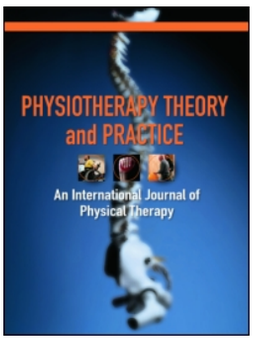

Taylor & Francis
Taylor & Francis Group

Physiotherapy Theory and Practice
An International Journal of Physical Therapy

ISSN: 0959-3985 (Print) 1532-5040 (Online) Journal homepage: http://www.tandfonline.com/loi/iptp20

Association between changes in function among patients with lumbar impairments classified according to the STarT Back Screening Tool and managed by McKenzie credentialed physiotherapists

Mark W. Werneke PT, MS, Dip. MDT, Susan Edmond PT, DSc, OCS, Michelle Young PT, Cred MDT, David Grigsby PT, Cert MDT, Brian McClenahan PT, MS, OCS, Dip. MDT, FAAOMPT & Troy McGill PT, PhD, SCS, Dip. MDT

To cite this article: Mark W. Werneke PT, MS, Dip. MDT, Susan Edmond PT, DSc, OCS, Michelle Young PT, Cred MDT, David Grigsby PT, Cert MDT, Brian McClenahan PT, MS, OCS, Dip. MDT, FAAOMPT & Troy McGill PT, PhD, SCS, Dip. MDT (2018): Association between changes in function among patients with lumbar impairments classified according to the STarT Back Screening Tool and managed by McKenzie credentialed physiotherapists, Physiotherapy Theory and Practice, DOI: 10.1080/09593985.2018.1490839

To link to this article: https://doi.org/10.1080/09593985.2018.1490839

Published online: 09 Jul 2018.

PHYSIOTHERAPY THEORY AND PRACTICE
https://doi.org/10.1080/09593985.2018.1490839

Taylor & Francis
Taylor & Francis Group

REPORT

Association between changes in function among patients with lumbar impairments classified according to the STarT Back Screening Tool and managed by McKenzie credentialed physiotherapists

Mark W. Werneke PT, MS, Dip. MDT[a], Susan Edmond PT, DSc, OCS[b], Michelle Young PT, Cred MDT[c], David Grigsby PT, Cert MDT[d], Brian McClenahan PT, MS, OCS, Dip. MDT, FAAOMPT[e], and Troy McGill PT, PhD, SCS, Dip. MDT[f]

[a]Private Practice, Physical Therapy, Waipahu, HI, USA; [b]Doctoral Programs in Physical Therapy, Rutgers The State University of New Jersey, Newark, NJ, USA; [c]Rehabilitation Services, Valley Health Wellness and Fitness Center, Winchester, VA, USA; [d]MidSouth Orthopaedic Rehabilitation, Physical Therapy, Cordova, TN, USA; [e]WellSpan Rehabilitation, Physical Therapy, Red Lion, PA, USA; [f]Providence Rehabilitation Services, Outpatient Therapies, Providence Alaska Medical Center, Anchorage, AK, USA

ABSTRACT

Purpose: To examine the association between functional status (FS) scores using a Patient Reported Outcome Measure (PROM) for patients with non-specific low back pain classified according to psychosocial risk using the STarT Back Screening Tool and managed by physiotherapists credentialed in McKenzie methods. Methods: Participants (n = 705) completed FS and STarT surveys at intake and discharge. Prevalence of STarT risk classifications and change in STarT risk was calculated. Regression models were developed to examine associations between baseline and change in STarT risk categories, and FS outcomes at discharge from rehabilitation services. Results: FS outcomes at discharge was not significantly different (p-values > 0.10) across baseline STarT risk subgroups after controlling for model covariates. Seventy-eight and 91.5% of medium and high-risk patients respectively decreased STarT risk. When compared with subjects whose STarT risk decreased, there was no significant difference in subjects whose STarT risk remained low. For subjects whose risk remained medium/high, or whose risk worsened, FS outcome scores were statistically significant (p < 0.001) and clinically relevant (−15.76 and −23.42 points respectively) compared to patients whose STarT risk decreased. Conclusions: Baseline STarT psychosocial risk stratifications should be interpreted cautiously to estimate the likelihood of good or poor FS outcomes at discharge from physiotherapy practice in the US when patients are managed by clinicians credentialed in McKenzie methods. Decreased STarT risk was associated with clinically important improvements in FS outcomes scores at discharge from McKenzie directed physiotherapy care.

ARTICLE HISTORY
Received 28 August 2017
Revised 19 May 2018
Accepted 24 May 2018

KEYWORDS
McKenzie; STarT; low back pain; function; patient reported outcome measure; physiotherapy

Introduction

Psychosocial screening and classification are considered a high priority for physiotherapists to stratify patients by risk level for elevated pain-related psychosocial distress. Psychosocial distress is a known precursor for increased risk of chronic disability, increased health care costs, and delayed treatment recovery (George and Beneciuk, 2015; Nicholas, Linton, Watson, and Main, 2011). To assist clinicians to accurately classify patients by psychosocial risk, Hill et al. (2008) developed a 9-item multidimensional biopsychosocial screening measure (i.e. Subgroups for Targeted Treatment (STarT) Back Screening Tool). One purpose for STarT risk classification is to target specific treatment strategies for each of the three STarT classification risk patient subgroups (i.e. low, medium, or high).

The three STarT classification treatment pathways consist of: (1) a single primary care visit for simple back education for the low-risk subgroup; (2) traditional physiotherapy according to best-practice guidelines for medium risk; and (3) management by a combination of physical and cognitive behavioral approaches (CBA) for the high-risk subgroup. The STarT classification approach for managing patients with non-specific low back pain (NSLBP) and for prescribing one of three matched treatment pathways, including a formal CBA training program, have been described in detail elsewhere (Foster et al., 2010; Hill et al., 2008; Main et al., 2012; Sowden et al., 2012).

In addition to classifying patients at intake according to the three STarT risk levels described above, we re-classified patients into three change STarT risk

CONTACT Mark W. Werneke PT, MS, Dip MDT mwsurf75@gmail.com Private Practice, Physical Therapy, Waipahu, HI, USA.

subgroups identified at discharge following previously published recommendations (Beneciuk, Fritz, and George, 2014). The three change STarT risk subgroups identified at discharge were: (1) improved (STarT risk changed from medium to low, high to low, or high to medium risk); (2) stable (STarT risk remained low or medium); or (3) worsened (STarT level changed from low to medium, low to high, medium to high, or remained high-risk).

The McKenzie method is another approach used by clinicians to classify patients to prescribe targeted-treatments and enhance estimates of the likelihood of good or poor treatment outcomes (May and Aina, 2012; May and Donelson, 2008; McKenzie and May, 2003). Managing NSLBP patients using the McKenzie method was associated with reduced downstream health care utilization (Machado et al., 2010) and an improved cost-effectiveness ratio (Manca et al., 2007). In addition, evidence supports that screening patients with NSLBP for symptom pain pattern subcategories, such as directional preference or centralization (CEN) observed during McKenzie assessments was associated with improved patient outcomes (Brennan et al., 2006; Browder, Childs, Cleland, and Fritz, 2007; Long, Donelson, and Fung, 2004; Long, May, and Fung, 2008; Werneke et al, 2011a). Unlike the STarT system, the McKenzie classification method does not include psychosocial screening or a formal CBA-targeted treatment pathway as recommended by STarT developers (Hill et al., 2008). Of interest, despite the absence of psychosocial risk stratification and a CBA treatment pathway in the McKenzie system, recent reviews have reported positive effects with the McKenzie system for decreasing elevated psychosocial distress and fear, while simultaneously improving FS treatment outcomes for patients with NSLBP receiving physiotherapy services (Al-Obaidi, Al-Sayegh, Ben Nakhi, and Skaria, 2013; Werneke et al., 2009; Werneke et al, 2011b).

There are no published studies that we are aware of which have examined the associations between outcomes for patients with NSLBP who are classified by baseline and change in STarT risk, and managed by physiotherapists credentialed in McKenzie methods. Therefore, our primary purposes were to examine whether there would be significant differences in changes in functional status (FS) outcomes assessed at discharge from rehabilitation services between patients who were: (1) initially classified by three STarT risk levels; and (2) re-classified at discharge by changes in STarT risk.

Methods

Design

A retrospective analysis of a longitudinal cohort was conducted. Patient characteristics and outcome data were collected continuously between February 2013 and June 2016 using Focus on Therapeutic Outcomes Inc. (FOTO), which is an international rehabilitation medical data management company (Swinkels et al., 2007). We selected this data management company because it provided observational data on a wide range of patient demographic and health characteristics to examine for associations with FS outcomes. The Rutgers University Institutional Review Board approved this project and determined these analyses to be research exempt from human subject review. All participants' rights and anonymity were protected throughout the course of the study. The study did not require any change in clinical practice, data documentation, or treatment; therefore patient informed consent was not required.

Clinicians

Nine physical therapists were recruited. The two criteria for a physical therapist to be eligible to participate were (1) credentialed by the McKenzie Institute International (McKenzie and May, 2003) and (2) agreed to follow a standardized data documentation protocol while participating with the FOTO system. Four clinicians were credentialed as Diplomats and five clinicians were certified for achieving basic competency in McKenzie methods by the International McKenzie Institute. The McKenzie Diploma refers to clinicians who have initially completed the Institute's Education postgraduate program (e.g. Parts A-D), including a pass in the McKenzie Institute International Credentialing Examination, as well as successfully completing an additional one semester in a Distance Learning Module and 360 hours of clinical practice supervised by a qualified McKenzie instructor. Those clinicians certified in McKenzie completed the Institute's Education program and passed the McKenzie Institute International Credentialing Examination.

Subjects

Subjects were included if patients were referred to physiotherapy with NSLBP, English was their primary language, and were aged 18 years or older. Our goal was to limit the exclusion criteria to replicate the diverse patient characteristics typically seen in

physiotherapy practices. Patients were excluded due to pregnancy or suspicion of serious spinal pathology.

Treatment

McKenzie classification and targeted-treatments were considered an initial approach regardless of the patient's baseline STarT risk level (i.e. either low, medium, or high). Each patient was treated by the evaluating therapist. Subsequent treatment visits were scheduled following the clinician's directive.

Types of exercises and manual techniques were implemented according to the different pain pattern responses (i.e. CEN or non-CEN) observed from repeated end-range movement tests and/or positioning techniques observed during the initial McKenzie examination (McKenzie and May, 2003). If CEN were observed during the initial visit, exercises, positioning techniques, and functional tasks based on the patient's directional preference were prescribed with the aim of facilitating CEN while avoiding movement patterns that peripheralized symptoms. After the patient's symptoms completely centralized and remained stable, the patient's rehabilitation program focused on recovery of function which may include additional types of exercises and functional tasks at the discretion of the treating therapist with the aim to achieve the patient's individual functional goals.

For patients not demonstrating CEN on the first visit, participating physiotherapists performed additional movement testing examining for CEN on subsequent visits (Werneke, Hart, and Cook, 1999). Although there is no consensus on the exact number of visits required to judge CEN, our research group agreed that three visits would be adequate to identify CEN. If CEN was not observed during the episode of care, the rehabilitation program included exercises, such as stretching and/or core stabilization and functional activities at the discretion of the treating therapist while avoiding exercises and movements that peripheralized symptoms.

Regardless of whether CEN was observed or not, the rehabilitation program emphasized return to function tailored to the patient's individual goals while avoiding movements and activities that peripheralized symptoms. All patients received the same educational approach to empower the patient to become actively involved in his or her own recovery to reduce fear of physical activity and movement intolerance. All treatments rendered by participating physiotherapists during the study were documented following a standardized intervention taxonomy (Werneke, Hart, Deutscher, and Stratford, 2011). Chance-corrected inter-rater agreement among credentialed McKenzie raters using this treatment taxonomy was good to excellent (Werneke, Hart, Deutscher, and Stratford, 2011). The present project is an observational study and, therefore, no attempt was made to standardize care beyond the above guidelines. Our study design does not allow analyses of the efficacy of interventions rendered.

Data collection

Patients completed STarT (Hill et al., 2008) and Lumbar Computer Adaptive Test (LCAT) FS measures (Hart et al., 2010) at intake and discharge. Discharge FS was the dependent variable. Intake and discharge outcome data collection was administered by the front office staff, however if office staff were unavailable the surveys were administered by the clinician. To decrease the possibility of influencing the patient's survey responses, participating clinicians were instructed on how to administer Patient Reported Outcome Measures (PROMs) in a neutral manner to their patients.

To evaluate for the potential for selection bias, we analyzed differences between subjects with complete versus missing discharge data across numerous case mix variables. The criteria we used for selection of these variables were based on: (1) prevalence of health conditions in secondary care (Iezzoni, 1994); (2) national clinical practice guidelines and registries for-record keeping (Meerhoff et al., 2017); and (3) variables known to influence patient outcomes independent of treatments prescribed (Resnik, Liu, Hart, and Mor, 2008). Using these criteria we identified the following variables that were available in the FOTO database: FS at admission (continuous); pain (0–10 Numeric Pain Rating Scale); age (four levels); gender (two levels); symptom duration (i.e. calendar days between date of condition onset to date of initial evaluation) (three levels); payer groups (six levels); medication use at intake for the lumbar impairment (two levels); exercise history (three levels); surgical history for lumbar impairment (two levels); prior treatment episode (i.e. none, any); condition complexity determined by the number of comorbid conditions (four levels based on 29 possible comorbid conditions); post graduate McKenzie credentials (two levels); and CEN category (three levels). These variables were included for consideration because previous research identified these factors as potential confounders influencing FS outcomes in large rehabilitation datasets (Deutscher et al., 2009; Werneke et al., 2016).

Psychosocial classification

We classified subjects at baseline using the three STarT risk categorization levels (i.e. low, medium, or high)

(Hill et al., 2008). Research findings demonstrated that STarT had acceptable test retest reliability (weighted kappa values of 0.76–0.79) and excellent validity for predicting long-term disability in a cohort of patients with LBP managed in a primary or secondary care setting (Beneciuk et al., 2013; Beneciuk, Fritz, and George, 2014; Foster et al., 2014; George and Beneciuk, 2015; Hill et al., 2008, 2011; Von Korff et al., 2014; Wideman et al., 2012). STarT risk levels were re-classified into three change STarT subgroups at discharge following previously published recommendations (Beneciuk, Fritz, and George, 2014). The three patient risk subgroups identified at discharge were improved, stable low or medium, and worsened including stable high category.

Outcome

The LCAT FS measure used in the study has been previously described in detail (Hart et al., 2012, 2011, 2010). FS scores estimated by the LCAT FS measure ranged from 0 (low functioning) to 100 (high functioning) on a linear metric. The items in the LCAT FS measure item bank have demonstrated internal consistency. The FS measure also demonstrated reliability, validity, sensitivity, and responsive- (minimal clinically important difference (MCID) averaged 5 points) (Hart et al., 2012, 2010).

Data analyses

We constructed two multivariable linear regression models to analyze differences in change in function, one model assessing baseline STarT risk categories and the second model assessing change in STarT risk levels at discharge. First, we included intake function and STarT risk category into the models in which FS at discharge was the dependent variable. To control for potential confounding by patient symptoms and clinician training levels, we than added two variables representing whether the patient demonstrated CEN at intake and whether the clinician was certified as having achieved basic competency or Diplomat. Our final step was to evaluate for collinearity by looking at the condition index. Condition indices less than 30 indicate no collinearity between covariates (SAS Institute Inc., SAS Online Doc®, Version 8, Cary, NC).

We examined the potential for patient selection bias related to missing FS outcome data at discharge across baseline STarT risk by comparing subjects with complete intake and discharge data on the aforementioned 13 variables with those missing discharge FS data. Completion rate was defined as the percentage of patients for whom FS outcomes was assessed both at admission and again at discharge (Deutscher et al., 2008).

Results

Descriptive statistics for 705 patients with complete and missing data are reported in Table 1. Completion rates were 76.0%, 70%, 72%, and 83% for entire sample, high STarT risk, medium risk, and low risk, respectively. We observed a higher loss to follow up among patients classified at baseline high STarT risk (30%) and medium risk (28%) compared to low risk levels (17%). Patients lost to follow up were younger and demonstrated CEN; both of which are potentially associated with improved FS outcomes. However, subjects lost to follow up had a slightly higher percent of patients classified into high or medium STarT risk levels compared to low risk and were not managed by McKenzie trained Diplomats, both of which are potentially associated with worse FS outcomes.

Forty percent, 45%, and 15% were classified into low, medium, or high STarT risk levels at baseline, respectively. A large majority of patients (91.5%) at high baseline STarT risk decreased to either low (67%) or medium risk (25%), whereas 8.5% remained at high STarT risk at discharge. Overall percentages in change in STarT risk at discharge were 49.4% improved, 39.0% remained stable low, 8.5% remained stable medium, and 3.1% remained high or worsened.

FS outcomes were not significantly different (p-values > 0.10; condition index 15.49) across baseline STarT risk when controlling for all variables in our model (Table 2). However, FS outcomes were significantly different when comparing subjects whose STarT risk levels improved, remained stable, or worsened when controlling for all variables in our model (Table 3). When compared with subjects whose STarT risk level decreased, there was no difference in subjects whose STarT risk remained low. Yet for subjects whose risk remained stable medium or high, or whose risk worsened, FS outcome scores were statistically significant ($p < 0.001$) and clinically relevant (−15.76 FS points and −23.42 FS points, respectively) compared to patients whose STarT risk level decreased during the episode of care. The condition index for this association was 13.86.

Discussion

Description of main findings

Our results demonstrated several important findings. First, intake STarT risk level was not

Table 1. Patient, therapist, and examination independent variables across patients classified as high, medium, and low baseline STarT risk categories (n = 705).

Variable	Low risk STarT category (n = 283 or 40%) N % or mean (SD)	Medium risk STarT category (n = 316 or 45%) N % or mean (SD)	High risk STarT category (n = 106 or 15%) N % or mean (SD)	p-Value for the difference in variable values across STarT categories
Intake function	n = 283 mean = 57.0 (10.8)	n = 316 mean = 47.8 (10.5)	n = 106 mean = 40.0 (9.5)	$p < 0.001$
Discharge function	n = 283 mean = 77.6 (14.2)	n = 316 mean = 71.4 (16.0)	n = 106 mean = 69.6 (18.1)	$p < 0.001$
Age	n = 69	n = 64	n = 35	$p = 0.06$
18 to < 45 years	24.4%	20.3%	33.0%	
45 to < 65 years	n = 97	n = 126	n = 42	
65 to < 75 years	34.3%	39.9%	39.6%	
75+ years	n = 81	n = 81	n = 20	
	28.6%	25.6%	18.9%	
	n = 36	n = 45	n = 9	
	12.7%	14.2%	8.5%	
Gender	n = 139	n = 138	n = 41	$p = 0.14$
Male	49.1%	43.7%	38.7%	
Female	n = 144	n = 178	n = 65	
	50.9%	56.3%	61.3%	
Symptom duration	n = 45	n = 79	n = 32	$p = 0.02$
0–21 days	15.9%	25.0%	30.2%	
22–90 days	n = 59	n = 78	n = 26	
> 90 days	20.8%	24.7%	24.5%	
	n = 179	n = 159	n = 48	
	63.3%	50.3%	45.3%	
Exercise history At least 3 times/week	n = 129	n = 124	n = 31	$p = 0.01$
	45.6%	39.2%	29.2%	
1–2 times/week	n = 72	n = 76	n = 25	
Seldom or never	25.4%	24.1%	23.6%	
	n = 82	n = 116	n = 50	
	29.0%	36.7%	47.2%	
Surgical history	n = 232	n = 260	n = 94	$p = 0.25$
None	82.0%	82.3%	88.7%	
Any	n = 51	n = 56	n = 12	
	18.0%	17.7%	11.3%	
Comorbid conditions	n = 15	n = 11	n = 4	$p = 0.02$
None	5.3%	3.4%	3.8	
1 or 2	n = 46	n = 32	n = 12	
3	16.3%	10.1%	11.3%	
4 or more	n = 97	n = 87	n = 30	
	34.3%	27.6%	28.3%	
	n = 125	n = 186	n = 60	
	44.2%	58.9%	56.6%	
MDT diplomat	n = 135	n = 166	n = 47	$p = 0.27$
No	47.7%	52.5%	44.3%	
Yes	N = 148	n = 150	n = 59	
	52.3%	47.5%	55.7%	
Centralization	n = 93	n = 128	n = 40	$p < 0.001$
Yes	32.9%	40.5%	37.7%	
No	n = 107	n = 130	n = 61	
Not classified	37.8%	41.1%	57.5%	
	n = 83	n = 58	n = 5	
	29.3%	18.4%	4.7%	
STarT risk at discharge	n = 275	n = 251	n = 71	$p < 0.001$
Low	97.2%	79.4%	67.0%	
Medium	n = 8	n = 60	n = 26	
High	2.8%	19.0%	24.5%	
	n = 0	n = 5	n = 9	
	0%	1.6%	8.5%	

STarT = Subgroups for Targeted Treatment Back Screening Tool (Hill et al., 2008).
Payer type: HMO = Health maintenance Organization, PPO = Preferred Provider Organization, MDT = Mechanical Diagnosis & Therapy.

associated with FS outcomes at discharge for patients with NSLBP. Second, although baseline STarT risk was not predictive of discharge FS outcomes, a substantial number of patients (49.4%) decreased STarT risk levels during treatment. A large majority (91.5%) of patients classified at high STarT risk at intake decreased STarT risk levels by discharge, and of those 67% and 25% reduced to low and medium risk respectively. Third, significant and clinically important improvements in FS outcomes

Table 2. Association between baseline STarT categories, and change in function during the physiotherapy episode of care.

	Baseline model (only intake function entered as an independent variable) Model r2 = 0.16 beta = 0.54 *p*-value < 0.001	Association between STarT category and change in function Model r2 = 0 .16 beta *p*-values	Association between STarT category and change in function controlling for risk adjustment variables Model r2 = 0.22 beta *p*-values
STarT 2 vs. 1 (reference)		−1.34 0.30	−1.81 0.15
STarT 3 vs. 1 (reference)		−0.01 1.00	−0.13 0.94

STarT = Subgroups for Targeted Treatment Back Screening Tool.
Baseline STarT 1 = low risk, Baseline STarT 2 = medium risk, Baseline STarT 3 = high risk.

Table 3. Association between change in STarT categories, and change in function during the physiotherapy episode of care.

	Baseline model	Association between STarT	Association between STarT
	Model r2 = 0.16 beta = 0.54 *p*-value < 0.001	Category and change in function Model r2 = 0.30 beta *p*-value	Category and change in function controlling for risk adjustment variables Model r2 = 0.34 beta *p*-value
STatT risk remained a 1 vs. STarT risk improved* (reference standard)		−0.86 0.47	−0.32 0.90
STarT risk remained a 2 vs. STarT risk improved*		−16.78 < 0.001	−15.76 < 0.001
STarT risk worsened subgroup 3 vs. STarT risk improved*		−23.62 < 0.001	−23.42 < 0.001

STarT = Subgroups for Targeted Treatment Back Screening Tool (Hill et al., 2008).
*Change in STarT subgroup improved i.e., STarT categorization changed from medium to low, high to low, or high to medium risk (reference standard), Change STarT subgroup: 1 = STarT categorization remained low, 2 = STarT categorization remained medium, 3 = STarT categorization worsened, i.e., changed from low to medium, low to high, medium to high, or remained high risk.

were observed for those subjects whose STarT risk improved over the episode of care compared to subjects whose STarT risk remained moderate or high and for those whose STarT risk levels worsened or remained stable high.

How the findings relate to the current literature

Prior studies have also examined the short- and/or long-term prognostic ability for using baseline STarT classification to estimate rehabilitation potential and associations with FS outcomes in physiotherapy secondary care settings in the US (Beneciuk, Fritz, and George, 2014; Fritz, Beneciuk, and George, 2011; Medeiros et al., 2017). Of those only one study examined short-term outcomes at discharge from rehabilitation (Fritz, Beneciuk, and George, 2011). In contrast to our findings, Fritz, Beneciuk, and George (2011) reported that baseline STarT risk stratification may provide important prognostic information for physiotherapists regarding FS outcomes. Comparing our results with this study is not straight forward due to differences in type and timing of the outcome variable, sample sizes, practice settings, and the number of patient case-mix variables controlled for in the analyses. For examples, in the study by Fritz, Beneciuk, and George (2011), the sample size was small (n = 214) and the sample consisted of subjects who were recruited from a single US health care practice system.

Large reductions in STarT risk levels were also observed by Beneciuk, Fritz, and George (2014) when patients with NSLBP were managed by physiotherapists trained in the Treatment-Based Classification (TBC) system (Delitto, Erhard, and Bowling, 1995). It is interesting that the participating clinicians in our study, as well as others (Beneciuk et al., 2013; Beneciuk, Fritz, and George, 2014; Fritz, Beneciuk, and George, 2011) did not receive formal cognitive behavioral training or followed STarT's targeted-treatment pathways as consistently recommended by researchers of the STarT program for successful management of patients with high psychosocial risk (Foster et al., 2010; Hill et al., 2008; Sowden et al., 2012). Previous evidence has demonstrated that specific classification-based physiotherapy approaches, such as McKenzie or TBC which are designed to match treatments to physical examination findings result in

better rehabilitation outcomes compared to nonspecific or passive physiotherapy intervention (Brennan et al., 2006; Fritz, Delitto, and Erhard, 2003; Long, Donelson, and Fung, 2004).

Strengths of the study

We choose an observational research design to examine our purposes. Observational design has been recommended to strengthen the external validity of results by reflecting actual data documentation and interpretation using PROM data by physiotherapists working during the real world of everyday clinical practice (Horn et al., 2005; Snyder et al., 2011). We also included data collection from participating clinicians working in multiple practice settings, i.e., military, hospital-based, and private practice. In addition, our sample size was more than 3× larger than earlier reviews examining associations between STarT risk levels and physiotherapy treatment FS outcomes (Beneciuk, Fritz, and George, 2014; Fritz, Beneciuk, and George, 2011; Medeiros et al., 2017). We believe the strengths of our study enhance the generalizability and clinical interpretation of our findings.

Limitations of the study

We did not track the type of treatment provided by physiotherapists, rather we reported that treatment was prescribed at the clinician's discretion. Although we could not directly examine the validity of targeted treatment approaches based on McKenzie methods, our results evaluating the FS outcomes of patients in different STarT risk categories in response to physiotherapy delivered under routine clinical practice should be considered for future research. While observational studies similar to ours are not ideal for determining cause and effect, our results may have important implications for designing RCT studies examining matched treatment-pathway efficacy in outpatient rehabilitation settings.

Our results need to be interpreted in light of missing data in that a higher loss to follow up rate was observed among patients classified at baseline high STarT risk (30%) and medium risk (28%) compared to low risk levels (17%). Although we observed a higher loss to follow up by patients with elevated risk, the large majority of these patients reduced their STarT risk level and that reduction in risk during the episode of care was associated with significant and clinically important improvements in FS scores at discharge from rehabilitation. Additionally, we could not verify if data from all potential and eligible subjects were entered into the FOTO system. We recommend that future observational studies verify participation rates using providers' electronic medical records.

Participating clinicians were not blinded to baseline STarT risk levels, therefore it is plausible that this may have influenced the physiotherapist's management of the patient. However, we believe it is unlikely that clinicians tailored treatment pathways by STarT stratification because of their interest and training in McKenzie treatment and classification methods. Finally, our results best apply to clinicians with certification in McKenzie methods. The generalizability of our results would be improved by including physiotherapists trained and not trained in MDT methods.

Clinical relevance

Our findings suggest that baseline STarT risk stratification does not differentiate FS outcomes at discharge from physiotherapy and that the majority of patients decreased risk or remained at low risk during the episode of care when managed by physiotherapists credentialed in McKenzie methods. Therefore, management of patients with NSLBP by McKenzie methods might attenuate some of the physical and psychosocial impairments identified by STarT. Our results also suggest that changes in STarT risk may be more informative clinically for managing patients with NSLBP compared with a pretreatment STarT risk classification assessment. Serial STarT risk assessments at different points during an episode of care may enhance treatment monitoring to identify patients who may continue to be at increased risk for poor outcomes and who may require additional psychosocial screening or prompt a referral to healthcare specialists trained in managing patients with recalcitrant psychosocial issues. Future studies are required to determine the optimal time to screen for changes in STarT risk during the physiotherapy episode of care.

Declaration of Interest

The authors report no conflict of interest.

References

Al-Obaidi SM, Al-Sayegh NA, Ben Nakhi H, Skaria N 2013 Effectiveness of McKenzie intervention in chronic low back pain: A comparison based on the centralization phenomenon utilizing selected bio-behavioral and physical

measures. International Journal Physical Medicine Rehabilitation 1: 128.

Beneciuk JM, Bishop MD, Fritz JM, Robinson ME, Asal NR, Nisenzon AN, George SZ 2013 The STarT Back Screening Tool and individual psychological measures: Evaluation of prognostic capabilities for low back pain clinical outcomes in outpatient physical therapy settings. Physical Therapy 93: 321–333.

Beneciuk JM, Fritz JM, George SZ 2014 The STarT Back Screening Tool for prediction of 6-month clinical outcomes: Relevance of change patterns in outpatient physical therapy settings. The Journal of Orthopaedic and Sports Physical Therapy 44: 656–664.

Brennan GP, Fritz JM, Hunter SJ, Thackeray A, Delitto A, Erhard RE 2006 Identifying subgroups of patients with acute/subacute "nonspecific" low back pain: Results of a randomized clinical trial. Spine 31: 623–631.

Browder DA, Childs JD, Cleland JA, Fritz JM 2007 Effectiveness of an extension-oriented treatment approach in a subgroup of subjects with low back pain: A randomized clinical trial. Physical Therapy 87: 1608–1618.

Delitto A, Erhard RE, Bowling RW 1995 A treatment-based classification approach to low back syndrome: Identifying and staging patients for conservative treatment. Physical Therapy 75: 470–485.

Deutscher D, Hart DL, Dickstein R, Horn SD, Gutvirtz M 2008 Implementing an integrated electronic outcomes and electronic health record process to create a foundation for clinical practice improvement. Physical Therapy 88: 270–285.

Deutscher D, Horn SD, Dickstein R, Hart DL, Smout RJ, Gutvirtz M, Ariel I 2009 Associations between treatment processes, patient characteristics, and outcomes in outpatient physical therapy practice. Archives of Physical Medicine and Rehabilitation 90: 1349–1363.

Foster NE, Mullis R, Hill JC, Lewis M, Whitehurst D, Doyle C, Konstantinou K, Main C, Somerville S, Sowden G, Wathall S, Young J, Hay EM 2014 Effect of stratified care for low back pain in family practice (IMPaCT Back): A prospective population-based sequential comparison. Annals of Family Medicine 12: 102–111.

Foster NE, Mullis R, Young J, Doyle C, Lewis M, Whitehurst D, Hay EM 2010 IMPaCT Back Study team IMPaCT Back study protocol. Implementation of subgrouping for targeted treatment systems for low back pain patients in primary care: A prospective population-based sequential comparison. BMC Musculoskeletal Disorders 11: 186.

Fritz JM, Beneciuk JM, George SZ 2011 Relationship between categorization with the STarT Back Screening tool and prognosis for people receiving physical therapy for low back pain. Physical Therapy 91: 722–732.

Fritz JM, Delitto A, Erhard RE 2003 Comparison of classification-based physical therapy with therapy based on clinical practice guidelines for patients with acute low back pain: A randomized clinical trial. Spine 28: 1363–1371.

George SZ, Beneciuk JM 2015 Psychological predictors of recovery from low back pain: A prospective study. BMC Musculoskeletal Disorders 16: 49.

Hart DL, Stratford PW, Werneke MW, Deutscher D, Wang YC 2012 Lumbar computerized adaptive test and modified oswestry low back pain disability questionnaire: relative validity and important change. Journal of Orthopaedic and Sports Physical Therapy 42: 541–551.

Hart DL, Werneke MW, Deutscher D, George SZ, Stratford PW, Mioduski JE 2011 Using intake and change in multiple psychosocial measures to predict functional status outcomes in people with lumbar spine syndromes: A preliminary analysis. Physical Therapy 91: 1812–1825.

Hart DL, Werneke MW, Wang YC, Stratford PW, Mioduski JE 2010 Computerized adaptive test for patients with lumbar spine impairments produced valid and responsive measures of function. Spine 35: 2157–2164.

Hill JC, Dunn KM, Lewis M, Mullis R, Main CJ, Foster NE, Hay EM 2008 A primary care back pain screening tool: identifying patient subgroups for initial treatment. Arthritis and Rheumatism 59: 632–641.

Hill JC, Whitehurst DG, Lewis M, Bryan S, Dunn KM, Foster NE, Konstantinou K, Main CJ, Mason E, Somerville S, Sowden G, Vohora K, Hay EM 2011 Comparison of stratified primary care management for low back pain with current best practice (STarT Back): A randomised controlled trial. Lancet 378: 1560–1571.

Horn SD, DeJong G, Ryser DK, Veazie PJ, Teraoka J 2005 Another look at observational studies in rehabilitation research: Going beyond the holy grail of the randomized controlled trial. Archives of Physical Medicine and Rehabilitation 86(12 Suppl 2): S8–S15.

Iezzoni LI 1994 Using risk-adjusted outcomes to assess clinical practice: An overview of issues pertaining to risk adjustment. Annals of Thoracic Surgery 58: 1822–1826.

Long A, Donelson R, Fung T 2004 Does it matter which exercise? A randomized control trial of exercise for low back pain. Spine 29: 2593–2602.

Long A, May S, Fung T 2008 The comparative prognostic value of directional preference and centralization: A useful tool for front-line clinicians? Journal of Manual Manipulative Therapy 16: 248–254.

Machado LA, Maher CG, Herbert RD, Clare H, McAuley JH 2010 The effectiveness of the McKenzie method in addition to first-line care for acute low back pain: A randomized controlled trial. BMC Medicine 8: 10.

Main CJ, Sowden G, Hill JC, Watson PJ, Hay EM 2012 Integrating physical and psychological approaches to treatment in low back pain: the development and content of the STarT Back trial's 'high-risk' intervention (StarT Back; ISRCTN 37113406). Physiotherapy 98: 110–116.

Manca A, Dumville JC, Torgerson DJ, Moffett K, Mooney MP, Eaton S 2007 Randomized trial of two physiotherapy interventions for primary care back and neck pain patients: cost effectiveness analysis. Rheumatology 46: 1495–1501.

May S, Aina A 2012 Centralization and directional preference: A systematic review. Manual Therapy 17: 497–506.

May S, Donelson R 2008 Evidence-informed management of chronic low back pain with the McKenzie methods. Spine Journal 8: 134–141.

McKenzie R, May S 2003 The Lumbar Spine: Mechanical Diagnosis and Therapy. 2nd ed. Waikanae, NZ: Spinal Publications New Zealand Ltd.

Medeiros FC, Pena Costa LO, Nemitalla MA, Salomao EC, Costa LC 2017 Serial monitoring of patients with chronic low back pain during physical therapy treatment using the STarT Back Screening Tool. Journal of Orthopaedic and Sports Physical Therapy 47: 314–323.

Meerhoff GA, Van Dulmen SA, Maas MJ, Heijblom K, Nijhuis-Van Der Sanden MW, Van Der Wees PJ 2017 Development

and evaluation of an implementation strategy for collecting data in a national registry and the use of patient-reported outcome measures in physical therapist practices: Quality improvement study. Physical Therapy 97: 837–851.

Nicholas MK, Linton SJ, Watson PJ, Main CJ 2011 "Decade of the Flags" Working G: early identification and management of psychological risk factors ("yellow flags") in patients with low back pain: a reappraisal. Physical Therapy 91: 737–753.

Resnik L, Liu D, Hart DL, Mor V 2008 Benchmarking physical therapy clinic performance: Statistical methods to enhance internal validity when using observational data. Physical Therapy 88: 1078–1087.

Snyder CF, Aaronson NK, Choucair AK, Elliott TE, Greenhalgh J, Halyard HY, Hess R, Miller DM, Reeve BB, Santana M 2011 Implementing patient-reported outcomes assessment in clinical practice: A review of the options and considerations. Quality Life Research 21: 1305–1314.

Sowden G, Hill JC, Konstantinou K, Khanna M, Main C, Salmon P, Somerville S, Wathall S, Foster NE 2012 IMPaCT Back study team. Targeted treatment in primary care for low back pain: The treatment system and clinical training programs used in the IMPaCT Back study. Family Practice 29: 50–62.

Swinkels IC, Van Den Ende CH, De Bakker D, Van Der Wees P, Hart DL, Deutscher D, Bosch WJ, Van Den, Dekker J 2007 Clinical databases in physical therapy. Physiotherapy Theory and Practice 23: 153–167.

Von Korff M, Shortreed SM, Saunders KW, Leresche L, Berlin JA, Stang P, Turner JA 2014 Comparison of back pain prognostic risk stratification item sets. Journal of Pain 15: 81–89.

Werneke MW, Edmond S, Deutscher D, Ward J, Grigsby D, Young M 2016 Effect of adding McKenzie syndrome, centralization, directional preference, and psychosocial classification variables to a risk-adjusted model predicting functional status outcomes for patients with lumbar impairments. Journal of Orthopaedic and Sports Physical Therapy 46: 726–741.

Werneke MW, Hart DL, Cook D 1999 A descriptive study of the centralization phenomenon: A prospective analysis. Spine 24: 676–683.

Werneke MW, Hart DL, Cutrone G, Oliver D, McGill T, Weinberg J, Grigsby D, Ward J 2011a Association between directional preference and centralization in patients with low back pain. Journal of Orthopaedic and Sports Physical Therapy 41: 22–31.

Werneke MW, Hart DL, Deutscher D, Stratford PW 2011 Clinician's ability to identify neck and low back interventions: An inter-rater chance-corrected agreement pilot study. Journal Manual Manipulative Therapy 19: 172–181.

Werneke MW, Hart DL, George SZ, Deutscher D, Stratford PW 2011b Change in psychosocial distress associated with pain and functional status outcomes in patients with lumbar impairments referred to physical therapy services. Journal of Orthopaedic and Sports Physical Therapy 41: 969–980.

Werneke MW, Hart DL, George SZ, Stratford PW, Matheson JW, Reyes A 2009 Clinical outcomes for patients classified by fear-avoidance beliefs and centralization phenomenon. Archives of Physical Medicine and Rehabilitation 90: 768–777.

Wideman TH, Hill JC, Main CJ, Lewis M, Sullivan MJ, Hay EM 2012 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 153: 2182–2191.