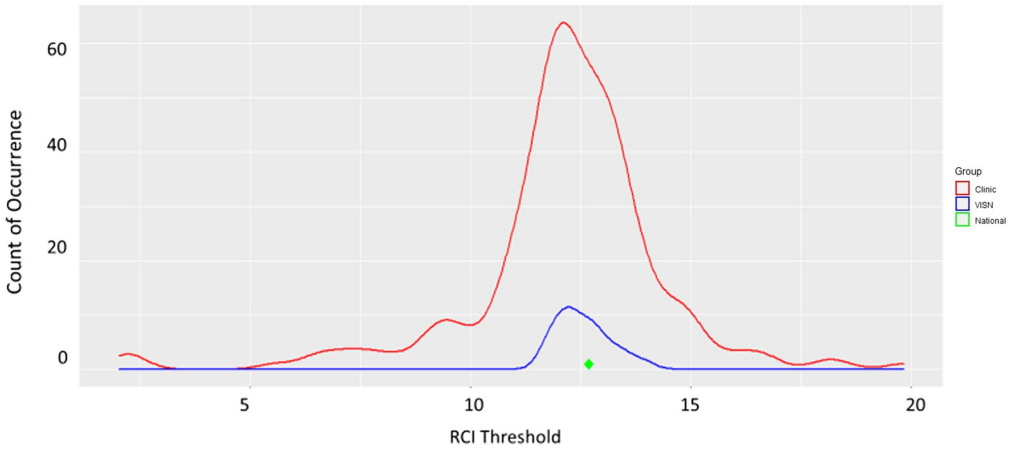
We compared the frequency of the four Jacobson & Truax (1991) outcomes between the EOC analytical and the modal cohorts, using both calculation methods, and at each level of locality. We estimated the prevalence difference using *Z* tests for difference of proportions, assessing statistical significance at the 0.05 level. All comparisons revealed significant differences, except for the difference between those who deteriorated at the VISN level, using the test–retest coefficient. Across both methods and all levels of locality, the modal cohort had fewer cases categorized as improved and probable recovery, and more categorized as unchanged and deteriorated.

Discussion

This national study was the first to assess the frequency of repeated measurement of PTSD outcome in the pre-COVID year, to generate benchmarks for clinically significant change based on sustained individual treatment (chiefly EBP), in an EOC, and to examine benchmarked clinically significant outcomes, using data from PTSD clinical teams in the VHA. Other studies have calculated J&T benchmarks using data from Veterans enrolled in efficacy trials, but these thresholds for clinically significant change may not generalize to real-world Veteran care, a point underscored by Marx et al. (2022). We aimed to determine whether benchmarks for clinically significant change based on real-world VHA data differed from the benchmarks generated by Marx et al. (2022) who used a combination of trial- and norm-referenced data, and whether level of

Figure 2

RCI Ranges at the Clinic and VISN Level, Using Internal Consistencies for EOC Analytical Cohort



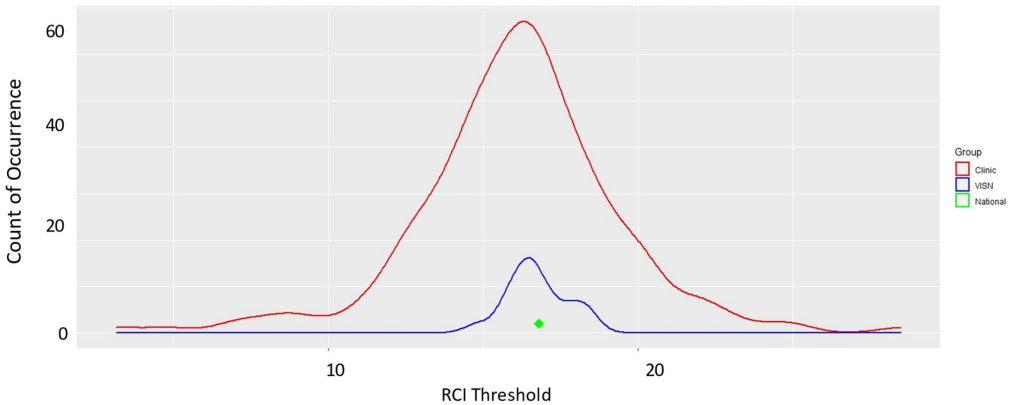
Note. Clinic: kurtosis = 8.29, *M* = 12.34, *Mdn* = 12.21, range = 2.01–19.83; VISN: kurtosis = 2.66, *M* = 12.47, *Mdn* = 12.38, range = 11.76–13.92; *N* clinics = 210; *N* VISNs = 18. RCI = reliable change index; VISN = Veteran Integrated Service Network; EOC = episode of care. See the online article for the color version of this figure.

locality (i.e., clinic, VISN, national), and the index of measurement error (i.e., internal consistency vs. test–retest coefficients) affected benchmarks.

Data used to generate clinically significant change benchmarks came from just 4.2% of all Veterans who received individual therapy in PTSD clinical teams in the pre-COVID-19 year (2.4% in an EOC, 1.8% in the modal cohort). This means that an extremely small number of cases in PTSD clinical teams get the extent of treatment that has been evaluated in the efficacy trials that have informed evidence-based practice in PTSD clinical teams (i.e., contiguous individual psychotherapy lasting ˜10–14 sessions). For our EOC cohort, the largest "cut" in creating the cohort was reducing the group who received any individual therapy (n = 103,389) to those who received at least eight sessions in 14 weeks (n = 6,697). Although not all contiguous care is EBP (approximately, 69% according to our results), other research examining the rates of EBP engagement using the Corporate Data Warehouse data has found similarly low rates, with approximately 20% of Veterans dropping out before the third session (Sayer et al., 2022) and nearly 70% receiving less than eight sessions (Hale et al., 2019). Our small n in the EOC is likely due to a mix of more patients receiving sporadic care as well as dropout, which is a known significant problem in PTSD treatment (Najavits, 2015). It will be important to ensure measurement-based care in the context of sporadic care so that the Corporate Data Warehouse can be populated and effectiveness can be examined. Finally, the cases included in our EOC and modal cohorts are by definition outliers and one of the unintended consequences of a lack of outcome assessment data in the Corporate Data Warehouse is that it is difficult to determine how the 96% of the PTSD clinical team cases differ from those that get chiefly EBPs.

Figure 3

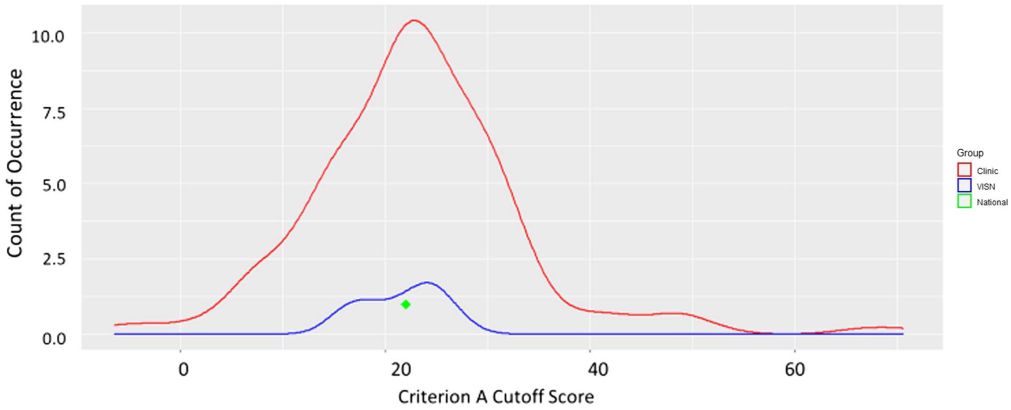
RCI Ranges at the Clinic and VISN Level, Using Test–Retest for EOC Analytical Cohort



Note. Clinic: kurtosis = 5.43, *M* = 16.05, *Mdn* = 15.92, range = 3.14–28.48; VISN: kurtosis = 2.56, *M* = 16.65, *Mdn* = 16.53, range = 14.79–18.42; *N* clinics = 210; *N* VISNs = 18. RCI = reliable change index; VISN = Veteran Integrated Service Network; EOC = episode of care. See the online article for the color version of this figure.

Figure 4

Criterion A Ranges at the Clinic and VISN Level for EOC Analytical Cohort



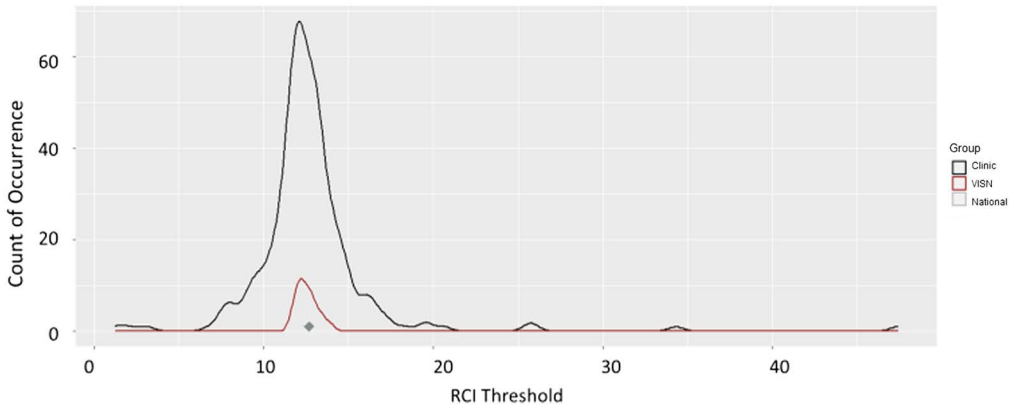
Note. Clinic: kurtosis = 6.48, *M* = 23.81, *Mdn* = 24.50, range = −6.54–70.60; *N* clinics = 210; VISN: kurtosis = 1.81, *M* = 22.27, *Mdn* = 23.75, range = 15.36–27.47; *N* VISNs = 18. VISN = Veteran Integrated Service Network; EOC = episode of care. See the online article for the color version of this figure.

Rates of PCL-5 administrations, even among those who engaged in consistent individual therapy, were strikingly low. Among those who received contiguous multisession individual therapy (eight sessions in 14 weeks) in PTSD clinical teams, only approximately 37% of Veterans had at least two PCL-5s in the Corporate Data Warehouse, and the average number of PCL-5s for the EOC cohort was three in the 14-week time period. PCL-5 administration was even lower for those who received sporadic care, which was modal; only 3% of the modal cohort of Veterans had at least two PCL-5s. Although two PCL-5s were a minimum requirement to generate sufficient data to calculate benchmarks for clinically significant change, *pre–postassessment is, in most cases, not measurement-based care*. Although an examination of change or the lack thereof when therapy has ended might inform shared decision-making about the need for more care or another type of care, it is too late to inform treatment planning *while therapy is ongoing*, which is the primary goal of measurement-based care. Measurement-based care requires sustained repeated assessments to track change and to guide decision-making. In the EOC analytical cohort, 38.5% of the Veterans had an individual psychotherapy session to PCL-5 administration ratio of greater than two, indicating that just over a third were receiving a PCL-5, on average, every other session. In the modal cohort, 22.3% had a session to PCL-5 ratio of greater than two. Findings are consistent with other investigations into PCL administration behavior culled from the Corporate Data Warehouse. Shiner et al. (2020) found that, in a cohort of Veterans who received at least eight EBP sessions, only 19% had at least two PCLs, and less than 1% of Veterans had at least four PCLs (Shiner et al., 2019).

We generated clinically significant change benchmarks using a variety of methods and levels of locality based on available EOC data. The RCI-based thresholds for clinically significant change were more conservative using the test–retest coefficient as the index

Figure 5

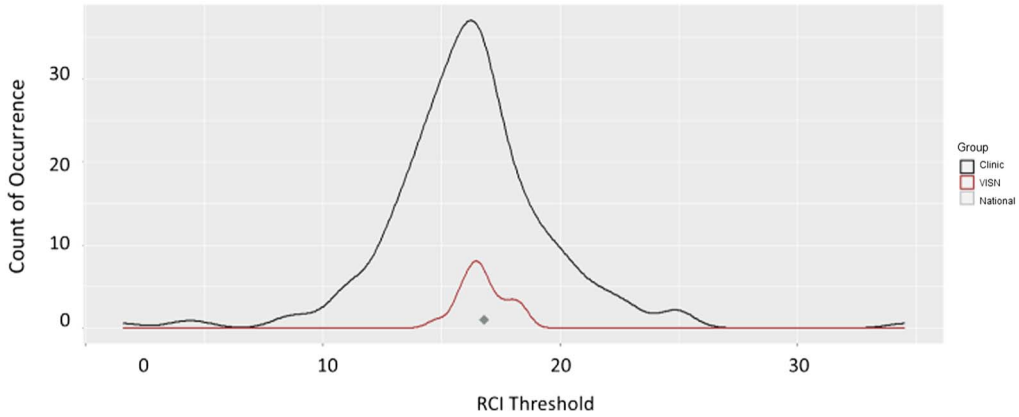
RCI Ranges at the Clinic and VISN Level, Using Internal Consistencies for Modal Cohort



Note. Clinic: kurtosis = 39.59, *M* = 12.39, *Mdn* = 12.32, range = 1.25–47.37; VISN: kurtosis = 2.66, *M* = 12.47, *Mdn* = 12.38, range = 11.76–13.92; *N* clinics = 220; *N* VISNs = 18. RCI = reliable change index; VISN = Veteran Integrated Service Network. See the online article for the color version of this figure.

Figure 6

RCI Ranges at the Clinic and VISN Level, Using Test–Retest for Modal Cohort



Note. Clinic: kurtosis = 7.91, *M* = 16.16, *Mdn* = 16.19, range = 1.57–34.50; VISN: kurtosis = 2.56, *M* = 16.65, *Mdn* = 16.53, range = 14.79–18.42; *N* clinics = 220; *N* VISNs = 18. RCI = reliable change index; VISN = Veteran Integrated Service Network. See the online article for the color version of this figure.

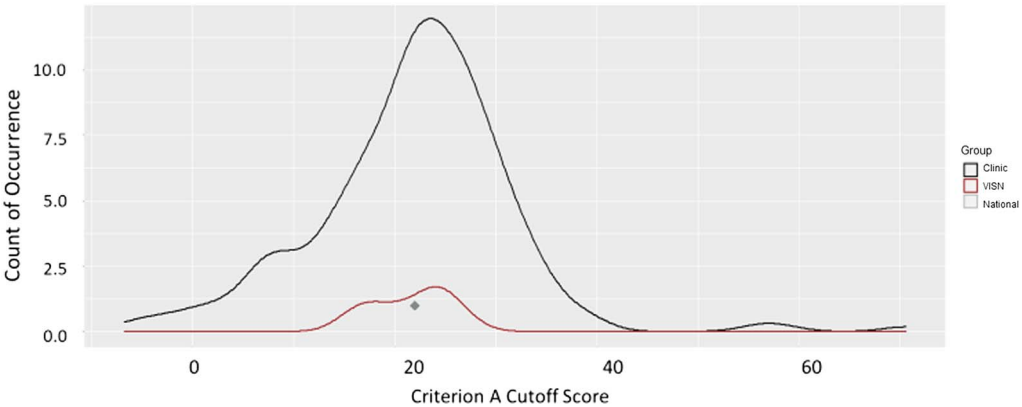
of measurement error, versus sample-specific internal consistency, regardless of the level of locality and the cohort type. The RCIs using test–retest were similar to the trial-referenced RCI from the clinical trial from Marx et al. (2022). The norm-referenced criterion cutoff for endpoint scores from Marx et al. (2022) was six points higher than the national VA population sample, and thus led to higher frequencies of probable recovery.

Regarding levels of locality, there was more variation in RCI and Criterion A at the clinic-level compared to the VISN-level, as would be expected, though there was also notable variation in outcomes across VISNs. However, some of the VISNs contained as few as one case, which constrains inferences about outcomes by VISN Despite more spread in the clinic-level benchmarks, level of locality did not matter. Across both EOC and modal cohorts, benchmarks were nearly identical for clinic, VISN, and national RCI using both test–retest, and internal consistency coefficients. Similarly, the RCI and Criterion A did not differ across cohorts, which is not surprising given that they generally used the same baseline data, aside from slightly more clinics included in the generation of modal clinic-level thresholds. Although there were some significant differences in the frequencies of good outcomes across levels of locality, these differences were generally small (<2%).

Overall, across methods and levels of locality, those in the modal cohort were more likely to have deteriorated (although clinically significant deterioration at each level of analysis was low) and to be unchanged than those in the EOC, and those in the EOC were more likely to have improved or be probably recovered. There are many possible reasons for these notable differences. First, those in the EOC, by definition, received more sessions, and thus may have experienced greater benefit from a higher dose of individual

Figure 7

Criterion A Ranges at the Clinic and VISN Level for Modal Cohort



Note. Clinic kurtosis = 6.48; *M* = 23.03; *Mdn* = 23.09; range = −6.83–70.60; *N* clinics = 220; VISN kurtosis = 1.81; *M* = 22.27; *Mdn* = 23.75; range = 15.36–27.47; *N* VISNs = 18. VISN = Veteran Integrated Service Network. See the online article for the color version of this figure.

psychotherapy. Second but related, differences in outcome could be at least partially due to the greater likelihood of EBP in the EOC compared to the modal cohort, given that the EOC had a minimum of eight sessions. Finally, it is likely that the patients who received an EOC differ from those who received modal care on characteristics that are associated with better treatment outcomes. Indeed, previous research has identified demographic (age, gender, race) and military history characteristics (combat exposure, trauma type, number of deployments) that are associated with greater rates of completion of EBPs and repeated measurement (Maguen et al., 2019; Shiner et al., 2019).

These results showed a serious lack of repeated measurement in PTSD clinical teams for individual psychotherapy, even in the context of EBPs, which greatly constrains any effort to leverage measurement-based care to improve outcome and efforts to leverage aggregated data for quality and outcome improvement (Litz, in press). We posit that charting total scores using benchmarks for clinically significant change will aid in shared decision-making with respect to treatment planning and measurement-based care and will foster greater use of repeated assessments because the results will be more actionable to Veterans. We believe that the benchmarks we generated can be used for these purposes, but they will need to be revised when there is much more data available. Benchmarks should be routinely updated over time to accurately reflect the symptom burden of Veterans in a given time period. Preliminary evidence suggests PTSD symptoms were impacted during COVID-19 for Veterans (Pedersen et al., 2021; Straus et al., 2022), and it is likely that rates of measurement, and how these data were stored, changed as providers were adapting assessments to a virtual format.

Our results revealed little variation in benchmarks or effectiveness based on locality, but the index for measurement error for the RCI mattered considerably. We suggest that the VHA use the test–retest coefficient to calculate the national RCI (16.80 based on available data in this study) because it was recommended originally by Jacobson and Truax (1991), and it is standardized and does not require constant updating of the internal consistencies from the baseline PCL-5s in the Corporate Data Warehouse. The RCI using the test–retest coefficient was very similar to the RCI identified in Marx et al. (2022), which suggests broad applicability. For determining the criterion cutoff for end-state scores, we recommend the cutoff of 28 that was identified in Marx et al. (2022), because they were able to use J&T's preferred Criterion C method, versus our use of the Criterion A approach. The Criterion C cutoff is indexed to a group that is not dysfunctional, which enhances the validity of the determination of whether an end-state score is clinically significant. The use of the national RCI with test–retest coefficient and norm-referenced criterion C was captured in our *combined* approach. Using this approach, half of the cases who received at least eight sessions in 14 weeks did not achieve clinically significant change, 16% were improved, and 19% were probably recovered. For those that received contiguous care (i.e., modal cohort), nearly 80% did not make clinically significant change gains.

In addition to the limited generalizability of our Corporate Data Warehouse-derived benchmarks and associated outcomes, a limitation of our findings is that we are unable to confidently say which of the benchmarks we calculated are the most valid. To do so, we would need to examine which benchmark had the greatest criterion-related validity, indexed to a gold standard, as was done with the benchmarks from Marx et al. (2022) in the original publication. Quality of life and functioning are good candidates for a gold standard (Hinton et al., 2021; Schnurr & Lunney, 2016), but measurement of these constructs is not routine in PTSD clinical teams. Ideally, routine measurement in PTSD clinical teams would include a brief, psychometrically sound and valid assessment of functioning and quality of life to validate benchmarks of clinically significant change in PTSD symptoms. We also acknowledge that there are likely many administered PCL-5s not captured in the Corporate Data Warehouse, given that chart data captured more assessments of PCLs than what is available in the Corporate Data Warehouse (Shiner et al., 2021). While this may have improved the picture of frequency of repeated measurement in PTSD clinical teams, it would not be meaningful because data need to be stored in a centralized data repository (i.e., the Corporate Data Warehouse) to be used to generate these benchmarks easily and continuously in order to be functional for measurement-based care. Thus, while findings on rates of repeated measurements may be an underestimate of actual clinician behavior, the Corporate Data Warehouse is alarmingly unsaturated with outcome data. We argue that once clinicians have a clinician- and patient-facing outcome tracking system with benchmarks for clinically significant change the treatment planning and shared decision-making benefits will promote measurement-based care and greater saturation (representativeness) of local, regional, and national metadata. This kind of data capture process could realize the promise of a learning health care system, by providing clinicians a means of aggregating data across caseloads to generate hypotheses about ways of improving quality and outcomes, and clinic and national leaders a means of examining and improving outcomes at scale.

References

American Psychiatric Association Division of Research. (2013). Highlights of changes from DSM-IV to DSM-5: Posttraumatic stress disorder. *Focus—American Psychiatric Publishing*, *11*(3), 358–361. https://doi.org/10.1176/appi.focus.11.3.358

Bovin, M. J., Kimerling, R., Weathers, F. W., Prins, A., Marx, B. P., Post, E. P., & Schnurr, P. P. (2021). Diagnostic accuracy and acceptability of the primary care posttraumatic stress disorder screen for the diagnostic and statistical manual of mental disorders (fifth edition) among US veterans. *JAMA Network Open*, *4*(2), Article e2036733. https://doi.org/10.1001/jamanetworkopen.2020.36733

Bovin, M. J., Marx, B. P., Weathers, F. W., Gallagher, M. W., Rodriguez, P., Schnurr, P. P., & Keane, T. M. (2016). Psychometric properties of the PTSD checklist for diagnostic and statistical manual of mental disorders–fifth edition (PCL-5) in veterans. *Psychological Assessment*, *28*(11), 1379–1391. https://doi.org/10.1037/pas0000254

Callaly, T., Hyland, M., Coombs, T., & Trauer, T. (2006). Routine outcome measurement in public mental health: Results of a clinician survey. *Australian Health Review*, *30*(2), 164–173. https://doi.org/10.1071/AH060164

Finley, E. P., Garcia, H. A., Ketchum, N. S., McGeary, D. D., McGeary, C. A., Stirman, S. W., & Peterson, A. L. (2015). Utilization of evidence-based psychotherapies in Veterans Affairs posttraumatic stress disorder outpatient clinics. *Psychological Services*, *12*(1), 73–82. https://doi.org/10.1037/ser0000014

Foa, E., Hembree, E., & Rothbaum, B. O. (2007). *Prolonged exposure therapy for PTSD: Emotional processing of traumatic experiences therapist guide*. Oxford University Press. https://doi.org/10.1093/med:psych/9780195308501.001.0001

Fortney, J. C., Unützer, J., Wrenn, G., Pyne, J. M., Smith, G. R., Schoenbaum, M., & Harbin, H. T. (2017). A tipping point for measurement-based care. *Psychiatric Services*, *68*(2), 179–188. https://doi.org/10.1176/appi.ps.201500439

Garland, A. F., Kruse, M., & Aarons, G. A. (2003). Clinicians and outcome measurement: What's the use? *The Journal of Behavioral Health Services & Research*, *30*(4), 393–405. https://doi.org/10.1007/BF02287427

Hale, A. C., Bohnert, K. M., Ganoczy, D., & Sripada, R. K. (2019). Predictors of treatment adequacy during evidence-based psychotherapy for PTSD. *Psychiatric Services*, *70*(5), 367–373. https://doi.org/10.1176/appi.ps.201800361

Hinton, M., O'Donnell, M., Cowlishaw, S., Kartal, D., Metcalf, O., Varker, T., McFarlane, A. C., Hopwood, M., Bryant, R. A., Forbes, D., Howard, A., Lau, W., Cooper, J., & Phelps, A. J. (2021). Defining post-traumatic stress disorder recovery in veterans: Benchmarking symptom change against functioning indicators. *Stress and Health*, *37*(3), 547–556. https://doi.org/10.1002/smi.3019

Institute of Medicine (US), Roundtable on Evidence-Based Medicine, Olsen, L., Aisner, D., & McGinnis, J. M. (Eds.). (2007). *The learning healthcare system: Workshop summary*. National Academies Press (US). https://doi.org/10.17226/11903

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, *59*(1), 12–19. https://doi.org/10.1037/0022-006X.59.1.12

Joint Commission. (2018). *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

King, P. R., Beehler, G. P., Wade, M., Buchholz, L. J., Funderburk, J. S., Lilienthal, K. R., & Vair, C. L. (2018). Opportunities to improve measurement-based care practices in mental health care systems: A case example of electronic mental health screening and measurement. *Families, Systems & Health*, *36*(4), 427–438. https://doi.org/10.1037/fsh0000379

Lambert, M. J., & Ogles, B. M. (2009). Using clinical significance in psychotherapy outcome research: The need for a common procedure and validity data. *Psychotherapy Research*, *19*(4–5), 493–501. https://doi.org/10.1080/10503300902849483

Litz, B. (in press). It is time to flip the script and leverage the point of care to discover ways of improving PTSD treatment outcomes: Commentary on "A Sobering Look at Treatment Effectiveness of Military-Related Posttraumatic Stress Disorder (Levi et al., 2021)". *Clinical Psychological Science*.

Maguen, S., Li, Y., Madden, E., Seal, K. H., Neylan, T. C., Patterson, O. V., DuVall, S. L., Lujan, C., & Shiner, B. (2019). Factors associated with completing evidence-based psychotherapy for PTSD among veterans in a national healthcare system. *Psychiatry Research*, *274*, 112–128. https://doi.org/10.1016/j.psychres.2019.02.027

Maguen, S., Madden, E., Holder, N., Li, Y., Seal, K. H., Neylan, T. C., Lujan, C., Patterson, O. V., DuVall, S. L., & Shiner, B. (2021). Effectiveness and comparative effectiveness of evidence-based psychotherapies for posttraumatic stress disorder in clinical practice. *Psychological Medicine*. Advance online publication. https://doi.org/10.1017/S0033291721001628

Martinovich, Z., Saunders, S., & Howard, K. (1996). Some comments on "assessing clinical significance". *Psychotherapy Research*, *6*(2), 124–132. https://doi.org/10.1080/10503309612331331648

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

Monson, C. M., Gradus, J. L., Young-Xu, Y., Schnurr, P. P., Price, J. L., & Schumm, J. A. (2008). Change in posttraumatic stress disorder symptoms: Do clinicians and patients agree? *Psychological Assessment*, *20*(2), 131–138. https://doi.org/10.1037/1040-3590.20.2.131

Najavits, L. M. (2015). The problem of dropout from "gold standard" PTSD therapies. *F1000prime Reports*, *7*, Article 43. https://doi.org/10.12703/P7-43

Norman, S. B., Trim, R., Haller, M., Davis, B. C., Myers, U. S., Colvonen, P. J., Blanes, E., Lyons, R., Siegel, E. Y., Angkaw, A. C., Norman, G. J., & Mayes, T. (2019). Efficacy of integrated exposure therapy vs integrated coping skills therapy for comorbid posttraumatic stress disorder and alcohol use disorder: A randomized clinical trial. *JAMA Psychiatry*, *76*(8), 791–799. https://doi.org/10.1001/jamapsychiatry.2019.0638

Oslin, D. W., Hoff, R., Mignogna, J., & Resnick, S. G. (2019). Provider attitudes and experience with measurement-based mental health care in the VA implementation project. *Psychiatric Services*, *70*(2), 135–138. https://doi.org/10.1176/appi.ps.201800228

Pedersen, E. R., Davis, J. P., Fitzke, R. E., Lee, D. S., & Saba, S. (2021). American Veterans in the era of COVID-19: Reactions to the pandemic, posttraumatic stress disorder, and substance use behaviors. *International Journal of Mental Health and Addiction*. Advance online publication. https://doi.org/10.1007/s11469-021-00620-0

Peterson, K., Anderson, J., & Bourne, D. (2019). *Evidence brief: Use of patient reported outcome measures for measurement based care in mental health shared decision-making* (VA evidence-based synthesis program reports). DC. Department of Veterans Affairs.

Reponen, E., Rundall, T. G., Shortell, S. M., Blodgett, J. C., Juarez, A., Jokela, R., Mäkijärvi, M., & Torkki, P. (2021). Benchmarking outcomes on multiple contextual levels in lean healthcare: A systematic review, development of a conceptual framework, and a research agenda. *BMC Health Services Research*, *21*(1), Article 161. https://doi.org/10.1186/s12913-021-06160-6

Resick, P. A., & Schnicke, M. K. (1992). Cognitive processing therapy for sexual assault victims. *Journal of Consulting and Clinical Psychology*, *60*(5), 748–756. https://doi.org/10.1037/0022-006X.60.5.748

Resick, P. A., Wachen, J. S., Dondanville, K. A., Pruiksma, K. E., Yarvis, J. S., Peterson, A. L., Mintz, J., Borah, E. V., Brundige, A., Hembree, E. A., Litz, B. T., Roache, J. D., Young-McCaughan, S., & the and the STRONG STAR Consortium. (2017). Effect of group vs individual cognitive processing therapy in active-duty military seeking treatment for posttraumatic stress disorder: A randomized clinical trial. *JAMA Psychiatry*, *74*(1), 28–36. https://doi.org/10.1001/jamapsychiatry.2016.2729

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., Matthieu, M. M., Wiltsey Stirman, S., Cook, J. M., Landes, S., Bernardy, N. C., Chard, K. M., Crowley, J., Eftekhari, A., Finley, E. P., Hamblen, J. L., Harik, J. M., Kehle-Forbes, S. M., Meis, L. A., Osei-Bonsu, P. E., Rodriguez, A. L., Ruggiero, K. J., Ruzek, J. I., Smith, B. N., … Watts, B. V. (2016). A review of studies on the system-wide implementation of evidence-based psychotherapies for posttraumatic stress disorder in the Veterans Health Administration. *Administration and Policy in Mental Health*, *43*(6), 957–977. https://doi.org/10.1007/s10488-016-0755-0

Sayer, N. A., Wiltsey-Stirman, S., Rosen, C. S., Bernardy, N. C., Spoont, M. R., Kehle-Forbes, S. M., Eftekhari, A., Chard, K. M., & Nelson, D. B. (2022). Investigation of therapist effects on patient engagement in evidence-based psychotherapies for posttraumatic stress disorder in the Veterans Health Administration. *Journal of Traumatic Stress*, *35*(1), 66–77. https://doi.org/10.1002/jts.22679

Schnurr, P. P., & Lunney, C. A. (2016). Symptom benchmarks of improved quality of life in PTSD. *Depression and Anxiety*, *33*(3), 247–255. https://doi.org/10.1002/da.22477

Shiner, B., Gui, J., Leonard Westgate, C., Schnurr, P. P., Watts, B. V., Cornelius, S. L., & Maguen, S. (2019). Using patient-reported outcomes to understand the effectiveness of guideline-concordant care for posttraumatic stress disorder in clinical practice. *Journal of Evaluation in Clinical Practice*, *25*(4), 689–699. https://doi.org/10.1111/jep.13158

Shiner, B., Levis, M., Dufort, V. M., Patterson, O. V., Watts, B. V., DuVall, S. L., Russ, C. J., & Maguen, S. (2021). Improvements to PTSD quality metrics with natural language processing. *Journal of Evaluation in*

Clinical Practice. Advance online publication. https://doi.org/10.1111/jep.13587

Shiner, B., Westgate, C. L., Gui, J., Cornelius, S., Maguen, S. E., Watts, B. V., & Schnurr, P. P. (2020). Measurement strategies for evidence-based psychotherapy for posttraumatic stress disorder delivery: Trends and associations with patient-reported outcomes. *Administration and Policy in Mental Health*, *47*(3), 451–467. https://doi.org/10.1007/s10488-019-01004-2

Sloan, D. M., Unger, W., Lee, D. J., & Beck, J. G. (2018). A randomized controlled trial of group cognitive behavioral treatment for veterans diagnosed with chronic posttraumatic stress disorder. *Journal of Traumatic Stress*, *31*(6), 886–898. https://doi.org/10.1002/jts.22338

Speer, D. C., & Greenbaum, P. E. (1995). Five methods for computing significant individual client change and improvement rates: Support for an individual growth curve approach. *Journal of Consulting and Clinical Psychology*, *63*(6), 1044–1048. https://doi.org/10.1037/0022-006X.63.6.1044

Sripada, R. K., Bohnert, K. M., Ganoczy, D., & Pfeiffer, P. N. (2018). Documentation of evidence-based psychotherapy and care quality for PTSD in the Department of Veterans Affairs. *Administration and Policy in Mental Health*, *45*(3), 353–361. https://doi.org/10.1007/s10488-017-0828-8

Straus, L. D., Dolsen, E. A., Nishimi, K., Neylan, T. C., & O'Donovan, A. (2022). Worse sleep, worsening post-traumatic stress disorder (PTSD): Poor sleep quality associated with increases in PTSD symptoms amongst those experiencing high threat of the COVID-19 pandemic. *Journal of Sleep Research*, *31*(5), Article e13568. https://doi.org/10.1111/jsr.13568

Tauscher, J. S., Cohn, E. B., Johnson, T. R., Diteman, K. D., Ries, R. K., Atkins, D. C., & Hallgren, K. A. (2021). What do clinicians want? Understanding frontline addiction treatment clinicians' preferences and priorities to improve the design of measurement-based care technology. *Addiction Science & Clinical Practice*, *16*(1), Article 38. https://doi.org/10.1186/s13722-021-00247-5

Tingey, R., Lambert, M., Burlingame, G., & Hansen, N. (1996). Assessing clinical significance: Proposed extensions to method. *Psychotherapy Research*, *6*(2), 109–123. https://doi.org/10.1080/10503309612331331638

Watts, B. V., Shiner, B., Zubkoff, L., Carpenter-Song, E., Ronconi, J. M., & Coldwell, C. M. (2014). Implementation of evidence-based psychotherapies for posttraumatic stress disorder in VA specialty clinics. *Psychiatric Services*, *65*(5), 648–653. https://doi.org/10.1176/appi.ps.201300176

Weathers, F. W., Bovin, M. J., Lee, D. J., Sloan, D. M., Schnurr, P. P., Kaloupek, D. G., Keane, T. M., & Marx, B. P. (2018). The Clinician-Administered PTSD Scale for DSM-5 (CAPS-5): Development and initial psychometric evaluation in military veterans. *Psychological Assessment*, *30*(3), 383–395. https://doi.org/10.1037/pas0000486

Weathers, F. W., Litz, B. T., Keane, T. M., Palmieri, P. A., Marx, B. P., & Schnurr, P. P. (2013). *The PTSD checklist for dsm-5 (pcl-5). Scale available from the National Center for PTSD*. www.ptsd.va.gov

Received July 20, 2022
Revision received October 17, 2022
Accepted October 25, 2022 ■