***Journal of Sport Rehabilitation*, 2018, 27, 469-479**
https://doi.org/10.1123/jsr.2017-0104
© 2018 Human Kinetics, Inc.

Human Kinetics
SYSTEMATIC REVIEW

Examination of Response Shift After Rehabilitation for Orthopedic Conditions: A Systematic Review

Cameron J. Powden, Matthew C. Hoch, and Johanna M. Hoch

Context: There is an increased emphasis on the need to capture and incorporate self-reported function to make clinical decisions when providing patient-centered care. Response shift (RS), or a change in an individual's self-evaluation of a construct, may affect the accurate assessment of change in self-reported function throughout the course of rehabilitation. A systematic review of this phenomenon may provide valuable information regarding the accuracy of self-reported function. ***Objectives:*** To systematically locate and synthesize the existing evidence regarding RS during care for various orthopedic conditions. ***Evidence Acquisition:*** Electronic databases (PubMed, MEDLINE, CINAHL, SPORTDiscus, and Psychology & Behavioral Sciences Collection) were searched from inception to November 2016. Two investigators independently assessed methodological quality using the modified Downs and Black Quality Index. The quality of evidence was assessed using the Strength-of-Recommendation Taxonomy. The magnitude of RS was examined through effect sizes. ***Evidence Synthesis:*** Nine studies were included (7 high quality and 2 low quality) with a median Downs and Black Quality Index score of 81.25% (range = 56.25%–93.75%). Overall, the studies demonstrated weak to strong effect sizes (range = −1.58–0.33), indicating the potential for RS. Of the 36 point estimates calculated, 22 (61.11%), 2 (5.56%), and 12 (33.33%) were associated with weak, moderate negative, and strong negative effect sizes, respectively. ***Conclusions:*** There is grade B evidence that a weak RS, in which individuals initially underestimate their disability, may occur in people undergoing rehabilitation for an orthopedic condition. It is important for clinicians to be aware of the potential shift in their patients' internal standards, as it can affect the evaluation of health-related quality of life changes during the care of orthopedic conditions. A shift in the internal standards of the patient can lead to subsequent misclassification of health-related quality of life changes that can adversely affect clinical decision making.

Keywords: health-related quality of life, self-reported function, orthopedics, shoulder, knee

Context

The evaluation of change in patient status throughout and after the cessation of orthopedic rehabilitation is a vital component of health care and is often captured from the patient's perspective by patient-based outcomes. Patient-based outcomes are used to assess the effect of the health condition on function at the personal and societal levels while examining concepts related to health-related quality of life (HRQL). There is an increased emphasis regarding the collection of patient-based outcomes to facilitate patient-centered care and quantify change in HRQL status from the patient's perspective.[1,2] HRQL is a broad, multidimensional concept that refers to the physical, psychological, spiritual, economic, and social domains of health that are affected by an individual's experiences, expectations, and perceptions.[2] Clinicians often measure HRQL through the utilization of a variety of patient-reported outcomes (PROs), which can be categorized as generic, region/disease specific, or dimension specific. Each of these instruments is constructed to measure different domains of HRQL and the effects of the health condition and interventions on these domains from the patient's perspective. The use of PROs to identify and categorize HRQL treatment responses is important because the measurement of patient-perceived change, or lack of change, is a key to the development and modification of treatment algorithms and the provision of patient-centered care.[3]

The increased emphasis on PROs to capture HRQL and make clinical decisions that incorporate the patient's perspective suggests that there is an increased demand to ensure accurate documentation of these outcomes. Because the concept of HRQL is rooted in the individual's perception, the commonly used measures automatically assume that the intraindividual standards remain stable throughout the rehabilitation process.[4,5] However, this may not be true, as it is reasonable to believe that patient values can change, particularly in cases where the condition is present for a prolonged period of time prior to intervention.[4,5] These changes characterize the phenomenon known as response shift (RS).[5,6] Response shift phenomenon is when an individual's self-evaluation of a construct is altered due to changes in internal standards of measurement (recalibration), changes in values (reprioritization), or a personal redefinition of the construct (reconceptualization).[5,6] Changes in self-evaluation may be a direct or indirect result of the rehabilitation that the patient is receiving due to their health condition. Changes in an individual's values, standards, or priorities throughout the rehabilitation process are hypothesized to lead to new conceptualization of the constructs in which the PROs are used to measure. If a patient shifts their responses on the PROs due to these changes, an inaccurate estimate of treatment effects may occur, which could impact clinical decision making.[5]

Response shift has been extensively evaluated in individuals with chronic, life-threatening conditions such as cancer.[6] Recently, there has been an increase in the number of studies that examine RS phenomenon in individuals with chronic musculoskeletal conditions. These studies have all demonstrated a degree of RS after surgical intervention and subsequent rehabilitation for patients

Powden is with the Post-Professional Doctorate in Athletic Training Program, Indiana State University, Terre Haute, IN. M.C. Hoch and J.M. Hoch are with the Division of Athletic Training, University of Kentucky, Lexington, KY. Powden (Cameron.Powden@indstate.edu) is corresponding author.

with arthritis,[7] spinal conditions,[8] rotator cuff tears,[9] and cartilage lesions in the knee,[10] utilizing the then-test method to quantify RS. However, formal synthesis of the aforementioned literature has not been completed to evaluate the magnitude of RS throughout the orthopedic rehabilitation. The completion of a systematic review of the literature would improve our understanding of RS' effect on the evaluation of HRQL following orthopedic rehabilitation. Thus, the purpose of this systematic review was to compile, critically appraise, and synthesize the published evidence that investigated the presence of RS following orthopedic rehabilitation.

Evidence Acquisition

Search Strategy

A systematic search was conducted based on the PRISMA guidelines[11] to locate studies that assessed RS after rehabilitation for an orthopedic condition. Online databases were searched with a combination of key words related to RS and self-reported outcomes (Table 1). Boolean operators "OR" and "AND" were utilized to combine search terms, and the search was limited to peer-reviewed, full-text manuscripts written in English. This systematic review was completed by 3 investigators who were experienced with the development and completion of systematic reviews.

Two investigators (C.J.P. and J.M.H.) derived the Boolean phrase and completed the systematic search. PubMed and EBSCOhost (CINAHL, MEDLINE, SPORTDiscus, and Psychology & Behavioral Sciences Collection) were searched from their inception through November 17, 2016. In addition, the reference lists of articles screened for inclusion were hand searched for publications that were not identified through the electronic search.

Eligibility Criteria

Two investigators (C.J.P. and J.M.H.) reviewed the articles identified by the systematic search for possible inclusion in the review. The titles and abstracts of all identified articles were screened for inclusion based on the criteria listed below. In cases of inclusion uncertainty, the full text of the manuscript was screened for inclusion.

Inclusion Criteria. The inclusion criteria used to select and screen the studies for inclusion into the systematic review were as follows:

- Studies that aimed to examine the presence of RS in individuals with orthopedic conditions after an intervention.
- Studies that utilized any method of evaluating RS (eg, then-test method, relative importance method, statistical approach, etc).
- Studies that included human participants who underwent rehabilitation and/or surgical interventions for an orthopedic condition.
- Studies that utilized PROs. No restrictions were made to the type of PRO used in the study.

Exclusion Criteria. The exclusion criteria used to screen studies for their suitability for exclusion were as follows:

- Articles that did not report or provide sufficient data to calculate the magnitude and direction of RS following an intervention.[12]
- Articles that included subjects whose rehabilitation was not for an orthopedic condition, such as spinal cord surgery, cancer treatment, or rheumatoid arthritis.[13,14]
- Articles that were not published in English.
- Articles that were case studies or case reviews.

Assessing Quality of Studies

Two investigators (C.J.P. and J.M.H.) independently assessed the quality of each of the included studies using a 16-item version of the original Downs and Black Quality Index (DBQI).[15,16] The DBQI was developed to critically appraise both randomized and nonrandomized studies.[15] The DBQI consists of questions that assess the internal and external validity as well as clarity in the reporting of the hypotheses, main outcomes, subject characteristics, and main findings. The DBQI has demonstrated acceptable reliability and internal consistency.[15] Disagreements between investigators were resolved by discussion and/or by a third reviewer (M.C.H.). Studies that met ≥60% of the criteria were deemed high quality and those that meet <60% were considered limited quality.[16]

Data Extraction

Two investigators (C.J.P. and J.M.H.) extracted data during the initial review that included study aims, study design, participant

Table 1 Search Strategy

Step	Search terms	Boolean operator	EBSCOhost	PubMed
1	Response shift	OR	3340	2095
	Recalibration			
	Reprioritization			
	Reconceptualization			
2	Health-related quality of life	OR	477,985	357,780
	Quality of life			
	Self-reported			
	Patient reported			
3	2 + 3	AND	263	302
Duplicates[a]				245
Hand search				1
Total identified				320

[a]Total number of duplicates between EBSCO and PubMed.

details, intervention details, outcome assessments, RS technique, statistical technique, and conclusions. Discussion or a third reviewer (M.C.H.) was used to resolve discrepancies in interpretations and achieve consensus. The evaluation of RS was further categorized based on type of PRO that was used to capture patient-perceived function and HRQL. The 3 categories of PROs used in the included studies were generic, region specific, and other. Generic outcomes are those designed to assess the patient's overall health and can be used to assess detriments to HRQL at the personal and societal level (eg, SF-36). Region-specific outcomes are designed to assess the effect of a health condition as it relates to function of a specific joint or region of the body (eg, International Knee Documentation Committee). Outcomes that fell outside the scope of region specific, dimension specific, and generic, or those for which it could not be determined what aspect of health was evaluated, were categorized as other.

Statistical Analysis

The magnitude of RS was examined through reported,[17] calculated Hedges' *g* effect sizes (ESs),[9,10,18–20] and standardized response mean ESs[7,21] with 95% confidence intervals. Hedges' *g* and standardized response mean ESs are unitless measures that represent the effect that exists on a parametric distribution.[22] Hedges' *g* was used in all cases where appropriate data were provided for analysis. If sufficient data were not provided, standardized response mean ESs were calculated. For all analyses, ESs were oriented, so positive ESs indicated participants estimated their disablement to be greater at their pretest compared with their then-test assessment. Conversely, negative ESs indicated participants estimated their disablement to be less at their pretest compared with their then-test assessment. ESs were interpreted as weak (≤0.40), moderate (0.41–0.69), or strong (≥0.70).[22] For studies where multiple subscales were reported, an average value was calculated to avoid excessive weighting from individual studies.[23] Averaging of values was completed for 2 studies.[18,21] For Finkelstein et al,[21] SF-36 ES estimates were averaged to create a summary effect for the physical component summary score and the mental component summary score for each time point. For Nagl and Farin,[18] ESs were averaged to create a single summary score. To synthesize ESs across studies, point estimates for overall RS as well as generic, region-specific, and other outcomes were examined descriptively using minimum, maximum, and categorical breakdowns.[23]

Level of Evidence

The quality of individual studies as well as the body of evidence was assessed using the Strength-of-Recommendation Taxonomy.[24] Each of the individual included studies was deemed as level 1, 2, or 3 evidence. Level 1 evidence was considered good-quality (DBQI score of ≥60%) patient-oriented evidence; level 2 evidence was considered limited-quality (DBQI score of <60%) patient-oriented evidence; and level 3 was considered other evidence.[24] To assess the collective body of evidence, the Strength-of-Recommendation Taxonomy assigns a strength of recommendation. The strength of recommendation for the Strength-of-Recommendation Taxonomy considers a grade of A as consistent, good-quality patient-oriented evidence; B as inconsistent or limited-quality patient-oriented evidence; and C as consensus evidence, disease-oriented evidence, and so forth.[24]

Sensitivity Analysis

The effect of methodologic quality criteria on the strength of recommendation was tested by subjecting the quality of evidence scores, as assessed using the DBQI, to changes of ±10%.[25] After the scores were subjected to this change, the potential modification in the strength of recommendation was determined to assess the sensitivity of the overall recommendation.

Evidence Synthesis

Literature Search

The flow of articles through the search and review process is illustrated in Figure 1. Of the 13 articles assessed for eligibility, 9[7,9,10,17–21,26] met the inclusion criteria for this systematic review. Of the 4 studies that were excluded, 1 study was excluded due to methodology that did not allow for RS ES calculation,[12] 1 was excluded as it was a clinical commentary,[27] and 2 were excluded because their subject populations did not undergo rehabilitation for orthopedic musculoskeletal conditions.[13,14] A summary of study characteristics for all included studies can be found in Table 2.

Methodological Quality

The results of the quality assessment can be found in Table 3. Individual DBQI scores can be found in the Appendix. The 2 investigators initially agreed on 124 out of 144 (86.11%) items on the DBQI. All disagreements were resolved by discussion among the 2 investigators and primarily pertained to items assessing blinding, accuracy of outcomes measures, and adjustment for

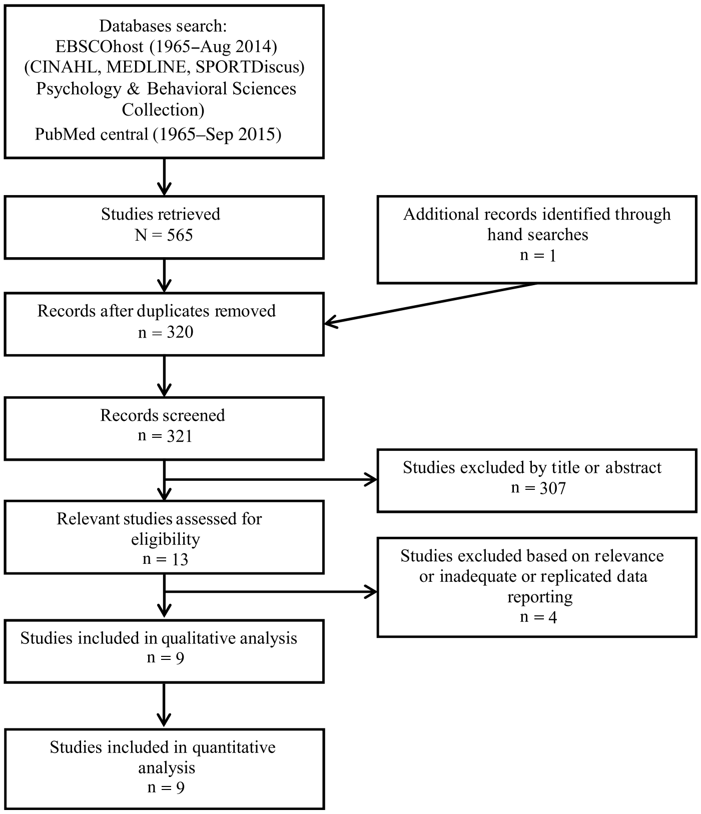
Databases search:
EBSCOhost (1965–Aug 2014)
(CINAHL, MEDLINE, SPORTDiscus)
Psychology & Behavioral Sciences Collection)
PubMed central (1965–Sep 2015)
Studies retrieved
N = 565
Additional records identified through hand searches
n = 1
Records after duplicates removed
n = 320
Records screened
n = 321
Studies excluded by title or abstract
n = 307
Relevant studies assessed for eligibility
n = 13
Studies excluded based on relevance or inadequate or replicated data reporting
n = 4
Studies included in qualitative analysis
n = 9
Studies included in quantitative analysis
n = 9

Figure 1 — Flowchart of literature review.

Table 2 Methodologic Summary of the Included Studies

Author	Sample	Subject characteristics	Intervention	Data collection time points	Response shift measurement	Outcome measures	Results
Hollman et al[26]	36 patients (16 males; 61.2 [8.22] y and 27.5 [3.72]) kg/m^2	Inclusion: traumatic or degenerative full-thickness tear of the supraspinatus or infraspinatus, or both as diagnosed with MRI. Symptoms for greater than 6 mo and unresponsive to at least 3 mo of conservative therapy Exclusion: partial-thickness tear, perioperative irreparable or partially reparable, revision surgery, rupture of the subscapularis tendon, glenohumeral osteoarthritis, adhesive capsulitis, body mass index >35 kg/m^2, fibromyalgia, current treatment with opiates, concomitant labral repair, lateral clavicle resection, and unable to understand Dutch	Arthroscopic rotator cuff repair in the beach chair position using a double-row or single-row suture bridge technique. Patients were immobilized after surgery and passive exercises were completed. Immobilization was phased out at 6 wk, starting with active-assisted and strengthening exercises.	Preoperation, 3 mo, and 1 y	Then-test (3 mo and 1 y)	WORC	No significant group-level response shift for the WORC between any of the time points. Patients rated their then-tests similarly to their baseline measures.
Howard et al[10]	48 patients (29 males; 35 [8.0] y, 180.7 [31.7] cm, and 92.4 [20.3] kg)	Inclusion: planned ACI surgery to the knee, willingness to participate, and no uncorrectable contraindications to ACI. Exclusion: undergoing concomitant meniscal transplant	Two-step ACI procedure, standardized rehabilitation protocol following surgery.	Preoperation, 6 mo, and 12 mo	Then-test (6 and 12 mo)	SF-36 PCS, WOMAC, IKDC, and Lysholm	No significant group-level response shift for any of the outcome measures.
Finkelstein et al[21]	169 patients (51.96 [16.46] y)	Inclusion: undergoing posterior lumbar spinal decompression surgery for spinal stenosis or disc herniation Exclusion: unable to complete questionnaires in English, visual or cognitive impairments, and disability that prevents independent completion	Posterior lumbar spinal decompression surgery.	Preoperation, 6 wk, and 3 mo	Then-test (6 wk and 3 mo)	ODI and 8 subscales of the SF-36	Significant differences from preoperation and 6 wk and 3 mo then-tests for all outcomes but SF-36 MHS.
Zhang et al[20]	74 patients (14 males; 68 [63–76] y)	Inclusion: undergoing total knee replacement Exclusion: cognitive impairment, unable to speak English or Mandarin, and undergoing additional surgery	Total knee replacement surgery.	Preoperation and 18 mo	Then-test (18 mo)	SF-6D and EQ-5D	There was a significant difference in then-test scores and preoperation scores for both the SF-6D and EQ-5D.
Nagl and Farin[18]	189 patients	Inclusion: chronic low back pain	Unspecified rehabilitation	Prerehabilitation and postrehabilitation	Then-test (postrehabilitation)	Self-devised questionnaire with 1 question to address pain, mobility, activities, emotional well-being, knowledge, cognitive coping, behavioral coping, behavior, family, and work Based on FESV, ODI, and SF-12	Then-test scores were significantly less then prerehab scores.

(continued)

Table 2 (continued)

Author	Sample	Subject characteristics	Intervention	Data collection time points	Response shift measurement	Outcome measures	Results
Razmjou et al[9]	107 patients (66 males; 57 ± 12 y)	Inclusion: patients returning for 2-y follow-up after rotator cuff surgery and unremitting pain in shoulder for 6 mo prior to surgery	Arthroscopic decompression, arthroscopic rotator cuff repair, or open rotator cuff repair.	Preoperation and 2 y	Then-test (2 y)	ASES (pain and ADL)	Significant difference in preoperation and then-test was identified in the pain domain of the ASES. No response shift occurred in the functional ability domain.
Razmjou et al[7]	236 patients (82 males; 67 ± 10 y)	Inclusion: candidates for total knee replacement arthroplasty	Total knee replacement arthroplasty	Baseline, 6 mo, and 1 y	Then-test (6 mo and 1 y)	WOMAC, SF36-PCS, and SF36 MCS	There were significant differences in baseline and then-test scores 6 mo and 1 y for the WOMAC and SF36-PCS. Significant differences were identified at 1 y for the SF36-MCS.
Balain et al[17]	53 patients (36 males; 42 [32–48] y)	Inclusion: undergoing knee microfracture surgery	Knee microfracture surgery	Pretest and >6 mo	Then-test (mean 34 mo)	VAS pain, Lysholm, IKDC-SA, and IKDC-S	Significant differences in pre-test and then-test were identified in each instrument.
Razmjou et al[19]	125 patients (34 males; 68 ± 9.5 y)	Inclusion: individuals undergoing total knee replacement between November 2004 and October 2005 Exclusion: previous total joint arthroplasty, required language translation, visual or cognitive problems, or were unable to complete questionnaires independently	Total knee replacement arthroplasty	Pretest and 6 mo	Then-test (6 mo)	WOMAC	Significant differences in pretest and then-test were identified for the overall WOMAC score as well as pain and physical function domains.

Abbreviations: ACI, Autologous Chondrocyte Implantation; ADL, Activities of Daily Living; ASES, American Shoulder and Elbow Surgeons; FESV, Questionnaire to Assess Pain Processing Fragebogen Zur Erfassung der Schmerzverarbeitug; EQ-5D, EuroQOL Five-Dimensional Questionnaire; IKDC, International Knee Documentation Committee; MCS, Mental Component Scale; ODI, Oswestry Disability Index; PCS, Physical Component Scale; RC, Rotator Cuff; S, symptom; SA, subjective assessment; SF-6D, Six-Dimensional Health State Short Form; SF-12, Short Form 12; SF-36, Short Form 36; VAS, visual analog scale; WOMAC, Western Ontario and McMaster Universities Osteoarthritis Index; WORC, Western Ontario Rotator Cuff Index.

Table 3 Downs and Black Quality Index for the Included Articles

Author	Quality index score, %	Reporting score, %	Internal validity score, %	External validity, %	Level of evidence
Hollman et al[26]	93.75 (15/16)	100.00 (7/7)	100 (7/7)	50.00 (1/2)	2b
Howard et al[10]	81.25 (13/16)	100.00 (7/7)	57.14 (4/7)	50.00 (1/2)	2b
Finkelstein et al[21]	75.00 (12/16)	100.00 (7/7)	57.14 (4/7)	0.00 (0/2)	2b
Zhang et al[20]	81.25 (13/16)	100.00 (7/7)	71.43 (5/7)	0.00 (0/2)	2b
Nagl and Farin[18]	56.25 (9/16)	71.43 (5/7)	42.86 (3/7)	0.00 (0/2)	4
Razmjou et al[9]	81.25 (13/16)	100.00 (7/7)	71.43 (5/7)	0.00 (0/2)	2b
Razmjou et al[7]	56.25 (9/16)	71.43 (5/7)	42.86 (3/7)	50.00 (1/2)	4
Balain et al[17]	75.00 (12/16)	85.71 (6/7)	71.43 (5/7)	0.00 (0/2)	2b
Razmjou et al[19]	87.50 (12/16)	100.00 (7/7)	85.71 (6/7)	0.00 (0/2)	2b

confounding. The overall quality scores of the included studies were a median of 81.25% and a range of 56.26% to 93.75%. Seven[9,10,17,19–21,26] high-quality (>60%) and 2[7,18] limited-quality studies were included. The reporting component of the DBQI had a median of 100.00% (71.43%–100.00%), the internal validity component had a median score of 71.43% (42.86%–100.00%), and the external validity had a median score of 0.00% (0.00%–50.00%).

Study Characteristics

The characteristics of the included studies are displayed in Table 2. All studies included in the review examined RS using the then-test method.[4] One study was identified in the initial search that utilized the relative importance method[12] in an orthopedic population. This study was excluded, however, due to a lack of data needed to calculate comparable ESs. Included subjects underwent surgical intervention and/or a rehabilitation program for an orthopedic condition. Interventions completed included autologous chondrocyte implantation,[10] total knee arthroplasty,[7,19,20] knee microfracture,[17] arthroscopic rotator cuff repair or decompression,[9] open rotator cuff repair,[9,26] lumbar spinal decompression surgery,[21] and unspecified rehabilitation for chronic low back pain.[18] The then-test method was used to evaluate RS for 6 weeks,[21] 3 months,[21,26] 6 months,[7,10,19] 12 months,[7,10,26] 18 months,[20] 24 months,[9] and an unspecified amount of time[17,18] after baseline. The type of PROs used to capture the patients' perception of their health and RS were categorized as generic,[7,10,20,21] region specific,[7,9,10,17,19,21,26] and other.[17,18] None of the included studies used dimension-specific PROs.

Overall the included studies demonstrated weak positive to strong negative ESs for RS with a range of −1.58 to 0.33. Of the 36 point estimates, 12 (33.33%) were strong negative ESs, 2 (5.56%) were moderate negative, and 22 (61.11%) were weak negative or positive. Furthermore, the results can also be examined for the individual types of instruments that were included. The generic instruments demonstrated weak positive to strong negative ESs for RS with a range of −1.31 to 0.19. Of the 12 generic point estimates, 6 (50.00%) were strong negative ESs, 1 (8.33%) was moderate negative, and 5 (41.67%) were weak negative or positive. These instruments most consistently demonstrated the largest ESs with the most individual point estimates that were interpreted as strong. In contrast, region-specific instruments demonstrated weak positive to strong negative ESs with a range of −1.58 to 0.33. Of the 22 region-specific point estimates, 5 (22.72%) were strong negative ESs, 1 (4.55%) was moderate negative, and 16 (72.72%) were weak negative or positive. Other instruments demonstrated weak to strong negative ESs with a range of –0.92 to –0.21. Of the 2 other point estimates, 1 (50.00%) was a strong negative ES and 1 (50.00%) was a weak negative. Individual ESs can be found in Table 4 presented by included study, time point, and outcome measure. When qualitatively examining the ES for each study across time points, there does not appear to be consistent change in interpretation of the individual point estimates as the time of the then-test application increases from baseline.

Level of Evidence

The results of the systematic review indicate there is grade B evidence that a weak RS, in which patients estimated their disability to be less at baseline compared with the then-test method, may occur in patients with orthopedic conditions undergoing care.[7,9,10,17–21,26] This recommendation is based on inconsistent and limited-quality evidence, as 7 studies[9,10,17,19–21,26] were considered high quality and 2 studies[7,18] were considered low quality, and the range of ESs consistently crossed 0. The results were further examined according to PROs type. There is grade B evidence that a strong RS, in which patients estimate their disability to be less at baseline compared with the then-test method, may occur in patients with orthopedic conditions undergoing care when measured using generic instruments.[7,10,20,21] In addition, when evaluating RS using region-specific PROs[7,9,10,17,19,21,26] or other instruments,[17,18] there is grade B evidence that a weak RS, in which patients initially estimate their disability to be less at baseline compared with the then-test method, may occur. However, these values should be interpreted with caution, as many of the ES estimates spanned from negative to positive and were associated with confidence intervals that crossed 0, indicating there may be no effect. This indicates that the patients included in these studies may inconsistently identify their then-test scores compared with their pretest scores.

Sensitivity Analysis

The sensitivity analysis, in which the criterion for study quality was subjected to a ±10%, did not affect the grade of recommendation for any of the analyses. This indicates that the included articles are primarily of high quality and that the findings of this review are not influenced by evidence that is on the lower end of the high-quality criteria.

Table 4 ES and 95% CI for Included Point Estimates

Author	Participants (no.)	Time points	Outcome measure	ES	95% CI
Generic patient-reported outcomes					
Howard et al[10]	ACI (48)	6 mo	SF-36 PCS	0.13	–0.31 to 0.56
	ACI (48)	12 mo	SF-36 PCS	0.19	–0.24 to 0.62
Finkelstein et al[21]	Lumbar decompression (169)	6 wk	SF-36 PCS summary	–0.98[a,c]	NA
	Lumbar decompression (169)	3 mo	SF-36 MCS summary	–0.96[a,c]	NA
	Lumbar decompression (169)	6 wk	SF-36 PCS summary	–0.81[a,c]	NA
	Lumbar decompression (169)	3 mo	SF-36 MCS summary	–0.86[a,c]	NA
Zhang et al[20]	Total knee replacement (74)	18 mo	SF-6D	–0.70	–1.04 to –0.36
	Total knee replacement (74)	18 mo	EQ-SD	–0.96	–1.32 to –0.61
Razmjou et al[7]	Total knee replacement (236)	6 mo	SF-36 PCS	–0.21[a]	NA
	Total knee replacement (236)	6 mo	SF-36 MCS	–0.04[a]	NA
	Total knee replacement (236)	12 mo	SF-36 PCS	–0.40[a]	NA
	Total knee replacement (236)	12 mo	SF-36 MCS	–0.31[a]	NA
Region-specific patient-reported outcomes					
Hollman et al[26]	RC surgery, full tear (36)	3 mo	WORC	0.23	NA
	RC surgery, full tear 36)	12 mo	WORC	0.27	NA
Howard et al[10]	ACI (48)	6 mo	WOMAC	0.25	–0.18 to 0.68
	ACI (48)	6 mo	IKDC	0.11	–0.32 to 0.54
	ACI (48)	6 mo	Lysholm	–0.29	–0.72 to 0.15
	ACI (48)	12 mo	WOMAC	0.11	–0.32 to 0.55
	ACI (48)	12 mo	IKDC	0.08	–0.35 to 0.51
	ACI (48)	12 mo	Lysholm	–0.17	–0.61 to 0.26
Finkelstein et al[21]	Lumbar Decompression (169)	6 wk	ODI	–1.22[a]	1.20 to 1.92
	Lumbar Decompression (169)	3 mo	ODI	–1.58[a]	1.41 to 2.13
Razmjou et al[9]	RC surgery, full tear (44)	24 mo	ASES-Pain	–1.26	0.80 to 1.72
	RC surgery, full tear (44)	24 mo	ASES-ADL	–0.13	–0.55 to 0.29
	RC surgery, partial tear (62)	24 mo	ASES-Pain	–0.95	0.57 to 0.29
	RC surgery, partial tear (62)	24 mo	ASES-ADL	–0.03	–0.38 to 0.32
Razmjou et al[7]	Total knee replacement (236)	6 mo	WOMAC	–0.32[a]	NA
	Total knee replacement (236)	12 mo	WOMAC	–0.40[a]	NA
Balain et al[17]	Knee microfracture surgery (53)	6 mo	Lysholm	0.33[b]	NA
	Knee microfracture surgery (53)	6 mo	IKDC-SA	–0.71[b]	NA
	Knee microfracture surgery (53)	6 mo	IKDC-S	–0.36[b]	NA
Razmjou et al[19]	Total knee preplacement (125)	6 mo	WOMAC-Pain	0.18	–0.05 to 0.45
	Total knee preplacement (125)	6 mo	WOMAC-stiffness	0.01	–0.24 to 0.25
	Total knee preplacement (125)	6 mo	WOMAC-physical function	0.22	–0.03 to 0.47
Other patient-reported outcomes					
Nagl and Farin[18]	Chronic low back pain (189)	Unknown	Custom Questionnaire Summary	–0.21[c]	NA
Balain et al[17]	Knee microfracture surgery (53)	6 mo	VAS	–0.92[b]	NA

Abbreviations: ASES, American Shoulder and Elbow Surgeons; CI, confidence interval; ES, effect size; EQ-5D, EuroQOL Five-Dimensional Questionnaire; IKDC, International Knee Documentation Committee; MCS, Mental Component Scale; NA, not available; PCS, Physical Component Scale; S, Symptom; SA, Subjective Assessment; VAS, Visual Analog Scale; WOMAC, Western Ontario and McMaster Universities Osteoarthritis Index; WORC, Western Ontario Rotator Cuff Index. Note: ESs calculated as Hedges' *g* unless otherwise noted. CI was not included due to a lack of reporting or insufficient data for them to be calculated.

[a]Standardized response mean was calculated. [b]ES reported by article. [c]Simple summary of ES estimates within the study.

Discussion

Summary of Results

The purpose of this systematic review was to critically synthesize the published evidence that investigated the presence of RS in patients with orthopedic conditions who underwent rehabilitation. The results of this review indicate there is grade B evidence that a weak RS (ES = –1.83 to 0.33) may occur in patients with orthopedic conditions undergoing rehabilitation. The nature of this RS indicates patient's prerehabilitation evaluation of HRQL may underestimate their baseline disability compared with their then-test evaluation of the baseline disability. This grade was indicated due to inconsistent findings from level 2 evidence. Furthermore, the presence of RS was strongest when using generic PROs. While no recommendation is being made to utilize a then-test method to assess RS in routine clinical practice, these findings indicate that clinicians should be cognizant of RS when capturing HRQL during the rehabilitation process for orthopedic conditions.

Methodological Considerations

A wide range of orthopedic patient populations undergoing various types of care were included within this systematic review. Care ranged from total knee replacement[7,19,20] and autologous chondrocyte implantation[10] to chronic back pain rehabilitation[18,21] and rotator cuff tear repair.[9,26] All but 1 study[18] evaluated RS following a care plan that included surgical intervention.[7,9,10,17,19–21] These articles primarily indicated that patients underestimated their initial disability prior to care. The 1 study[18] that investigated RS during conservative care primarily reported weak ES, indicating that a recalibration may not have occurred within their chronic low back pain population. It is believed that for RS to occur a catalyst is needed to change an individual's condition.[4,5] This may indicate that conservative rehabilitation alone was not a substantial enough catalyst to initiate RS. Further research is needed to understand the impact of care type and to examine if RS occurs following conservative care. Additional consideration should be made to the length of symptoms prior to care or surgical intervention. It is possible the length of symptoms prior to conservative care or surgical intervention could play a role in the RS phenomenon.

Regardless of the PRO type used to evaluate HRQL, there was a trend toward orthopedic patients underestimating their disability prior to rehabilitation. Overall, a larger RS was demonstrated when HRQL was evaluated using generic PROs compared with region-specific and other PROs. This was indicated by a greater number of strong ES for generic PROs than for region-specific and other PROs. From these findings, it is reasonable to hypothesize that specific PRO types may be more susceptible to RS.[19] This may be due to the constructs evaluated within the varying PRO types. Generic PROs often focus on societal and personal factors of HRQL, whereas region-specific PROs focus on physical function of a specific body part. The focused concepts of region-specific PROs may provide greater context for patients, reducing room for varying interpretations and in turn reducing the potential for RS when compared with the global nature of the questions found on generic instruments. Future investigations should look to examine differences in RS phenomenon across different types of PROs, including generic, region-specific, and dimension-specific instruments, in their investigations.

Practical Implications

The results of this systematic review indicated that RS occurs in patients with orthopedic conditions undergoing rehabilitation after surgery. This was reflected by mostly moderate to large ES supporting the idea that individuals initially underrate their HRQL deficits prior to orthopedic rehabilitation. The notion of underestimating HRQL deficits was most notable when using generic instruments, most commonly the SF-36, and some region-specific PROs. The presence of RS can inhibit a clinician's ability to accurately identify improvement or deterioration in HRQL and make the appropriate adjustments to the care provided.[28] Clinicians should be cognizant that RS has the potential to confound the determination of HRQL changes, and should employ strategies to combat its effects.[6] Howard et al[28] suggested that clinicians should evaluate an individual's frame of reference over the course of care to assess RS that may alter a patient's frame of reference. This could be completed through continual reevaluation of patient goals and expectations to provide a standardized frame of reference throughout the rehabilitation process.[28] The implementation of then-tests, as used within the included studies,[7,9,10,17–21] may help clinicians identify potential confounding due to RS and make proper clinical decisions.[28] Finally, clinicians can also employ an external measure of quality of life (performance testing, clinical findings, family member rating, etc) to aid in the determination of confounding due to RS.[29] When there is a high degree of discrepancy between an individuals' rating of HRQL and the external measure, this may be an indication of RS.[29] However, further research is needed to develop and validate clinical strategies to mitigate the potential effect of RS and thereby enhance the ability to use PRO data in clinical decision making.

Limitations of Review

This systematic review was not without limitations. The electronic search was conducted within databases thought to be most relevant to RS and orthopedics. It is possible that articles relevant to this review were not located within these databases and subsequently failed to be identified during the search. Furthermore, our search only yielded articles that assessed RS using the then-test method. The limitation of the then-test method is that it is subject to recall bias[30]; thus, the results of this review should be interpreted within the limitations of the then-test method. To combat the issue of recall bias, future investigations should utilize minimal detectable change and minimally clinical important difference values in addition to traditional statistical analysis for the individual PROs. These can be used to determine if the magnitude of RS exceeds the measurement error and/or clinically meaningful change. In addition, future investigations that examine RS using methods other than then-test should ensure proper data reporting to allow for the comparison of studies.

The articles included primarily focused on chronic orthopedic conditions undergoing a surgical intervention and a lengthy rehabilitation program. Because of this, no recommendation can be made regarding the potential for RS during the conservative care of chronic or acute orthopedic conditions. Furthermore, factors such as length of symptoms, rehabilitation type, and the length of rehabilitation may all influence the potential for RS. Because of limitations in the reporting of these factors, we were unable to assess the impact of these factors on RS in the included studies. In addition, there was a lack of consistency in the data reported by the included studies, which limited the ability to complete a unified synthesis of the data. Future RS studies should place emphasis on providing consistent data reporting to facilitate comparisons between investigations. Finally, there was a lack of literature regarding RS when HRQL was captured using dimension-specific PROs. Future research should examine the potential for RS within HRQL concepts such as fear and avoidance beliefs to examine the presence of RS within a multidimensional profile of HRQL.

Conclusions

The results of this systematic review indicate there is grade B evidence that patients who complete surgery and rehabilitation for an orthopedic condition may experience a weak RS. The results further indicate when patients do experience an RS as measured by the then-test method, the prerehabilitation or baseline evaluation of HRQL may have underestimated their disability compared with then-test evaluations of their baseline. The magnitude of RS was largest when HRQL was evaluated using generic PROs that are designed to assess a patient's overall health as well as detriments to HRQL at the personal and societal level. It is important for clinicians to be aware of the potential shift in their patients' internal standards, as it can affect the evaluation of HRQL changes during the care of orthopedic conditions. Misclassification of HRQL changes can in turn adversely affect clinical decision making. Clinicians can consider the use of a frame of reference standard when implementing the instruments in practice to abate some of these changes. At this time, there is need for further research pertaining to RS to provide clinicians with the tools to identify and disentangle the influence of RS on HRQL assessment. Clinicians should continue to utilize both PROs and clinician-based outcomes when determining treatment effectiveness and making clinical decisions.

References

1. Snyder AR, Parsons JT, Valovich McLeod TC, Curtis Bay R, Michener LA, Sauers EL. Using disablement models and clinical outcomes assessment to enable evidence-based athletic training practice, part I: disablement models. *J Athl Train*. 2008;43(4): 428–436. PubMed ID: 18668176 doi:10.4085/1062-6050-43.4.428
2. Lavallee DC, Chenok KE, Love RM, et al. Incorporating patient-reported outcomes into health care to engage patients and enhance care. *Health Aff*. 2016;35(4):575–582. PubMed ID: 27044954 doi: 10.1377/hlthaff.2015.1362
3. Valovich McLeod TC, Snyder AR, Parsons JT, Curtis Bay R, Michener LA, Sauers EL. Using disablement models and clinical outcomes assessment to enable evidence-based athletic training practice, part II: clinical outcomes assessment. *J Athl Train*. 2008;43(4): 437–445. PubMed ID: 18668177 doi:10.4085/1062-6050-43.4.437
4. Schwartz CE, Sprangers MA. Methodological approaches for assessing response shift in longitudinal health-related quality-of-life research. *Soc Sci Med*. 1999;48(11):1531–1548. PubMed ID: 10400255 doi:10.1016/S0277-9536(99)00047-7
5. Sprangers MA, Schwartz CE. Integrating response shift into health-related quality of life research: a theoretical model. *Soc Sci Med*. 1999;48(11):1507–1515. PubMed ID: 10400253 doi:10.1016/S0277-9536(99)00045-3
6. Schwartz CE, Bode R, Repucci N, Becker J, Sprangers MA, Fayers PM. The clinical significance of adaptation to changing health: a meta-analysis of response shift. *Qual Life Res*. 2006;15(9):1533–1550. PubMed ID: 17031503 doi:10.1007/s11136-006-0025-9
7. Razmjou H, Schwartz CE, Yee A, Finkelstein JA. Traditional assessment of health outcome following total knee arthroplasty was confounded by response shift phenomenon. *J Clin Epidemiol*. 2009; 62(1):91–96. PubMed ID: 19095168 doi:10.1016/j.jclinepi.2008.08.004
8. Schwartz CE, Finkelstein JA. Understanding inconsistencies in patient-reported outcomes after spine treatment: response shift phenomena. *Spine J*. 2009;9(12):1039–1045. PubMed ID: 19574107 doi:10.1016/j.spinee.2009.05.010
9. Razmjou H, Schwartz CE, Holtby R. The impact of response shift on perceived disability two years following rotator cuff surgery. *J Bone Joint Surg Am*. 2010;92(12):2178–2186. PubMed ID: 20844160 doi:10.2106/JBJS.I.00990
10. Howard JS, Mattacola CG, Mullineaux DR, English RA, Lattermann C. Influence of response shift on early patient-reported outcomes following autologous chondrocyte implantation. *Knee Surg Sports Traumatol Arthrosc*. 2014;22(9):2163–2171. doi:10.1007/s00167-013-2654-1
11. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ*. 2009;339:b2535. PubMed ID: 19622551 doi:10.1136/bmj.b2535
12. Schwartz CE, Sajobi TT, Lix LM, Quaranto BR, Finkelstein JA. Changing values, changing outcomes: the influence of reprioritization response shift on outcome assessment after spine surgery. *Qual Life Res*. 2013;22(9):2255–2264. PubMed ID: 23519975 doi:10.1007/s11136-013-0377-x
13. Kievit W, Hendrikx J, Stalmeier PF, van de Laar MA, Van Riel PL, Adang EM. The relationship between change in subjective outcome and change in disease: a potential paradox. *Qual Life Res*. 2010; 19(7):985–994. PubMed ID: 20454862 doi:10.1007/s11136-010-9665-x
14. Robertson C, Langston AL, Stapley S, et al. Meaning behind measurement: self-comparisons affect responses to health-related quality of life questionnaires. *Qual Life Res*. 2009;18(2):221–230. PubMed ID: 19142744 doi:10.1007/s11136-008-9435-1
15. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *J Epidemiol Community Health*. 1998;52(6):377–384. PubMed ID: 9764259 doi:10.1136/jech.52.6.377
16. Munn J, Sullivan SJ, Schneiders AG. Evidence of sensorimotor deficits in functional ankle instability: a systematic review with meta-analysis. *J Sci Med Sport*. 2010;13(1):2–12. PubMed ID: 19442581 doi:10.1016/j.jsams.2009.03.004
17. Balain B, Ennis O, Kanes G, et al. Response shift in self-reported functional scores after knee microfracture for full thickness cartilage lesions. *Osteoarthritis Cartilage*. 2009;17(8):1009–1013. doi:10.1016/j.joca.2009.02.007
18. Nagl M, Farin E. Response shift in quality of life assessment in patients with chronic back pain and chronic ischaemic heart disease. *Disabil Rehabil*. 2012;34(8):671–680. PubMed ID: 22013979 doi: 10.3109/09638288.2011.619616
19. Razmjou H, Yee A, Ford M, Finkelstein JA. Response shift in outcome assessment in patients undergoing total knee arthroplasty. *J Bone Joint Surg Am*. 2006;88(12):2590–2595. PubMed ID: 17142408 doi:10.2106/JBJS.F.00283
20. Zhang XH, Li SC, Xie F, et al. An exploratory study of response shift in health-related quality of life and utility assessment among patients with osteoarthritis undergoing total knee replacement surgery in a tertiary hospital in Singapore. *Value Health*. 2012; 15(1)(suppl):S72–S78. PubMed ID: 22265071 doi:10.1016/j.jval.2011.11.011
21. Finkelstein JA, Quaranto BR, Schwartz CE. Threats to the internal validity of spinal surgery outcome assessment: recalibration response shift or implicit theories of change? *Appl Res Qual Life*. 2014; 9(2):215–232. doi:10.1007/s11482-013-9221-2
22. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum; 1988.
23. Turner HM, Bernard RM. Calculating and synthesizing effect sizes. *Contemp Issues Commun Sci Disord*. 2006;33(1):42–55.
24. Ebell MH, Siwek J, Weiss BD, Woolf SH, Susman J, Ewigman B, Bowman M. Strength of recommendation taxonomy (SORT): a patient-centered approach to grading evidence in the medical literature. *J Am Board Fam Pract*. 2004;17(1):59–67. PubMed ID: 15014055 doi:10.3122/jabfm.17.1.59
25. Gorgos KS, Wasylyk NT, Van Lunen BL, Hoch MC. Inter-clinician and intra-clinician reliability of force application during joint

mobilization: a systematic review. *Man Ther*. 2014;19(2):90–96. PubMed ID: 24405786 doi:10.1016/j.math.2013.12.003

26. Hollman F, Wessel RN, Wolterbeek N. Response shift of the Western Ontario Rotator Cuff Index in patients undergoing arthroscopic rotator cuff repair. *J Shoulder Elbow Surg*. 2016;25(12): 2011–2018. PubMed ID: 27424250 doi:10.1016/j.jse.2016.05.012
27. Finkelstein JA, Razmjou H, Schwartz CE. Response shift and outcome assessment in orthopedic surgery: is there a difference between complete and partial treatment? *J Clin Epidemiol*. 2009; 62(11):1189–1190. PubMed ID: 19679446 doi:10.1016/j.jclinepi.2009.03.022
28. Howard JS, Mattacola CG, Howell DM, Lattermann C. Response shift theory: an application for health-related quality of life in rehabilitation research and practice. *J Allied Health*. 2011;40(1): 31–38. PubMed ID: 21399850
29. Rapkin BD, Schwartz CE. Toward a theoretical model of quality-of-life appraisal: implications of findings from studies of response shift. *Health Qual Life Outcomes*. 2004;2(1):14. doi:10.1186/1477-7525-2-14
30. Schwartz CE. Applications of response shift theory and methods to participation measurement: a brief history of a young field. *Arch Phys Med Rehabil*. 2010;91(9)(suppl):S38–S43. PubMed ID: 20801278 doi:10.1016/j.apmr.2009.11.029

Appendix

Table 1 Individual Downs and Black Quality Index Items for the Included Articles

	Howard et al[10]	Finkelstein et al[21]	Zhang et al[20]	Nagl and Farin[18]	Razmjou et al[9]	Razmjou et al[7]	Balain et al[17]	Razmjou et al[19]
Is the hypothesis/Aim/objective clearly described?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Are the main outcomes clearly described in the intro or methods?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Are the characteristics of subjects included and clearly described?	Yes	Yes	Yes	Yes	Yes	No	No	Yes
Are the distributions of principle confounders in each group of subjects clearly described?	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
Are the main findings of the study clearly described?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Does the Study provide estimates of variability?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Have actual *P* values been reported?	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Were asked subjects representative of the entire population form which they were recruited?	Yes	Unable to Determine	Unable to Determine	Unable to Determine	Unable to Determine	Unable to Determine	Unable to Determine	Unable to Determine
Were prepared to participate representative of the entire population from which they were recruited?	Unable to Determine	Unable to Determine	Unable to Determine	Unable to Determine	Unable to Determine	Yes	Unable to Determine	Unable to Determine
Was an attempt made to blind those measuring the main outcomes?	No	Yes	Unable to Determine	Unable to Determine	Unable to Determine	Yes	Unable to Determine	Yes
If any of the results were based on data dredging was this made clear?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Were the statistical tests used to assess the main outcomes appropriate?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Were the main outcomes measures used accurate (valid and reliable)?	Yes	Yes	Yes	No	Yes	Unable to Determine	Yes	Unable to Determine
Were the subjects (both groups) recruited from the same population?	Yes	Unable to Determine	Yes	Yes	Yes	Unable to Determine	Yes	Yes
Were the subjects (both groups) recruited over the same period of time?	Unable to Determine	Unable to Determine	Unable to Determine	Unable to Determine	Unable to Determine	Unable to Determine	Unable to Determine	Yes
Was there adequate adjustment for confounding in the analyses from which the main findings were drawn?	Unable to Determine	Unable to Determine	Yes	No	Yes	Unable to Determine	Yes	Yes
Total Score	13	12	13	9	13	9	12	14
%	81.25	75.00	81.25	56.25	81.25	56.25	75.00	87.50