The STarT Back Screening Tool and Individual Psychological Measures: Evaluation of Prognostic Capabilities for Low Back Pain Clinical Outcomes in Outpatient Physical Therapy Settings

Jason M. Beneciuk, Mark D. Bishop, Julie M. Fritz, Michael E. Robinson, Nabih R. Asal, Anne N. Nisenzon, Steven Z. George

Background. Psychologically informed practice emphasizes routine identification of modifiable psychological risk factors being highlighted.

Objective. The purpose of this study was to test the predictive validity of the STarT Back Screening Tool (SBT) in comparison with single-construct psychological measures for 6-month clinical outcomes.

Design. This was an observational, prospective cohort study.

Methods. Patients (n=146) receiving physical therapy for low back pain were administered the SBT and a battery of psychological measures (Fear-Avoidance Beliefs Questionnaire physical activity scale and work scale [FABQ-PA and FABQ-W, respectively], Pain Catastrophizing Scale [PCS], 11-item version of the Tampa Scale of Kinesiophobia [TSK-11], and 9-item Patient Health Questionnaire [PHQ-9]) at initial evaluation and 4 weeks later. Treatment was at the physical therapist’s discretion. Clinical outcomes consisted of pain intensity and self-reported disability. Prediction of 6-month clinical outcomes was assessed for intake SBT and psychological measure scores using multiple regression models while controlling for other prognostic variables. In addition, the predictive capabilities of intake to 4-week changes in SBT and psychological measure scores for 6-month clinical outcomes were assessed.

Results. Intake pain intensity scores (β=.39 to .45) and disability scores (β=.47 to .60) were the strongest predictors in all final regression models, explaining 22% and 24% and 43% and 48% of the variance for the respective clinical outcome at 6 months. Neither SBT nor psychological measure scores improved prediction of 6-month pain intensity. The SBT overall scores (β=.22) and SBT psychosocial scores (β=.25) added to the prediction of disability at 6 months. Four-week changes in TSK-11 scores (β=−.18) were predictive of pain intensity at 6 months. Four-week changes in FABQ-PA scores (β=−.21), TSK-11 scores (β=−.20) and SBT overall scores (β=−.18) were predictive of disability at 6 months.

Limitations. Physical therapy treatment was not standardized or accounted for in the analysis.

Conclusions. Prediction of clinical outcomes by psychology-based measures was dependent upon the clinical outcome domain of interest. Similar to studies from the primary care setting, initial screening with the SBT provided additional prognostic information for 6-month disability and changes in SBT overall scores may provide important clinical decision-making information for treatment monitoring.
In the United States, low back pain (LBP) point prevalence estimates range from 17% to 26%, with chronic LBP estimated at approximately 10%.¹⁻³ Recent North Carolina data indicate that the prevalence of acute and chronic LBP has increased 44% and 162%, respectively, over a 14-year period.⁴ Furthermore, approximately 25% of the caseload in outpatient physical therapist practice is LBP related.⁵⁻⁷ Direct costs associated with health care services for LBP are enormous, with physical therapy services accounting for a large portion of these costs.⁸⁻⁹ However, only a small percentage of patients with chronic LBP may account for a large fraction of associated costs.¹⁰⁻¹¹

Early risk factor screening has been implicated as one strategy to identify patients who may be at risk for poor clinical outcomes and as a potential method to improve the efficiency and effectiveness of care.¹²⁻¹⁴ In physical therapy, “psychologically informed practice” has been presented as a secondary prevention approach for chronic LBP that integrates both biomedical (focused on pathology or physical impairments) and cognitive-behavioral (focused on psychological distress or behavior) principles.¹⁵ The primary goal of a psychologically informed approach is to prevent future LBP-associated disability, with routine identification of modifiable psychological risk factors being emphasized. Specific to this current study, determining the validity and clinical utility of commonly used psychological screening measures has been indicated as a top future research priority for psychologically informed practice.¹⁵

There are unresolved issues relevant to psychologically informed practice involving the value of information obtained during the assessment of modifiable psychological factors. First, the design of self-report psychological questionnaires ranges from those focusing on a single construct that measure a specific psychological factor to those that utilize a multiple-construct approach and measure overall psychological distress. Each of these approaches is associated with strengths and weaknesses. For example, a potential weakness in using single-construct questionnaires is that they do not provide information beyond the targeted psychological factors of interest. Although more feasible for use in clinical settings, multiple-construct questionnaires are frequently used as prognostic indicators for future outcomes based on overall psychological distress, which is a potential strength; however, they do not provide detailed information on specific psychological factors that may serve as behavioral treatment targets.

Second, information obtained during psychological assessment has potential to vary based on the timing (eg, pretreatment, over an episode of care, posttreatment) and number of repeated assessments.¹⁶⁻²¹ with some studies indicating that changes in psychological risk factors may improve the prediction of clinical outcomes compared with pretreatment assessments.¹⁸⁻²¹ The clinical relevance of these findings is that changes in psychological risk factors have potential not only for use as prognostic indicators administered at initial patient encounter, but also for treatment monitoring during a patient encounter, specifically associated with psychologically oriented interventions.²²⁻²³ Although potentially appealing, using a single measure for multiple purposes (eg, prognostic screening and treatment monitoring) is not always appropriate, as many screening measures were developed to be used only as brief triage instruments.

Finally, the prediction of clinical outcomes by psychologically based...
measures is likely to be dependent on the clinical outcome domain of interest.24,25 Combined, these are important issues because information about specific psychological factors may be needed to implement psychologically oriented interventions that have potential to reduce the likelihood of developing chronic LBP. However, assessing psychological factors may increase the burden on clinicians, staff, and patients. An optimal clinical scenario, therefore, would consist of early screening with a brief multiple-construct risk prediction instrument to determine which patients require further detailed assessment via individual single-construct questionnaires.13,15,23

Hill et al26 developed the STarT Back Screening Tool (SBT), a 9-item screening measure used to identify subgroups of patients with LBP in primary care settings based on the presence of physical or modifiable psychosocial prognostic factors (or both), which may be useful in matching patients with targeted interventions. The SBT has demonstrated predictive validity for long-term disability outcomes for patients with LBP in primary care settings.26 Furthermore, recent findings from a clinical trial have indicated that a stratified primary care management approach for LBP based on SBT subgroup allocation was associated with improved clinical and economic outcomes compared with current best practice.27 The SBT was developed and intended for use in primary care settings, so information about utilizing the SBT in outpatient physical therapy settings is limited.28 Moreover, the SBT was designed as a screening tool and has previously been implemented in primary care26 and outpatient physical therapy28 settings. Its usefulness for treatment monitoring, however, has only recently been evaluated in primary care settings.29 Therefore, the specific purpose of our study were: (1) to test the predictive validity of the SBT in comparison with commonly used single-construct psychological screening measures for 6-month clinical outcomes and (2) to test the value of psychological information measured at different points in time (ie, intake or 4-week changes) in predicting 6-month clinical outcomes. Investigation of these purposes will allow us to make recommendations about optimal screening methods for identifying psychological distress in patients with LBP seeking treatment in outpatient physical therapy settings.

Method
Data for this observational, prospective cohort study were collected between December 14, 2009, and February 5, 2012, from 4 outpatient physical therapy clinics of Brooks Rehabilitation located in Jacksonville, Florida, and 2 outpatient physical therapy clinics of Shands Rehabilitation located in Gainesville, Florida.
Participants
Consecutive patients seeking treatment for LBP at 6 participating outpatient physical therapy clinics were screened for study eligibility by a physical therapist. Potential study participants met both of the following criteria before being enrolled into this study: (1) adults between the ages of 18 and 65 years seeking physical therapy for LBP (defined as having symptoms at T12 or lower, including radiating pain into the buttocks and lower extremity) and (2) the ability to read and speak the English language. These broad inclusion criteria were chosen to allow for a cohort that was applicable to clinical practice. Potential study participants were ineligible to participate in this study if any of the following criteria were met: (1) the presence of systemic involvement related to metastatic or visceral disease, (2) recent spinal fracture, (3) osteoporosis, or (4) pregnancy. Physical therapists provided all patients who met study eligibility criteria with a brief explanation of the study and a study advertisement with primary investigator contact information. Clinicians emphasized to patients that participating in this study would not dictate the treatment they received for their LBP and that if they elected not to participate, they would receive the same treatment. Informed consent was obtained in compliance with the University of Florida’s Institutional Review Board.

Demographic and Historical Information
Study participants were asked to complete a standardized self-report questionnaire consisting of questions related to age, sex, race, ethnicity, education, household income, and marital and employment status. Additionally, information involving LBP clinical characteristics (ie, prior surgery, symptom duration, symptom onset, symptom location, work-related LBP) was obtained.

Psychological Measures
Data for psychological measures were collected at baseline, at 4 weeks, and at 6 months. These measures are described in more detail below.

SBT. The primary measure of interest for this study was the SBT. The SBT contains items related to physical and psychosocial factors that have been identified as strong independent predictors for persistent disabling LBP. The SBT overall scores (ranging from 0 to 9) are determined by summing all positive responses, and the SBT psychosocial subscale scores (ranging from 0 to 5) are determined by summing items related to bothersomeness, fear, catastrophizing, anxiety, and depression. Based on overall and psychosocial subscale scoring, the SBT categorizes patients as “high risk” (psychosocial subscale scores ≥4), in which high levels of psychosocial prognostic factors are present with or without physical factors present; “medium risk” (overall score >3, psychosocial subscale score <4), in which physical and psychosocial factors are present but not a high level of psychosocial factors; or “low risk” (overall score 0–3), in which few prognostic factors are present.

Utilization of SBT overall and psychosocial subscale scoring has demonstrated acceptable test-retest reliability and internal consistency.26 The SBT subgroup cutoff scores have demonstrated good predictive validity for dichotomized 6-month disability outcomes in low-risk (16.7%), medium-risk (53.2%), and high-risk (78.4%) subgroups in primary care settings.26 The SBT overall scores have demonstrated acceptable to outstanding discriminant validity for physical reference standards (eg, disability, referred leg pain), and the SBT psychosocial subscale scores best discriminated psychosocial reference standards (eg, catastrophizing, fear, depression).26 The SBT has demonstrated concurrent validity in comparison with the Örebro Musculoskeletal Pain Screening Questionnaire, in which both multiple-construction instruments displayed similar subgroup characteristics and the ability to discriminate for disability, catastrophizing, fear, comorbid pain, and time off work reference standards.31

Fear-Avoidance Beliefs Questionnaire (FABQ). Fear-avoidance beliefs specific to LBP were assessed with the FABQ.32 The FABQ consists of a 4-item FABQ physical activity scale (FABQ-PA, scores potentially ranging from 0 to 24) and a 7-item FABQ work scale (FABQ-W, scores potentially ranging from 0 to 42), with higher scores indicating higher levels of fear-avoidance beliefs for both FABQ scales. Both FABQ scales have been found to have acceptable reliability32–35 and demonstrated internal consistency.32,35–38 The FABQ-W has demonstrated predictive validity for disability and work loss in patients with LBP.35,59–41

Pain Catastrophizing Scale (PCS). The PCS was used to assess the degree of catastrophic cognitions due to LBP.42 Pain catastrophizing has been broadly defined as an exaggerated negative orientation toward actual or anticipated pain experiences.42 The PCS is a 13-item questionnaire with a potential range of scores from 0 to 52, with higher scores indicating higher levels of pain catastrophizing. The PCS has been found to have good test-retest reliability42 and internal consistency.43–45 The PCS also has been found to demonstrate several different types of validity.42–45
The STarT Back Screening Tool and Individual Psychological Measures

11-Item Version of the Tampa Scale of Kinesiophobia (TSK-11). The TSK-11 was used to assess the degree of fear of movement and injury or reinjury in individuals with LBP.46 The TSK-11 is an 11-item questionnaire with a potential range of scores from 11 to 44, with higher scores indicating greater fear of movement and injury or reinjury due to pain. The TSK-11 has been found to have good test-retest reliability and internal consistency.46 Predictive and concurrent validity also have been reported for the TSK-11.46

9-Item Patient Health Questionnaire (PHQ-9). The PHQ-9 was used to assess the degree to which depressive symptoms have on patients with LBP. The PHQ-9 is a 9-item questionnaire with a potential range of scores from 0 to 27, with higher scores indicating elevated depressive symptoms. The PHQ-9 has demonstrated various types of validity in different health care settings.47,48 and has been used in studies involving patients with LBP.49

Outcome Measures
Data for outcome measures were collected at baseline, at 4 weeks, and at 6 months. These measures are described in more detail below:

Numerical pain rating scale (NPRS). Pain intensity was rated using an NPRS, ranging from “0” (no pain) to “10” (worst pain imaginable).50–52 Participants were asked to rate their current pain intensity, as well as their best and worst levels of pain intensity over the past 24 hours. These 3 pain ratings were averaged and used as the NPRS variable in this study.53

Revised Oswestry Disability Questionnaire (ODQ). Disability related to LBP was assessed with a revised version of the ODQ, which has 10 items that assess how LBP affects common daily activities.54,55

The ODQ has a range of 0% (“no disability due to LBP”) to 100% (“completely disabled due to LBP”), with higher scores indicating higher disability from LBP. The ODQ has been found to have high levels of test-retest reliability, internal consistency, validity, and responsiveness.55–57

Data Analysis
All data analyses were performed using SPSS version 20.0 (SPSS Inc, Chicago, Illinois). Means and standard deviations were calculated for all baseline continuous variables, and frequency counts with percentages were calculated for categorical variables. These descriptive statistics are presented for the entire study sample at each assessment time point (ie, intake, 4 weeks, and 6 months). The distributions of baseline continuous variables were examined by visual inspection of histograms and by calculating skewness and kurtosis statistics. Nonparametric estimates were used for variables with significant deviations from a normal distribution based on visual inspection and skewness and kurtosis statistics. Univariate relationships between SBT (overall and psychosocial) scores and single-construct psychological measures (using either intake scores or 4-week change scores) are presented as Pearson correlation coefficients (r) for continuous variables.

Predictive Validity Using 4-Week Change Scores: SBT and Single-Construct Psychological Measures for 6-Month Clinical Outcomes
For these analyses, we used similar regression models as previously described, with the only exception being that we used intake to 4-week change scores for SBT and individual single-construct psychological measures instead of their respective intake scores alone. Specifically, for multiple regression analyses, model 3 added either SBT (overall or psychosocial) or individual single-construct psychological measure (FABQ-PA, FABQ-W, PCS, TSK-11, or PHQ-9) intake to 4-week change scores to model 2. Regression diagnostics were performed to assess for multicollinearity between predictor variables in all multiple regression analyses.

Predictive Validity Using Intake Scores: SBT and Single-Construct Psychological Measures for 6-Month Clinical Outcomes
Univariate analyses consisted of separate simple linear regression models that were used to evaluate the independent contribution of intake SBT (overall or psychosocial) and individual single-construct psychological measure (FABQ-PA, FABQ-W, PCS, TSK-11, or PHQ-9) scores as predictors of NPRS and ODQ scores at 6 months. Then separate multiple regression models were used to evaluate the relative contributions of intake SBT and individual single-construct psychological measures as predictors of NPRS and ODQ scores at 6 months. Three separate models were created for each SBT and individual single-construct psychological measure. Model 1 accounted for intake NPRS or ODQ scores depending upon the outcome of interest, model 2 added demographic and clinical variables (age, sex, symptom duration, and surgery for current condition) to model 1, and model 3 added either SBT (overall or psychosocial score) or individual single-construct psychological measure (FABQ-PA, FABQ-W, PCS, TSK-11, or PHQ-9) scores to model 2. Regression diagnostics were performed to assess for multicollinearity between predictor variables in all multiple regression analyses.
Results
During the study period, 275 patients were screened for eligibility criteria (Figure). Of these patients, 123 were excluded from study participation, with the most common reason being that they were greater than 65 years of age (n=47). The remaining 152 patients provided informed consent and were enrolled into the study. Of these patients, 6 were not able to complete the study due to personal reasons. Therefore, baseline data was obtained from 146 patients, 4-week follow-up data were obtained from 128 patients (88%) either in person (n=107) or through mail (n=21), and 6-month follow-up data were obtained from 111 patients (76%) either in person (n=61) or through mail (n=50).

Baseline demographic data for entire sample are presented in Table 1. Approximate normal distribution for initial pain intensity, disability, and individual single-construct psychological measures was suggested based on visual inspection of histograms and normality plots.

The average number of physical therapy visits at 4 weeks was 6.91 (SD=2.64, range=1-13) where 102 participants (79.7%) were still receiving physical therapy, 16 (12.5%) completed physical therapy and were discharged, and 10 (7.8%) elected not to continue with physical therapy. Independent-samples t tests were used to compare participants who completed the 4-week follow-up assessment (n=128) with those who did not (n=18) on demographic, clinical, and psychological characteristics at initial evaluation. Results indicated that completers were older (mean difference=9.4 years, standard error=3.3) and reported higher PHQ-9 scores (mean difference=2.9, standard error=1.1) compared with noncompleters (P<.05). The average number of physical therapy visits at 6 months was 11.40 (SD=6.51, range=2-35) where no participants were still receiving physical therapy, 72 (66.7%) completed physical therapy and were discharged, and 36 (33.3%) elected not to continue with physical therapy. Independent-samples t tests also were used to compare participants that completed the 6-month follow-up assessment

| Patients with low back pain screened for eligibility during this study period (n=275) |
| Patients enrolled in study after obtaining informed consent (n=152) |
| Patients included in intake analysis (n=146) |
| Patients included in analysis of 4-week outcomes (n=128)  
  • in-person (n=107)  
  • mail (n=21) |
| Patients included in analysis of 6-month outcomes (n=111)  
  • in-person (n=61)  
  • mail (n=50) |
| Patients screened and not eligible due to:  
  • >65 years of age (n=47)  
  • language barrier or illiterate (n=2)  
  • cancer (n=3)  
  • pregnant (n=2) |
| Patients screened and eligible but not enrolled due to:  
  • refusal to participate (n=19)  
  • never completed informed consent process (n=24)  
  • never returned to physical therapy (n=26) |
| Patients dropped from study due to personal reasons (n=6) |

Figure.
Flow diagram of patients through the study.
(n=111) with those who did not (n=35) on demographic, clinical, and psychological characteristics at initial evaluation. Results indicated that completers were older (mean difference=8.9 years, standard error=2.7) compared with noncompleters \((P<.05)\). Based on initial SBT risk category status, there were no differences in number of visits, follow-up rates, methods of follow-up, or current physical therapy status at 4 weeks or 6 months \((P>.05)\).

Predictive Validity Using Intake Scores: SBT and Single-Construct Psychological Measures for 6-Month Clinical Outcomes

Intake univariate associations.

As expected, initial single-construct psychological measures were signifi-
The STarT Back Screening Tool and Individual Psychological Measures

Table 2.
Prediction of 6-Month Pain Intensity (NPRS) Scores: Multiple Regression Models

<table>
<thead>
<tr>
<th>Psychological Variable (Intake Score)</th>
<th>Intake NPRS Scores</th>
<th>Model 1: Demographic and LBP Clinical Variables</th>
<th>Change From Model 1 to Model 2: Additional Variance Explained by Demographic and LBP Clinical Variables</th>
<th>Model 2: Intake Psychological Measure</th>
<th>Change From Model 2 to Model 3: Additional Variance Explained by Intake Psychological Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>FABQ-PA</td>
<td>% R² = 18.2</td>
<td>% R² = 22.3, adjusted % R² = 17.4</td>
<td>% R² = 4.0, adjusted % R² = 17.4, p &lt; .05</td>
<td>% R² = 22.3, adjusted % R² = 17.4, β = .01</td>
<td>% R² = 0.0, adjusted % R² = 17.4, p &lt; .05</td>
</tr>
<tr>
<td>FABQ-W</td>
<td>% R² = 18.2</td>
<td>% R² = 22.3, adjusted % R² = 17.4</td>
<td>% R² = 4.0, adjusted % R² = 17.4, p &lt; .05</td>
<td>% R² = 23.5, adjusted % R² = 18.9, β = .12</td>
<td>% R² = 1.2, adjusted % R² = 17.4, p &lt; .05</td>
</tr>
<tr>
<td>PCS</td>
<td>% R² = 18.2</td>
<td>% R² = 22.3, adjusted % R² = 17.4</td>
<td>% R² = 4.0, adjusted % R² = 17.4, p &lt; .05</td>
<td>% R² = 22.3, adjusted % R² = 17.4, β = .02</td>
<td>% R² = 0.0, adjusted % R² = 17.4, p &lt; .05</td>
</tr>
<tr>
<td>TSK-11</td>
<td>% R² = 18.2</td>
<td>% R² = 22.3, adjusted % R² = 17.4</td>
<td>% R² = 4.0, adjusted % R² = 17.4, p &lt; .05</td>
<td>% R² = 22.5, adjusted % R² = 17.8, β = .04</td>
<td>% R² = 0.2, adjusted % R² = 17.4, p &lt; .05</td>
</tr>
<tr>
<td>PHQ-9</td>
<td>% R² = 19.1</td>
<td>% R² = 22.7, adjusted % R² = 18.8</td>
<td>% R² = 3.6, adjusted % R² = 18.8, p &lt; .05</td>
<td>% R² = 23.3, adjusted % R² = 18.6, β = .09</td>
<td>% R² = 0.6, adjusted % R² = 18.6, p &lt; .05</td>
</tr>
<tr>
<td>SBT overall score</td>
<td>% R² = 18.2</td>
<td>% R² = 22.3, adjusted % R² = 17.4</td>
<td>% R² = 4.0, adjusted % R² = 17.4, p &lt; .05</td>
<td>% R² = 22.4, adjusted % R² = 17.7, β = .03</td>
<td>% R² = 0.1, adjusted % R² = 17.7, p &lt; .05</td>
</tr>
<tr>
<td>SBT psychosocial score</td>
<td>% R² = 18.2</td>
<td>% R² = 22.3, adjusted % R² = 17.4</td>
<td>% R² = 4.0, adjusted % R² = 17.4, p &lt; .05</td>
<td>% R² = 22.8, adjusted % R² = 8.2, β = .08</td>
<td>% R² = 0.5, adjusted % R² = 8.2, p &lt; .05</td>
</tr>
</tbody>
</table>

*NPRS = numerical pain rating scale, LBP = low back pain, FABQ-PA = Fear-Avoidance Beliefs Questionnaire physical activity scale, FABQ-W = Fear-Avoidance Beliefs Questionnaire work scale, PCS = Pain Catastrophizing Scale, TSK-11 = 11-item version of the Tampa Scale for Kinesiophobia, PHQ-9 = 9-item Patient Health Questionnaire, SBT = STarT Back Screening Tool.

...correlated with each other (r = .28 to .57, p < .01) and with initial SBT overall (r = .28 to .63, p < .01) and SBT psychosocial (r = .20 to .62, p < .01) scores. Initial single-construct psychological measure scores that were most strongly correlated with intake SBT scores were obtained for the PCS, TSK-11, and PHQ-9.

**Pain intensity.** In the univariate analysis, initial SBT (overall and psychosocial) and single-construct psychological measure scores accounted for between 4.5% and 8.4% of the variance in 6-month NPRS scores (p < .05), with the exception of FABQ-PA and TSK-11 scores (p > .05) (results not shown). In the multiple regression analysis, for each linear regression model, intake NPRS scores explained a large amount of variability in 6-month NPRS scores (range = 18.2%–19.1%) (Tab. 2, model 1). The addition of demographic and clinical variables did not significantly explain any additional variability in 6-month NPRS scores (Tab. 2, change from model 1 to model 2). The addition of intake SBT or single-construct psychological measure scores also did not significantly explain any additional variability in 6-month NPRS scores (Tab. 2, change from model 2 to model 3).

**Disability.** In the univariate analysis, initial SBT (overall and psychosocial) and single-construct psychological measure scores accounted for between 5.5% and 26.6% of the variance in 6-month ODQ scores (p < .05) (results not shown). In the multiple regression analysis, for each linear regression model, intake ODQ scores explained a large amount of variability in 6-month ODQ scores (range = 38.3%–38.5%) (Tab. 3, model 1). The addition of demographic and clinical variables did not significantly explain any additional variability in 6-month ODQ scores (Tab. 3, change from model 1 to model 2). Only the addition of intake SBT overall (2.5%) or psychosocial (4.5%) scores explained an additional amount of variability in 6-month ODQ scores (Tab. 3, change...
Table 3.
Prediction of 6-Month Disability (ODQ) Scores: Multiple Regression Models

<table>
<thead>
<tr>
<th>Psychological Variable (Intake Score)</th>
<th>Intake ODQ Scores</th>
<th>Model 1 + Demographic and LBP Clinical Variables</th>
<th>Additional Variance Explained by Demographic and LBP Clinical Variables</th>
<th>Model 2 + Intake Psychological Measure</th>
<th>Additional Variance Explained By Intake Psychological Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>FABQ-PA</td>
<td>% $R^2$=38.3</td>
<td>% $R^2$=43.0</td>
<td>% $R^2$=4.7 Adjusted % $R^2$=40.2 β=.05</td>
<td>% $R^2$=43.0 Adjusted % $R^2$=39.6 β=.02</td>
<td>% $R^2$=0.0 Adjusted % $R^2$=.05</td>
</tr>
<tr>
<td>FABQ-W</td>
<td>% $R^2$=38.3</td>
<td>% $R^2$=43.0</td>
<td>% $R^2$=4.7 Adjusted % $R^2$=40.2 β=.05</td>
<td>% $R^2$=44.8 Adjusted % $R^2$=41.4 β=.15</td>
<td>% $R^2$=1.7 Adjusted % $R^2$=.05</td>
</tr>
<tr>
<td>PCS</td>
<td>% $R^2$=38.3</td>
<td>% $R^2$=43.0</td>
<td>% $R^2$=4.7 Adjusted % $R^2$=40.2 β=.05</td>
<td>% $R^2$=44.6 Adjusted % $R^2$=41.2 β=.15</td>
<td>% $R^2$=1.6 Adjusted % $R^2$=.05</td>
</tr>
<tr>
<td>TSK-11</td>
<td>% $R^2$=38.3</td>
<td>% $R^2$=43.0</td>
<td>% $R^2$=4.7 Adjusted % $R^2$=40.2 β=.05</td>
<td>% $R^2$=43.8 Adjusted % $R^2$=40.4 β=.10</td>
<td>% $R^2$=0.8 Adjusted % $R^2$=.05</td>
</tr>
<tr>
<td>PHQ-9</td>
<td>% $R^2$=38.5</td>
<td>% $R^2$=43.3</td>
<td>% $R^2$=4.8 Adjusted % $R^2$=40.4 β=.05</td>
<td>% $R^2$=44.6 Adjusted % $R^2$=41.2 β=.13</td>
<td>% $R^2$=1.3 Adjusted % $R^2$=.05</td>
</tr>
<tr>
<td>SBT overall score</td>
<td>% $R^2$=38.9</td>
<td>% $R^2$=43.0</td>
<td>% $R^2$=4.7 Adjusted % $R^2$=40.2 β=.05</td>
<td>% $R^2$=45.6 Adjusted % $R^2$=42.3 β=.22</td>
<td>% $R^2$=2.5 Adjusted % $R^2$=.05</td>
</tr>
<tr>
<td>SBT psychosocial score</td>
<td>% $R^2$=38.3</td>
<td>% $R^2$=43.0</td>
<td>% $R^2$=4.7 Adjusted % $R^2$=40.2 β=.05</td>
<td>% $R^2$=47.5 Adjusted % $R^2$=44.3 β=.25</td>
<td>% $R^2$=4.5 Adjusted % $R^2$=.05</td>
</tr>
</tbody>
</table>

*ODQ = Oswestry Disability Questionnaire, LBP = low back pain, FABQ-PA = Fear-Avoidance Beliefs Questionnaire physical activity scale, FABQ-W = Fear-Avoidance Beliefs Questionnaire work scale, PCS = Pain Catastrophizing Scale, TSK-11 = 11-item version of the Tampa Scale for Kinesiophobia, PHQ-9 = 9-item Patient Health Questionnaire, SBT = STarT Back Screening Tool.

from model 2 to model 3). None of the intake single-construct psychological measure scores significantly explained any additional variability in 6-month ODQ scores.

Predictive Validity Using 4-Week Change Scores: SBT and Single-Construct Psychological Measures for 6-Month Clinical Outcomes

Four-week univariate associations. Four-week changes in single-construct psychological measures were significantly correlated with each other ($r=.26$ to $.43$, $P<.01$), with the exceptions being that FABQ-PA changes were not related to FABQ-W, PCS, or PHQ-9 changes and FABQ-W changes were not related to PCS or PHQ-9 changes ($P>.05$). Four-week changes in single-construct psychological measures also were significantly correlated with 4-week changes in SBT overall ($r=.23$ to $.49$, $P<.01$) and SBT psychosocial ($r=.39$ to $.50$, $P<.01$) scores, with the only exception being that FABQ-W changes were not related to either SBT overall ($r=.03$, $P>.05$) or psychosocial ($r=-.02$, $P>.05$) change scores.

Pain intensity. In the univariate analysis, among psychological measures, only TSK-11 change scores explained variance (5.9%, $P=.01$) in 6-month NPRS scores (results not shown). In the multiple regression analysis, for each linear regression model, intake NPRS scores explained a large amount of variability in 6-month NPRS scores (range=18.8%–20.2%). The addition of demographic and clinical variables did not significantly explain any additional variability in 6-month NPRS scores. Among psychological measure change scores, only TSK-11 change scores significantly explained additional variability (2.9%, $P=.05$) in 6-month NPRS scores (results not shown).

Disability. In the univariate analysis, among psychological measures, only TSK-11 change scores explained variance (7.7%, $P=.01$) in 6-month ODQ scores (results not shown). In the multiple regression analysis, for each linear regression model, intake
ODQ scores explained a large amount of variability in 6-month ODQ scores (range = 41.9%–44.4%). The addition of demographic and clinical variables did not significantly explain any additional variability in 6-month ODQ scores. Among psychological measure change scores, only SBT overall (2.8%, \(P < .05\)), FABQ-PA (4.0%, \(P < .01\)), and TSK-11 (3.8%, \(P < .01\)) change scores significantly explained additional variability in 6-month ODQ scores (results not shown).

Exploratory Analyses and Results
To be consistent with previous studies assessing the SBT, we ran additional analyses where SBT was entered as a categorical variable (ie, low, medium, and high SBT risk). The results from those analyses indicated similar results (not shown). Therefore, we elected to keep SBT scores as a continuous measure because it was a better fit for the regression modeling techniques used in this study and it allowed for direct comparisons with other psychological measures that do not have established risk categories.

Discussion
The primary aim of this study was to determine whether using the SBT or individual psychological measure scores provided more valuable prog nostic information for 6-month clinical outcomes. Our results from multiple regression analyses suggested that: (1) intake SBT scores were valuable in predicting 6-month ODQ scores, but not NPRS scores; (2) SBT change scores following 4 weeks of physical therapy were valuable in predicting 6-month ODQ scores; (3) TSK-11 change scores were valuable in predicting 6-month NPRS and ODQ scores; and (4) FABQ-PA change scores were valuable in predicting 6-month ODQ scores. Although the SBT and individual psychological factors appeared to add predictive value for outcome determinination, intake NPRS and ODQ scores were the strongest predictors for all respective 6-month clinical outcome regression models. This finding was not unexpected and is consistent with previous reviews and empirical studies. An unexpected finding was that neither SBT nor individual psychological measures were predictive of 6-month NPRS scores. This specific finding highlights 2 main concepts: (1) psychological distress is not always predictive of pain intensity and (2) the prediction of clinical outcomes by psychological-based measures varies, which is why multiple outcome domains should be assessed.

SBT
Previous studies have indicated that multiconstruct prediction instrument intake scores are useful in predicting long-term clinical outcomes based on dichotomized recovery criteria. However, previous suggestions indicate that although using dichotomized cutoff scores with continuous data may improve interpretation of results, they also may be associated with potential clinical limitations. For example, using dichotomized cutoff scores may lead to unrealistic results, with some individuals close to but on opposite sides of the cutpoint characterized as having very different rather than very similar outcomes. In this study, we used clinical outcomes that were measured on a continuous scale and arrived at similar results for initial SBT scores as prognostic indicators for 6-month ODQ scores. To our knowledge, there is one previous study that has evaluated relationships between changes in SBT scores and 4-month clinical outcomes. In a primary care setting, Wideman et al29 found that 4-month changes in SBT scores were associated with 4-month changes in treatment-related outcomes and, therefore, suggested that the SBT has potential for use as a treatment monitoring measure. The results of our analyses in a physical therapy setting showed that 4-week changes in SBT overall scores were predictive of 6-month ODQ scores. Collectively, these studies suggest that the SBT also may have potential use as a treatment monitoring measure for LBP disability-related outcomes in outpatient physical therapy settings.

Potential strengths of this study in comparison with a previous study that incorporated the SBT in outpatient physical therapy settings include comparisons with several individual single-construct psychological measures. Specifically, this study is an extension of our previous work because we were able to provide additional information about the diagnostic capabilities of the SBT for 6-month clinical outcomes and make comparisons with several individual single-construct psychological measures via multiple regression analyses. Furthermore, we compared the value SBT scores at initial evaluation in our primary analyses with changes in SBT scores following 4 weeks of physical therapy in our exploratory analyses so that we could further speculate on the SBT’s clinical importance as a prognostic indicator and speculate on its potential usefulness as a treatment monitoring measure in outpatient physical therapy settings. Therefore, the results of this study have potentially important clinical implications for physical therapists in outpatient settings relevant to initial and follow-up screening procedures based on the prognostic value of the SBT.

Single-Construct Psychological Measures
Following 4 weeks of physical therapy, changes in single-construct psychological measures were more valuable in predicting 6-month clinical outcomes in comparison with values obtained at initial evaluation. Interestingly, initial single-construct psy-
The STarT Back Screening Tool and Individual Psychological Measures

Psychological measures were not predictive of 6-month NPRS scores when administered at intake, which was unexpected considering our previous analyses linking PCS and FABQ-PA scores to LBP-related pain intensity outcomes in outpatient physical therapy settings. However, changes in TSK-11 scores were predictive of 6-month NPRS and ODQ scores and changes in FABQ-PA scores were predictive of 6-month ODQ scores. These findings are interesting because specific studies comparing the prognostic value of intake and changes in pain-related fear and fear-avoidance beliefs during an episode of physical therapy care for clinical outcomes are limited in the LBP literature. Previous studies have indicated that repeated psychological assessments provide more valuable prognostic information for clinical outcomes when compared with only relying on initial assessments. Therefore, clinically there appears to be potential value in using the TSK-11 and FABQ-PA as treatment monitoring measures for self-report LBP-related pain intensity and disability.

Strengths and Limitations

This study has several strengths. First, this was only the second study, to our knowledge, that has incorporated SBT data from outpatient physical therapy settings in the United States. Second, in an attempt to meet previous research priorities, the SBT was compared with individual single-construct psychological measures using a study design with standardized timing of follow-up assessments and measurement of multiple outcome domains. Finally, this study focused on modifiable psychological risk factors, many of which can be addressed through direct physical therapy intervention, which may have important implications for future studies involving matched treatment approaches in outpatient physical therapy settings.

This study also had several limitations. First, our psychological questionnaires were largely aligned with fear-avoidance models and we did not measure the influence that other potentially important psychological factors (eg, self-efficacy, preference, expectation) have on LBP outcomes. Second, physical therapy treatment was not standardized in this study. Specifically, clinicians participating in this study (n = 16) were not required to review or tailor treatment based on SBT categorization or individual single-construct psychological measure scores, which may have influenced our results because psychological factors were not targeted with treatment if indicated. Although specific types of treatment were not tracked during this study, many of the participating physical therapists were graduates of orthopedic residency programs where the Treatment-Based Classification System is highlighted as an initial management approach for patients experiencing LBP during didactic and clinical training. However, adherence to this system or any other system was not measured as part of this study. Finally, the use of self-report outcome measures may have biased our results. Therefore, future studies should incorporate both quantitative and self-report outcome measures.

Future Research

This was an observational study; therefore, participating clinicians were not educated on SBT scoring method and associated stratified management approaches for LBP. Furthermore, clinicians were not required to interpret or supplement treatment based on individual single-construct measures. Recent clinical trial findings have indicated that a stratified management care approach (ie, STarT Back approach) for LBP was associated with significantly greater beneficial changes in disability at 4 and 12 months compared with current best practice in primary care settings, with similar findings also reported for generic health benefit and cost-savings at 12 months. Although the findings of that study have potential implications for primary care settings, further testing is needed before the findings can be translated to physical therapy settings. Furthermore, there is the important question of when psychological distress warrants referral to another health care provider (eg, clinical psychologist). That specific question was beyond the purpose of the current study, but we speculate that patients who remain in the SBT high-risk category at 4-week reassessment may be those who warrant referral, and this hypothesis could be tested in future clinical studies. Finally, the lack of prediction for 6-month NPRS scores using intake psychologically based measures indicates that those interested in predicting future pain intensity may need to consider other measures. For example, future studies can test using proxy measures of central sensitization that are feasible for use in clinical settings as predictors of future pain intensity.

Conclusions and Clinical Implications

The American Pain Society and the American College of Physicians have recommended routine assessment of psychological risk factors for patients with LBP. However, a previous review has indicated that this type of assessment may be lacking from many physical therapy clinical decision tools. In this study, we aimed to provide direction for an optimal psychological assessment process specific to physical therapists in outpatient settings while considering the burden placed on clinicians, staff, and patients. These data provide an important reminder that prediction of clinical outcomes...
by psychology-based measures is dependent upon the clinical outcome domain of interest.24-25 Therefore, based on the findings of this study, we suggest a 2-phase psychological assessment process consisting of: (1) initial screening with the SBT to provide prognostic information for 6-month disability outcomes only and (2) treatment monitoring at 4 weeks with the TSK-11 and the FABQ-PA to allow for an indication of 6-month pain and disability outcomes, particularly for patients allocated to the SBT high-risk category at initial assessment.

Dr Beneciuk, Dr Fritz, Dr Robinson, Dr Asal, Dr Nisenzon, and Dr George provided concept/idea/research design. Dr Beneciuk, Dr Bishop, Dr Robinson, Dr Asal, and Dr George provided writing. Dr Beneciuk and Dr Nisenzon provided data collection. Dr Beneciuk, Dr Bishop, Dr Robinson, and Dr Nisenzon provided data analysis. Dr Beneciuk provided project management. Dr Beneciuk, Dr Robinson, and Dr George provided funding procurement. Dr Bishop, Dr Fritz, Dr Robinson, Dr Asal, Dr Nisenzon, and Dr George provided consultation (including review of manuscript before submission).

The authors acknowledge the following clinicians and personnel from Brooks Rehabilitation in Jacksonville, Florida (Ryan Reed, M. Brian Hagist, Tim Shreve, Jason Krat, Matthew Stafford, Michael Spigel, Holly Morris, Flo Singletary, Amanda Osborne, and Robert Rowe), and the University of Florida Orthopaedics and Sports Medicine Institute and Shands Rehabilitation Center at Magnolia Parke in Gainesville, Florida (Giorgio Zeppepi, Josh Barabas, Debi Jones, Derek Miles, Zack Sutton, Dalton Reed, Michael Hodges, Shannon Long, Tim Shay, Amy Borut, and Yvette Silvey). Dr Robinson’s laboratory (Calia Torres and Laura Wandner) assisted with data collection and management.

This study was approved by the University of Florida’s Institutional Review Board.

This study was funded by a 2009 award from the Brooks Health System to the University of Florida Foundation (Dr Beneciuk and Dr George). Dr Beneciuk was supported by a National Institutes of Health T32 Neuromuscular Plasticity Research Training Fellowship Grant (T32 HD043730).

References


