Psychological Services

In the public domain
ISSN: 1541-1559

2024, Vol. 21, No. 2, 235–246
https://doi.org/10.1037/ser0000761

Expected Symptom Change Trajectories for the Early Identification of Probable Treatment Nonresponse in VA PTSD Specialty Care Clinics: A Proof-of-Concept

Benjamin C. Darnell[1], Natasha Benfer[1], Maya Bina N. Vannini[1], Breanna Grunthal[1], Luke Rusowicz-Orazem[1, 2], Elliot Fielstein[3, 4, 5], and Brett T. Litz[1, 6]

[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] Department of Biomedical Informatics, Vanderbilt University
[4] Department of Psychiatry, Vanderbilt University
[5] U.S. Department of Veterans Affairs, Washington, DC, United States
[6] Department of Psychiatry, Boston University Chobanian & Avedisian School of Medicine

The purpose of measurement-based care (MBC) is to detect treatment nonresponse sufficiently early in treatment to adjust treatment plans and prevent failure or dropout. Thus, the potential of MBC is to provide the infrastructure for a flexible, patient-centered approach to evidence-based care. However, MBC is underutilized across the Department of Veterans Affairs (VA) posttraumatic stress disorder (PTSD) specialty clinics, likely because no actionable, empirically determined guidelines for using repeated measurement effectively are currently available to clinicians. With data collected as part of routine care in VA PTSD specialty clinics across the United States in the year prior to COVID-19 (n = 2,182), we conducted a proof-of-concept for a method of generating session-by-session benchmarks of probable patient nonresponse to treatment, which can be visualized alongside individual patient data using the most common measure of PTSD symptoms used in VA specialty clinics, the PTSD Checklist for *Diagnostic and Statistical Manual of Mental Disorders, 5th Edition* (PCL-5). Using survival analysis, we first identified the probability of cases reaching clinically significant change at each session, as well as any significant moderators of treatment response. We then generated a multilevel model with initial symptom burden predicting the trajectory of PCL-5 scores across sessions. Finally, we determined the slowest changing 50% and 60% of all cases to generate benchmarks at each session for each level of the predictor(s) and then assessed the accuracy of these benchmarks at each session for classifying treatment responders and nonresponders. The final models were able to accurately identify nonresponders as early as the sixth session of treatment.

Impact Statement

The models we generated were able to accurately identify treatment nonresponders as early as the sixth session. Should these models prove similarly accurate in specific clinics and for specific providers and patients in practice, then they may be applicable as guidelines for clinicians to make repeated measurement clinically useful and actionable, supporting measurement-based care goals in VA PTSD specialty care.

Keywords: measurement-based care, evidence-based practice, posttraumatic stress disorder, Veterans, patient-centered

Supplemental materials: https://doi.org/10.1037/ser0000761.supp

Current evidence-based practice (EBP) for posttraumatic stress disorder (PTSD) is largely limited to packages of serially ordered treatment sessions with prescribed content intended to be applied in the manner tested in efficacy trials, namely with high fidelity per the treatment manual (e.g., Foa et al., 2007; Resick et al., 2008). However, efficacy trials across medicine and mental health typically do not translate well to routine care (they sacrifice external validity for superb internal validity; e.g., Diener et al., 2022; Hansen et al., 2002), and at best, providers should only expect to see the mean effect sizes of an efficacy trial if a given patient's characteristics and the quality of the care are similar to that of the trial. Modally, clinicians should expect variable effectiveness and be prepared to address it (e.g., Benfer et al., 2022; Maguen et al., 2021; Sripada et al., 2020). This is particularly pertinent to the care of Veterans with PTSD in that the clinical significance of the effects of EBPs in psychotherapy trials of service members and Veterans with PTSD are lower than civilian trials

This article was published Online First April 6, 2023.

Benjamin C. Darnell https://orcid.org/0000-0002-3962-0544

Correspondence concerning this article should be addressed to Brett T. Litz, Department of Psychiatry, Boston University Chobanian & Avedisian School of Medicine, 720 Harrison Avenue, Boston, MA 02130, United States. Email: litzb@bu.edu

(e.g., Litz et al., 2019; Steenkamp et al., 2015). Moreover, EBP is more than adherence to a manual supported by efficacy trials evidence with fidelity (Sackett et al., 1996). Proper *person-centered* care requires that providers meet the idiographic needs of diverse patient presentations (Bensing, 2000; Engle et al., 2021), and providers need to be prepared to utilize a shared decision-making framework to collaboratively discuss treatment options whenever a selected approach is not working, to avoid treatment failure and dropout. Outcome tracking in an actionable manner is the backbone of measurement-based care (MBC; e.g., Joint Commission, 2018), yet actionable MBC is only possible if providers (and their patients) can identify the need for a shift in approach.

Unfortunately, even though repeated assessments of PTSD is recommended by EBPs (e.g., Foa et al., 2007), there is very little evidence-based guidance about how to use repeated assessments of outcome to avoid treatment failures. Without actionable results (and guidance about what to do), the use of measurement in treatment is limited and de facto not MBC. Across the Department of Veterans Affairs (VA) as a whole, only about a half of the mental health providers reported using at least one measurement for at least half of their patients (58%; Oslin et al., 2019). In VA PTSD specialty care clinics (i.e., PTSD clinical teams [PCTs]), only 44% of patients had an administration of at least one PTSD measure during treatment documented by their provider (Maguen et al., 2021); far fewer documented repeated assessment of PTSD severity over the course of treatment (0.5% used multiple measurements; Shiner et al., 2018). Benfer et al. (2022) examined data captured and stored in the VA's corporate data warehouse in the year before COVID-19 as part of routine individual therapy across all PCTs to assess the prevalence of repeated assessments of outcome. Of those patients who received an episode of care (EoC) defined as at least eight sessions within 14 weeks, only 37% received at least two documented (i.e., data available in the corporate data warehouse) measurements, representing just 2.4% of all Veterans who received care within a PCT that year. Of those patients who received treatment most typical of PCTs (i.e., noncontiguous sporadic treatment sessions), only 3% received at least two documented measurements (Benfer et al., 2022).

For mental health care outside of the VA (i.e., in civilian hospitals/clinics), even when providers use multiple measurements to track outcomes, the majority report little, if any, impact of these measurements on their decisions about care (e.g., Garland et al., 2003). Such limited impact may be due to difficulty visualizing and interpreting tracked data (e.g., Callaly et al., 2006; Tauscher et al., 2021). The potential of MBC is chiefly to avoid treatment failures and promote a more flexible, personalized, and patient-centered approach to EBP. However, for this potential to be realized, clinicians and patients need tools and guidance to generate *actionable* information from tracked outcome data (Peterson et al., 2019). Despite the current limited utility and application of MBC across VA broadly, and PCTs specifically, the limited patient outcome data available within the corporate data warehouse is sufficient to create initial clinically actionable benchmarks for the most commonly used measure of PTSD symptoms, the PTSD Checklist for *Diagnostic and Statistical Manual of Mental Disorders, 5th Edition* (PCL-5; Weathers et al., 2013). Although the available data is a convenience sample representing only the small fraction of cases seen in PCTs with repeated measurement of outcome, the hope is that initial benchmarks will improve the clinical utility of MBC and increase its adoption, thus resulting in greater data generation and capture and allowing for further improvements of benchmarks.

Typically, in clinical trials and some practice contexts, the methods recommended by Jacobson and Truax (1992) have been used to establish statistical benchmarks for clinically significant change (CSC). This method generates thresholds to indicate sufficient improvement or deterioration at each assessment point post-baseline. Lambert et al. (2001) later generated a benchmarking method that may further improve shared decision-making in MBC. Their method generates individual-level expected trajectories of scores based on the best predictors of change in a given setting to determine if progress is less than expected and if continued treatment is unlikely to lead to CSC after a given session number.

Lambert et al. (2001) used data from their national database of civilian outpatient psychotherapy providers and clinics to generate a survival curve that represented the relationship between CSC on the Outcome Questionnaire-45 (a measure of general behavioral health; Lambert et al., 2004) and session number. Treatment site and initial symptom burden (i.e., baseline Outcome Questionnaire-45 total score) significantly moderated the curve. Lambert et al. (2001) then calculated recovery trajectories for each level of initial symptom burden (i.e., bands of initial Outcome Questionnaire-45 scores representing 2% increments of the sample, 90 total curves), with two-tailed "tolerance intervals" generated around the curves at the 68th and 80th percentiles. These tolerance intervals were used as guideposts in determining if a patient's actual data trajectory throughout treatment is within limits of the expected recovery curve for patients with similar initial symptom burden. These percentiles are characteristic of the 10%–15% of patients with the slowest response to treatment, chosen to ensure capture of the ~10% of patients likely to deteriorate during psychotherapy (see Lambert, 2013, for a review). When a patient's outcome falls outside their predicted tolerance intervals, this may predict treatment failure, thus allowing providers and patients to take action and make changes to the treatment plan. Initial testing of the utility of this method in clinical practice suggested that cases with providers who were given feedback using this system were more likely to improve on the Outcome Questionnaire-45, and less likely to deteriorate, than cases with providers who did not receive that feedback (Lambert et al., 2001). Later replications of this study confirm that the use of this feedback system reliably promotes improved patient outcomes and decreased dropout compared to treatment as usual and other feedback methods (see Lambert et al., 2018 for a meta-analysis).

In this study, we attempted to replicate the methods proposed by Lambert et al. (2001) using data culled from the VA's corporate data warehouse of PCL-5 data collected in the year prior to COVID-19. We reasoned that COVID-19 led to limitations in collecting assessment data and that the year prior to COVID-19 would represent an optimal interval to capture a cohort with an EoC and at least two PCL-5s. We calculated a survival curve for CSC and assessed the moderating effects of initial symptom burden, region (i.e., Veteran Integrated Service Network [VISN]), and their interaction (to assess for differences in the moderating effect of initial symptom burden by region). Similar to Lambert et al. (2001), we expected that both initial symptom burden and treatment location would moderate the survival curve. Unlike Lambert, the availability of data at any specific treatment site was too limited to assess differences at the site level, thus treatment sites were grouped by region (i.e., VISN) to approximate geographic areas where treatment sites are likely to have similar culture and access to resources (i.e.,

differences between treatment sites are increasingly complex with greater geographic distance; Reponen et al., 2021). Expected PTSD symptom trajectories and tolerance intervals were generated for every level of each identified moderator and/or interaction between moderators, which could include initial symptom burden, treatment location, or their interaction term. Our aim was to generate a working set of trajectories of treatment response for Veterans initiating care in PCTs). We planned to use these trajectories to generate benchmarks for determining when a Veteran is offtrack for expected change and at-risk for treatment failure.

Method

Participants and Procedure

This project was determined not to require oversight by the VA Boston Research and Development Committee.

Generating the Cohort

Actionable MBC requires a useful benchmark for CSC at each assessment/session cross-section and an a priori expectation that, if change has not been demonstrated after a given session number of a treatment approach, an expected positive trajectory of CSC is improbable (i.e., a dose–response should be observed if treatment is working as intended; e.g., Hansen et al., 2002). The statistical methods for generating benchmarks for CSC require considerable data within treatment settings. To generate a cohort of cases with sufficient outcome data from the VA's corporate data warehouse, we first identified cases with PCL-5 data captured in the year before the start of COVID-19. We then kept the data of those who received contiguous individual therapy conservatively defined as an EoC of at least eight sessions in 14 weeks (with no other sessions in the preceding 10 weeks to ensure capture of new EoCs). Of these cases, we retained the 37% who received at least two PCL-5's during treatment, which is the minimum required to calculate indices of CSC (this is 2.4% of all Veterans who received some kind PTSD specialty care in VA nationally in the year prior to COVID-19). We relied on repeated measurements to accomplish our aims, yet available PCL-5 data in the CDW is not explicitly linked to a specific session. However, the administration date of the first PCL-5 score for each case was within at least 1 week of the date of their first recorded session for that EoC and the date of their last PCL-5 was no earlier than the fourth session, thus allowing for more accurate estimates of treatment response (for a more detailed description of the method for culling EoC data, see Benfer et al., 2022). To approximate PCL-5 administrations at least at every other session, thus minimizing cases with sporadic measurement and more closely tying PCL-5 scores with specific session numbers, we extracted a repeated-measures (RM) subsample of Veterans who, in addition to the above criteria, had recorded PCL-5 administrations equal to at least 50% of their total number of sessions during the 14-week window.

Measures

PCL-5

The PCL-5 is a 20-item self-report measure of PTSD symptoms, per *Diagnostic and Statistical Manual of Mental Disorders, 5th Edition* (American Psychiatric Association, 2013). Participants indicate the extent to which they have been bothered by each symptom on a scale of 0 (*Not at all*) to 4 (*Extremely*), all tied to their index (Criterion A) event. The PCL-5 is traditionally anchored to past month distress, but in clinical care, a weekly version is often used. Due to inconsistencies in clinicians using both the weekly and monthly versions to track weekly change, both the PCL-5 weekly and PCL-5 monthly were extracted. The PCL-5 has demonstrated excellent psychometric properties in Veterans (Bovin et al., 2016).

Treatment Site

For Veterans who met our subsample criteria, we also extracted the VISN in which they received care. There are 18 VISNs across the United States, which are systems of VA medical centers defined by regional boundaries, and thus, are a proxy for potential differences in geography, demographics, and VA resources.

Data Analysis Plan

Clinically Significant Change

CSC was calculated according to Jacobson and Truax (1992). Reliable change indices (RCIs) were calculated for each case at each time point as the difference between an individual's PCL-5 total score at that time point and their initial PCL-5 total score, divided by the standard error of measurement of the difference. Standardized difference is calculated using an indicator of reliability of the measure—typically either the internal consistency or test–retest coefficient—and standard deviation of the measure. It has been previously demonstrated that generating RCIs from national-level data is appropriate (Benfer et al., 2022) and preferable to generating them from regional or site-specific data, as it allows for simpler comparisons across cases. We followed the recommendations of Benfer et al. (2022) who generated Jacobson and Truax (1992) RCIs using national corporate data warehouse data and the published test–retest coefficient of $r = .84$ (Bovin et al., 2016) to index measurement error. An RCI of ≥ 1.96 is indicative of CSC (Jacobson & Truax, 1992).

Survival Analysis

Following the procedures outlined in Lambert et al. (2001), survival analysis was used to identify the percentage of patients who meet criteria for CSC, based on the reliable change index, at each session. Survival analysis is a nonparametric procedure that can be used for tracking the occurrence or nonoccurrence of a dichotomous event over multiple time points. This type of analysis examines data across time for a predetermined "terminal event" to answer the question "how long until X occurs in this sample?" We used survival analysis to track the occurrence or nonoccurrence of CSC for the RM subsample. A terminal event was considered to have occurred at the session when a patient met the criteria for CSC and did not change from that outcome at a later session, including if data collection terminated at that point (e.g., a patient who reaches CSC at their last session and a patient who reaches CSC and maintains it for several more sessions would both be considered to have a terminal event).

Survival analysis incorporates censored data (i.e., when the conditions for a terminal event are not met for a given patient and no further data for that patient is available) using a likelihood-based approach, arguably the most effective method for adjusting for

censored data (Turkson et al., 2021), assuming that censored data occurs at random. Therefore, unlike other methods for analyzing longitudinal data (e.g., repeated-measures multiple analysis of variance), survival analysis does not require that all patients have the same number of sessions; if a patient has too few sessions and does not meet the terminal event, their data are treated as censored and are still included in analysis following likelihood adjustment. We used survival analysis to estimate a survival curve representing the number of cases who met the conditions for a terminal event (i.e., CSC) at each session, starting from initiation of an EoC.

Many variables may impact the number of sessions required to reach CSC for patients who are responsive to a given treatment approach. Lambert et al. (2001) used data collection site and pretreatment symptom levels as moderators of their survival curve; therefore, we similarly assessed the effect of treatment region (i.e., VISNs; a proxy for site), initial symptom burden, and their interaction term using Cox Proportional Hazards and backwards removal of insignificant moderators. Significant moderators were then used in subsequent multilevel analysis to generate clinically actionable treatment response curves that may be used to assess when individual patients are likely not responding to treatment as expected, thus indicating the need to shift approach.

Multilevel Modeling

Longitudinal data are often hierarchically clustered, such that within-person, time-varying variables (e.g., an individual's PTSD symptoms across an EoC; Level 1) are influenced by higher order, between-person variables such as diagnostic categories or clinics (Level 2). Multilevel modeling is a statistical method that can be used for examining individual-level change as a function of group-level variables. This method models within-person data for the Level 1 analysis, then uses the Level 1 estimates as the dependent variables in the Level 2 analysis, with between-person predictors. Further, estimates can be calculated for every individual regardless of number of sessions in their EoC; as such, fixed effects for session number are not as influenced by differences in measurement/treatment duration or missing data, as in other repeated-measures approaches (Kahn, 2011; Lutz et al., 1999).

Using the RM subsample, significant predictor variables identified from the Cox Proportional Hazards (i.e., initial PCL-5 symptom burden, defined as eight ranges or "bands" of initial PCL-5 total scores; see Supplemental Material, for how initial symptom bands were determined) were entered as Level 2 between-person predictors in a multilevel model predicting Level 1 within-person change in PTSD symptoms over time (site was ultimately not included as a Level 2 predictor in the multilevel model due to low sample sizes in several VISNs, despite statistically significant moderation in the survival analysis). The final multilevel within-person model is as follows (see MacCallum et al., 1997):

$$Y_{it} = \pi_{0i} + \pi_{1i}(\text{SesLog}_{it}) + e_{it}, \quad (1)$$

where Y_{it} is the PCL-5 total score for individual i at time (i.e., session number) t, π_{0i} is the random intercept for individual i at Time 0, π_{1i} is the random slope of change over time for individual i (with time calculated as a log-linear transformation of session number for individual i at time t; i.e., SesLog_{it}), and e_{it} is a residual error term for individual i's actual PCL-5 score deviation from their predicted score. The between-person model is as follows:

$$\pi_{0i} = \beta_{00} + \beta_{01}(\text{Initial Symptom Burden}) + r_{0i}, \quad (2)$$

$$\pi_{1i} = \beta_{10} + \beta_{11}(\text{Initial Symptom Burden}) + r_{1i}, \quad (3)$$

where π_{0i} is the within-person random intercept for individual i at Time 0, β_{00} is the average intercept, β_{01} is the slope of change in intercept by initial PCL-5 total score (i.e., band),

r_{0i} is a residual error term for individual i's actual intercept's deviation from their predicted intercept, π_{1i} is the within-person random slope of change over time for individual i, β_{10} is the average slope of change over time, β_{11} is the slope of change in slope by initial PCL-5 total score (i.e., Band), and r_{1i} is a residual error term for individual i's actual slope's deviation from their predicted slope.

Tolerance Intervals and Predictive Accuracy (Exploratory)

Lambert et al. (2001) generated tolerance intervals at the two-tailed 68th and 80th percentiles (i.e., one-tailed 85th and 90th percentiles, respectively) to capture the ~10%–15% of cases who will likely *deteriorate* (e.g., Lambert & Vermeersch 1994). However, our goal was to identify all treatment nonresponders, which included 61.7% of the RM subsample and up to 50% of cohorts in other studies (Resick et al., 2017; Schottenbauer et al., 2008). Therefore, we generated tolerance intervals at the one-tailed 40th and 50th percentiles. Ideally, CSC trajectories and tolerance intervals should be generated in one sample (i.e., the training sample) and tested in an independent sample (i.e., the testing sample); however, we did not have sufficient data to split our data set into training and testing subsets. Instead, to assess the predictive sensitivity (i.e., the probability at each session that a patient who ultimately *does not* reach CSC is to be identified as *offtrack*) and specificity (i.e., the probability at each session that a patient who ultimately *does* reach CSC is to be identified as *on-track*) of these trajectories within the RM subsample, we categorized each case into their appropriate band (e.g., any case with an initial score of 52 was categorized into Band 4) and their PCL-5 score at each session according to its relative position to the associated expected trajectory for cases within that band (i.e., if at any given session a case's PCL-5 score was above the 50th percentile tolerance interval for the expected trajectory of their initial score band, it was coded as offtrack).

Predictive sensitivity at each session was determined by the proportion of cases that were correctly coded as offtrack (i.e., they were coded as offtrack at that session and ultimately did not meet for CSC by the end of their EoC); predictive specificity was determined by the proportion of cases that were correctly coded as on-track (i.e., they were coded as on-track at that session and ultimately did meet for CSC by the end of their EoC). Higher predictive sensitivity at a session would indicate that the method accurately categorized treatment nonresponders as offtrack, while higher predictive specificity would indicate that the method is not miscategorizing treatment responders as offtrack. As such, it is important to reach a balance of predictive sensitivity and specificity for a method to be clinically useful at a given session for distinguishing probable responders from nonresponders, with the

benchmark for a balanced method being a combined sensitivity and specificity of at least 1.5 (Power et al., 2013).

Results

Preliminary Analysis

The RM subsample was comprised of 2,182 patients and included only those cases with a ratio of at least 1:2 of PCL-5 data to sessions within an EoC, to approximate measurements for at least every other session. Total initial PCL-5 scores ranged from 17 to 80 (M = 51.98, SD = 13.56) and were normally distributed. The number of PCL-5 administrations ranged from 4 to 36 (M = 9.01, SD = 3.55). The number of cases per VISN ranged from 25 to 293, with an average of 121.22 (SD = 57.69). The threshold for clinically significant improvement based on the RCI was a decrease in PCL-5 scores of at least 15 points, per J&T calculations. Although our threshold was higher than estimates generated using the original *Diagnostic and Statistical Manual of Mental Disorders, 4th Edition*-based PCL (i.e., a 10 point change; Monson et al., 2008), our threshold is similar to previously published CSC thresholds for the PCL-5 that were benchmarked against changes in psychosocial functioning (i.e., a 15–18 point decrease in PCL-5 total score; Marx et al., 2022), as well as recent efforts to generate clinically actionable population-based benchmarks for meaningful change across locality levels using a similar sample culled from the VA corporate data warehouse (i.e., a 12.34–16.80 point change; Benfer et al., 2022). For the RM subsample, 38.3% of Veterans improved by the end of their EoC (n = 836). There were no differences in significant results between analyses run with and without outliers (i.e., cases with a total session number greater than 3x the interquartile range beyond the 1st/3rd quartile for total session number for the full sample; n = 14), therefore all reported results include outliers to preserve power.

Primary Analysis

Survival Analysis

The effect of initial PTSD symptom burden and treatment location (by VISN) on the probability of reaching CSC at each session (i.e., the survival curve) was assessed using Cox Proportional Hazards (see Table 1). Initial PCL-5 total score, region (i.e., VISN), and an interaction term (initial PCL-5 score × VISN) were included as predictors in the model. The interaction term was not significant (p = .96) and was thus removed. The overall goodness-of-fit of the final model was significant, $\chi^2(18) = 371.34$, $p < .001$. The overall variable for initial PTSD symptom burden was significant, Wald $\chi^2(1) = 21.49$, $p < .001$, such that, as initial symptom burden increased, the predicted probability of a treatment responder reaching CSC at a given session also increased. Compared to the reference VISN (i.e., VISN 1; VA New England HealthCare System), 10 of the other 17 VISNs significantly predicted a lower probability of reaching CSC at each session (p's < .05), one predicted a higher probability (p = .01), and six were not significant (p's > .05).

Holding moderators constant (i.e., for patients with average initial symptom burden at the average VISN), treatment responders had a >10% probability of reporting CSC by the sixth PCL-5 administration (i.e., approximately Session 6), a >25%

Table 1

Summary of Cox Proportional Hazards Regression Comparing Effect of the Variables VISN and Initial PCL-5 Total Score on Overall Survival

Variable	β	*SE*	Wald	*df*	*R*	Exp(β)
Initial PCL-5 score	0.01	.00	21.49***	1	.06	1.01
VISN						
1. (New England)[a]			341.71***	17	.01	
2. (New York/New Jersey)	−0.87	.20	19.94***	1	.01	0.42
4. (Pennsylvania, Delaware)	−0.21	.17	1.46	1	—	—
5. (West Virginia, Maryland)	−0.82	.14	33.51***	1	.01	0.44
6. (Mid-Atlantic)	−0.63	.13	25.04***	1	.00	0.53
7. (Southeast)	−0.41	.13	9.66**	1	.03	0.66
8. (Florida, Puerto Rico, Virgin Islands)	0.08	.12	0.49	1	—	—
9. (Mid-South)	−1.09	.13	76.01***	1	.03	0.34
10. (Ohio, Indiana, Michigan)	−0.39	.11	12.74***	1	.02	0.68
12. (Great Lakes)	−0.11	.14	0.62	1	—	—
15. (Heartland)	−0.14	.15	0.87	1	—	—
16. (South-Central)	−0.27	.12	5.55*	11	.02	0.76
17. (Texas)	−0.16	.14	1.29	1	—	—
19. (Rocky Mountain)	0.30	.12	6.67**	1	.06	1.35
20. (Northwest)	−0.52	.14	13.46***	1	.00	0.60
21. (Sierra Pacific)	−0.47	.12	15.35***	1	.00	0.63
22. (Desert Pacific)	−0.38	.12	9.65**	1	.02	0.69
23. (Midwest)	−0.12	.13	0.78	1	—	—

Note. N = 2,182. VISN = Veterans Integrated Services Network; PCL-5 = PTSD Checklist for *DSM-5*; *SE* = standard error; *df* = degrees of freedom; PTSD = posttraumatic stress disorder; *DSM-5* = *Diagnostic and Statistical Manual of Mental Disorders, 5th Edition*; CSC = clinically significant change. For Exp(β), values >1 have increased chance to reach CSC and values <1 have decreased chance to reach CSC.

[a] Reference group.

* $p < .05$. ** $p < .01$. *** $p < .001$.

probability by the ninth, >50% by the 13th, >75% by the 22nd, and >88% by the 30th (see Figure 1). As session number increased (and holding moderators constant), the probability of a given treatment responder reaching CSC also increased, with increasingly larger increases in probability until Session 9 (i.e., an accelerating increase in probability). After Session 9, the increase in probability steadily decreased in magnitude (i.e., a decelerating increase in probability; Session 9 represented the point of diminishing returns for treatment) until more than half of the sample had met for CSC at around session 14 and the change in the probability of reaching CSC between subsequent sessions reached an asymptote.

Multilevel Models

Detailed information regarding the multilevel modeling can be found in the Supplemental Materials. Like Lambert et al. (2001), prior to building the model, we used a log-linear transformation by session number to normalize the symptom score distributions across time. Further, and also similar to Lambert et al. (2001), we grouped initial PCL-5 scores into ranges (i.e., bands) of severity (eight total bands; see Table 2). Based on the results of the Cox Proportional Hazards (see Table 1), both VISN and initial symptom burden were significant predictors of treatment response. However, due to limited availability of outcome data for several VISNs, only initial symptom burden was included in the multilevel models.

Using the full RM subsample, the full model was constructed by first generating an unconditional means model (Model 1; all models are presented in Table 3) as the initial reference model, then unconditional growth models with a fixed slope for session number (Model 2_f) and random slopes (Model 2_r), and finally a conditional growth model (Model 3). Models were compared in order of increasing complexity to the previously best fitting model. Overall, Model 3 was the best fitting model, accounting for 93.54% of the proportion reduction in unexplained variance in the relationship between session number (log) and PCL-5 total score (i.e., at Level 1 of the analysis), compared to the previously best fitting model. The Level 2 proportion reduction in unexplained variance was −3.44%, indicating that the effect of initial PCL-5 score band was small (for two-level models, a negative proportion reduction in unexplained variance is not considered indicative of reduced model fit; see Roberts et al., 2011). Fixed effects indicated a significant intercept, $b = 28.83$, $p < .001$, 95% CI [28.83, 29.88], a significant effect of session number (log), $b = -12.61$, $p < .001$, 95% CI [−13.37, −11.86], and a significant effect of initial PCL-5 band, $b = 5.56$, $p < .001$, 95% CI [5.45, 5.66]. Random effects suggest significant differences between individuals in the variability of their PCL-5 scores across sessions, 95% CI [3.04, 3.73] and their length of treatment (log), 95% CI [15.40, 16.65]. The interclass coefficient (ICC) indicated little unexplained clustering within the data (ICC = .20).

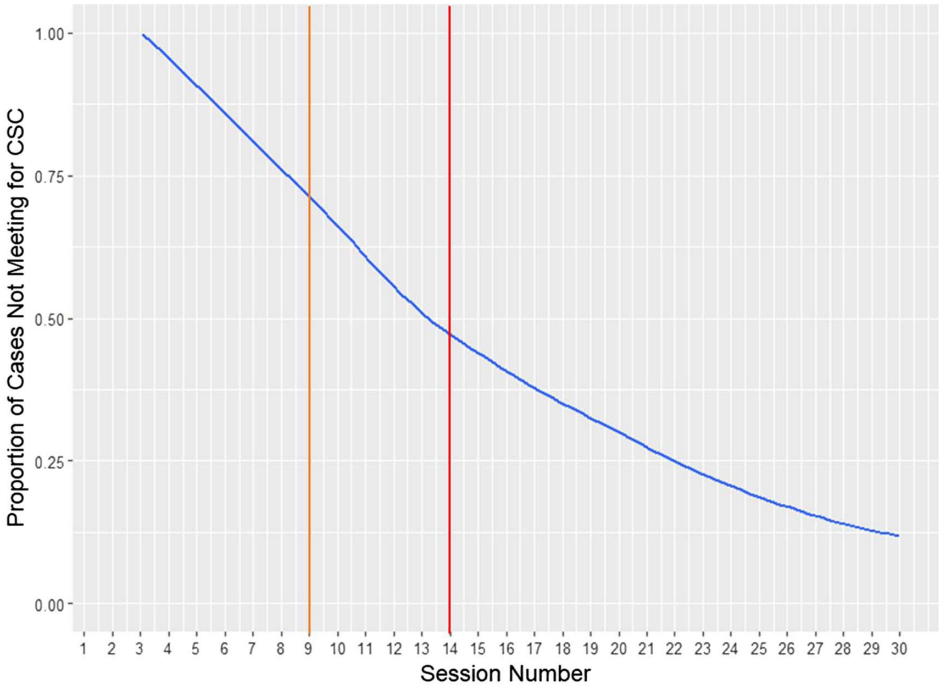
Exploratory Analyses

Survival Analyses by Band

Considering that a conditional growth model with random slopes was the best fitting model and the effect of initial PCL-5

Figure 1

Survival Curve of Patients Who Met for CSC by the End of an EoC



Note. $N = 2{,}182$. CSC = clinically significant change; EoC = episode of care; Left vertical line = the session number at which the rate of patients meeting for CSC begins to decelerate; Right vertical line = the session number at which the rate of patients meeting for CSC begins to flatten. See the online article for the color version of this figure.

Table 2
Session Number to Reach CSC

Band	*n* (CSC/ no CSC)	25%	50%	75%	90%	Point of diminishing returns
1. (17–38)	(96/247)	10	21	—	—	9
2. (39–43)	(82/114)	9	11	16	25	9
3. (44–48)	(106/156)	8	12	18	23	10
4. (49–53)	(123/185)	9	13	23	29	10
5. (54–58)	(135/199)	8	12	20	23	11
6. (59–63)	(117/165)	9	13	22	25	12
7. (64–68)	(88/135)	9	16	—	—	9
8. (69–80)	(89/145)	9	16	—	—	9

Note. Band = range of initial PCL-5 total scores; CSC = clinically significant change; CSC/no CSC = sample sizes of patients within that range who did/did not meet for CSC by the end of an episode of care; PCL-5 = PTSD Checklist for *DSM-5*; PTSD = posttraumatic stress disorder; *DSM-5* = *Diagnostic and Statistical Manual of Mental Disorders, 5th Edition*. Each percent column represents the probability that a case that reaches CSC will reach it at that session. The Point of Diminishing Returns is the session number at which the rate of cases reaching CSC begins to decelerate.

band on the rate of change was significant but small, exploratory survival analyses were conducted for cases within each band to parse this effect (see Table 2). For Band 1 (initial PCL-5 scores ranging from 17 to 38), Band 7 (64–68), and Band 8 (69–80), the estimated probability of reaching CSC by the last available session for cases who report CSC prior to the end of an EoC was 67.3%, 54.6%, and 73.5%, respectively. These probabilities are lower than all other bands, with each reaching a minimum of 90% by the last available session. The point of diminishing returns was Session 9 for Bands 1, 2, 7, and 8, Session 10 for Bands 3 and 4, Session 11 for Band 5, and Session 12 for Band 6. See Figure 2 for example trajectories and benchmarks.

Predictive Sensitivity and Specificity

Using the 40th percentile tolerance interval, the model reached a combined sensitivity and specificity of 1.55 by Session 6, with a sensitivity of 80% and a specificity of 74% (see Table 4, for results of all sensitivity and specificity analyses). Following Session 6, high sensitivity and specificity was maintained across sessions (e.g., at Session 9, sensitivity was 89% and specificity was 79%). Using the 50th percentile tolerance interval, the model reached a combined sensitivity and specificity of 1.51 by Session 6, with a sensitivity of 68% and a specificity of 83%. Following Session 6, high sensitivity and specificity was maintained across sessions (e.g., at Session 9, sensitivity was 76% and specificity was 89%).

Discussion

We leveraged available VA data to generate initial benchmarks for tracking and predicting patient progress to avoid treatment failure. The results provide preliminary support for predicting treatment nonresponse, which may prove clinically useful for improving Veteran outcomes in PTSD specialty care. Further, our findings confirm the results of many clinical trials (see Litz et al., 2019) that have shown that, contrary to the expectation for a regression to the mean, Veterans with higher initial PCL-5 total scores change at a slower rate and/or are more likely to be treatment nonresponders within a typical EoC than Veterans with lower initial scores. Contrarily, our planned survival analysis indicated that, as initial PTSD symptom burden increased, the probability of reaching CSC in an EoC also increased. However, this finding may be due to a floor effect, as many cases with low initial scores may not be expected to demonstrate much change. This interpretation is consistent with the exploratory survival analyses by band (as well as the final multilevel model), which demonstrated that the individuals in our sample who reported CSC by the end of an EoC whose initial PCL-5 scores were in the lowest scoring band (i.e., Band 1) or two highest scoring bands (i.e., Bands 7 and 8) had much lower probabilities of reaching CSC at higher session numbers compared to all other bands.

In light of a probable floor effect for cases in Band 1, a potential pattern is recognizable in the rest of these results; as initial symptom burden increases, more sessions are generally required to reach CSC, with cases initially scoring above 64 on the PCL-5 requiring more sessions than are represented in our data set to achieve the same probability of success and reach a point of diminishing returns as cases with lower initial scores. In practice, the latter point translates to Veterans who score in the highest two ranges (i.e., Bands 7 and 8) having a <75% probability of reaching CSC by the last available session (Session 26 and 20, respectively). The most parsimonious explanation for this is that cases with higher baseline scores are more complex and multifaceted and may require longer or more intensive treatment options, or referral to follow-up services (Ruscio & Holohan, 2006).

For many of those in the RM subsample initially scoring 64 or higher on the PCL-5, the number of individual sessions required *may be 20 or more*. This number is substantially higher than the recommended course of most EBPs (8–12 sessions; Foa et al., 2007; Resick et al., 2008) or the modal number of sessions in an EoC for cases in the RM subsample (mode = 11; 59% had 11 or fewer sessions, 71% had 12 or fewer, and 91% had fewer than 20), suggesting that providers currently practicing in VA PCTs should expect higher scoring patients to require substantially more treatment to reach CSC than is typical during an EoC. The rest of the bands in our cohort received enough sessions to achieve a 75% probability of reaching CSC. The probability of reaching CSC

Table 3

Multilevel Model of Within-Person PCL-5 Scores by Session Number Predicted by Initial PCL-5 Score Band

Model fit information	Model 1	Model 2_f	Model 2_r	Model 3
Log-likelihood	−74816.03	−73024.46	−70864.52	−68919.59
ICC	0.72	0.77	0.80	0.20
DF (Lvl 1)	17,481	17,480	17,480	17,480
DF (Lvl 2)	—	—	—	2,180

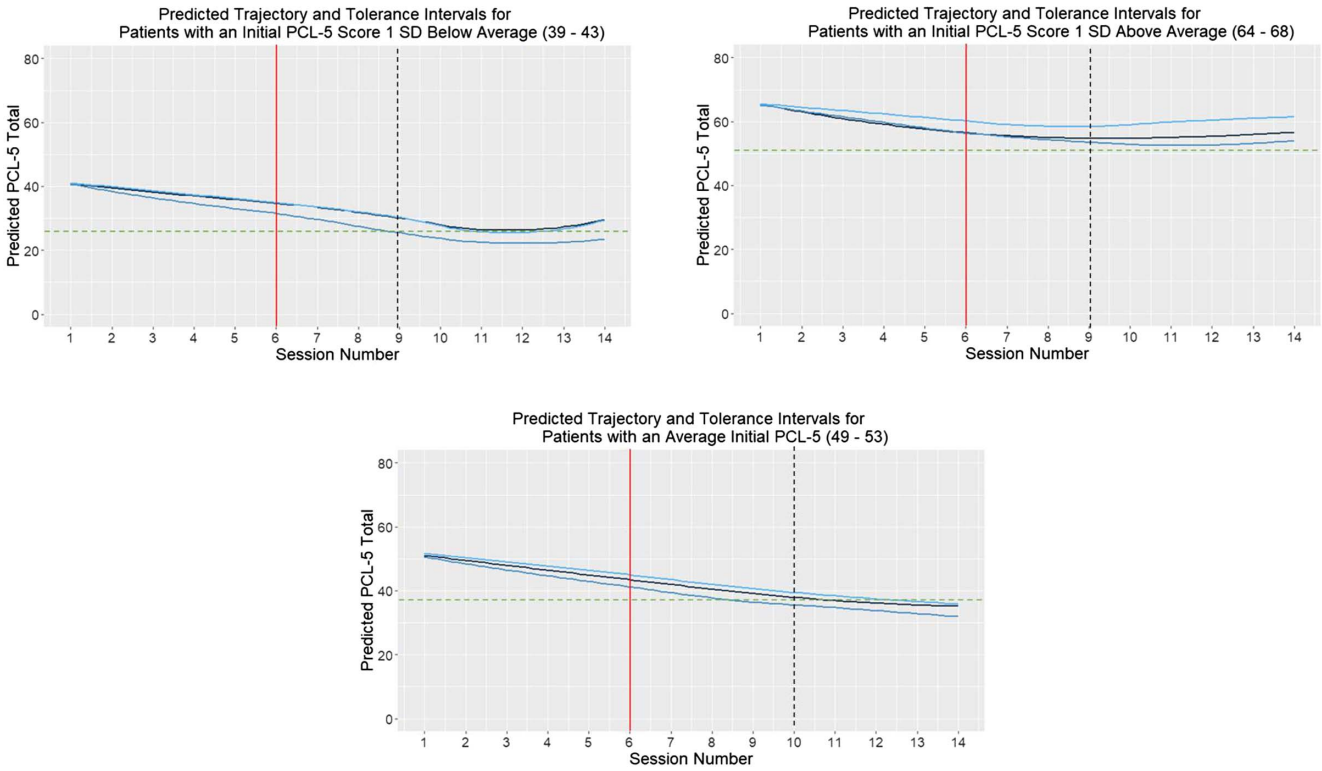
	Random Effects	Fixed Effects			Random Effects	Fixed Effects			Random Effects	Fixed Effects			Random Effects	Fixed Effects		
Effect	[95% CI]	β	*p* value	[95% CI]	[95% CI]	β	*p* value	[95% CI]	[95% CI]	β	*p* value	[95% CI]	[95% CI]	β	*p* value	[95% CI]
Intercept	[14.34, 15.26]	45.98	<.001	[45.35, 46.62]	[14.46, 15.38]	53.41	<.001	[52.73, 54.08]	[12.79, 13.70]	53.48	<.001	[52.88, 54.07]	[3.04, 3.73]	28.83	<.001	[28.83, 29.88]
Session Number	—	—	—	—	—	−12.37	<.001	[−12.76, −11.99]	[15.15, 16.36]	−12.50	<.001	[−13.25, −11.76]	[15.40, 16.65]	−12.61	<.001	[−13.37, −11.86]
Band	—	—	—	—	—	—	—	—	—	—	—	—	—	5.56	<.001	[5.45, 5.66]

	Comparisons		
Model fit information	Model 1 versus Model 2_f	Model 2_f versus Model 2_r	Model 2_r versus Model 3
Log-ratio	3583.15	4319.86	3889.88
p value	<.001	<.001	<.001
DF	4	6	7
PRV Lvl 1	−1.5%	21.14%	93.54%
PRV Lvl 2	—	—	−3.45%

Note. $N = 2{,}182$; PCL-5 = PTSD Checklist for *DSM-5*; ICC = intraclass coefficient; DF = degrees of freedom; Lvl = level; CI = confidence interval; PRV = proportion reduction in unexplained variance; PTSD = posttraumatic stress disorder; *DSM-5* = *Diagnostic and Statistical Manual of Mental Disorders, 5th Edition*.

Figure 2

Expected Trajectories and Tolerance Intervals at the 40th and 50th Percentiles for Cases at, 1 SD Below, and 1 SD Above the Mean Initial PCL-5 Score



Note. SD = standard deviation; PCL-5 = PTSD Checklist for *DSM-5*; PTSD = posttraumatic stress disorder; *DSM-5* = *Diagnostic and Statistical Manual of Mental Disorders, 5th Edition*; Top horizontal curve = 50th percentile tolerance interval; Bottom horizontal curve = 40th percentile tolerance interval. Dotted horizontal line indicates clinically significant change (CSC). At the left (solid) vertical line (session 6), 68% of cases who do not ultimately meet for CSC are above the 50th percentile (83% of those who do ultimately reach CSC will be below) and 80% will be above the 40th percentile (74% of those who do not reach CSC will be below). The right vertical (dashed) line represents the point of diminishing returns. See the online article for the color version of this figure.

was 50% at session 12 or higher for all but Band 2 (which was at Session 11), suggesting that greater than 50% of Veterans receiving treatment in VA PCTs may require more sessions to achieve CSC than is currently typical of care, with higher scoring, potentially more complex patients requiring a greater number of sessions.

With respect to the overall survival curve, the point of diminishing returns was Session 9, however, further exploration revealed that it steadily increased from Session 9 to Session 12 from Band 2 to Band 6, suggesting that more sessions had a greater benefit as initial symptom burden increased. The highest two bands (i.e., Bands 7 and 8) both had points of diminishing returns at Session 9, which may be an artifact of having an insufficient number of sessions for a large proportion of these cases. Until corporate data warehouse data saturation improves estimates (especially for high severity cases), a point of diminishing returns of Sessions 9–12 may be useful as a benchmark for CSC (depending on initial symptom severity), such that the probability of a treatment responder reaching CSC increases at an accelerating rate at each session as treatment approaches the point of diminishing returns, and that rate decelerates at each session following. If this benchmark were to be used in clinical practice, the point of diminishing returns could be considered the session where, if a patient has not met CSC and is not at least "on-track" for expected change, continued treatment without changing approach will unlikely result in CSC. Notably, 9–12 sessions of treatment is the total length of care for the majority of Veterans in our sample and the standard recommended length of frontline EBPs for PTSD (e.g., prolonged exposure therapy; Foa et al., 2007; cognitive processing therapy; Resick et al., 2008). As such, this benchmark may prove most useful for informing the course of extended care, either extended variable length EBPs for PTSD (e.g., up to 15 sessions of prolonged exposure therapy; Morland et al., 2020; up to 24 sessions of cognitive processing therapy; Resick et al., 2021) or an alternative treatment plan. Given the time-limited nature of treatment in PCTs, in practice, information gleaned from actionable MBC may also need to be shared collaboratively with referral sites (e.g., VA general mental health clinics), with the ideal being continued capture and tracking of patient data across clinics.

However, although the point of diminishing returns can be used as a benchmark for when to expect CSC to occur (i.e., CSC is most likely to have occurred by Sessions 9–12), session-by-session benchmarks indicating whether a patient is expected to experience CSC at any point during treatment may be more clinically useful. Session-by-session benchmarks may be useful at earlier sessions

Table 4
Sensitivity and Specificity Analyses Predicting CSC by the End of an EoC

Session Number	40th percentile			50th percentile		
	Sensitivity	Specificity	Combined	Sensitivity	Specificity	Combined
1	.48	.49	0.97	.38	.58	0.96
2	.66	.56	1.22	.55	.66	1.21
3	.70	.62	1.33	.60	.70	1.31
4	.74	.66	1.40	.64	.75	1.39
5	.77	.70	1.47	.66	.79	1.45
6	.80	.74	1.55[a]	.68	.83	1.51[a]
7	.84	.77	1.61[a]	.71	.85	1.56[a]
8	.86	.79	1.65[a]	.73	.86	1.60[a]
9	.89	.79	1.68[a]	.76	.89	1.65[a]
10	.87	.79	1.67[a]	.76	.87	1.63[a]
11	.85	.77	1.62[a]	.77	.87	1.63[a]
12	.83	.76	1.58[a]	.72	.87	1.59[a]
13	.80	.78	1.58[a]	.71	.89	1.60[a]
14	.77	.77	1.54[a]	.70	.87	1.56[a]

Note. CSC = clinically significant change; EoC = episode of care. *Sensitivity* is the probability that a patient who is unlikely to change is to be identified as offtrack. *Specificity* is the probability that a patient who is likely to change is to be identified as on-track.
[a] Combined sensitivity and specificity of at least 1.50 (Power et al., 2013).

to determine whether a patient's score at that session is consistent with that of those who report CSC by the end of treatment. A patient's data falling outside a given session's benchmark would be indicative of probable treatment failure, even before a patient has reached the session representing their point of diminishing returns.

The model we used to generate benchmarks for determining whether a patient is "on-track" for expected change was a good fit to the data, accounting for nearly all clustering and having significant between-person differences in within-person treatment response predicted by initial PTSD symptom severity. The eight expected trajectories (with eight associated initial symptom ranges, see Table 2) we generated were used at each session to predict each case in the RM subsample's ultimate response to treatment, with tolerance intervals based on previously reported estimates of treatment nonresponse (e.g., ~50%, Resick et al., 2017) as well as the nonresponse rate in this sample (~60%). Both intervals had adequate sensitivity and specificity by Session 6, thus suggesting they could be useful as benchmarks for expected change by as early as halfway through a modal EoC or EBP package for PTSD. Unsurprisingly, the interval based on the data set was more sensitive for predicting nonresponse (80% vs. 68%), although the 50th percentile interval miscategorized fewer treatment responders at each session (specificity of 83% vs. 74%). Considering our goal of capturing treatment nonresponders, the 40th percentile tolerance interval may provide more useful benchmarks. For example, an 8-point decrease in PCL-5 total scores by Session 6 may be useful as a benchmark indicative of a patient being "on-track" for reaching CSC in an EoC (i.e., using the 40th percentile tolerance interval, 80% of nonresponders had a <8-point decrease by Session 6). However, more research is needed to assess whether these benchmarks have clinical utility in more specific samples.

There are several noteworthy limitations to this study. First, we were unable to test the accuracy of these trajectories on an independent data set, which may inflate the significance of our findings. A second limitation is that PCL-5 data in the corporate data warehouse is not tied to a specific session (e.g., the second PCL-5 administration may or may not have actually occurred at Session 2). Although the first administration was within 1 week of the first session, the last administration was at minimum four sessions later, and data were only used from the cases that had PCL-5 data for at least half of their sessions, estimates of session numbers were based on a proxy (i.e., PCL-5 administration number), thus potentially introducing unexplained variance to our models. Third, relative to the number of cases that get specialty care for PTSD in VA, the corporate data warehouse has a small percentage of EoC and repeated assessments of outcome. To assess if the models are sufficient for early identification of treatment nonresponse across contexts, or if there are additional considerations that need to be made when predicting treatment response, the predictive accuracy of these models must be tested in an independent data set across a variety of potential moderators (e.g., types of trauma-focused EBP, Schnurr et al., 2022; treatment schedule, Ehlers et al., 2014; age of the patient, Dewar et al., 2020).

Relatedly, VISN was found to be a significant predictor of the survival curve, but many VISNs had very limited outcome data that precluded including VISN in all other analyses. Within VA, VISNs are geographic regions which may be useful as a proxy for regional differences in demographics and resources. These contextual differences are likely to influence PTSD treatment, and previous research has demonstrated treatment location is a significant moderator of the dose–response relationship (Lambert et al., 2001). As such, the applicability of all generated benchmarks may initially depend on where they are implemented until such a time when the corporate data warehouse houses sufficient data. Further, therapist-level data were not available for this study, thus the effect of clustering within therapist could not be determined or accounted for in the final model.

Finally, the wide range of variability in initial scores for the lowest and highest ranges (i.e., Bands 1 and 8) affects the validity of these trajectories. For example, a patient with an initial score of 17 or 80 may be underrepresented by the applied trajectory; predicted trajectories and tolerance intervals may be somewhat lower or higher due to the inclusion of relatively few cases with exceptionally low or high scores. Thus, the prediction for cases using

these wide bands may be less accurate than for the narrower bands in the middle of the symptom severity spectrum (e.g., the proportion of cases that are misidentified as likely/unlikely to meet CSC may be skewed to scores on one side of the initial PCL-5 score bands). As routine outcome monitoring becomes more routine and a higher percentage of cases with repeated assessments are captured in the corporate data warehouse, we can generate valid expected trajectories in the narrower bands with greater sensitivity and specificity for predicting change. We decided not to exclude exceptionally low (i.e., "subthreshold") or high cases from this report because we did not want to ignore these bands and inadvertently discourage or de-incentivize providers to track patient progress and capture those data in these instances. If that happened, there would be limited data to refine trajectories for cases on the lower or higher end of the symptom severity spectrum, but who are still seeking and receiving treatment from PCTs, as evidenced by the existence of such data in the corporate data warehouse.

While it remains to be seen if these benchmarks have predictive utility in specific samples or clinics, or if providing this information to clinicians in PCTs will improve patient outcomes (and patient satisfaction), the results of this study provide a foundation for the continued refinement of a feedback system for PTSD specialty care, and thus begin to realize the full potential of MBC. Research has demonstrated robust improvements in outcomes across studies when measurement feedback systems are used in practice (e.g., Lambert et al., 2018; Lutz et al., 2019). The VA needs to generate and test a feedback system designed to meet the needs of Veterans receiving specialty care for PTSD. Once our findings are extended and replicated, expected trajectories of change could be fielded to improve shared decision-making and to improve outcome.

References

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)*.

Benfer, N., Rusowicz-Orazem, L., Fielstein, E. M., Darnell, B. C., Frankfurt, S. B., Mignogna, J., & Litz, B. T. (2022). Leveraging available metadata in VA PTSD clinics and generating benchmarks for clinically significant change. *Journal of Consulting and Clinical Psychology*. Advance online publication. https://doi.org/10.1037/ccp0000779

Bensing, J. (2000). Bridging the gap. The separate worlds of evidence-based medicine and patient-centered medicine. *Patient Education and Counseling*, *39*(1), 17–25. https://doi.org/10.1016/S0738-3991(99)00087-7

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

Dewar, M., Paradis, A., & Fortin, C. A. (2020). Identifying trajectories and predictors of response to psychotherapy for post-traumatic stress disorder in adults: A systematic review of literature. *Canadian Journal of Psychiatry*, *65*(2), 71–86. https://doi.org/10.1177/0706743719875602

Diener, E., Northcott, R., Zyphur, M. J., & West, S. G. (2022). Beyond Experiments. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *17*(4), 1101–1119. https://doi.org/10.1177/17456916211037670

Ehlers, A., Hackmann, A., Grey, N., Wild, J., Liness, S., Albert, I., Deale, A., Stott, R., & Clark, D. M. (2014). A randomized controlled trial of 7-day intensive and standard weekly cognitive therapy for PTSD and emotion-focused supportive therapy. *The American Journal of Psychiatry*, *171*(3), 294–304. https://doi.org/10.1176/appi.ajp.2013.13040552

Engle, R. L., Mohr, D. C., Holmes, S. K., Seibert, M. N., Afable, M., Leyson, J., & Meterko, M. (2021). Evidence-based practice and patient-centered care: Doing both well. *Health Care Management Review*, *46*(3), 174–184. https://doi.org/10.1097/HMR.0000000000000254

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

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*, 393–405. https://doi.org/10.1007/BF02287427

Hansen, N. B., Lambert, M. J., & Forman, E. M. (2002). The psychotherapy dose–response effect and its implications for treatment delivery services. *Clinical Psychology: Science and Practice*, *9*(3), 329–343. https://doi.org/10.1093/clipsy.9.3.329

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

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

Kahn, J. H. (2011). Multilevel modeling: Overview and applications to research in counseling psychology. *Journal of Counseling Psychology*, *58*(2), 257–271. https://doi.org/10.1037/a0022680

Lambert, M. J. (2013). Outcome in psychotherapy: The past and important advances. *Psychotherapy*, *50*(1), 42–51. https://doi.org/10.1037/a0030682

Lambert, M. J., Gregersen, A. T., & Burlingame, G. M. (2004). The Outcome Questionnaire-45. In M. E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcomes assessment: Instruments for adults* (pp. 191–234). Lawrence Erlbaum.

Lambert, M. J., Hansen, N. B., & Finch, A. E. (2001). Patient-focused research: Using patient outcome data to enhance treatment effects. *Journal of Consulting and Clinical Psychology*, *69*(2), 159–172. https://doi.org/10.1037/0022-006X.69.2.159

Lambert, M. J., Whipple, J. L., & Kleinstäuber, M. (2018). Collecting and delivering progress feedback: A meta-analysis of routine outcome monitoring. *Psychotherapy*, *55*(4), 520–537. https://doi.org/10.1037/pst0000167

Lambert, M. J., & Vermeersch, D. A. (1994). The effectiveness of psychotherapy. In M. Hersen & W. Sledge (Eds.), *Encyclopedia of psychotherapy* (Vol. 1, pp. 709–714). Acedemic Press.

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

Lutz, W., Martinovich, Z., & Howard, K. I. (1999). Patient profiling: An application of random coefficient regression models to depicting the response of a patient to outpatient psychotherapy. *Journal of Consulting and Clinical Psychology*, *67*(4), 571–577. https://doi.org/10.1037/0022-006X.67.4.571

Lutz, W., Rubel, J. A., Schwartz, B., Schilling, V., & Deisenhofer, A.-K. (2019). Towards integrating personalized feedback research into clinical practice: Development of the Trier Treatment Navigator (TTN). *Behaviour Research and Therapy*, *120*, Article 103438. https://doi.org/10.1016/j.brat.2019.103438

MacCallum, R. C., Kim, C., Malarkey, W. B., & Kiecolt-Glaser, J. K. (1997). Studying multivariate change using multilevel models and latent curve models. *Multivariate Behavioral Research*, *32*(3), 215–253. https://doi.org/10.1207/s15327906mbr3203_1

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*, *53*(2), 419–428. https://doi.org/10.1017/S0033291721001628

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

Morland, L. A., Mackintosh, M. A., Glassman, L. H., Wells, S. Y., Thorp, S. R., Rauch, S. A. M., Cunningham, P. B., Tuerk, P. W., Grubbs, K. M., Golshan, S., Sohn, M. J., & Acierno, R. (2020). Home-based delivery of variable length prolonged exposure therapy: A comparison of clinical efficacy between service modalities. *Depression and Anxiety*, *37*(4), 346–355. https://doi.org/10.1002/da.22979

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

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*. Department of Veterans Affairs.

Power, M., Fell, G., & Wright, M. (2013). Principles for high-quality, high-value testing. *Evidence-Based Medicine*, *18*(1), 5–10. https://doi.org/10.1136/eb-2012-100645

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., Monson, C. M., & Chard, K. M. (2008). *Cognitive processing therapy. Veteran/Military Version*. Department of Veterans' Affairs.

Resick, P. A., Wachen, J. S., Dondanville, K. A., LoSavio, S. T., Young-McCaughan, S., Yarvis, J. S., Pruiksma, K. E., Blankenship, A., Jacoby, V., Peterson, A. L., Mintz, J., & the STRONG STAR Consortium. (2021). Variable-length Cognitive Processing Therapy for posttraumatic stress disorder in active duty military: Outcomes and predictors. *Behaviour Research and Therapy*, *141*, Article 103846. https://doi.org/10.1016/j.brat.2021.103846

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 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

Roberts, J. K., Monaco, J. P., Stovall, H., & Foster, V. (2011). Explained variance in multilevel models. In J. J. Hox & J. K. Roberts (Eds.), *Handbook of advanced multilevel analysis* (pp. 219–230). Routledge/Taylor & Francis Group.

Ruscio, A. M., & Holohan, D. R. (2006). Applying empirically supported treatments to complex cases: Ethical, empirical, and practical considerations. *Clinical Psychology: Science and Practice*, *13*(2), 146–162. https://doi.org/10.1111/j.1468-2850.2006.00017.x

Sackett, D. L., Rosenberg, W. M., Gray, J. A., Haynes, R. B., & Richardson, W. S. (1996). Evidence based medicine: What it is and what it isn't. *The BMJ*, *312*(7023), 71–72. https://doi.org/10.1136/bmj.312.7023.71

Schnurr, P. P., Chard, K. M., Ruzek, J. I., Chow, B. K., Resick, P. A., Foa, E. B., Marx, B. P., Friedman, M. J., Bovin, M. J., Caudle, K. L., Castillo, D., Curry, K. T., Hollifield, M., Huang, G. D., Chee, C. L., Astin, M. C., Dickstein, B., Renner, K., Clancy, C. P., … Shih, M. C. (2022). Comparison of prolonged exposure vs cognitive processing therapy for treatment of posttraumatic stress disorder among US veterans: A randomized clinical trial. *JAMA Network Open*, *5*(1), Article e2136921. https://doi.org/10.1001/jamanetworkopen.2021.36921

Schottenbauer, M. A., Glass, C. R., Arnkoff, D. B., Tendick, V., & Gray, S. H. (2008). Nonresponse and dropout rates in outcome studies on PTSD: Review and methodological considerations. *Psychiatry*, *71*(2), 134–168. https://doi.org/10.1521/psyc.2008.71.2.134

Shiner, B., Leonard Westgate, C., Simiola, V., Thompson, R., Schnurr, P. P., & Cook, J. M. (2018). Measuring use of evidence-based psychotherapy for PTSD in VA residential treatment settings with clinician survey and electronic medical record templates. *Military Medicine*, *183*(9–10), e539–e546. https://doi.org/10.1093/milmed/usy008

Sripada, R. K., Ready, D. J., Ganoczy, D., Astin, M. C., & Rauch, S. A. M. (2020). When to change the treatment plan: An analysis of diminishing returns in VA patients undergoing prolonged exposure and cognitive processing therapy. *Behavior Therapy*, *51*(1), 85–98. https://doi.org/10.1016/j.beth.2019.05.003

Steenkamp, M. M., Litz, B. T., Hoge, C. W., & Marmar, C. R. (2015). Psychotherapy for military-related PTSD: A review of randomized clinical trials. *JAMA*, *314*(5), 489–500. https://doi.org/10.1001/jama.2015.8370

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

Turkson, A. J., Ayiah-Mensah, F., & Nimoh, V. (2021). Handling censoring and censored data in survival analysis: A standalone systematic literature review. *International Journal of Mathematics and Mathematical Sciences*, *2021*, Article 9307475. https://doi.org/10.1155/2021/9307475

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)*. U.S. Department of Veterans Affairs. https://www.ptsd.va.gov/

Received September 6, 2022
Revision received February 6, 2023
Accepted February 9, 2023 ■