Critical Appraisal of Clinical Prediction Rules That Aim to Optimize Treatment Selection for Musculoskeletal Conditions

Tasha R. Stanton, Mark J. Hancock, Christopher G. Maher and Bart W. Koes

PHYS THER. 2010; 90:843-854.
Originally published online April 22, 2010
doi: 10.2522/ptj.20090233

Critical Appraisal of Clinical Prediction Rules That Aim to Optimize Treatment Selection for Musculoskeletal Conditions

Tasha R. Stanton, Mark J. Hancock, Christopher G. Maher, Bart W. Koes

Background. Clinical prediction rules (CPRs) for treatment selection in musculoskeletal conditions have become increasingly popular.

Purpose. The purposes of this review are: (1) to critically appraise studies evaluating CPRs and (2) to consider the clinical utility and stage of development of each CPR.

Data Sources. Pertinent databases were searched up to April 2009. Studies aiming to develop or evaluate a CPR for treatment response in musculoskeletal conditions were included. Two independent reviewers assessed eligibility and extracted methodological data, stage of development, and effect size information.

Study Selection/Data Extraction and Synthesis. Eighteen studies, evaluating 15 separate CPRs, were included. Fourteen CPRs were at the derivation stage, and all CPRs had been evaluated using a single-arm trial design, thus it is not possible to determine whether the CPRs identify prognosis (regardless of treatment) or specifically response to treatment. The CPR at the validation stage investigated spinal manipulative therapy (SMT) for low back pain and had been evaluated in 2 separate well-conducted randomized controlled trials. The first trial demonstrated a clinically meaningful effect of the SMT CPR; the additional effect from SMT in patients "positive-on-the-rule" was 15 Oswestry disability units at week 1 and 9 units at week 4. The second trial showed that the CPR did not generalize to a different clinical setting, including a modified treatment.

Limitations. Due to differences in methods of reporting and journal publication restraints (eg, word count restrictions), some quality assessment items may have been completed in the included studies, but not captured in this review.

Conclusions. There is, at present, little evidence that CPRs can be used to predict effects of treatment for musculoskeletal conditions. The principal problem is that most studies use designs that cannot differentiate between predictors of response to treatment and general predictors of outcome. Only 1 CPR has been evaluated within an RCT designed to predict response to treatment. Validation of these rules is imperative to allow clinical application.

T.R. Stanton, BScPT, MScRS, is a PhD candidate, Musculoskeletal Division, The George Institute for International Health, University of Sydney, PO Box M201, Missenden Road, Sydney, New South Wales, Australia 2111. Address all correspondence to Ms Stanton at: tstanton@george.org.au.

M.J. Hancock, BAppSc, PhD, is Lecturer, University of Sydney.

C.G. Maher, BAppSc, PhD, is Director, Musculoskeletal Division, The George Institute for International Health, University of Sydney.

B.W. Koes, MSc, PhD, is Professor, Erasmus Medical Centre, Rotterdam, the Netherlands.

[Stanton TR, Hancock MJ, Maher CG, Koes BW. Critical appraisal of clinical prediction rules that aim to optimize treatment selection for musculoskeletal conditions. *Phys Ther*. 2010;90:843–854.]

Many treatments for musculoskeletal disorders have shown small effects when tested in randomized controlled trials (RCTs).[1–5] One possible reason for these small effects is that trials include heterogeneous groups of patients,[6] some of whom respond to the treatment and some of whom do not.[7] Identifying patients who respond best to certain treatments has been spotlighted as a research priority.[8,9] As a result, there has been increased interest and research activity regarding characteristics of patients who respond best to certain interventions.

Identifying patients who respond best to treatment can be done using a single patient characteristic[10–13]; however, a combination or cluster of patient characteristics may be more informative than single features. With a clinical prediction rule (CPR), various components of the patient's history, physical examination, and basic laboratory results are combined to determine the diagnosis, prognosis, or likely response to treatment of that individual.[14] The development of a CPR involves the following stages: derivation (analyzing a data set to establish a rule with predictive power), narrow validation (evaluating the rule in a similar clinical setting and population), broad validation (evaluating the rule in multiple clinical settings), and impact analysis (determining whether the rule changes clinicians' behavior, improves patient outcomes, or reduces costs).[14] It is suggested that CPRs undergo full validation before being recommended for clinical practice.[14,15]

The recent surge in research activity in the area of CPRs to select treatment for musculoskeletal conditions provides the rationale for a systematic review to locate, appraise, and synthesize the primary studies. At present, 2 systematic reviews[16,17] exist, but both have limitations. The review by Beneciuk et al[16] focused on intervention studies but judged quality using a scale[18] designed for prognosis studies. The review by May and Rosedale[17] included intervention, prognosis, diagnosis, and construct validity and judged the quality of these very different study designs with the same scale. These methods produced a counterintuitive result where 2 RCTs, the optimal design for assessing treatment effect modification,[19,20] both received lower scores than a study with a single-arm design, which cannot measure treatment effect modification.

The 2 existing reviews also do not provide a clear understanding of the clinical utility of each CPR. Issues such as the stage of development of each CPR, the components of the CPR, and the specific treatment outcome the rules aim to predict were not considered. Consequently, the purposes of this article are: (1) to give a comprehensive summary of existing CPRs and (2) to critically appraise the research evaluating CPRs used to select treatment for musculoskeletal conditions in primary care.

Available With This Article at ptjournal.apta.org

- eAppendix 1: Database Search Strategies
- eAppendix 2: Populations in Which Clinical Prediction Rules Have Been Tested
- The Bottom Line Podcast
- Audio Abstracts Podcast

This article was published ahead of print on April 22, 2010, at ptjournal.apta.org.

Method

Data Sources and Searches

Potential studies were identified via a literature search of the following databases: MEDLINE, EMBASE, CINAHL, AMED, PubMed, and PEDro (up to April 1, 2009). PubMed was searched (in addition to MEDLINE) to include articles currently published electronically, but not yet available in MEDLINE or PreMEDLINE. The following key words were used: "clinical prediction rule" or "clinical prediction tool" or "prediction tool" or "clinical decision rule" or "clinical decision tool' or "decision tool" or "decision model" combined with musculoskeletal disorders and pain terminology. See eAppendix 1 (available at ptjournal.apta.org) for full search strategies for all databases.

Study Selection

The following eligibility criteria were applied to each study to determine inclusion status:

- The study was published in a peer-reviewed journal.
- An explicit aim of the study was to develop or evaluate a CPR.
- The CPR aims to assist treatment selection for patients with musculoskeletal conditions seen in primary care. We considered primary care to be a clinical setting, such as a general practice, physical therapy, or chiropractic clinic, where no referral is required.
- The criteria in the CPR must be easily obtained in primary care (eg, patient history, assessment findings, simple laboratory results). These criteria do not include invasive procedures such as nerve blocks.
- The CPR comprises >1 criterion.

One author (T.R.S.) examined the titles, key words, and abstracts of the results from the electronic database search and excluded clearly ineligible studies. Full reports of the re-

maining records were independently assessed for inclusion by 2 authors drawn from a panel of 3 authors (T.R.S., C.G.M., and B.W.K.). Any disagreements were resolved through consensus or, if not possible, through consultation with all authors. Citation tracking of included studies identified 2 additional references.

Data Extraction and Quality Assessment

The following data were extracted:

- **General information on the CPR:** the musculoskeletal condition for which the CPR was created, the target treatments, the components of the rule, the scoring schema (and weighting, if applicable), whether the rationale for predictors was described, the number of studies available, and the stage in the process of development of the rule (derivation, validation [narrow/broad], or impact analysis stage).[14]
- **Derivation studies:** the sample, number of candidate variables tested, what constitutes "positive on the rule" (eg, 4 out of 5 predictors present or not applicable), whether analyses were specified *a priori*, effect size information (eg, positive likelihood ratios, significance of the interaction test), whether the outcome measure was dichotomized or continuous, the definition of successful outcome if dichotomized (eg, 50% reduction in disability), the length of follow-up, the proportion of sample meeting the rule, the number of participants not meeting the study inclusion criteria, and whether an internal validity investigation was completed (eg, bootstrapping, split-half, jackknife).
- **Validation studies:** in addition to the information for derivation studies, we extracted the study design (single arm or controlled), the type of validation (eg, narrow, broad), and, in RCTs only, whether an interaction test was performed, whether the sample size was calculated for the interaction, and the methodological quality (PEDro score)[21] of the articles.

Data Synthesis and Analysis

Statistical pooling of the results of included studies was not completed, as the purpose of this study was to summarize and evaluate the CPRs currently published. Due to the lack of a validated quality assessment scale for CPRs that assist in treatment selection, quality was assessed using previous recommendations on methodological issues for treatment CPRs.[14,15,19,20,22–24] Using 7 methodological quality assessment items taken from the literature (Tab. 1), all included studies were evaluated, and criteria were scored as "present," "absent," "unclear," or "not applicable."

Results

The search retrieved 574 records with 18 studies, representing 15 separate CPRs considered eligible.[25–42] The search results are shown in the Figure. Assessment of methodological quality items for all included studies is presented in Table 1.

The general characteristics of the CPRs are shown in Table 2 (the sample populations are shown in eAppendix 2, available at ptjournal.apta.org). The musculoskeletal conditions and number of CPRs developed were: low back pain (LBP)—4 rules[25,30,31,33]; neck pain—4 rules[27,36,38,42]; patellofemoral pain—4 rules[34,35,37,39]; knee osteoarthritis (OA)—1 rule[29]; ankle sprain—1 rule[41]; and lateral epicondylalgia—1 rule.[40] The CPRs were developed to inform selection of: manipulative and manual therapy—9 rules[27,29–31,34,38,40–42]; exercise—4 rules[33,36,40,41]; traction—2 rules[25,36]; taping—1 rule[35]; and orthoses—2 rules.[37,39] Of the 15 CPRs included, only 1 was at the validation stage of development,[26,28,30,32] with all others at a derivation level.[25,27,29,31,33–42]

The Bottom Line

What do we already know about the topic?

Clinical prediction rules (CPRs) that aim to select the most effective treatment for an individual patient are becoming increasingly common. It is recommended that CPRs are not applied clinically until they are validated.

What new information does this study offer?

This study found that, currently, there is little evidence that published CPRs can be used to predict effects of treatment for musculoskeletal conditions. Most studies use designs that cannot differentiate between predictors of response to treatment and predictors of outcome regardless of treatment.

If you're a patient, what might these findings mean for you?

You and your clinician should base treatment decisions on evidence of what helps most patients with a similar condition (eg, evidence from a well-conducted clinical trial).

Table 1.
Methodological Quality Assessment Items Met by Each Included Study[a]

Study	Design[b]	Rationale[c]	Sample Size[d]	Analysis[e]	Interaction Test[f]	Powered for Interaction[g]	Internal Validity[h]
Derivation-level studies							
Cai et al,[25] 2009	X	X	X	X	X	X	X
Cleland et al,[27] 2007	X	X	X	X	X	X	X
Currier et al,[29] 2007	X	X	X	X	X	X	X
Flynn et al,[30] 2002	X	X	X	X	X	X	X
Fritz et al,[31] 2005	X	X	√	X	X	X	X
Hicks et al,[33] 2005	X	X	X	X	X	X	X
Iverson et al,[34] 2008	X	X	X	X	X	X	X
Lesher et al,[35] 2006	X	X	X	X	X	X	X
Raney et al,[36] 2009	X	X	?	X	X	X	X
Sutlive et al,[37] 2004	X	X	X	X	X	X	X
Thiel et al,[42] 2008	X	X	√	X	X	X	X
Tseng et al,[38] 2006	X	X	?	X	X	X	X
Vicenzino et al,[40] 2008	X	X	X	X	X	X	X
Vicenzino et al,[39] 2008	X	X	?	X	X	X	X
Whitman et al,[41] 2009	X	X	X	X	X	X	X
Validation-level studies							
Childs et al,[26] 2004	√	X	n/a	X	√	√	X
Cleland et al,[28] 2006	X	X	n/a	X	X	X	X
Hancock et al,[32] 2008	√	X	n/a	√	√	X	X

[a] √ = present, X = absent, ? = unclear, n/a = not applicable.
[b] Appropriate study design used (a controlled study design is recommended to assess treatment effect modification).
[c] Rationale provided for predictors (predictors with no logical rationale may represent spurious findings).
[d] Appropriate sample size (at least 10 outcome events per candidate variable recommended).
[e] Analysis specified *a priori*.
[f] Interaction test performed in controlled studies to determine predictors.
[g] Sample size powered for the interaction test.
[h] Internal validity investigation completed (techniques such as boot-strapping, split-half, and jackknife should be performed to ensure internal validity).

Table 3 provides information on the characteristics of the derivation studies. All 14 CPRs at a derivation level had been generated from data from single-arm studies or from data from 1 arm of an RCT.[39,40] Outcomes typically were measured in the short term (eg, after a single treatment session); only 2 studies included a follow-up of greater than 3 weeks.[33,39] No study provided a rationale for the candidate variables considered for inclusion in the CPR, and in 33% (5/15) of the studies, it was unclear how many total candidate variables were used in the analysis.[35–37,39,42] Counterintuitively, 1 CPR reported the same 4 variables to predict both improvement and worsening of neck symptoms with cervical manipulation.[42] All studies reported positive findings (found predictors significantly related to outcome).

Table 4 presents characteristics specific to the validation studies. Three validation studies were performed,[26,28,32] all in respect to the same CPR on manipulation for LBP.[30] Two of these studies, 1 single-arm trial[28] and 1 RCT,[26] looked at narrow validation (same patient population and treatment), and the third validation study, an RCT,[32] investigated the broad validation of this CPR (different sample patient population and modified treatment). Only 1 study specified the analyses *a priori*,[32] and only 1 of the 2 RCTs calculated the sample size for the interaction between rule status and treatment effect.[26] Clinically important effect sizes were found in the RCT validation study of spinal manipulative therapy for LBP.[26] The additional effect from spinal manipulative therapy in patients who were positive on the rule was 15 Oswestry disability units at week 1 and 9 Oswestry disability units at week 4.[26] The second RCT validation study (in which sample size was not calculated for the interaction) did not find a significant

effect of rule status (positive or negative) on response to treatment (manipulation versus placebo).[32] The 2 validation RCTs both used longer, more clinically relevant maximum follow-up times of 12 weeks[32] and 6 months.[26]

Discussion

We found that all derivation-level CPRs used single-arm study designs to derive predictors, raising doubts about the veracity of current CPRs. In total, 15 CPRs to aid selection of a range of treatments for musculoskeletal conditions were found. However, only 1 CPR for selecting spinal manipulation for LBP[30] had reached the validation stage of CPR development, with the CPR predicting response to treatment in a narrow validation study but not treatment effects in a broad validation study.[26,28,32]

Although single-arm study designs can be a preliminary step in developing prediction rules by identifying potential candidate variables, they are not able to differentiate between predictors of response to treatment and predictors of outcome regardless of treatment. These studies do not include a control group, so they cannot provide information on treatment effects or on factors that modify treatment effects. The predictive factors identified in these studies, therefore, have a higher risk of being merely nonspecific predictors of outcome or prognostic factors. Neither of the 2 previous systematic reviews[16,17] on this topic discussed this issue of use of a single-arm trial design at the derivation level. One review stated that "most of the derivation studies were of high quality,"[17(p40)] a claim we would argue is potentially misleading, considering no derivation studies used a study design that allows specific identification of treatment effect modifiers.

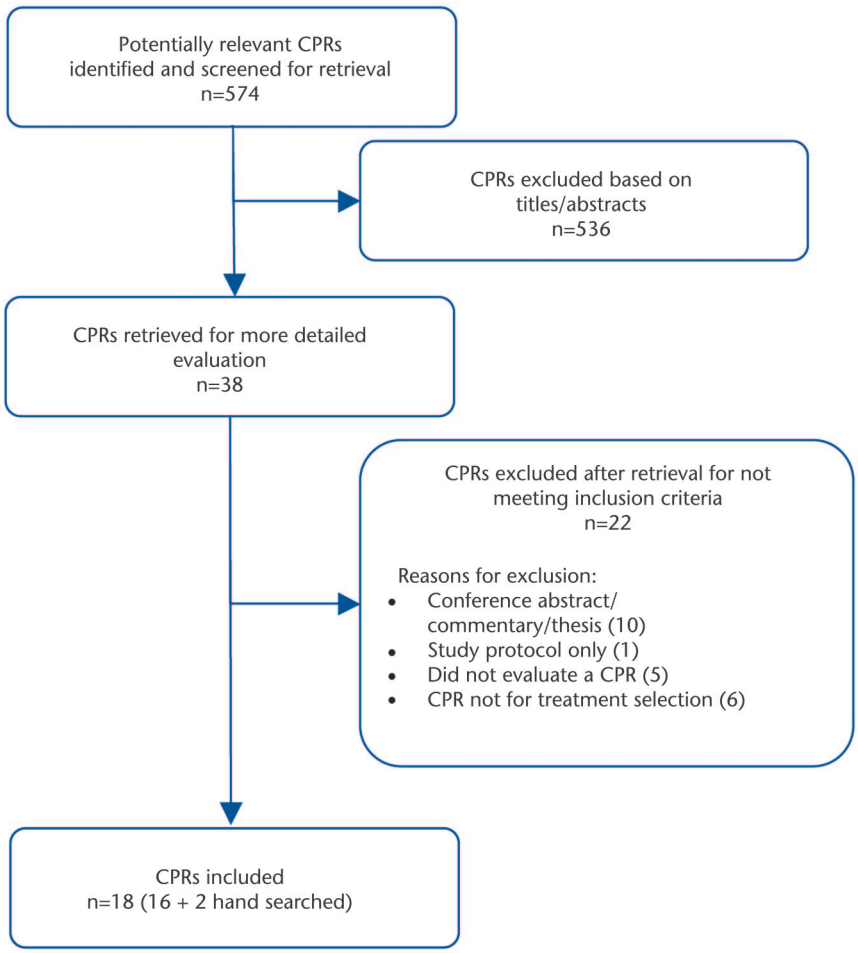

Figure.
Flow chart describing the results of the literature search. CPR=clinical prediction rule.

One potential justification for using prognostic factors identified in single-arm trials to develop a treatment CPR is that the prognostics factors also may be treatment effect modifiers. Although this can happen,[30] there are cases where it does not.[43] There also are examples where the same clinical feature predicts poor prognosis yet predicts a good response to treatment.[44] This uncertain relationship makes it essential to carefully interpret the results of studies reporting a treatment CPR from single-arm trials. Although these studies are hypothesis generating, variables identified in single-arm trial designs run a greater risk of not being significant in a subsequent controlled study. Moreover, for some CPR candidate variables, there are many existing data sets from RCTs that would provide a better evaluation of the variable as an effect modifier than a single-arm study. For example, 5 of the CPRs included age as a variable,[25,33,36,39,40] and as age is almost always measured in RCTs, a more robust evaluation of age as a treatment effect modifier would be possible from secondary analysis of individual trials or from the pooled data from several trials using a meta-regression approach. Having said that, to properly develop a CPR, an RCT designed specifically for the purpose of CPR development, with appropriate sample size and *a priori* analysis, is necessary.[22]

Table 2.

General Characteristics of the Clinical Prediction Rule[a]

Clinical Prediction Rule (Original Study)	Musculoskeletal Condition/ Treatment	Features of Rule	What Constitutes "Positive on the Rule"?	Stage of Development
Cai et al,[25] 2009	LBP/mechanical traction	1. FABQW score <21 2. Absence of neurological deficit 3. Age >30 y 4. Noninvolvement in manual work	4 of 4 variables	Derivation
Cleland et al,[27] 2007	Neck pain/thoracic spine manipulation	1. Symptoms <30 d 2. No symptoms distal to the shoulder 3. Looking up does not aggravate symptoms 4. FABQPA score <12 5. Diminished upper thoracic spine kyphosis 6. Cervical extension range of motion <30°	≥3 out of 6 variables Weighting of variables using relative size of the β coefficients	Derivation
Currier et al,[29] 2007	Painful knee osteoarthritis/hip mobilization	1. Pain with ipsilateral hip distraction 2. Ipsilateral knee passive flexion <122° 3. Ipsilateral hip passive medial (internal) rotation <17° 4. Pain or paresthesia in ipsilateral hip or groin 5. Ipsilateral anterior thigh pain	Any 2 of the 5 variables	Derivation
Flynn et al,[30] 2002 (original 5-item rule)	LBP/lumbar spine manipulation	1. Duration of symptoms <16 d 2. At least one hip with >35° medial rotation 3. Lumbar hypomobility with spring test ≥1 level 4. No symptoms distal to the knee 5. FABQW score <19	≥4 out of 5 variables	Derivation, narrow validation (Childs et al,[26] 2004; Cleland et al,[28] 2006); broad validation (Hancock et al,[32] 2008)
Fritz et al,[31] 2005 (2-item Flynn rule)	LBP/lumbar spine manipulation	1. Duration of symptoms <16 d 2. No symptoms distal to the knee	2 out of 2 variables	Derivation
Hicks et al,[33] 2005	LBP/stabilization exercise	1. Positive prone instability test 2. Aberrant movements present 3. Average straight leg raise >91° 4. Age <40 y	≥3 out of 4 variables	Derivation
Iverson et al,[34] 2008	Patellofemoral pain/ lumbopelvic manipulation	1. Side-to-side difference in hip medial rotation >14° 2. Ankle dorsiflexion (knee flexed) >16° 3. Navicular drop >3 mm 4. No self-reported stiffness with sitting >20 min 5. Squatting reported as most painful activity	Presence of hip medial rotation asymmetry ≥14° or, if hip medial rotation asymmetry not present, ≥3 out of 5 predictors	Derivation
Lesher et al,[35] 2006	Patellofemoral pain syndrome/patellar taping	1. Positive patellar tilt test 2. Tibial varum >5°	Either 1 of 2 variables	Derivation
Raney et al,[36] 2009	Nonspecific neck pain/cervical traction and exercise	1. Peripheralization with lower cervical spine (C4–C7) mobility testing 2. Positive shoulder abduction test 3. Age ≥55 y 4. Positive upper-limb tension test A 5. Positive neck distraction test	≥3 out of 5 variables or ≥4 out of 5 variables	Derivation
Sutlive et al,[37] 2004	Patellofemoral pain syndrome/foot orthosis use and modified activity	1. Forefoot valgus alignment ≥2° 2. Great toe extension of ≤78° 3. Navicular drop ≤3 mm	Any 1 of 3 variables No combination of variables led to positive likelihood ratio >2.0	Derivation

(*Continued*)

Table 2.
Continued

Clinical Prediction Rule (Original Study)	Musculoskeletal Condition/ Treatment	Features of Rule	What Constitutes "Positive on the Rule"?	Stage of Development
Thiel et al,[42] 2008	Patients receiving a cervical spine manipulation seen by chiropractors/ cervical manipulation	Improving: 1. Neck pain 2. Shoulder, arm pain 3. Reduced neck, shoulder, arm movement, stiffness 4. Headache 5. Upper, mid back pain 6. ≤1 presenting symptom Worsening: 1. Neck pain 2. Shoulder, arm pain 3. Headache 4. Numbness, tingling in upper limbs 5. Upper, mid back pain 6. Fainting, dizziness, light-headedness	Improvement: any 4 variables Worsening: any 4 of 6 variables Global improvement: not able to be used	Derivation
Tseng et al,[38] 2006	Neck pain (radiculopathy, disk herniation, myofascial pain syndrome, and cervicogenic headache)/cervical manipulation	1. Neck Disability Index initial score <11.50 2. Having bilateral involvement 3. Not performing sedentary work >5 h/d 4. Feeling better while moving the neck 5. Without feeling worse while extending the neck 6. Diagnosis of spondylosis without radiculopathy	≥3 out of 6 variables or ≥4 out of 6 variables	Derivation
Vicenzino et al,[39] 2008	Patellofemoral pain/ foot orthoses	1. Age >25 y 2. Height <165 cm 3. Worst pain (VAS) <53.25 mm 4. Midfoot width difference from WB to NWB >10.96 mm	Not given	Derivation
Vicenzino et al,[40] 2008	Lateral epicondylalgia/ manual therapy and exercise	1. Age <49 y 2. Affected limb pain-free grip >112 N 3. Unaffected limb pain-free grip <336 N	≥3 out of 4 variables	Derivation
Whitman et al,[41] 2009	Ankle sprain/manual therapy and general mobility	1. Symptoms worse when standing 2. Symptoms worse in evening 3. Navicular drop ≥5.0 mm 4. Distal tibial fibular joint hypomobility	3 out of 4 variables, not 4 out of 4 variables	Derivation

[a] LBP=low back pain, FABQW=Fear-Avoidance Beliefs Questionnaire–Work Subscale, FABQPA=Fear-Avoidance Beliefs Questionnaire–Physical Activity Subscale, VAS=visual analog scale, WB=weight bearing, NWB=non–weight bearing.

Two derivation studies used data from RCTs to develop CPRs but ignored data from the control group.[39,40] This approach effectively results in a single-arm study that has the same risks as mentioned above. The genesis of 1 CPR[40] for management of lateral epicondylalgia was particularly unusual because it was preceded by a meta-regression study[45] based upon pooled data from 2 RCTs[46,47] (n=383). The conclusion of the meta-regression study was "patient characteristics play only a small role in predicting treatment outcomes,"[45(p1601)] with only baseline pain intensity predicting response to physical therapy treatment. Inexplicably, the CPR derivation study[40] considered only data from the physical therapy arm of 1 RCT (n=64) and created a CPR that did not include baseline pain. It seems erroneous to revert to a weaker single-arm design to develop a CPR, and the pitfalls of this approach were well illustrated when the one treatment effect modifier identified in the meta-regression study was not identified in the single-arm design study.

In contrast to previous reviews in this area, we examined the proportion of patients suitable for rule application and the proportion that were rule positive, as these factors are important to the generalizability and clinical importance of the CPR. The proportion of potential participants excluded from the studies ranged from 20% to 71%,[26–29,32,41] suggesting that some rules have limited application. For example, in a CPR on hip mobilization for painful knee OA, only 35% of patients seeking care for their OA met the inclu-

Table 3.
Methodological Characteristics of Derivation Studies Evaluating Clinical Prediction Rules[a]

Clinical Prediction Rule (Original Study)	No. of Variables Tested	Predictive Ability[b]	Definition of Successful Outcome	Length of Follow-up	Percentage of Sample Meeting the Rule	No. of Participants Ineligible at Baseline
Cai et al,[25] 2009	44	4 LR+=9.4 (3.1–28.0) ≥3 LR+=3.0 (2.0–4.5) ≥2 LR+=1.8 (1.5–2.2) ≥1 LR+=1.1 (0.99–1.2)	>50% reduction in Oswestry score Success rate[c]=19% (25/129)	Maximum: 9 d	13/129 (10%)	Not reported
Cleland et al,[27] 2007[d]	34	≥5 LR+=infinite ≥4 LR+=12 (2.3–70.8) ≥3 LR+=5.5 (2.7–12.0) ≥2 LR+=2.09 (1.5–2.5) ≥1 LR+=1.2 (1.1–1.2) Weighted: 3.5 points out of a total of 10 points; LR+=5.9 (2.6–13.0)	Score of +5 or greater on the −7 to +7 global rating of change scale Success rate=54% (42/78)	2–8 d	≥3=37/78 (47%)	25/105 (24%)
Currier et al,[29] 2007[d]	≥73	2 LR+=12.9 (0.8–205.6) 1 LR+=5.1 (1.8–14.6) More than 2 variables did not improve the LR+ (no participants were positive on >3 variables)	30% reduction in pain (NRS) during 2 functional tasks or global rating of change score of +3 or greater on a −7 to +7 scale Success rate=68% (41/60)	2 d	Unclear	110/170 (65%)
Flynn et al,[30] 2002 (original 5-item rule)	61	5 LR+=infinite ≥4 LR+=24.4 (4.6–139.4) ≥3 LR+=2.6 (1.8–4.2) ≥2 LR+=1.2 (1.1–1.4) ≥1 LR+=1.0 (1.0–1.2)	50% reduction in Oswestry score Success rate=45% (32/71)	2–8 d	4=15/71 (21%)	Not reported
Fritz et al,[31] 2005 (2-item Flynn rule)[e]	2	Both criteria present; LR+=7.2 (3.2–16.1)	50% reduction in Oswestry score Success rate=45% (63/141)	2–8 d	41/141=29%	Not applicable
Hicks et al,[33] 2005	43	≥3 LR+=4.0 (1.6–10.0) ≥2 LR+=1.9 (1.2–2.9) ≥1 LR+=1.3 (1.0–1.6)	50% reduction in Oswestry score Success rate=33% (18/54)	8 wk	Not reported	Not reported
Iverson et al,[34] 2008	39	5 LR+=infinite ≥4 LR+=infinite ≥3 LR+=18.4 (3.6–105.3) ≥2 LR+=2.1 (1.3–2.9) ≥1 LR+=1.1 (0.9–1.3)	50% reduction in pain (NRS) during 3 functional tasks or global rating of change score of +4 or greater on a −7 to +7 scale Success rate=45% (22/49)	Immediate (after the treatment session)	5=2/49 (4%) 4=5/49 (10%) 3=9/49 (18%) 2=16/49 (33%) 1=11/49 (2%) 0=6/49 (12%) ≥3=16/49 (33%)	Not reported
Lesher et al,[35] 2006	31 (unclear)	Either 1 of 2 variables; LR+=4.4 (1.3–12.3)	≥50% reduction in mean pain (NRS) during 3 functional tasks or global rating of change score of +4 or higher on a −7 to +7 scale Success rate=52% (26/50)	Immediate (after the treatment session)	Positive patellar tilt test=17/50 (34%) Tibial varum >5°=10 (unclear)/50 (20%) or 15 (unclear)/50 (30%)	Not reported
Raney et al,[37] 2009[d]	Unclear	≥4 LR+=23.1 (2.5–227.9) ≥3 LR+=4.8 (2.2–11.4) ≥2 LR+=1.4 (1.1–2.0) ≥1 LR+=1.2 (0.97–1.4)	Global rating of change score of +6 or greater on a −7 to +7 scale Success rate=44% (30/68)	3 wk	5=0 4=9/68 (13%) 3=15/68 (22%)	Not reported
Sutlive et al,[37] 2004	37 (unclear)	Forefoot alignment ≥2° of valgus; LR+=4.0 (0.7–21.9) Great toe extension ≤78°; LR+=4.0 (0.7–21.9) Navicular drop test; LR+=2.3 (1.3–4.3)	≥50% improvement in pain on visual analog scale Success rate=60% (27/45)	3 wk	Not reported	Not reported

(*Continued*)

Table 3.
Continued

Clinical Prediction Rule (Original Study)	No. of Variables Tested	Predictive Ability[b]	Definition of Successful Outcome	Length of Follow-up	Percentage of Sample Meeting the Rule	No. of Participants Ineligible at Baseline
Thiel et al,[42] 2008	≥22	Immediately improving: 5 LR+=0.75 (0.65–0.86) 4 LR+=6.3 (5.2–7.5) 3 LR+=4.3 (3.9–4.7) 2 LR+=2.6 (2.5–2.8) 1 LR+=1.5 (1.5–1.6) Immediately worsening: 6 LR+=1.0 (0.67–1.6) 5 LR+=3.2 (1.9–5.2) 4 LR+=3.6 (2.9–4.3) 3 LR+=2.4 (2.1–2.7) 2 LR+=1.4 (1.3–1.6) 1 LR+=1.0 (0.9–1.1) Global improvement: 2 LR+=0.94 (0.9–0.98) 1 LR+=1.0 (1.0–1.1)	"Immediate improvement" vs "no immediate improvement" "Immediate worsening" vs "no immediate worsening" At 7 days, global improvement was measured: "much better and a noticeable chance that has made a real difference" vs "no global improvement" Success rate=70% (20,083/28,807 treatment consultations)	Immediate (after treatment session) and 7 d following	Not reported	Not reported
Tseng et al,[38] 2006	≥18	≥5 LR+=infinite ≥4 LR+=5.3 (1.7–16.5) ≥3 LR+=1.9 (1.0–3.7) ≥2 LR+=0.2 (0.08–0.49) ≥1 LR+=0.07 (0.01–0.5)	≥50% reduction in pain (NRS) or global rating of change score of +4 or higher on a −7 to +7 scale or satisfaction with treatment rated as "very satisfied" (5-point scale) Success rate=60% (60/100)	Immediate (after the treatment session)	6=0 5=4/100 (4%) 4=27/100 (27%) 3=35/100 (35%)	Not reported
Vicenzino et al,[40] 2008[d,e]	12	MWMT: 1 LR+=1.8 (1.1–3.0) 2 LR+=3.7 (1.0–13.6) 3 LR+=infinite Wait and see: 1 LR+=1.0 (0.08–13.6) 2 LR+=3.1 (0.42–23.0) 3 LR+=1.2 (0.29–5.0)	Score of 0, 1, or 2 on a 0 ("completely recovered") to 5 ("much worse") global perceived effect Success rate=79% (49/62)	3 wk	MWMT: 3=4/62 (6%) 2=34/62 (55%) 1=57/62 (92%) 0=5/62 (8%) Wait and see: 3=2/57 (4%) 2=30/57 (53%) 1=14/57 (25%) 0=7/57 (12%)	Not reported
Vicenzino et al,[39] 2008[e]	Unclear	4=unable to calculate, no patients met 4 ≥3 LR+=8.8 (1.2–66.9) ≥2 LR+=2.2 (1.1–4.2) ≥1 LR+=1.6 (1.2–2.1)	Patients reporting "marked improvement" on a 5-point global improvement scale Success rate=40% (17/42)	12 wk	≥3=7/42 (17%)	Not reported
Whitman et al,[41] 2009[d]	45	4 LR+=0.43 (0.11–1.8) 3 LR+=5.9 (1.1–41.6) 2 LR+=1.2 (0.67–2.0) 1 LR+=0.33 (0.11–1.0)	Global rating of change score of +5 or greater on a −7 to +7 scale Success rate=75% (64/85)	2–8 d	3=19/85 (22%)	85/125 (68%)

[a] Outcome measures of improvement/success of treatment were dichotomized for all studies. LR+=positive likelihood ratio, Oswestry score=modified Oswestry Disability Questionnaire score, MWMT=mobilization with movement treatment arm, NRS=numerical rating scale.
[b] Predictive ability is expressed as the likelihood of a positive outcome for each score on the clinical prediction rule. Values in parentheses are 95% confidence intervals.
[c] Success rate was defined as the % of participants considered to have a successful intervention based on the definition of a successful outcome.
[d] Cutoff for success determined *a priori*.
[e] *Post hoc* analysis of one arm of a randomized controlled trial.

sion criteria of the study.[29] Similarly, of those patients who do enter the study, if only a very small proportion of patients, or almost everyone, meets the rule, the rule will have limited usefulness in clinical practice. In the studies included in this review, the number of patients from the sample who were positive on the rule ranged from 10% to 47%.[25,27,29–31,33–39,41] Generally, these proportions seem to be in the range of clinical importance, although Cai and colleagues' CPR, in which only 10% of the patients met the rule for mechanical traction for LBP,[25] is perhaps of questionable impact.

Table 4.
Methodological Characteristics of Validation Studies Evaluating Clinical Prediction Rules[a]

Clinical Prediction Rule (Original Study)	Interaction Test/Sample Size Calculated for the Interaction	Effect Size	Definition of Successful Outcome	Length of Follow-up	Percentage of Sample Meeting the Rule	No. of Participants Ineligible for the Study
Childs et al,[26] 2004[b,c,d] PEDro score: 8/10	Yes/yes	SMT × rule status (Oswestry score): 1 wk: 15 units[e] 4 wk: 9 units[e] 6 mo: 3 units[e]	50% reduction in Oswestry score (dichotomized)	1 wk, 4 wk, 6 mo	≥4/5=24/131 (18%)	386/543 (71%)
Cleland et al,[28] 2006[b,d,f]	Not applicable	11 of the 12 achieved a successful outcome	50% reduction in Oswestry score (dichotomized)	1 wk	Not applicable (meeting the rule was part of inclusion criteria)	3/15 (20%)
Hancock et al,[32] 2008[c,d,g,h] PEDro score: 9/10	Yes/no	SMT × rule status: 1 wk: NRS=0.31,[i] RMQ=1.77[i] 2 wk: NRS=0.11,[i] RMQ=2.38[i] 4 wk: NRS=−0.22,[i] RMQ=1.08[i] 12 wk: NRS=0.051,[i] RMQ=2.31[i]	Pain (NRS) Disability (RMQ) (continuous)	1, 2, 4, and 12 wk	Unable to tell	80/320 (25%)

[a] All studies evaluated Flynn and colleagues' original 5-item rule.[30] Oswestry score=modified Oswestry Disability Questionnaire score, SMT=spinal manipulation therapy, NRS=numerical rating scale, RMQ=Roland-Morris Disability Questionnaire.
[b] Narrow validation.
[c] Randomized controlled trial.
[d] Cutoff for success determined *a priori.*
[e] Significant at $P<.05$.
[f] Single-arm trial.
[g] Broad validation.
[h] Main analysis determined *a priori.*
[i] Significant at $P\geq.05$.

Related to the proportion of patients to whom the rule is applicable is the possibility of spectrum bias in the reviewed studies. Spectrum bias can occur when a study draws preferentially from a limited portion of the patient spectrum.[48] Although the purpose of a CPR is to identify a limited portion of the spectrum of patients who will respond to a certain treatment, starting with a baseline group that is limited in spectrum decreases the generalizability of the CPR. Many CPR studies have recruited a specific group of patients (eg, army recruits), and although not a point of criticism, this approach requires attention and supports the importance of broad validation studies.

Similar to previous reviews, we found that short-term follow-ups commonly were used when determining "success" with a treatment (eg, immediately after treatment, 2–8 days following treatment), meaning that what is predicted in many of the CPRs is of questionable importance. Only 2 derivation studies had a follow-up greater than 3 weeks posttreatment.[33,39] Furthermore, very few studies clearly reported the candidate variables (and the number of candidate variables) entered into their analysis. Only one study[32] was based upon a registered trial with a published protocol. This makes it difficult to judge the potential for type I error within a study, as we are unsure whether the authors failed to report nonsignificant predictors. With regard to the last issue, we note that there was only 1 negative study[32] among the 18 studies we located. Last, the risk of response bias is more of concern in single-arm studies compared with RCTs, as patients are not blinded to the treatment received or expected outcome.

None of the included studies presented a rationale for why the variables included in the CPR would be expected to predict response to treatment. Several authors have warned about spurious treatment effect modifiers and the need to be mindful of this issue when no logical rationale for a predictor exists.[14,15,23] The chance of spurious findings is classically illustrated by a study finding that subgrouping patients based on astrological birth sign predicted response to aspirin therapy following a myocardial infarction.[49] Accordingly, the CPR must make clinical sense.[14,15] For example, absence of nerve root signs predicted a better

response to mechanical traction for LBP,[25] which is directly opposite to what clinical lore advocates.[50] Furthermore, because in many studies[25,27,29,30,33,41] a large number of predictor variables were used in the analyses (eg, ≥73 potential predictors),[29] the likelihood that some variables will be significant by chance alone (type I error) is greatly increased.[22] No studies tested the internal validity of their analysis using techniques such as bootstrapping.

Although in some CPRs the predictor variables demonstrated face validity, 1 CPR found quite counterintuitive results. In a CPR by Thiel and Bolton[42] looking for predictors of response to manipulation in patients with neck symptoms, the presence of 4 symptoms (neck pain, shoulder/arm pain, headache, and upper/mid back pain) were predictors of improvement (versus staying the same) and worsening (versus staying the same). This pattern of results does not make sense and is perhaps an artifact of the way the analysis was set up. It may have been more informative to set up the analysis to predict improvement (versus staying the same or deteriorating).

To date, only 1 CPR on manipulation as a treatment for LBP[30] has undergone validation testing. Of the studies that did perform validation analyses, only 1 study was powered for the interaction.[26] It is recommended that CPR studies have a sample large enough to detect differences between the interaction of the rule status (positive or negative on the rule) and the treatment given (treatment or placebo/alternate treatment).[22,23] If a validation study is not powered for this interaction, it may not find statistically significant differences between those positive and negative on the CPR when these differences truly exist. The remaining RCT validation study was not powered for the interaction[32]; however, it exhibited tight confidence intervals, suggesting that it was not underpowered. Interestingly, this manipulation CPR demonstrated narrow validation[26] but not broad validation. Failure of broad validation of this CPR could have occurred for numerous reasons. First, the broad validation RCT used a treatment different from that of the original CPR study (provided mainly mobilization treatment for LBP instead of manipulation only). Alternatively, the possibility of spectrum bias influencing results cannot be ignored. Hancock et al[32] studied a population of patients with LBP in community primary care, whereas Childs et al[26] studied a population of primarily army recruits.

The paucity of validation studies for other CPRs and the absence of impact analysis investigations are likely related to the recent use of CPRs to determine response to treatment. However, based on the number of derivation studies this review found, it appears researchers are creating new rules but not validating the existing ones.

Conclusion

In contrast to previous systematic reviews, we found that all of the CPRs included in this study were derived using single-arm study designs. The results of these studies must be interpreted with caution, as these CPRs run a greater risk of identifying prognostic factors rather than factors that modify the effect of a treatment. Other important limitations of many of the included studies are use of short-term outcomes only, arguably trivial findings, and limited rule application potential. Only 1 CPR on spinal manipulation for LBP underwent validation in a controlled trial and can be considered for clinical application (and only in a population similar to that tested).

All authors provided concept/idea/research design, data analysis, and consultation (including review of manuscript before submission). Ms Stanton, Dr Hancock, and Dr Maher provided writing. Ms Stanton and Dr Maher provided data collection.

Data from this article were presented at the Australian Physiotherapy Conference (Musculoskeletal Physiotherapy Association Conference); October 1, 2009; Sydney, New South Wales, Australia.

Ms Stanton is supported by the University of Sydney International Research Scholarship for funding of her PhD candidacy. Professor Maher is supported by Australia's National Health and Medical Research Council for funding of his research fellowship.

This article was received July 14, 2009, and was accepted February 24, 2010.

DOI: 10.2522/ptj.20090233

References

1 Hancock MJ, Maher CG, Latimer J, et al. Assessment of diclofenac or spinal manipulative therapy, or both, in addition to recommended first-line treatment for acute low back pain: a randomised controlled trial. *Lancet*. 2007;370: 1638–1643.

2 Lund H, Weile U, Christensen R, et al. A randomized controlled trial of aquatic and land-based exercise in patients with knee osteoarthritis. *J Rehabil Med*. 2008;40: 137–144.

3 Skillgate E, Vingard E, Alfredsson L. Naprapathic manual therapy or evidence-based care for back and neck pain: a randomized, controlled trial. *Clin J Pain*. 2007;23: 431–439.

4 Staples MP, Forbes A, Ptasznik R, et al. A randomized controlled trial of extracorporeal shock wave therapy for lateral epicondylitis (tennis elbow). *J Rheumatol*. 2008; 35:2038–2046.

5 Veenhof C, Koke AJ, Dekker J, et al. Effectiveness of behavioral graded activity in patients with osteoarthritis of the hip and/or knee: a randomized clinical trial. *Arthritis Rheum*. 2006;55:925–934.

6 Delitto A. Research in low back pain: time to stop seeking the elusive "magic bullet." *Phys Ther*. 2005;85:206–208.

7 Kent P, Keating J. Do primary-care clinicians think that nonspecific low back pain is one condition? *Spine (Phila Pa 1976)*. 2004;29:1022–1031.

8 Borkan JM, Cherkin DC. An agenda for primary care research on low back pain. *Spine (Phila Pa 1976)*. 1996;21: 2880–2884.

9 Borkan JM, Koes B, Reis S, Cherkin DC. A report from the Second International Forum for Primary Care Research on Low Back Pain: reexamining priorities. *Spine (Phila Pa 1976)*. 1998;23:1992–1996.

10 Baldwin ML, Butler RJ, Johnson WG, Cote P. Self-reported severity measures as predictors of return-to-work outcomes in occupational back pain. *J Occup Rehabil*. 2007;17:683–700.

11 Fishbain DA, Lewis JE, Cutler R, et al. Does smoking status affect multidisciplinary pain facility treatment outcome? *Pain Med*. 2008;9:1081–1090.

12 Ritter MA, Wing JT, Berend ME, et al. The clinical effect of gender on outcome of total knee arthroplasty. *J Arthroplasty*. 2008;23:331–336.

13 Schmidt I, Rechter L, Hansen VK, et al. Prognosis of subacute low back pain patients according to pain response. *Eur Spine J*. 2008;17:57–63.

14 McGinn TG, Guyatt GH, Wyer PC, et al; Evidence-Based Medicine Working Group. Users' guides to the medical literature, XXII: how to use articles about clinical decision rules. *JAMA*. 2000;284:79–84.

15 Laupacis A, Sekar N, Stiell IG. Clinical prediction rules: a review and suggested modifications of methodological standards. *JAMA*. 1997;277:488–494.

16 Beneciuk JM, Bishop MD, George SZ. Clinical prediction rules for physical therapy interventions: a systematic review. *Phys Ther*. 2009;89:114–124.

17 May S, Rosedale R. Prescriptive clinical prediction rules in back pain research: a systematic review. *J Man Manip Ther*. 2009;17:36–45.

18 Kuijpers T, van der Windt DA, van der Heijden GJ, Bouter LM. Systematic review of prognostic cohort studies on shoulder disorders. *Pain*. 2004;109:420–431.

19 Beattie P, Nelson R. Clinical prediction rules: what are they and what do they tell us? *Aust J Physiother*. 2006;52:157–163.

20 Hancock MJ, Herbert RD, Maher CG. A guide to interpretation of studies investigating subgroups of responders to physical therapy interventions. *Phys Ther*. 2009;89:698–704.

21 Maher CG, Sherrington C, Herbert RD, et al. Reliability of the PEDro scale for rating quality of randomized controlled trials. *Phys Ther*. 2003;83:713–721.

22 Klebanoff MA. Subgroup analysis in obstetrics clinical trials. *Am J Obstet Gynecol*. 2007;197:119–122.

23 Pocock SJ, Assmann SE, Enos LE, Kasten LE. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. *Stat Med*. 2002;21:2917–2930.

24 Toll DB, Janssen KJ, Vergouwe Y, Moons KG. Validation, updating, and impact of clinical prediction rules: a review. *J Clin Epidemiol*. 2008;61:1085–1094.

25 Cai C, Pua YH, Lim KC. A clinical prediction rule for classifying patients with low back pain who demonstrate short-term improvement with mechanical lumbar traction. *Eur Spine J*. 2009;18:554–561.

26 Childs J, Fritz J, Flynn T, et al. A clinical prediction rule to identify patients with low back pain most likely to benefit from spinal manipulation: a validation study. *Ann Intern Med*. 2004;141:920–928.

27 Cleland JA, Childs JD, Fritz JM, et al. Development of a clinical prediction rule for guiding treatment of a subgroup of patients with neck pain: use of thoracic spine manipulation, exercise, and patient education. *Phys Ther*. 2007;87:9–23.

28 Cleland JA, Fritz JM, Whitman JM, et al. The use of a lumbar spine manipulation technique by physical therapists in patients who satisfy a clinical prediction rule: a case series. *J Orthop Sports Phys Ther*. 2006;36:209–214.

29 Currier L, Froehlich P, Carow S, et al. Development of a clinical prediction rule to identify patients with knee pain and clinical evidence of knee osteoarthritis who demonstrate a favorable short-term response to hip mobilization. *Phys Ther*. 2007;87:1106–1119.

30 Flynn T, Fritz J, Whitman J, et al. A clinical prediction rule for classifying patients with low back pain who demonstrate short-term improvement with spinal manipulation. *Spine (Phila Pa 1976)*. 2002; 27:2835–2843.

31 Fritz JM, Childs JD, Flynn TW. Pragmatic application of a clinical prediction rule in primary care to identify patients with low back pain with a good prognosis following a brief spinal manipulation intervention. *BMC Fam Pract*. 2005;6:29.

32 Hancock MJ, Maher CG, Latimer J, et al. Independent evaluation of a clinical prediction rule for spinal manipulative therapy: a randomised controlled trial. *Eur Spine J*. 2008;17:936–943.

33 Hicks GE, Fritz JM, Delitto A, McGill SM. Preliminary development of a clinical prediction rule for determining which patients with low back pain will respond to a stabilization exercise program. *Arch Phys Med Rehabil*. 2005;86:1753–1762.

34 Iverson CA, Sutlive TG, Crowell MS, et al. Lumbopelvic manipulation for the treatment of patients with patellofemoral pain syndrome: development of a clinical prediction rule. *J Orthop Sports Phys Ther*. 2008;38:297–309; discussion 309–212.

35 Lesher J, Sutlive T, Miller G, et al. Development of a clinical prediction rule for classifying patients with patellofemoral pain syndrome who respond to patellar taping. *J Orthop Sports Phys Ther*. 2006; 36:854–866.

36 Raney NH, Petersen EJ, Smith TA, et al. Development of a clinical prediction rule to identify patients with neck pain likely to benefit from cervical traction and exercise. *Eur Spine J*. 2009 Jan 4 [Epub ahead of print].

37 Sutlive TG, Mitchell SD, Maxfield SN, et al. Identification of individuals with patellofemoral pain whose symptoms improved after a combined program of foot orthosis use and modified activity: a preliminary investigation. *Phys Ther*. 2004; 84:49–61.

38 Tseng Y, Wang WT, Chen W, et al. Predictors for the immediate responders to cervical manipulation in patients with neck pain. *Man Ther*. 2006;11:306–315.

39 Vicenzino B, Collins N, Cleland J, McPoil T. A clinical prediction rule for identifying patients with patellofemoral pain who are likely to benefit from foot orthoses: a preliminary determination. *Br J Sports Med*. 2008 Dec 3 [Epub ahead of print].

40 Vicenzino B, Smith D, Cleland J, Bisset L. Development of a clinical prediction rule to identify initial responders to mobilisation with movement and exercise for lateral epicondylalgia. 2008 Oct 1 [Epub ahead of print].

41 Whitman JM, Cleland JA, Mintken P, et al. Predicting short-term response to thrust and nonthrust manipulation and exercise in patients post inversion ankle sprain. *J Orthop Sports Phys Ther*. 2009; 39:188–200.

42 Thiel HW, Bolton JE. Predictors for immediate and global responses to chiropractic manipulation of the cervical spine. *J Manipulative Physiol Ther*. 2008;31: 172–183.

43 Underwood MR, Morton V, Farrin A, UK BEAM Trial Team. Do baseline characteristics predict response to treatment for low back pain: secondary analysis of the UK BEAM dataset [ISRCTN32683578]. *Rheumatology (Oxford)*. 2007;46:1297–1302.

44 Behrendt CE, Gehan EA. Treatment-subgroup interaction: an example from a published, phase II clinical trial. *Contemp Clin Trials*. 2009;30:279–281.

45 Bisset L, Smidt N, Van der Windt DA, et al. Conservative treatments for tennis elbow do subgroups of patients respond differently? *Rheumatology (Oxford)*. 2007;46: 1601–1605.

46 Bisset L, Beller E, Jull G, et al. Mobilisation with movement and exercise, corticosteroid injection, or wait and see for tennis elbow: randomised trial. *BMJ*. 2006; 333:939.

47 Smidt N, van der Windt DA, Assendelft WJ, et al. Corticosteroid injections, physiotherapy, or a wait-and-see policy for lateral epicondylitis: a randomised controlled trial. *Lancet*. 2002;359:657–662.

48 Mulherin SA, Miller WC. Spectrum bias or spectrum effect: subgroup variation in diagnostic test evaluation. *Ann Intern Med*. 2002;137:598–602.

49 ISIS-2 Collaborative Group. Randomised trial of intravenous streptokinase, oral aspirin, both, or neither among 17,187 cases of suspected acute myocardial infarction: ISIS-2. *Lancet*. 1988;332:349–360.

50 Delitto A, Erhard RE, Bowling RW. A treatment-based classification approach to low back syndrome: identifying and staging patients for conservative management. *Phys Ther*. 1995;75:470–489.

Critical Appraisal of Clinical Prediction Rules That Aim to Optimize Treatment Selection for Musculoskeletal Conditions

Tasha R. Stanton, Mark J. Hancock, Christopher G. Maher and Bart W. Koes

PHYS THER. 2010; 90:843-854.
Originally published online April 22, 2010
doi: 10.2522/ptj.20090233