Prescriptive Clinical Prediction Rules in Back Pain Research: A Systematic Review

STEPHEN MAY, MSc¹; RICHARD ROSEDALE, PT²

Clinical prediction rules (CPRs) use a minimal set of variables from the history and physical examination to make a diagnosis, to predict an outcome, or to match patients to optimal interventions¹. CPRs facilitate the simultaneous consideration of several factors to determine the prognosis of individual patients and to assist with clinical decision-making². Sometimes validated CPRs are referred to as clinical decision rules. While clinical experience provides an intuitive sense of what aspects of a patient's presentation might be important in determining outcome, CPRs at-

tempt to formally test, simplify, and increase the accuracy of diagnostic and prognostic decisions^{3,4}. CPRs are an attempt to move away from intuitive guessing about which patient has a particular diagnosis or will respond to a particular intervention, and replace this with an evidence-based list of clinical characteristics to guide diagnostic or management decisions. CPRs are becoming increasingly common throughout the medical literature⁴. A CPR has been defined as a *decision-making tool that is derived from original research . . . and incorporates three or more variables*

*from the history, physical examination, or simple tests*⁴.

CPRs have been derived and validated for a number of musculoskeletal topics. The Ottawa ankle rule describes certain clinical features to identify patients needing ankle or foot radiography following ankle trauma seen in accident and emergency departments⁵⁻⁸. The Ottawa knee rules were designed for a similar patient population following knee trauma to determine the need for radiography⁹⁻¹². The Canadian C-spine rules were developed to determine the need for radiography in alert patients following cervical trauma¹³, and a CPR for diagnosing carpal tunnel syndrome has been derived¹⁴. CPRs have also featured in low back pain research, for instance being used to identify potential responders to manipulation and stabilization exercises¹⁵⁻²⁰. Given the increasing popularity of CPRs and use to advocate the prescription of treatment and investigation options within physical therapy, an examination of quality would seem appropriate. The aim of this systematic review was to examine the validity of CPRs advocated for the prescription of specific treatments for patients with low back pain against established criteria.

Methods

Medline was searched from 1966 to June 2008. The search terms were as follows:

ABSTRACT: Prescriptive clinical prediction rules (CPRs) are a way of using a small selection of clinical findings to match patients to optimal interventions. A number of CPRs have been developed for use with back pain patients, but these have not been systematically reviewed. The purpose of this review was to evaluate existing CPRs against established criteria to determine the quality of the studies and the overall development of the CPR against a set number of stages. Medline was searched up until June 2008, and 16 studies were reviewed that related to 9 different CPRs. These studies investigated and attempted to find clinical characteristics for responders to manipulation, stabilization exercise, physical therapy, chiropractic, traction, rehabilitation, usual care, and zygapophyseal joint injections. Eleven of these studies related to the derivation stage and five to the validation stage. The manipulation and stabilization CPRs had been the most studied. The derivation studies were mostly high quality, whereas none of the validation studies were. Some of the validation studies did not provide evidence that validated the CPR. Most of these CPRs need further evaluation before they can be applied clinically; most did not pass the lowest level of evidence hierarchy. As regards the manipulation CPR, evidence to date for its clinical utility is limited and contradictory. For the stabilization CPR, there was limited evidence that it may be considered but only with caution and in similar patients. Overall, there is limited evidence to support the general application of spinal CPRs.

KEYWORDS: Back Pain Research, Clinical Decision-Making, Clinical Prediction Rules

¹Faculty Health and Wellbeing, Sheffield Hallam University, Sheffield, UK

²Physiotherapist, London Health Sciences Centre, London, Ontario, Canada

Address all correspondence and requests for reprints to: Stephen May, s.may@shu.ac.uk

predicts OR clinical OR outcome OR risk; AND back pain OR low back pain. These search terms for CPRs were found to have the highest sensitivity (98%) for identifying clinical prediction studies compared to a gold-standard hand search²¹. The reference lists of all retrieved studies and the authors' databases were also searched. To be included, studies had to report some stage in the development of a CPR or set of clinical criteria to be used in the selection of conservative management of low back pain. The development had to be with patients with low back pain, it had to be based on prospective data collection, and the set of criteria had to suggest an intervention based on a limited set of clinical characteristics. Papers that looked at prognostic factors for treatment success or failure but that did not attempt to construct a set of criteria that suggested a specific course of action were not included. Papers that sought to explore prognostic clinical characteristics retrospectively were also not included.

Quality criteria for evaluating individual studies were first described by Wasson et al²²; these were later augmented^{4,23}. Stiell et al⁴ described a checklist of standards for the six stages in the development of CPR. This checklist was used to evaluate the quality of included studies, with additional detail provided by a recent text on evidence-based clinical epidemiology². As the standards were developed to assess CPR in use in emergency departments, they were slightly adapted as appropriate for back pain research and to clarify operational definitions. These adaptations were very minor. For instance, in the original, "clinically important outcomes" were defined as fractures or survival⁴; and because of the emphasis placed on "outcome and predictor variables being recorded blind to each other," we made this a separate item. For the validation criteria, although adequate sample size was referred to in the text, it was not a specific item in the original table⁴ but we made this an item. Furthermore, we defined reliability of interpretation of CPR as $\geq$ kappa 0.60, while a specific kappa value was not mentioned in the original description⁴. We felt these minor adap-

tations were not capricious and were supported by the original textual description⁴ and other relevant literature².

The first stage addresses this question: Is there a need for a prediction rule for this topic⁴? The need is evaluated by the following criteria: Is the prevalence of the condition high? Are common diagnostic tests inefficient? Is there variation in practice? Do clinicians believe that a CPR would be useful? and Are clinicians able to predict the outcome on clinical findings alone? We think low back pain meets all these criteria and will return to this topic in more detail in the discussion below.

The next five stages relate to the derivation, validation, implementation, evaluation, and dissemination of the CPR, and the quality criteria for these different stages have been described (Figure 1)^{22,23}. The second stage lists the key methodological issues in deriving or building a CPR. Once the rule is created, it should be refined and validated (Stage 3), then successfully implemented into clinical practice (Stage 4). Stages 5 and 6 relate to the economic evaluation of the rule and the strategy for its wider dissemination and implementation. The methodological standards for each stage were used as the quality checklist for CPRs that had reached that stage in the developmental process.

In this review, studies were judged to be of high quality if they scored 60% or more on the quality criteria. Overall judgment about the stage of development of the CPR was based on a pre-established hierarchy of evidence (see foot of results Table 1)^{3,24}. As these stages are sequential, a CPR would have to pass the earlier stages to be considered for the next level of evidence. Potentially relevant studies were reviewed by both researchers who independently decided which studies met the inclusion criteria; included studies were then quality-scored independently. Individual studies were scored according to the criteria outlined in Figure 2 (1A to 1E depending on the stage of the CPR development). For instance, for derivation studies, the criteria in 1A was used, and a total of 9 points was possible. Then the overall level of evidence of the CPR was determined according to the criteria in

Table 1, which included all studies in its development and validation process. At all stages, disagreement was resolved by consensus.

Results

The initial search netted 663 references; from scanning titles and abstracts, 39 were deemed potentially relevant and the articles were reviewed (Figure 2). Twelve of these matched our inclusion criteria and another four studies were located from the reference lists of these articles, which also met our inclusion criteria. Finally 16 studies were included in this review (Table 2). Eleven studies related to the derivation of a CPR^{17-20,25-30,32}, and five were deemed to be validation studies^{15,16,31,33,34}. Seven of the studies related to derivation or validation of CPRs for identifying patients who did or did not improve with manipulation^{15-19,33,34}. Three studies related to the derivation and construct validation of a stabilization CPR^{15,20,31}. Other CPR derivation studies sought to identify criteria for determining which patients would improve: with physiotherapy or chiropractor³⁰, with chiropractor^{25,28}, with usual care or minimal intervention²⁷, with zygapophyseal joint injections²⁹, which patients with sciatica would improve with traction²⁶, and which chronic back pain patients would improve with multidisciplinary rehabilitation³². Brennan et al¹⁵ conducted a trial in which patients were randomized to manipulation, stabilization exercises, or directional preference exercises, but their main intent was to investigate outcomes according to whether patients received their matched sub-grouping based on initial clinical presentation. The matching was decided by criteria from previously derived CPRs. Most CPRs were successful in that they supported a set of clinical characteristics that addressed the aim of the rule developers, but two were not^{32,34}.

Regarding the quality of the derivation studies, 10 met our standard of high quality^{17-20,25-29,32}. The most common weaknesses were lack of outcome and predictor variables being recorded blind to each other, lack of reliable predictor variables, and small sample size (Table

A. Methodological standards for derivation of clinical prediction rules

1. Clearly defined and clinically important outcome.
2. Clearly defined predictor variables collected prospectively.
3. Outcome and predictor variables recorded blind to each other.
4. Reliability of predictor variables ≥ 0.6 kappa.
5. Clear description of subjects—inclusion criteria, method of selection, clinical and demographic characteristics, and study settings.
6. Sample size: at least 10 outcome events per predictor variable in the CPR.
7. Data analysis should include univariate and multivariate analysis of significant predictor variables.
8. Sensibility of the rule—clinically reasonable, comprehensive, easy to use; suggests a course of action.
9. Accuracy of rule reported—sensitivity, specificity, negative predictive value, positive predictive value, likelihood ratios, with respective 95% confidence intervals.

B. Methodological standards for validation of clinical prediction rules

10. Prospective validation of CPR in a different patient population with different clinicians and in a different setting.
11. Unbiased selection of patients representing a wide spectrum of clinical conditions.
12. Application of rule is adequately taught.
13. Criterion standard applied to all patients to determine true outcome.
14. Accuracy of rule reported—sensitivity, specificity, negative predictive value, positive predictive value, likelihood ratios, with respective 95% confidence intervals.
15. Sample size: at least 10 outcome events per predictor variable in the CPR.
16. Reliability of interpretation of CPR kappa ≥ 0.6 .
17. Accuracy of interpretation of CPR.
18. Refinement of CPR, which is then re-validated in a new patient set.
19. Calculation of potential effect if CPR implemented into practice.

C. Standards for successful implementation into clinical practice

20. Implementation with a clinical trial of CPR, ideally a randomized controlled trial, and shown to affect clinicians' behavior.
21. Effect of CPR on resource use.
22. Careful follow-up to determine accuracy of the rule.
23. Survey of clinicians and patients to determine acceptability.

D. Economic analysis to determine if CPR is cost-effective

E. Plan for dissemination and implementation of CPR

F. For criteria for development of a CPR (first stage) see discussion

Source: Haynes et al²; Laupacis et al²³; Stiell & Wells⁴; Wasson et al²².

FIGURE 1. Checklist of standards for 6 stages in the development of a CPR

TABLE 1. Level of evidence for spinal CPR (level 5 = lowest level of evidence³)

Spinal CPR	Reference	Level 5	Level 4	Level 3	Level 2	Level 1
Manipulation CPR ¹	Flynn et al ¹⁷ Childs et al ¹⁶ Brennan et al ¹⁵ Cleland et al ³³ Hancock et al ³⁴	X*	X ⁵ X ⁵ X* X ⁶			
Manipulation CPR ^{1a}	Fritz et al ¹⁹	X*				
Manipulation CPR ²	Fritz et al ¹⁸	X*				
Stabilization CPR ³	Hicks et al ²⁰ Brennan et al ¹⁵ Teyhen et al ³¹	√	√ √			
Traction CPR ⁴	Fritz et al ²⁶	√				
PT/chiropractic	Skargren et al ³⁰ Leboeuf-Y de C et al ²⁸ Axen et al ²⁵	√ X* X*				
ZJ injection	Laslett et al ²⁹	√				
Rehabilitation	van der Hulst et al ³²	X ⁷				
Usual care	Jellema et al ²⁷	√				

X = failed to attain that level of evidence; √ = attained that level of evidence

* No blinded collection of outcome measures

CPR¹⁻⁴—relate to table 2

⁵ CPR did not pass first level of evidence.

⁶ Did not support CPR.

⁷ No rule was formulated.

Level 1—CPR that can be used in varied settings with confidence that they can change clinical behavior and improve patient outcomes

- Prospective validation in varied settings and impact analysis in various settings demonstrating change in clinical behavior and improved outcomes

Level 2—CPR that can be used in similar settings with confidence that they can change clinical behavior and improve patient outcomes

- One prospective validation in a different population and one impact analysis demonstrating change in clinical behavior and improved outcomes

Level 3—CPR that can be used in various settings with confidence

- Demonstrated accuracy in either one large prospective study or validated in several smaller settings that differ from each other; requires impact analysis

Level 4—CPR that clinicians may consider using with caution and only in similar patients

- Rules that have only been validated in another narrow prospective sample with blinded assessment of outcomes

Level 5—CPR that needs further evaluation before it can be applied clinically

- Derived but not validated or only validated in split samples, large retrospective databases or statistical techniques; with blinded assessment of outcomes

Source: McGinn et al³; Reilly & Evans²⁴

3). According to the criteria used, none of the validation studies were high quality, although two studies only just failed our standard for good quality^{16,34} (Table 4). After the first round of blinded quality scoring, overall agreement between the authors was 93% on 149 items, kappa = 0.86. Differences were resolved by consensus.

Regarding overall evidence in the development process of building a CPR for clinical use, some CPRs were at level 5, but most failed to make even this ini-

tial level as outcomes had not been gathered blind, which was a requirement²⁵ (Table 1). Most CPRs had simply been derived in one population but not validated in another. The manipulation CPR¹⁷ had not used blind outcome data-gathering, so it did not pass the first level of evidence. While it had been validated in another population¹⁶, its validation was compromised by its failure to identify manipulation-responders in a different country and healthcare setting³⁴. Overall, it must be concluded that the

evidence for this rule was limited and contradictory. The stabilization CPR²⁰ has been supported by a construct validation study³¹ and validated in another population¹⁵, and so this had the most evidence in its favor.

Discussion

Regarding the necessary criteria for the development of a CPR⁴ (high prevalence, inefficient diagnostic tests, variation in practice, potential development

FIGURE 2. Flowchart of search findings

and value of a CPR), back pain meets most of these requirements. Lifetime prevalence is between 54 and 84%³⁵. The reliability of most physical examination procedures used in low back pain assessment is weak³⁶. Surveys from the US, the Netherlands, and the UK show that physical therapy management is variable and non-standardized³⁷⁻⁴⁰, and back surgery rates vary considerably worldwide⁴¹. The popularity with which CPRs have been embraced in the US would suggest that there is a perceived need for them. Some elements of the physical examination have demonstrated predictive validity, such as centralization⁴²⁻⁴⁵ or a set of clinical criteria to identify patients with acute back pain who will respond well to manipulation¹⁶⁻¹⁹. Thus, it could be argued that the first stage in the development of a CPR—Is there a need for such rules in back pain research?—is mostly fulfilled.

CPRs for back pain have become increasingly popular in the last decade as it has been recognized that treatment based on classification is superior to non-sub-grouped treatment^{15,46,47}. CPRs are a way of sub-grouping responders to certain interventions based on clinical characteristics. For our systematic review, we found 16 studies relating to the derivation and validation of 9 different rules or sets of criteria to guide treatment, of which 3 rules and 7 studies related to responders and non-responders to manipulation. Other rules attempted to determine responders to physical therapy, chiropractic, stabilization exercises, traction, rehabilitation, usual care, and zygapophyseal joint injections. All of the studies were able to determine a set of clinical characteristics that identified responders except two^{32,34}.

Most of the derivation studies were of high quality, whereas the validation

studies were weaker, according to the criteria we used to evaluate quality. The most common weaknesses in the derivation studies were lack of blinding, lack of reliable predictor variables, and small sample sizes. The common weaknesses in the validation studies were small sample sizes, lack of reliable predictor variables, failing to establish accuracy of the interpretation of the CPR, lack of refinement of the CPR, and lack of calculation of the effect of the CPR if implemented into practice. In terms of overall development of the rules, all were at an early stage of development or did not make the first level, according to previously published levels of evidence²⁴. Two CPRs, for manipulation and stabilization exercises, had validation studies, but one of these related to construct validation³¹ and one challenged the validity of the CPR for manipulation³⁴. So most CPRs need further evaluation

TABLE 2. Details of included studies

Reference	Aim of CPR	Stage of CPR	Methods / Patients	Outcome
Skargren et al ¹³⁰	To identify predictor factors for physiotherapy and chiropractic	Derivation	RCT: 323 patients with LBP or NP randomized to physiotherapy or chiropractic	Duration < 1 week, no h/o LBP, high Oswestry score better with chiropractic treatment. Duration > 1 month, h/o LBP, low Oswestry score better with physiotherapy
Flynn et al ¹⁷	To identify patients who have > 50% improvement in Oswestry score with manipulation	Derivation	71 patients with LBP given standardized manipulation	CPR ¹ (≥4/5 criteria) sensitivity, 63%, specificity 97%, +LR 24.38
Fritz et al ¹⁸	To identify patients who fail to improve (< 50% change in Oswestry score) with manipulation	Derivation	71 LBP patients (same data set as Flynn et al., 2002)	CPR ² identified 63% of variance in manipulation outcome
Childs et al ¹⁶	To validate CPR ¹ for patients who improve with manipulation	Validation	RCT: 131 patients with LBP randomized to manipulation or exercises	CPR ¹ identified responders to manipulation (disability, $P = < 0.001$)
Leboeuf-Y de C et al ²⁸	To identify predictor factors for chiropractic in patients with persistent LBP	Derivation	875 patients with minimum LBP 30 days in previous year: 799, 598 and 512 followed up at 4 th visit and at 3 and 12 months	Female, social benefit, high pain severity, duration > 6 month, neck pain identified poor responders (sensitivity 10%, specificity 98%, -OR 2.2 43)
Axen et al ²⁵	To identify predictor factors for chiropractic in patients with LBP for status at 4 th visit	Derivation / refinement of previous model	1057 LBP patients mostly treated with manipulation	5 models tested: sensitivity 0–19%, specificity 96–100% identified poor responders
Fritz et al ¹⁹	To identify if patients with CPR ^{1A} are responders to manipulation	Derivation	141 LBP patients (data from Flynn et al., 2002 and Childs et al., 2004)	Sensitivity 56%, specificity 92%, +LR 7.2
Hicks et al ²⁰	To identify patients who have > 50% improvement in Oswestry score with stabilization exercises	Derivation	54 patients with LBP given standardized stabilization exercises	CPR ³ (≥3/4 criteria) sensitivity 56%, specificity 86%, +LR 4.0, -LR 0.52
Brennan et al ¹⁵	CPR ^{1A,3,1,2} plus a CPR for DP exercises used to classify patients in a RCT	Validation	123 patients with LBP randomized to SMT, SE, or DP exercises	Analysis by randomization and by classification sub-group = NS; but if randomized to appropriate classification group $P = 0.013$
Cleland et al ³³	To identify if patients with CPR ¹ respond to alternate manipulation technique	Validation	Case series of 12 patients	11/12 had > 57% reduction in Oswestry score
Jellema et al ²⁷	To identify predictor factors for response to usual care and minimal intervention	Derivation	Re-analysis of data from previous RCT: 314 patients randomized to usual care or minimal intervention	Usual care: model with 4 criteria area under the curve (AUC) 0.77. Minimal intervention: model with 6 criteria AUC 0.78.

(continued)

TABLE 2. Details of included studies (*continued*)

Reference	Aim of CPR	Stage of CPR	Methods / Patients	Outcome
Laslett et al ²⁹	To identify who would respond to ZJ injections (> 75% pain reduction)	Derivation	120 patients with chronic LBP who received a ZJ injection	5 models identified negative response to ZJ injection: sensitivity 85–100%, specificity 26–91%, PPV 13–55%, NPV 98–100%, +LR 1.4–9.7, –LR 0.0–0.17
Fritz et al ²⁶	To identify patients who have > 50% improvement in Oswestry score with traction	Derivation	RCT: 64 patients with LBP, leg pain and signs of nerve root compression randomized to extension exercises with or without traction	CPR ⁴ identified patients who responded to traction ($P < 0.05$)
Teyhen et al ³¹	To identify kinematic construct validity of CPR ³ for stabilization exercises	Construct validation	Spinal motion using fluoroscopic video in 20 patients positive to CPR compared to 20 healthy controls	10-variable model identified CPR membership; ≥ 5 present: sensitivity 80%, specificity 85%, +LR 5.3, –LR 0.2
Hancock et al ³⁴	Independent evaluation of manipulation CPR ¹	Validation	Pre-planned analysis of previous RCT: 239 patients with acute LBP randomized to manipulation or placebo	CPR ¹ did not identify patients who responded to manipulation (pain, $P = 0.805$; disability, $P = 0.600$)
van der Hulst et al ³²	To identify predictor factors for multidisciplinary rehabilitation (MDR)	Derivation	Reanalysis of data from previous RCT: 163 patients with chronic LBP randomized to MDR or usual care	2 models tested explained 18.5% to 42.3% of outcome variance

CPR = clinical prediction rule; LBP = low back pain; NP = neck pain; RCT = randomized controlled trial; DP = directional preference exercises; ZJ = zygapophyseal joint; +LR = positive likelihood ratio; –LR = negative likelihood ratio; PPV = positive predictive value; NPV = negative predictive value; SMT = spinal manipulative therapy; SE = stabilization exercises; NS = non-significant; FABQW = fear avoidance beliefs questionnaire-work

CPR¹: duration LBP < 16 days, no symptoms below knee, FABQW < 19 points, ≥ 1 segment hypomobile, ≥ 1 hi > 35° medial rotation.

CPR^{1A}: duration LBP < 16 days, no symptoms below knee.

CPR²: longer duration LBP, symptoms buttock or leg, absence hypomobility, less hip range of movement, left-right hip medial rotation similar, negative Gaenslen sign.

CPR³: age < 40 years, average straight leg raise > 91°, aberrant movement, positive prone instability test.

CPR⁴: peripheralization with extension movement, crossed straight leg raise.

before they can be applied clinically. The evidence for the manipulation CPR is limited and contradictory, and the stabilization CPR should only be used with caution and only in similar patients²⁴. Our findings suggest that none of the CPRs can be used in various settings with confidence²⁴.

A number of clinical and research implications emanate from the conclusions from this review. Spinal CPRs have very limited support from available literature; if this is to be improved upon, several further stages are required in the development of these CPRs. Blinded assessment of outcomes is required in dif-

ferent populations. CPRs need to demonstrate accuracy in either one large prospective study or several settings that differ from each other. CPRs need prospective validation in a different population and an impact analysis demonstrating change in clinical behavior and improved outcomes. CPRs also need a prospective validation in varied settings and impact analysis in varying settings to demonstrate that change in clinical behavior results in improved outcomes.

To our knowledge, no previous systematic reviews of CPRs have been published. Certain potential weaknesses were revealed during our review. We en-

deavored to be as systematic and comprehensive as possible; however, the search strategy, although recommended as optimal for CPR²¹, failed to locate all the studies that we subsequently located through a review of the reference lists of other studies. The checklist we used did not appear to use a consensus method thus, key ingredients might be missing. The test checklist has also not been formally validated. Furthermore, we made some minor modifications to it that were not validated but were in line with previous commentators on the topic²⁴. Language and publication bias may potentially affect all reviews, but as the

TABLE 3. Quality score—Derivation studies

Reference	1	2	3	4	5	6	7	8	9	Total
Skargren et al ³⁰	0	1	0	0	1	0	0	1	0	3
Flynn et al ¹⁷	1	1	0	0	1	0	1	1	1	6
Fritz et al ¹⁸	1	1	0	0	1	0	1	1	1	6
Leboeuf-Y de C et al ²³	1	1	0	0	1	1	1	1	1	7
Axen et al ²⁵	1	1	0	0	1	1	1	1	1	7
Fritz et al ¹⁹	1	1	0	1	1	1	0	1	1	7
Hicks et al ²⁰	1	1	1	1	1	0	1	1	1	8
Jellema et al ²⁷	1	1	0	0	1	0	1	1	1	6
Laslett et al ²⁹	1	1	1	0	1	0	1	1	1	7
Fritz et al ²⁶	1	1	1	1	1	1	1	1	0	8
van der Hulst et al ³²	1	1	1	1	1	0	1	0	0	6

60% = 5.4

1. Clearly defined and clinically important outcome.
2. Clearly defined predictor variables collected prospectively.
3. Outcome and predictor variables recorded blind to each other.
4. Reliability of predictor variables ≥ 0.6 kappa.
5. Clear description of subjects—inclusion criteria, method of selection, clinical and demographic characteristics and study settings.
6. Sample size: at least 10 outcome events per predictor variable in the CPR.
7. Data analysis should include univariate and multivariate analysis of significant predictor variables.
8. Sensibility of the rule—clinically reasonable, comprehensive, easy to use, suggests a course of action.
9. Accuracy of rule reported—sensitivity, specificity, negative predictive value, positive predictive value, likelihood ratios, with respective 95% confidence intervals.

TABLE 4. Quality score—Validation studies

Reference	10	11	12	13	14	15	16	17	18	19	Total
Childs et al ¹⁶	1	1	1	1	1	0	0	0	0	0	5
Brennan et al ¹⁵	1	1	0	1	0	0	0	0	0	0	3
Cleland et al ³³	1	1	0	1	0	0	0	0	0	0	3
Hancock et al ³⁴	1	1	1	1	0	1	0	0	0	0	5
Teyhen et al ³¹	0	1	0	0	1	0	0	0	0	0	2

60% = 6

10. Prospective validation of CPR in a different patient population with different clinicians and in a different setting.
11. Unbiased selection of patients representing a wide spectrum of clinical condition.
12. Application of rule is adequately taught.
13. Criterion standard applied to all patients to determine true outcome.
14. Accuracy of rule reported—sensitivity, specificity, negative predictive value, positive predictive value, likelihood ratios, with respective 95% confidence intervals.
15. Sample size: at least 10 outcome events per predictor variable in the CPR.
16. Reliability of interpretation of CPR kappa ≥ 0.6 .
17. Accuracy of interpretation of CPR.
18. Refinement of CPR, which is then re-validated in a new patient set.
19. Calculation of potential effect if CPR implemented into practice.

evidence overall was very weak for all the CPRs, this would be unlikely to change our review conclusions.

Conclusion

Predictive CPRs are a method of using a few clinical characteristics to identify potential responders to different interventions. Following a systematic review, we identified 16 studies developing 9 different predictive CPRs for low back pain treatments. Most were successful in identifying a set of characteristics, and the derivation studies were mostly of high quality. However, the validation studies were weaker and did not always validate the CPRs to which they applied. Overall development of all the CPRs was limited and, from our findings, none should be applied in the general low back pain population with confidence. The evidence to date is not yet compelling enough to support the general use of these CPRs.

REFERENCES

- Childs JD, Cleland JA. Development and application of clinical prediction rules to improve decision making in physical therapy practice. *Phys Ther* 2006;86:122–131.
- Haynes RB, Sackett DL, Guyatt GH, Tugwell P. *Clinical Epidemiology: How to do Clinical Practice Research*. 3rd ed. Philadelphia, PA: Lippincott, Williams & Wilkins, 2006.
- McGinn T, Guyatt G, Wyer P, Naylor CD, Stiell I. How to use articles about clinical prediction rules. *JAMA* 2000;284:79–84.
- Stiell IG, Wells GA. Methodological standards for the development of clinical decision rules in emergency medicine. *Ann Emerg Med* 1999;33:437–447.
- Stiell IG, Greenberg GH, McKnight RD, Nair RC, McDowell I, Worthington JR. A study to develop clinical decision rules for the use of radiography in acute ankle injuries. *Ann Emerg Med* 1992;21:384–390.
- Stiell IG, McKnight RD, Greenberg GH, et al. Implementation of the Ottawa ankle rules. *JAMA* 1994;271:827–832.
- Auleley GR, Ravaud P, Giraudeau B, Kerboull L, Nizard R, Massin P. Implementation of the Ottawa ankle rules in France: A multicenter randomized controlled trial. *JAMA* 1997;277:1935–1939.
- Verbeek PR, Stiell IG, Herbert G, Sellens C. Ankle radiograph utilization after learning a decision rule: A 12-month follow-up. *Acad Emerg Med* 1997;4:776–779.
- Stiell IG, Greenberg GH, Wells GA, McKnight RD, Cwinn AA, Cacciotti T. Derivation of a decision rule for the use of radiography in acute knee injuries. *Ann Emerg Med* 1995;26:405–413.
- Stiell IG, Greenberg GH, Wells GA, McDowell I, Cwinn AA, Smith NA. Prospective validation of a decision rule for the use of radiography in acute knee injuries. *JAMA* 1996;275:611–615.
- Stiell IG, Wells GA, Hoag RH, et al. Implementation of the Ottawa knee rule for the use of radiography in acute knee injuries. *JAMA* 1997;278:2075–2079.
- Nichol G, Stiell IG, Wells GA, Juergensen LS, Laupacis A. An economic analysis of the Ottawa knee rule. *Ann Emerg Med* 1999;34:438–447.
- Stiell IG, Wells GA, Vandemheen KL, Clement CM, Lesiuk H, De Maio VJ. The Canadian C-spine rule for radiography in alert and stable trauma patients. *JAMA* 2001;286:1841–1848.
- Wainner RS, Fritz JM, Irrgang JJ, Delitto A, Allison S, Boninger ML. Development of a clinical prediction rule for the diagnosis of carpal tunnel syndrome. *Arch Phys Med Rehabil* 2005;86:609–618.
- Brennan GP, Fritz JM, Hunter SJ, Thackeray A, Delitto A, Erhard RE. Identifying subgroups of patients with acute/subacute “non-specific” low back pain. *Spine* 2006;31:623–631.
- Childs JD, Fritz JM, Flynn TW, et al. A clinical prediction rule to identify patients with low back pain most likely to benefit from spinal manipulations: A validation study. *Ann Intern Med* 2004;141:920–928.
- Flynn T, Fritz J, Whitman J, Wainner R, Magel J, Rendeiro D. A clinical prediction rule for classifying patients with low back pain who demonstrate short-term improvement with spinal manipulation. *Spine* 2002;27:2835–2843.
- Fritz JM, Whitman JM, Flynn TW, Wainner RS, Childs JD. Factors related to the inability of individuals with low back pain to improve with a spinal manipulation. *Phys Ther* 2004;84:173–190.
- 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. Available at: <http://www.biomedcentral.com/1471-2296/6/29>
- Hicks GE, Fritz JE, 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.
- Ingui BJ, Rogers MAM. Searching for clinical prediction rules in MEDLINE. *J Am Med Inform Assoc* 2001;8:391–397.
- Wasson JH, Sox HC, Neff RK, Goldman L. Clinical application rules: Application and methodological standards. *N Engl J Med* 1985;313:793–799.
- Laupacis A, Sekar N, Stiell IG. Clinical prediction rules: A review and suggested modifications of methodological standards. *JAMA* 1997;277:488–494.
- Reilly BM, Evans AT. Translating clinical research into clinical practice: Impact of using prediction rules to make decisions. *Ann Intern Med* 2006;144:201–209.
- Axen I, Jones JJ, Rosenbaum A, et al. The Nordic back pain subpopulation program: Validation and improvement of a predictive model for treatment outcome in patients with low back pain receiving chiropractic treatment. *J Manipulative Physiol Ther* 2005;28:381–385.
- Fritz JM, Lindsay W, Matheson JW, et al. Is there a subgroup of patients with low back pain likely to benefit from mechanical traction? Results of a randomized clinical trial and subgrouping analysis. *Spine* 2007;32:E793–E800.
- Jellema P, van der Horst HE, Vlaeyen JW, Stalman WA, Bouter LM, van der Windt DA. Predictors of outcome in patients with (sub) acute low back pain differ across treatment groups. *Spine* 2006;31:1699–1705.
- Leboeuf-Y de C, Gronstvedt A, Borge JA, et al. The Nordic back pain subpopulation program: Demographic and clinical predictors for outcome in patients receiving chiropractic treatment for persistent low-back pain. *J Manipulative Physiol Ther* 2004;27:493–502.
- Laslett M, McDonald B, Aprill CN, Tropp H, Oberg B. Clinical predictors of screening lumbar zygapophyseal joint blocks: Development of clinical prediction rules. *Spine J* 2006;16:370–379.
- Skargren EI, Carlsson PG, Oberg BE. One-year follow-up comparison of the cost and effectiveness of chiropractic and physio-

- therapy as primary management for back pain: Subgroup analysis, recurrence, and additional health care utilization. *Spine* 1998; 23:1875–1883.
31. Teyhen DS, Flynn TW, Childs JD, Abraham LD. Arthokinematics in a subgroup of patients likely to benefit from a lumbar stabilization exercise program. *Phys Ther* 2007; 87:313–325.
32. van der Hulst M, Vollenbroek-Hutten MM, Groothuis-Oudshoorn K, Hermens HJ. Multidisciplinary rehabilitation treatment of patients with chronic low back pain: A prognostic model for its outcome. *Clin J Pain* 2008;24:421–430.
33. Cleland JA, Fritz JM, Whitman JM, Childs JD, Palmer JA. 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.
34. Hancock MJ, Maher CG, Latimer J, Herbert RD, McAuley JH. Independent evaluation of a clinical prediction rule for spinal manipulative therapy: A randomized controlled trial. *Eur Spine J* 2008; 17:936–943.
35. Dionne CE. Low back pain. In Crombie IK, Croft PR, Linton SJ, LeResche L, & Von Korff M, eds. *Epidemiology of Pain*. Seattle, WA: IASP Press, 1999.
36. May S, Littlewood C, Bishop A. Reliability of procedures used in the physical examination of non-specific low back pain: A systematic review. *Aus J Physiother* 2006;52:91–102.
37. Foster NE, Thompson KA, Baxter GD, Allen JM. Management of non-specific low back pain by physiotherapists in Britain and Ireland: A descriptive questionnaire of current clinical practice. *Spine* 1999;24:1332–1342.
38. Jette MA, Delitto A. Physical therapy treatment choices for musculoskeletal impairments. *Phys Ther* 1997;77:145–154.
39. Mielenz TJ, Carey TS, Dyrek DA, Harris BA, Garrett JM, Darter J. Physical therapy utilization by patients with acute low back pain. *Phys Ther* 1997;77:1040–1051.
40. van Baar ME, Dekker J, Bosveld W. A survey of physical therapy goals and interventions for patients with back and knee pain. *Phys Ther* 1998;78:33–42.
41. Cherkin DC, Deyo RA, Loeser JD, Bush T, Waddell G. An international comparison of back surgery rates. *Spine* 1994;19:1201–1206.
42. Werneke M, Hart DL, Cook D. A descriptive study of the centralization phenomenon. *Spine* 1999;24:676–683.
43. Werneke M, Hart DL. Centralization phenomenon as a prognostic factor for chronic low back pain and disability. *Spine* 2001; 26:758–765.
44. Werneke M, Hart DL. Discriminant validity and relative precision for classifying patients with non-specific neck and back pain by anatomic pain patterns. *Spine* 2003; 28:161–166.
45. Werneke M, Hart DL. Categorizing patients with occupational low back pain by use of the Quebec Task Force Classification system versus pain pattern classification procedures: Discriminant and predictive validity. *Phys Ther* 2004;84:243–254.
46. Fritz JM, Delitto A, Erhard RE. Comparison of classification-based physical therapy with therapy based on clinical practice guidelines for patients with acute low back pain: A randomized clinical trial. *Spine* 2003;28:1363–1371.
47. Long A, Donelson R, Fung T. Does it matter which exercise? A randomized control trial of exercise for low back pain. *Spine* 2004;29:2593–2602.