ACRM
AMERICAN CONGRESS OF REHABILITATION MEDICINE

Archives of Physical Medicine and Rehabilitation

journal homepage: www.archives-pmr.org

Archives of Physical Medicine and Rehabilitation 2020;■:■■■-■■■

ORIGINAL RESEARCH

Longitudinal Monitoring of Pain Associated Distress With the Optimal Screening for Prediction of Referral and Outcome Yellow Flag Tool: Predicting Reduction in Pain Intensity and Disability

Steven Z. George, PT, PhD, FAPTA,[a,b] Cai Li, PhD,[c] Sheng Luo, PhD,[a,c] Maggie E. Horn, DPT, MPH, PhD,[b,d] Trevor A. Lentz, PT, PhD, MPH[a,b]

From the [a]Duke Clinical Research Institute, Duke University, Durham, North Carolina; [b]Department of Orthopaedic Surgery, Duke University, Durham, North Carolina; [c]Biostatistics & Bioinformatics, Duke University, Durham, North Carolina; and [d]Division of Physical Therapy, Duke University, Durham, North Carolina

Abstract

Objective: To investigate the Optimal Screening for Prediction of Referral and Outcome Yellow Flag (OSPRO-YF) tool for longitudinal monitoring of pain associated distress with the goal of improving prediction of 50% reduction in pain intensity and disability outcomes.

Design: Cohort study with 12-month follow-up after initial care episode.

Setting: Ambulatory care, participants seeking care from outpatient physical therapy clinics.

Participants: Participants (N=440) were seeking care for primary complaint of neck, low back, knee, or shoulder pain. This secondary analysis included 440 subjects (62.5% female; mean age, 45.1±17y) at baseline with n=279 (63.4%) providing follow-up data at 12 months.

Interventions: Not applicable.

Main Outcome Measures: A 50% reduction (baseline to 12-mo follow-up) in pain intensity and self-reported disability.

Results: Trends for prediction accuracy were similar for all versions of the OSPRO-YF. For predicting 50% reduction in pain intensity, model fit met the statistical criterion for improvement ($P<.05$) with each additional time point added from baseline. Model discrimination improved statistically when the 6-month to 12-month change was added to the model (area under the curve=0.849, $P=.003$). For predicting 50% reduction in disability, there was no evidence of improvement in model fit or discrimination from baseline with the addition of 4-week, 6-month, or 12-month changes ($P>.05$).

Conclusions: These results suggested that longitudinal monitoring improved prediction accuracy for reduction in pain intensity but not for disability reduction. Differences in OSPRO-YF item sets (10 vs 17 items) or scoring methods (simple summary score vs yellow flag count) did not affect predictive accuracy for pain intensity, providing flexibility for implementing this tool in practice settings.

Archives of Physical Medicine and Rehabilitation 2020;■:■■■-■■■

Musculoskeletal (MSK) pain is the largest subset of chronic pain conditions.[1] Chronic MSK pain states are often characterized by persistent pain, prolonged disability, and/or work absenteeism; risk of these outcomes can be established at the initial episode of care seeking.[2] The Federal Pain Research Strategy has placed a priority on assessment tools that estimate risk for the development of persistent pain or prolonged disability.[3] Examples of such tools reported in the literature include the Orebro Musculoskeletal Pain Questionnaire,[4] STarT Back Screening Tool (SBST),[5] and the Optimal Screening for Prediction of Referral and Outcome Yellow Flag (OSPRO-YF) tool.[6]

Each of these tools assesses pain-associated distress and has demonstrated predictive validity. However, studies suggest that the accuracy of prediction can be improved by considering changes in

Supported by the 2013 Clinical Research Network grant from the Academy of Orthopaedic Physical Therapy (formerly known as the Orthopaedic Section of the American Physical Therapy Association). The funding body had no role in the design of the study, analysis, and interpretation of the data or in writing this article.

Disclosures: none.

0003-9993/20/$36 - see front matter © 2020 by the American Congress of Rehabilitation Medicine
https://doi.org/10.1016/j.apmr.2020.05.025

pain associated distress. For example, changes from baseline to 4 weeks later in SBST risk status[7] and OSPRO-YF scores[8] improved the prediction of 6- and12-month self-report of disability, respectively. These studies indicate that prediction accuracy may be improved with repeated assessment of pain-associated distress beyond baseline assessment only.

Repeat longitudinal monitoring is referred to as *treatment monitoring* when done in conjunction with individual response to intervention.[9] Treatment monitoring incorporates assessment of time varying prognostic measures as a way to further refine baseline risk determination. However, there are important considerations of implementing longitudinal monitoring approaches for improving risk assessment. One such issue is the extra administrative burden to the patient and provider. Instead of a risk assessment tool being administered only at baseline, it is administered multiple times throughout the episode of care. Another consideration is determining the best timing for longitudinal monitoring. Theoretically, improved accuracy can be obtained from longitudinal risk assessment until the temporal outcome of interest is collected (eg, meeting definition for chronic pain). However, such an approach is not practically feasible in most clinical settings; therefore, establishing the ideal longitudinal monitoring time points that balance prediction accuracy with reasonable burden can guide implementation efforts. Finally, if shortened versions of risk assessment tools are available (eg, Orebro Musculoskeletal Pain Questionnaire[10] and OSPRO-YF[6]) and demonstrate reasonable accuracy, this aids clinical implementation efforts and should be considered.

Therefore, the purpose of these analyses was to investigate the OSPRO-YF, a validated risk assessment tool, for longitudinal monitoring of pain-associated distress in a cohort on individuals receiving physical therapy care for MSK pain.[11] The OSPRO-YF tool was the focus of this analysis because it has been validated for many common MSK pain conditions.[8] The aims of this analysis were to determine how (1) the timing of OSPRO-YF collection (baseline, 4wk, and 6mo), (2) the length of OSPRO-YF items sets (10- vs 17-item), and (3) OSPRO-YF scoring method (simple summary score vs yellow flag count) influenced prediction accuracy for 12-month reduction in pain intensity and disability.

Methods

Overview

This is a secondary analysis from the Optimal Screening for Prediction of Referral and Outcome (OSPRO) validation cohort, which has been previously described.[11] A convenience sample was gathered from December 2014 and December 2015 by participating Orthopaedic Physical Therapy Investigator Network clinics (n=9), representing the United States Mideast, Southeast, Great Lakes, Rocky Mountain States, and Far West. This study was approved by an institutional review board, and all participants provided informed consent.

List of abbreviations:

AIC	Akaike information criterion
AUC	area under the curve
MSK	musculoskeletal
OSPRO	Optimal Screening for Prediction of Referral and Outcome
OSPRO-YF	Optimal Screening for Prediction of Referral and Outcome Yellow Flag
SBST	STarT Back Screening Tool

Participants and procedures

Participants were recruited from Orthopaedic Physical Therapy Investigator Network clinics when seeking ambulatory physical therapy treatment for a primary complaint of neck, low back, knee, or shoulder pain. Eligibility criteria are summarized in table 1. Eligible participants were directed to a secure University of Florida–hosted website for the informed consent process and baseline assessment. All assessments were self-report and completed electronically by the participant in a deidentified manner. Follow-up time points were at 4 weeks, 6 months, and 12 months, and participants were notified of a pending assessment via e-mail.

Predictive measures

Participants completed a standard intake form previously used in our clinical studies to capture specific predictive domains.[19,20] These domains are summarized in table 1, and all variables were used in the prediction models.

The primary predictor of interest was the OSPRO-YF tool, which includes items from pain vulnerability (negative affect and fear avoidance) and pain resilience domains (positive affect and self-efficacy) and is available in a 10- and 17-item version.[6] In analyses, the OSPRO-YF was represented by a "simple summary score" (sum of individual item responses) and a "yellow flag count" (number of full-length parent questionnaire estimates in 75th percentile).[6] Additional information on OSPRO-YF scoring is located in supplemental appendix S1 (available online only at http://www.archives-pmr.org/).

Outcome measures

Pain intensity and region-specific disability were the outcomes of interest (see table 1). These measures were selected because they are commonly used as primary endpoints in clinical studies, and the region-specific measures were transformed to z scores because of different scaling, consistent with our prior analyses.[8,21]

These outcomes were dichotomized into 12-month 50% reduction in pain intensity and disability. A 50% reduction exceeds commonly recommended minimal change criterion (eg, 30% reduction[22]). Furthermore, a 50% reduction is more likely to be indicative of a successful outcome[23] and closer in magnitude to patient centered definitions for pain and disability recovery.[24-26]

Data analysis

We developed logistic regression models to examine the relationship between a 50% reduction in pain intensity and disability outcomes and the longitudinal change in OSPRO-YF. All models included 16 variables for Demographic, Clinical History, and Other Body Systems in addition to the baseline outcome measure (see table 1). Then the OSPRO-YF was sequentially added for each of the following time points to determine if predictive accuracy was improved: baseline, baseline-4 weeks; 4 weeks-6 months; 6-12 months. The different OSPRO-YF versions (10- or 17-item) and scoring methods (simple summary or yellow flag count) were analyzed separately, leaving 4 final regression models

Table 1 Methodological summary of the Optimal Screening for Prediction of Referral and Outcome cohort

Eligibility	
Inclusion Criteria	**Exclusion Criteria**
1) Seeking outpatient physical therapy treatment for musculoskeletal pain 2) Primary complaints involving the cervical spine, lumbar spine, shoulder, or knee 3) Ability to read and comprehend English language (this criterion was necessary because of the large number of self-report forms)	1) Widespread chronic pain syndrome (eg, fibromyalgia or irritable bowel syndrome) 2) Neuropathic pain syndrome (eg, complex regional pain syndrome or diabetic neuropathy) 3) Psychiatric history (currently in care of mental health care provider or taking ≥2 prescription psychiatric medications) 4) Cancer (currently receiving treatment for active cancer) 5) Neurologic disorder (eg, stroke, spinal cord injury, or traumatic brain injury)

Predictors			
Demographic (n=8)	**Clinical History (n=5)**	**Other Body Systems (n=3)**	**OSPRO-YF (n=4)**
1) Age 2) Sex 3) Race 4) Income 5) Employment 6) Education 7) Medical insurance 8) Geographic region	1) Pain location 2) Pain duration 3) Pain onset 4) Previous episode same location 5) History of surgery	1) Charlson Comorbidity Index 2) Functional Comorbidity Index 3) OSPRO Review of Systems Tool	1) Baseline 2) Baseline-4 wk 3) 4 wk-6 mo 4) 6 mo-12 mo

Outcomes	
Pain intensity (average [worst, best, current], 12-mo 50% reduction)	**Disability (*z* score transformed, 12-mo 50% reduction)**
1) Numeric pain rating scale[12-14]	1) Neck Disability Index[15] 2) Quick Disability of Arm, Hand, and Shoulder[16] 3) Oswestry Disability Questionnaire[17] 4) International Knee Documentation Committee Subjective Knee Form[18]

with 21 predictor variables each (20 predictor variables listed in table 1 plus baseline outcome measure).

Residual deviance, Akaike information criterion (AIC), and area under the curve (AUC) determined relative model fit and predictive performance, respectively. Residual deviance is the measure for model fitting when incorporating later time points, so a smaller value indicates better goodness-of-fit with later follow-ups. AIC was used for full model selection; the model with the minimum AIC among others is the best model statistically. AUC was used for assessing prediction performance in logistic regression and comparing model performance based on length and scoring method. A larger AUC was indicative of better discrimination ability. Finally, a sensitivity analysis was completed with a 30% reduction in pain intensity and disability outcomes.[22]

Sample size

There are no uniform standards for determining sample size in cohort studies. In the OSPRO validation cohort sample size estimates were initially based on precision of prediction estimates.[11] In this secondary analysis the number of subjects per predictor variables was relevant for determining the appropriateness of fit for regression models. We had 279 subjects with completed 12-month data for 21 predictor variables, a ratio that exceeds the recommended 10 subjects per variable.[27]

Results

A total of 440 participants completed baseline measures, and 279 participants (63.4%) completed the 12-month follow-up (tables 2 and 3). There were no differences in follow-up rates by anatomic region, but those that completed the 12-month follow-up were more likely to be younger, have higher income, and completed higher levels of education ($P<.05$). Those that completed follow-up also had lower simple summary scores on OSPRO-YF, neck disability, pain intensity, and composite z score for region-specific disability ($P<.05$). All of these variables had already been planned for the prediction models, so no action was taken based on differences in follow-up rates. Models for 50% (planned analysis) and 30% (sensitivity analysis) were similar, so models for 50% reductions are presented. Model results are fully reported in tables 4 and 5, with a brief summary in the article text. Additional text reporting on the influence of item numbers and scoring methods are reported in supplemental appendix S1.

Longitudinal monitoring for pain intensity reduction

The change in predictive accuracy for 50% Pain Intensity Reduction for 4-week, 6-month, and 12-month OSPRO-YF assessments is reported in table 4 (horizontal comparisons). The 17-item summary score model had the smallest AIC. In this model for AUC, there were incremental model improvements from baseline noted for 4-week and 6-month changes (AUC from 0.783 to 0.799 to 0.808, respectively), with none meeting the statistical criterion for an improvement ($P>.05$). The addition of the 12-month time point improved the AUC statistically (AUC, 0.849; $P=.003$) from that of the 6-month time point.

A similar trend was noted in residual deviance for the models, with this decreasing consistently with each additional time point added from baseline (from 309.0 to 299.9 to 294.9 to 265.8 from baseline to 4-wk to 6-mo to 12-mo change, respectively). The decrease in residual deviance met statistical criterion for improvement ($P<.05$) with each additional monitoring time point. The largest statistical improvements were noted with the addition of the 6- to 12-month change score to the 4-week change score ($P<.0001$).

Table 2 Descriptive summary of Optimal Screening for Predication of Referral and Outcome Validation cohort

Variable/Category	Baseline Characteristics (N=440)
Age (y), mean ± SD	45.1±15.8
Age (y), median (IQR)	45 (27)
Age (y), range	18-75
Sex, n (%)	
Male	164 (37.3)
Female	275 (62.5)
Prefer not to answer	1 (0.2)
Race, n (%)	
American Indian/Alaska Native	3 (0.7)
Asian	25 (5.7)
Black	62 (14.1)
White	343 (78.0)
Don't know/prefer not to answer	7 (1.6)
Ethnicity, n (%)	
Hispanic or Latino	31 (7.0)
Not Hispanic or Latino	376 (85.5)
Don't know/prefer not to answer	33 (7.5)
Employment, n (%)	
Full-time employed (reference for analyses)	237 (53.9)
Part-time employed	62 (14.1)
Unemployed	61 (13.9)
Retired	58 (13.2)
Prefer not to answer	22 (5.0)
Education, n (%)	
Less than high school	11 (2.5)
Graduated from high school	38 (8.6)
Some college	112 (25.5)
Graduated from college (reference for analyses)	120 (27.3)
Some postgraduate course work	56 (12.7)
Completed postgraduate degree	97 (22.0)
Prefer not to answer	6 (1.4)
Insurance, n (%)	
Private (reference for analyses)	273 (62.0)
Medicare	52 (11.8)
Medicaid	19 (4.3)
Worker's compensation	14 (3.2)
Disability	4 (0.9)
Uninsured	7 (1.6)
Other	45 (10.2)
Unknown/prefer not to answer	26 (5.9)
Anatomic region, n (%)	
Neck	98 (22.3)
Low back	118 (26.8)
Shoulder	107 (24.3)
Knee (reference for analyses)	117 (26.6)
Pain duration (d), mean ± SD	398.6±1715.8
Pain duration (d), median (IQR)	90 (270)
Onset of symptoms, n (%)	
Gradual	239 (54.3)
Sudden	138 (31.4)
Traumatic	63 (14.3)
Previous episodes over the past year, n (%)	
Yes (reference for analyses)	224 (50.9)
No	185 (42.0)
Do not remember	31 (7.0)
Work-related symptoms, n (%)	
Yes	63 (14.3)
No	345 (78.4)
Do not know	32 (7.3)
Surgery for primary complaint, n (%)	
Yes	83 (18.9)
No	357 (81.1)
OSPRO-ROS score, mean ± SD	3.8±3.7
OSPRO-ROS score, median (IQR)	3 (5)
Comorbidity counts, n (%)	
0	296 (68.8)
1	76 (17.7)
2	19 (4.4)
3+	39 (9.1)

(*continued on next column*)

Abbreviations: IQR, interquartile range; OSPRO-ROS, Optimal Screening for Prediction of Referral and Outcome Review of Systems tool.

Longitudinal monitoring for disability reduction

The change in predictive accuracy for 50% disability reduction for 4-week, 6-month, and 12-month OSPRO-YF assessments is reported in table 5 (horizontal comparisons). The 17-item summary score model is reported in the text to allow for direct comparisons to the pain intensity reduction model. For AUC, there was no evidence of model improvement from baseline with the addition of 4-week, 6-month, and 12-month changes (AUC from 0.551 to 0.551 to 0.554 to 0.553, respectively). The same trend was noted in residual deviance, with this staying relatively stable with each additional time point added from baseline (AUC from 319.0 to 319.0 to 318.0 to 317.8 from baseline to 4-wk to 6-mo to 12-mo change, respectively). None of the changes in residual deviance met statistical criterion for improvement (P>.05).

Discussion

This analysis will add to the existing literature by being directly responsive to Federal Pain Research Strategy priorities of

Table 3 Baseline and 12-month summary for OSPRO-YF tool, pain intensity, and disability measures

Variable/Category	Baseline Characteristics (N=440)
OSPRO-YF17, mean ± SD	32.4±11.2
OSPRO-YF17, median (IQR)	32 (16)
OSPRO-YF10, mean ± SD	17.4±6.7
OSPRO-YF10, median (IQR)	17 (9)
Pain intensity, mean ± SD	4.2±2.0
Pain intensity, median (IQR)	4.0 (2.3)
Neck Disability Index, mean ± SD	28.6±16.1
Neck Disability Index, median (IQR)	24 (20)
Oswestry, mean ± SD	28.7±18.2
Oswestry, median (IQR)	26 (24)
QuickDASH, mean ± SD	38.8±20.1
QuickDASH, median (IQR)	34.1 (27.3)
IKDC total score, mean ± SD	38.5±15.4
IKDC total score, median (IQR)	37 (22)
Variable/Category	12-Month Follow-up (n=279)
OSPRO-YF17, mean ± SD	28.6±12.5
OSPRO-YF17, median (IQR)	28 (18)
OSPRO-YF10, mean ± SD	16.2±7.5
OSPRO-YF10, median (IQR)	15 (11)
Pain intensity, mean ± SD	1.5±1.9
Pain intensity, median (IQR)	0.7 (2.3)
Neck Disability Index, mean ± SD	19.2±16.6
Neck Disability Index, median (IQR)	14 (21)
Oswestry, mean ± SD	16.4±15.6
Oswestry, median (IQR)	10.0 (22.5)
QuickDASH, mean ± SD	18.6±19.3
QuickDASH, median (IQR)	13.6 (20.5)
IKDC total score, mean ± SD	58.1±17.5
IKDC total score, median (IQR)	59.5 (25.8)

Abbreviations: DASH, Disability of Arm, Shoulder, and Hand questionnaire; IKDC, International Knee Documentation Committee questionnaire; IQR, interquartile range.

identifying methods to improve prediction of chronic pain states.[3] The American College of Physicians and the Centers for Disease Control and Prevention guidelines emphasize nonpharmacologic approaches for acute and chronic MSK pain.[28,29] As such, we expect increased interest in predicting long-term reductions in pain intensity and disability following treatment from nonpharmacologic providers. This interest will be driven by it being an indication of successful treatment outcome[23] and decreased chance of developing or maintaining chronic pain. This analysis highlighted the potential of longitudinal monitoring to improve baseline risk assessment for predicting reductions in pain intensity and disability. The primary findings of our analyses were (1) prediction accuracy for pain intensity reduction could be incrementally improved with longitudinal monitoring, but improvements in risk estimation lack clinical relevance after 4 weeks and (2) prediction accuracy for disability reduction was poor and the risk prediction did not improve with longitudinal monitoring.

The number of items (10 vs 17 items) or the scoring approach (simple summary score vs yellow flag count) used for the OSPRO-YF tool did not strongly affect predictive accuracy in our models. The implication of these findings is that selection of OSPRO tool item version and scoring approach can be flexible based on what is most feasible for a given setting and patient population. The yellow flag count scoring method provides readily interpretable information with high clinical relevance, but its calculation is complicated and should be automated.[6] In settings that can readily implement these measures in the electronic health record, generating yellow flag counts is likely feasible. In settings that cannot modify or do not have access to the electronic health record, the simple summary score is a viable option because it does not involve complicated scoring algorithms.[6]

Timing of longitudinal monitoring improved prediction accuracy for pain intensity reduction. However, there are 2 caveats when interpreting these results. First, including the 6- to 12-month OSPRO-YF change provided the best model fit statistically but has limited clinical applicability because it is not current standard of care to collect information after discharge. Second, these findings suggest that any value imparted by longitudinal monitoring are realized in the 4-week change, with limited additional value in additional monitoring beyond that time. Therefore, a preliminary recommendation for a pragmatic approach to increase accuracy of predicting 12-month pain intensity reduction would be to collect the 10-item OSPRO-YF at intake and then again 4 weeks later for longitudinal monitoring.

There is not extensive literature on longitudinal monitoring of psychosocial factors. Wideman et al[9] investigated 4-month changes in SBST risk status and found that changes in SBST risk status predicted individuals with clinically meaningful changes in 12-month disability outcomes. Similarly, Beneciuk et al[7] reported that 4-week change in SBST risk status added to the prediction of 6-month disability outcomes, while Medeiros et al[30] found high responsiveness for global perceived effect, disability, and pain intensity at 6 weeks for medium- and high-risk SBST designation and for all SBST risk designations at 6 months. These results are similar to the current analysis in indicating improvements with longitudinal monitoring but with one notable inconsistency. In the prior cited SBST studies and the primary analysis for the OSPRO-YF tool, longitudinal monitoring added to prediction of 12-month disability scores.[8] However, that was not the case in the current analysis, and this difference is likely because of how the outcomes were defined. In prior studies the disability scores were kept in continuous metric, while for this analysis they were dichotomized to a 50% reduction at 12 months. The use of a binary outcome could explain the reason why longitudinal monitoring did not improve predictive accuracy for disability outcomes in this analysis compared with previous studies.

There are some areas in which this current analysis advances the field of risk assessment. The prior cited studies included individuals with low back pain, while we included individuals with neck, shoulder, low back, and knee pain as a primary complaint. Therefore, findings suggest that longitudinal monitoring with the OSPRO-YF tool is appropriate for multiple MSK pain conditions. The prior cited studies included 1 additional monitoring time point prior to the outcome of interest (eg, 4-mo,[9] 4-wk,[7] or 6-wk change[30]), while the current analysis included multiple time

Table 4 Model summary of treatment monitoring for predicting 12-month pain intensity reduction .

Model No.		Model 2 = Model 1+Δ OSPRO-YF Baseline-4 wk	Model 3 = Model 2+ Δ OSPRO-YF 4 wk-6 mo	Model 4 = Model 3+ Δ OSPRO-YF 6-12 mo	Full Model AIC
OSPRO-YF simple summary score (17 items)					
AUC (*P* value)	0.783	0.799 (.123)	0.808 (.246)	0.849 (.003)	385.85
Residual deviance (*P* value)	309.0	299.9 (.002)	294.9 (.025)	265.8 (<.001)	
OSPRO-YF simple summary score (10 items)					
AUC (*P* value)	0.781	0.797 (.107)	0.803 (.393)	0.833 (.024)	395.89
Residual deviance (*P* value)	309.6	302.2 (.006)	297.5 (.029)	275.8 (<.001)	
OSPRO-YF yellow flag count (17 items)					
AUC (*P* value)	0.783	0.800 (.078)	0.806 (.435)	0.834 (.037)	396.62
Residual deviance (*P* value)	309.2	300.0 (.002)	295.3 (.031)	276.6 (<.001)	
OSPRO-YF yellow flag count (10 items)					
AUC (*P* value)	0.781	0.790 (.239)	0.800 (.264)	0.832 (.023)	396.37
Residual deviance (*P* value)	309.4	304.3 (.020)	297.8 (.010)	276.3 (<.001)	

NOTE. Model 1 = Demographic+Clinical History+Other Body Systems+Baseline OSPRO-YF.

points before 12-month outcome. Therefore, these findings identified a "window of opportunity" for longitudinal monitoring when improvement in prediction accuracy occurred and when improvement in prediction accuracy plateaued.

Study limitations

The overriding limitations of the OSPRO validation cohort have been described in the cohort profile article,[11] and these include convenience sampling and not capturing parameters to characterize treatment episodes. Other limitations are lack of specific medical diagnoses in the predictive models and a 12-month 63% follow-up rate that was lower than anticipated. The current findings should be interpreted as a "completers only" analysis, which we believe is appropriate because loss to follow-up had a minimal effect in other analyses from this cohort in which we imputed missing data or used inverse probability weighting.[8,31] A final limitation is that all outcomes for this analysis were self-reported, with no corresponding physical performance measures.

Future research

Future research could compare predictive accuracy of existing tools in determining how longitudinal monitoring is used within a psychologically informed approach that targets individuals with high levels of pain-associated distress.[32-34] Because determining dosing of psychologically informed treatments can be difficult, longitudinal monitoring could be used to assess response to behavioral strategies and confirm that favorable changes are indicative of better clinical outcomes. There is additional implementation-focused research that is needed to determine how best to get longitudinal monitoring into existing health care systems that focus almost entirely on intake data capture for risk determination. Capturing longitudinal data in real-world settings is challenging, and additional methods of data collection (eg, wearables, text messaging, etc) will need to be tested for effectiveness in increasing follow-up rates. The aforementioned lack of prediction accuracy for disability reduction was an unexpected finding, and future research is necessary to determine how longitudinal monitoring with the OSPRO-YF tool is affected when there are changes in whether the outcome of interest is defined in continuous metric or converted to a binary outcome. Determining whether longitudinal monitoring improves prediction of treatment episodes reaching patient-defined levels of successful outcomes is an important area for future research because these levels often exceed a 50% reduction.[24-26]

Conclusions

This analysis investigated the OSPRO-YF as a longitudinal monitoring tool of pain-associated distress. These results suggested that prediction accuracy for pain intensity reduction could be improved with longitudinal monitoring but not for disability reduction. Furthermore, differences in OSPRO-YF item sets (10 vs 17 items) or scoring methods (simple summary score vs yellow flag count) did not affect predictive accuracy for pain intensity reduction, providing flexibility for implementing the tool in practice settings.

Keywords

Chronic pain; Prognosis; Rehabilitation

Corresponding author

Steven George, PT, PhD, FAPTA, 200 Morris St, Durham, NC 27001. *E-mail address:* steven.george@duke.edu.

Acknowledgments

We thank Sam Wu, PhD, and Yunfeng Dai, MS, for their contributions to the development of the Yellow Flag Tool.

Table 5 Model summary of longitudinal monitoring for predicting 12-month disability reduction

Model Version	Model 1	Model 2 = Model 1+ Δ OSPRO-YF Baseline-4 wk	Model 3 = Model 2+ Δ OSPRO-YF 4 wk-6 mo	Model 4 = Model 3+ Δ OSPRO-YF 6-12 mo	Full Model AIC
OSPRO-YF simple summary score (17 items)					
AUC (*P* value)	0.551	0.551 (.123)	0.554 (.246)	0.553 (.003)	437.83
residual deviance (*P* value)	319.0	319.0 (.985)	318.0 (.301)	317.8 (.667)	
OSPRO-YF simple summary score (10 items)					
AUC (*P* value)	0.551	0.545 (.026)	0.530 (.096)	0.544 (<.001)	435.62
residual deviance (*P* value)	318.5	318.3 (.619)	318.0 (.563)	315.6 (.122)	
OSPRO-YF yellow flag count (17 items)					
AUC (*P* value)	0.556	0.550 (.004)	0.534 (.293)	0.550 (.003)	432.58
residual deviance (*P* value)	317.7	313.5 (.041)	312.8 (.387)	312.5 (.623)	
OSPRO-YF yellow flag count (10 items)					
AUC (*P* value)	0.547	0.527 (.008)	0.527 (.095)	0.532 (<.001)	435.36
Residual deviance (*P* value)	318.1	317.7 (.500)	316.9 (.379)	315.3 (.208)	

NOTE. Model 1 = Demographic + Clinical History + Other Body Systems + Baseline OSPRO-YF.

The Orthopaedic Physical Therapy-Investigator Network included University of Florida: Joel Bialosky; UF Health: Giorgio Zeppieri Jr, Daniel Broome, Marty Huegel, Debi Jones, Steve Emery, Mike Hodges, Derek Miles, Jodi Davis, Charlene Stubbington, Mike Darcy; ATI Physical Therapy: Ellen Shanley, Thomas Denninger, Jenna Bartsokas, Elise Harris, Jordan Floyd, Wade Harrell; University of Southern California: Lori Michener, Amy Pomrantz, Brooks Rehabilitation: Jason Beneciuk, Raine Osborne, Nata Salvatori, John Leschitz, Brian Hagist, Laura Langer, Tim Shreve, Nando Malaman, Michael Bourassa, Justin Zych, Tasha Mouton Shanklin; University of Illinois at Chicago: Aaron Keil, Brad Myers, Deb Davey, Justin Payette, Adam Wielechowski, Richard Severin, Erik Martinez; Indiana State University: Ryan Hanigan, Carolina Valencia, Danielle Jena, Nicole Woodard; Arcadia University: Angela Tate; Life's Work Physical Therapy: Sandra Stryker, Aaron Leonard, Erin Courtney, Brandon Little, Kathryn Jankord, Brad Simpson, Charleen Hall, Paige Nixon, Julia Neufeld; University of Colorado, Denver: Paul Mintken, Virginia Arnette, Andrea Barsch.

References

1. Institute of Medicine (US) Committee on Advancing Pain Research, Care, and Education. Relieving pain in America: a blueprint for transforming prevention, care, education, and research. Washington (DC): National Academies Press (US); 2011.
2. Karran EL, McAuley JH, Traeger AC, et al. Can screening instruments accurately determine poor outcome risk in adults with recent onset low back pain? A systematic review and meta-analysis. BMC Med 2017;15:13.
3. National Institute of Health Intraagency Pain Research Coordinating Committee. Federal pain research strategy. Available at: https://iprcc.nih.gov/sites/default/files/iprcc/FPRS_Research_Recommendations_Final_508C.pdf. Accessed October 31, 2019.
4. Linton SJ, Boersma K. Early identification of patients at risk of developing a persistent back problem: the predictive validity of the Orebro Musculoskeletal Pain Questionnaire. Clin J Pain 2003;19:80-6.
5. Hill JC, Dunn KM, Lewis M, et al. A primary care back pain screening tool: identifying patient subgroups for initial treatment. Arthritis Rheum 2008;59:632-41.
6. Lentz TA, Beneciuk JM, Bialosky JE, et al. Development of a yellow flag assessment tool for orthopaedic physical therapists: results from the Optimal Screening for Prediction of Referral and Outcome (OSPRO) cohort. J Orthop Sports Phys Ther 2016;46:327-43.
7. Beneciuk JM, Fritz JM, George SZ. The STarT Back Screening Tool for prediction of 6-month clinical outcomes: relevance of change patterns in outpatient physical therapy settings. J Orthop Sports Phys Ther 2014;44:656-64.
8. George SZ, Beneciuk JM, Lentz TA, et al. Optimal Screening for Prediction of Referral and Outcome (OSPRO) for musculoskeletal pain conditions: results from the validation cohort. J Orthop Sports Phys Ther 2018;48:460-75.
9. Wideman TH, Hill JC, Main CJ, Lewis M, Sullivan MJ, Hay EM. Comparing the responsiveness of a brief, multidimensional risk screening tool for back pain to its unidimensional reference standards: the whole is greater than the sum of its parts. Pain 2012;153:2182-91.
10. Linton SJ, Nicholas M, MacDonald S. Development of a short form of the Orebro Musculoskeletal Pain Screening Questionnaire. Spine 2011;36:1891-5.
11. George SZ, Lentz TA, Beneciuk JM, Wu SS. The Optimal Screening for Prediction of Referral and Outcome (OSPRO) in patients with musculoskeletal pain conditions: a longitudinal validation cohort from the USA. BMJ Open 2017;7:e015188.
12. Bolton JE. Accuracy of recall of usual pain intensity in back pain patients. Pain 1999;83:533-9.

13. Childs JD, Piva SR, Fritz JM. Responsiveness of the numeric pain rating scale in patients with low back pain. Spine 2005;30:1331-4.
14. Jensen MP, Turner JA, Romano JM, Fisher LD. Comparative reliability and validity of chronic pain intensity measures. Pain 1999;83:157-62.
15. Vernon H. The Neck Disability Index: state-of-the-art, 1991-2008. J Manipulative Physiol Ther 2008;31:491-502.
16. Beaton DE, Wright JG, Katz JN. Development of the QuickDASH: comparison of three item-reduction approaches. J Bone Joint Surg Am 2005;87:1038-46.
17. Fritz JM, Irrgang JJ. A comparison of a modified Oswestry Low Back Pain Disability Questionnaire and the Quebec Back Pain Disability Scale. Phys Ther 2001;81:776-88.
18. Irrgang JJ, Anderson AF, Boland AL, et al. Development and validation of the international knee documentation committee subjective knee form. Am J Sports Med 2001;29:600-13.
19. Beneciuk JM, Bishop MD, Fritz JM, et al. The STarT Back Screening Tool and individual psychological measures: evaluation of prognostic capabilities for low back pain clinical outcomes in outpatient physical therapy settings. Phys Ther 2013;93:321-33.
20. George SZ, Zeppieri G Jr, Cere AL, et al. A randomized trial of behavioral physical therapy interventions for acute and sub-acute low back pain (NCT00373867). Pain 2008;140:145-57.
21. Butera KA, Lentz TA, Beneciuk JM, George SZ. Preliminary evaluation of a modified STarT Back Screening Tool across different musculoskeletal pain conditions. Phys Ther 2016;96:1251-61.
22. Ostelo RW, Deyo RA, Stratford P, et al. Interpreting change scores for pain and functional status in low back pain: towards international consensus regarding minimal important change. Spine 2008;33:90-4.
23. Fritz JM, Hebert J, Koppenhaver S, Parent E. Beyond minimally important change: defining a successful outcome of physical therapy for patients with low back pain. Spine 2009;34:2803-9.
24. Thorne FM, Morley S. Prospective judgments of acceptable outcomes for pain, interference and activity: patient-determined outcome criteria. Pain 2009;144:262-9.
25. Robinson ME, Brown JL, George SZ, et al. Multidimensional success criteria and expectations for treatment of chronic pain: the patient perspective. Pain Med 2005;6:336-45.
26. Zeppieri G Jr, George SZ. Patient-defined desired outcome, success criteria, and expectation in outpatient physical therapy: a longitudinal assessment. Health Qual Life Outcomes 2017;15:29.
27. Austin PC, Steyerberg EW. The number of subjects per variable required in linear regression analyses. J Clin Epidemiol 2015;68:627-36.
28. Qaseem A, Wilt TJ, McLean RM, Forciea MA. Noninvasive treatments for acute, subacute, and chronic low back pain: a clinical practice guideline from the American College of Physicians. Ann Intern Med 2017;166:514-30.
29. Dowell D, Haegerich TM, Chou R. CDC guideline for prescribing opioids for chronic pain–United States, 2016. JAMA 2016;315:1624-45.
30. Medeiros FC, Costa LOP, Oliveira IS, Costa L. A responsiveness analysis of the Subgroups for Targeted Treatment (STarT) Back Screening Tool in patients with nonspecific low back pain. J Orthop Sports Phys Ther 2019;49:725-35.
31. Lentz TA, Beneciuk JM, George SZ. Prediction of healthcare utilization following an episode of physical therapy for musculoskeletal pain. BMC Health Serv Res 2018;18:648.
32. Main CJ, George SZ. Psychologically informed practice for management of low back pain: future directions in practice and research. Phys Ther 2011;91:820-4.
33. Keefe FJ, Main CJ, George SZ. Advancing psychologically informed practice for patients with persistent musculoskeletal pain: promise, pitfalls, and solutions. Phys Ther 2018;98:398-407.
34. Beneciuk JM, Ballengee LA, George SZ. Treatment monitoring as a component of psychologically informed physical therapy: a case series of patients at high risk for persistent low back pain related disability. Musculoskelet Sci Pract 2019;41:36-42.