Systematic review

Clinical prediction rules in the physiotherapy management of low back pain: A systematic review

Robin Haskins*, Darren A. Rivett, Peter G. Osmotherly

School of Health Sciences, The University of Newcastle, NSW 2308, Australia

A R T I C L E   I N F O

Article history:
Received 16 January 2011
Received in revised form 28 April 2011
Accepted 9 May 2011

Keywords:
Low back pain
Physical Therapy (Specialty)
Decision making
Probability

A B S T R A C T

Objective: To identify, appraise and determine the clinical readiness of diagnostic, prescriptive and prognostic Clinical Prediction Rules (CPRs) in the physiotherapy management of Low Back Pain (LBP).

Data sources: MEDLINE, EMBASE, CINAHL, AMED and the Cochrane Database of Systematic Reviews were searched from 1990 to January 2010 using sensitive search strategies for identifying CPR and LBP studies.

Study selection: Two independent reviewers used a two-phase selection procedure to identify studies that explicitly aimed to develop one or more CPRs involving the physiotherapy management of LBP. Diagnostic, prescriptive and prognostic studies investigating CPRs at any stage of their development, derivation, validation, or impact-analysis, were considered for inclusion using a priori criteria. 7453 unique records were screened with 23 studies composing the final included sample.

Data extraction: Two reviewers independently extracted relevant data into evidence tables using a standardised instrument.

Data synthesis: Identified studies were qualitatively synthesized. No attempt was made to statistically pool the results of individual studies. The 23 scientifically admissible studies described the development of 25 unique CPRs, including 15 diagnostic, 7 prescriptive and 3 prognostic rules. The majority (65%) of studies described the initial derivation of one or more CPRs. No studies investigating the impact phase of rule development were identified.

Conclusions: The current body of evidence does not enable confident direct clinical application of any of the identified CPRs. Further validation studies utilizing appropriate research designs and rigorous methodology are required to determine the performance and generalizability of the derived CPRs to other patient populations, clinicians and clinical settings.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

A Clinical Prediction Rule (CPR) is “a clinical tool that quantifies the individual contributions that various components of the history, physical examination and basic laboratory results make towards the diagnosis, prognosis, or likely response to treatment in an individual patient” (McGinn et al., 2008). These tools aim to facilitate clinical decision-making in the assessment and treatment of individual patients (Beattie and Nelson, 2006) and are thought to be of greatest potential when they are developed and utilised for clinical conditions that involve complex clinical decision making.

Low Back Pain (LBP) is a common and costly complaint specifically identified as an ideal target for CPRs due to its heterogeneous population and numerous treatment alternatives (Fritz, 2009). Clinical trials (Fritz et al., 2003; Long et al., 2004; Brennan et al., 2006) have highlighted the benefits of LBP classification systems that aim to ‘match’ interventions according to the particular sub-group of patients. Concordantly, there has been a surge in the number of publications that discuss the development and application of CPRs that are relevant to the assessment and treatment of LBP (Beneciuk et al., 2009; May and Rosedale, 2009; Stanton et al., 2010). However, before a CPR can be confidently incorporated into clinical practice, it must undergo a process of development that investigates the rule’s performance, generalizability, and influence upon clinical outcomes and/or resource consumption.

Numerous publications have discussed the common methodological standards that should apply to the development of CPRs (Wasson et al., 1985; Laupacis et al., 1997; Randolph et al., 1998; Stiell and Wells, 1999; McGinn et al., 2000, 2008; Beattie and...
Nelson, 2006; Childs and Cleland, 2006; Cook, 2008), although the specific criteria often differ between studies. It is, however, commonly accepted that a hierarchical process of rule development is utilised (McGinn et al., 2000), initially commencing with derivation of the rule, and then progressing to a process of validation and then subsequent investigation of its clinical impact.

CPRs that have been derived, but not yet validated are not considered ready for clinical use (McGinn et al., 2000, 2008; Reilly and Evans, 2006). Even rigorously derived rules may reflect chance associations between variables and the target condition or outcome, or they may be unique to the studied population or other characteristics of that clinical setting (McGinn et al., 2008). This is reflected in the finding that most CPRs perform less accurately in subsequent studies involving different patients (Toll et al., 2008).

Despite these limitations, it has been suggested that derived CPRs may inform clinical practice by providing clinicians with an understanding of some of the most important predictors of a given target condition or outcome (McGinn et al., 2008).

The process of validation investigates a rule's performance and generalizability to other patient populations, clinicians and clinical settings. Importantly, the validation of a CPR cannot be accomplished by a single study, but requires a process involving a series of studies that test the internal and external validity of a rule across a broad range of clinical environments (Hancock et al., 2009a). Narrow validation of a CPR involves investigating the performance of the rule in a similar patient population and similar clinical setting to the derivation study. A CPR that has been demonstrated to perform well in such a setting is considered to be ready for cautious clinical application to patients that are representative of the studied population (McGinn et al., 2000, 2008).

Confidence in the rule's accuracy improves as it is progressively investigated in various other settings comprising different clinicians and patients with differing prevalence of disease or injury and with differing responsiveness to treatment. CPRs that demonstrate consistent and strong performance in this process of broad validation are considered ready to be applied in clinical practice with confidence in their accuracy (McGinn et al., 2000).

It is not inappropriate, however, to assume that the clinical application of a rigorously-validated rule will result in improved clinical care. Impact-analysis is the process of CPR development that involves testing a rule's ability to positively influence clinical outcomes and/or resource consumption, and change clinicians' behaviour (McGinn et al., 2008). Ideally, this involves a direct comparison to usual clinical care or judgement (Toll et al., 2008).

Rules that are demonstrated to be highly accurate and perform well across multiple clinical environments may actually be no more accurate, or even worse, than unassisted clinician judgement. Rigorously-validated CPRs that have been demonstrated to produce beneficial clinical consequences via impact-analysis can be confidently incorporated into clinical practice (McGinn et al., 2000, 2008; Reilly and Evans, 2006).

Before clinicians can consider incorporating the growing number of CPRs into their practice, a determination of their readiness for clinical application is required. Previous systematic reviews of CPRs relevant to physiotherapy (Beneciuk et al, 2009; May and Rosedale, 2009; Stanton et al., 2010) have focused upon recognising that methodological standards for the development of CPRs differ to that of other types of studies (Stanton et al., 2009). Other reviews (Wisnivesky et al., 2005; Dahri and Loewen, 2007; May and Rosedale, 2009; Stanton et al., 2010) have developed checklists based upon previously proposed methodological standards. A potential problem with this approach is that the proposed methodological standards differ between texts, leading to the possible inclusion of extraneous criteria or the possible exclusion of important criteria dependent upon the text(s) selected. For example, although Stiell and Wells (1999) highlight the importance of a representative sample in the derivation phase of a rule's development, this criterion is omitted from other well-cited texts (Laupacis et al., 1997; McGinn et al., 2000).

The aim of the present review was to identify, appraise and determine the clinical readiness of CPRs in the physiotherapy management of LBP.

2. Methods

2.1. Data sources and searches

A systematic literature search of MEDLINE, EMBASE, CINAHL, AMED and the Cochrane Database of Systematic Reviews from 1990 to January 2010 limited to articles available in English was conducted. A sensitive search strategy for CPRs (Ingui and Rogers, 2001) that has been used in previous systematic reviews (Dahri and Loewen, 2007; Beneciuk et al., 2009; May and Rosedale, 2009) was employed in combination with the search strategy recommended by the Cochrane Back Group (2009) for identifying articles relevant to LBP (Appendix 1). Citation tracking and hand-searching of relevant journals were used as supplemental search strategies.

2.2. Study selection

For a study describing the development of a CPR to be included in the review it had to meet the following criteria:

1. Studies needed to explicitly aim to develop one or more CPRs involving the physiotherapy management of LBP. The operational definition of a CPR for this study was that defined by McGinn et al. (2008). Although it has been suggested that there should be a minimum of three variables in a CPR (Laupacis et al., 1997; Stiell and Wells, 1999), previous systematic reviews (Tamariz et al., 2004; Wisnivesky et al., 2005) have included studies with two or more predictor variables. To ensure all relevant studies were identified, this review used the more liberal definition of a CPR as that containing two or more predictor variables.

2. Substantial practice variation between low back pain treatment providers (Kent and Keating, 2005; Werner et al., 2005) including marked differences in the methods chosen to assess this condition (Kent et al., 2008) makes it arguably inappropriate to assume that the selection and assessment of potential predictor variables will generalise across disciplines. Thus, it was determined a priori that for a study to be included, the assessment of potential predictor variables was required to be performed by a physiotherapist to ensure their direct relevance to the primary research aim.
3. Consistent with the definition of a CPR employed in this review (McGinn et al., 2008), predictor variables were required to be independently meaningful.
4. Diagnostic, prescriptive and prognostic studies investigating CPRs at any stage of their development (McGinn et al., 2000), derivation, validation, or impact-analysis, were included.

No restriction was placed upon the type of potential predictor variables (eg. history items, imaging modalities, physical examination items, psychological variables etc) under investigation in the studies considered for inclusion. Further, no restriction was placed upon the clinical setting or the type of patients with LBP under investigation in studies considered for eligibility in this review.

Identified studies were downloaded into an electronic reference management system (EndNote, version X2.0.11) and duplicates were removed.

Two reviewers performed the first-stage screening of titles and abstracts based upon the stated eligibility criteria. Any study denoted eligible by either reviewer was progressed to the second-stage of eligibility screening. Additionally, studies identified by citation tracking and hand-searching of relevant journals were progressed to the second-stage. The full-text of included studies was obtained and examined by two reviewers. During this second-stage of screening, concordance between reviewers determined inclusion, with disagreements resolved by consensus, or if needed by a third reviewer.

2.3. Data extraction and quality assessment

A standardised instrument was used for data extraction. Information collected from each study included the country of origin, the number of rules developed, study design, stated objective, and details of the patient population. The reviewers also investigated whether included studies specifically used the term "clinical prediction rule". The hierarchy of evidence for CPRs (McGinn et al., 2000) was initially employed to determine which stage of CPR development an article was describing. Studies were subsequently defined as derivation, validation or impact-analysis.

Consistent with the aim of the present review, the quality of the included studies were evaluated against the well-cited methodological standards that are employed by researchers in the development of all forms of CPRs. These criteria reflect the necessary methodological requirements to develop any form of a CPR and should be considered as an extension to the various methodological requisites that are specific to the underlying study design. In the absence of an appropriate standardised tool and to avoid the limitations of unsystematically selecting criteria from previous reports, we initially identified the key texts describing the methodological standards common to the development of all forms of CPRs, including those used in previous systematic reviews. From these texts, five (Laupacis et al., 1997; Stell and Wells, 1999; McGinn et al., 2000; Beattie and Nelson, 2006; Childs and Cleland, 2006) were selected based upon their inclusion in previous reviews, their number of citations in MEDLINE and EMBASE and their relevance to the research aim. Criteria that were represented in two or more of the five selected texts were included in the methodological appraisal of the included studies. This review employed definitions of the accepted CPR quality criteria that have been previously published (Laupacis et al., 1997; Stell and Wells, 1999; McGinn et al., 2000; Beattie and Nelson, 2006; Childs and Cleland, 2006). A checklist was subsequently developed for each of the three phases of rule development. The research designs of the included studies were anticipated to be extensively heterogeneous ranging from randomised controlled and observational intervention studies, to cross-sectional diagnostic investigations and longitudinal prognostic studies. Consequently, no attempt was made to appraise and contrast the included studies against the methodological standards that are specific to their unique underlying research design.

Two reviewers independently appraised the methodological quality of the included studies. Each criterion was evaluated independently with concordance between examiners determining the appropriate outcome. Disagreement was resolved by consensus and if needed, by a third reviewer. For a criterion to be marked as being met, studies must have entirely fulfilled the requirements of that criterion with no occasions of disparity. For example, in studies that aimed to develop two or more CPRs, all rules within the study must have achieved the requirements of that criterion for it to be considered met. Criteria marked as ‘unclear’ or ‘not met’ were consolidated to enable the dichotomisation of each criterion as ‘met’ or ‘not met’.

The research design of studies investigating predictors of responsiveness to intervention were specifically evaluated for their ability to identify treatment-effect modifiers. These variables, also known as ‘moderators’, are the baseline characteristics that identifies subgroups of patients with differing treatment-effect sizes for a given intervention (Kraemer et al., 2002, 2006; Turner et al., 2007; MacKinnon and Lueckne, 2008; Kraemer and Gibbons, 2009). Recent commentary in the rehabilitation literature (Hancock et al., 2009a) has highlighted the inadequacy of single-arm research designs in identifying the variables that influence a patient’s responsiveness to an intervention. Controlled trials are required in all stages of prescriptive CPR development to discriminate between the non-specific prognostic factors associated with clinical outcome, and the specific treatment-effect modifying variables that help further guide clinical decision making. The distinction between single-arm prescriptive CPR studies and prognostic CPR studies was determined by the stated clinical aim of the CPR in each study.

2.4. Data synthesis and analysis

Due to the anticipated heterogeneity of the included studies, no attempt was made to statistically pool the results of individual studies.

Between-rater agreement was evaluated for each stage of the screening process and for the methodological appraisal of the included studies. The absolute and chance-corrected degrees of agreement (κ) with 95% confidence intervals were calculated for both stages of the screening procedure. Between group comparisons were analysed following exploratory data analysis and relevant parametric or non-parametric tests were applied. All statistical analyses were conducted using Stata 11.0.

3. Results

3.1. Study selection

The database search strategy yielded 10 202 studies. Another twelve studies were identified via hand-searching of relevant journals and citation-tracking of included studies. Following the removal of duplicate records, 7453 records were screened via title.
and abstract with 381 records progressing to the second stage of screening. The full-text copies of these studies were located and reviewed with 23 studies composing the final included sample. The reasons for exclusion are highlighted in Fig. 1.

The absolute agreement between raters for the first and second-round screening procedures was 96.6% and 94% respectively. The chance-corrected degree of agreement was observed to be "moderate" (Sackett et al., 1991) for both procedures with $\kappa = 0.49$ (95% CI 0.43–0.55) for the screening by titles and abstracts, and $\kappa = 0.53$ (95% CI 0.35–0.72) for the screening by full-text. All but one episode of disagreement between raters was resolved by consensus, with the remaining study ruled to be included by the third reviewer.

3.2. Characteristics of included studies

The majority of included studies ($n = 15$) originated from the USA. Three studies were conducted in Australia and two in The Netherlands. The remaining three studies were conducted in Singapore, Spain and the United Kingdom. Although the search strategy enabled the inclusion of studies from 1990, the earliest year of publication of the included sample was 2002. The majority of included studies developed just one CPR, although some studies investigated up to five rules in one publication.

Fifteen derivation and eight validation studies compose the included sample. No studies investigating the impact phase of rule development were identified. Fourteen studies describe CPRs used to influence treatment decision-making. Ten (43%) of the included studies relate to the prediction of clinical outcome with the use of spinal manipulation. Seven studies concern diagnosis and only two prognostic studies were included. Across the 23 included publications, 25 unique CPRs are described including 15 diagnostic, 7 prescriptive and 3 prognostic rules. Appendices 2a, 2b and 2c detail the identified CPRs and the relevant studies that have contributed to their development.

3.3. Qualitative appraisal of included studies

Quality scoring for the derivation and validation studies is provided in Tables 1 and 2 respectively. "Substantial" (Sackett et al., 1991) between-rater agreement was observed for the quality scoring with an absolute degree of agreement of 88.7% ($\kappa = 0.74$, 95% CI 0.66–0.81). Three episodes of disagreement required resolution by a third reviewer, with the remaining disagreements being resolved by consensus.

Five of the 14 publications (36%) concerning prescriptive CPRs used a randomised controlled-study design that would permit the identification of treatment-effect modifiers.

Although all included studies satisfied the operational definition of a CPR, not all articles specifically used the term. Of the 23 included studies, only 15 (65%) explicitly used the term "clinical prediction rule" when describing the clinical tool being developed. It was more common for prescriptive studies to use the term "clinical prediction rule", compared to diagnostic and prognostic studies ($p < 0.001$).

4. Discussion

There has been a rapid growth in the number of studies reporting upon the development of CPRs in the physiotherapy literature. This trend mirrors that seen in Medicine, particularly in the fields of Emergency and Intensive Care and may be reflective of a progressive move towards models of clinical decision-making that are increasingly data-driven and firmly founded upon the process of scientific enquiry. The quest to identify meaningful subgroups of patients will have important implications for clinical practice, particularly for presentations, such as LBP, which are confounded by their degree of heterogeneity and numerous treatment alternatives.

To our knowledge, the present review is the first to systematically locate, appraise and determine the clinical readiness of diagnostic, prescriptive and prognostic CPRs involving the physiotherapy management of LBP in all phases of their development. Twenty-five unique CPRs were identified encompassing a diverse range of factors. While the growth in this research is arguably important for LBP treatment providers, this observed large variation in CPR themes may reflect the current lack of understanding of clinicians’ priorities for CPRs. Investigation of the areas of perceived clinical need for CPRs would facilitate the development of rules with the greatest potential to positively influence clinical practice (Eagles et al., 2008).

Previous systematic reviews of CPRs in the physical rehabilitation literature (Beneciuk et al., 2009; May and Rosedale, 2009; Stanton et al., 2010) have included four studies involving the physiotherapy management of LBP which were excluded in the present review. Two studies (Fritz et al., 2004, 2007) included in earlier reviews have investigated the characteristics that are associated with treatment outcomes. However, as both studies did not develop a clinical tool that may be applied to an individual patient they did not meet the present review’s eligibility criteria. One excluded study (Brennan et al., 2006) was determined to have investigated a classification system while the other excluded study (Teyhen et al., 2007) was limited to describing the arthrokinematic characteristics of a sub-group that were positive on a previously derived CPR.

4.1. Summary of evidence

Based upon the findings of the present review, the available evidence does not support the direct clinical application of any of the identified CPRs for LBP at this time. Of the 25 unique CPRs
Table 1
Derivation study quality appraisal.

<table>
<thead>
<tr>
<th>Diagnostic</th>
<th>Fritz et al., 2005b</th>
<th>Henschke et al., 2006a</th>
<th>Laslett et al., 2005</th>
<th>Laslett et al., 2006b</th>
<th>Laslett et al., 2003</th>
<th>Laslett et al., van der Wurff et al., 2006</th>
<th>Alonso-Blanco et al., 2009</th>
<th>Cai et al., 2009</th>
<th>Flynn et al., 2002</th>
<th>Fritz et al., 2005a</th>
<th>Hicks et al., 2005</th>
<th>May et al., 2008</th>
<th>George et al., Hancock et al., 2009b</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prospective design</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Outcome clinical important</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Blinded outcome assessment</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>All important predictors included</td>
<td>No</td>
<td>✔</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>✔</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Predictive variables clearly defined</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Blinded predictor assessment</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Assessment of the reliability of the predictive variables</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Important patient characteristics described</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Representative sample</td>
<td>No</td>
<td>✔</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>✔</td>
<td>No</td>
<td>✔</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Study site described</td>
<td>No</td>
<td>✔</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>✔</td>
<td>No</td>
<td>✔</td>
<td>No</td>
<td>✔</td>
</tr>
<tr>
<td>Justification for the number of study subjects</td>
<td>No</td>
<td>✔</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>≥10 outcome events per independent variable in the rule</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Mathematical techniques described</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Multivariable analysis</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Results of the rule described</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Clinically sensible/reasonable</td>
<td>No</td>
<td>✔</td>
<td>No</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Easy to use</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Probability of diagnosis or outcome described</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Course of action described</td>
<td>No</td>
<td>✔</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>✔</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Specifically uses the term “Clinical prediction rule”</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

a Secondary analysis of prospectively derived data.
b Single diagnostic injection not consistent with current SIJ diagnostic criterion standard (Szadek et al., 2009).
c Multivariable regression not performed for all prediction rules.
d More than one rule presented. Not all rules satisfy criterion.
identified, only two have progressed to the process of validation and no rule has been investigated for its ability to positively influence clinical outcomes and/or resource consumption.

The 5-item spinal manipulation CPR derived by Flynn et al. (2002) in a single-arm study design is one of the CPRs that has been further investigated in a series of validation studies. Recent commentary in the literature (Allison, 2009; Hancock et al., 2009a; Hancock et al., 2009b) have discussed the limitations of single-arm study designs in the development of prescriptive CPRs. The lack of a control group enables the identification of non-specific prognostic variables but is unable to investigate the moderators of treatment-effect. Controlled-study designs utilizing tests of interactions are required to identify on whom and under what circumstances treatments produce different outcomes (Kraemer et al., 2002; Hancock et al., 2009a). Accordingly, it has been suggested that the subsequent study undertaken by Childs et al. (2004) is most appropriately considered a derivation study and not a validation study. This is because it was the first controlled-study that enabled the investigation of the CPR as a treatment response modifier, in contrast to a non-specific prognostic factor (Hancock et al., 2009a).

Of the remaining validation studies that have aimed to develop the 5-item spinal manipulation CPR in new cohorts of patient populations, only two (Hancock et al., 2008b; Cleland et al., 2009) have used a controlled-study design. Cleland et al. (2009) aimed to examine the generalizability of the CPR to different thrust and non-thrust manipulative techniques. The generalizability of a CPR to other procedures is most appropriately determined by controlled-study designs that investigate if a patient’s status on the rule significantly moderates the effect-size of an intervention (Assmann et al., 2000; Kraemer et al., 2002, 2006; Turner et al., 2007; MacKinnon and Luecken, 2008). However, as the patient population in this study were all positive on the spinal manipulation CPR, the performance of the rule in identifying those with a difference in treatment responsiveness remained untested. Finally, in the well-designed validation study by Hancock et al. (2008b), the spinal manipulation CPR was found to perform no better than chance in identifying patients likely to respond to this intervention. Positive status on the rule, however, was found to be a non-specific prognostic factor. One of the many possible explanations for the observed findings noted by these researchers (Hancock et al., 2008a, 2008b) and others (Hebert and Perle, 2008) is the difference in treatment provided in this study compared to the original derivation studies (Flynn et al., 2002; Childs et al., 2004), with high-velocity thrust manipulative techniques only being used on a very small proportion of the patients in this study.

The 2-item pragmatic spinal manipulation CPR derived by Fritz et al. (2005a) was based upon the collated results of two previous studies (Flynn et al., 2002; Childs et al., 2004) used to develop the 5-item rule. This abbreviated form of the spinal manipulation CPR was found to strongly identify those patients with a good outcome following treatment. However, as no control group was included in the derivation, the variables may represent prognostic factors that may have no specific relationship with the intervention provided. Two subsequent studies (Fritz et al., 2006; Hallegraeff et al., 2009) attempting to validate this rule restricted their patient populations to only those that were positive on the pragmatic spinal manipulation CPR. As previously noted, without the inclusion of patients that are also negative on the rule, a prescriptive CPR’s performance is unable to be rigorously investigated. Consequently, the body of evidence does not yet enable confidence in the direct clinical application of either the 5-item or 2-item spinal manipulation CPRs in identifying subgroups of patients with differences in responsiveness to this intervention.

The 23 rules that have been derived, but not yet proceeded to validation may inform clinical practice by providing clinicians with an understanding of some of the most important predictors of a given target condition or outcome (McGinn et al., 2008). However, even in this limited application clinicians must exercise due caution as predictor variables may simply reflect chance associations or unique characteristics of the studied population or setting. Further, prescriptive predictor variables identified through single-arm study designs may not identify the relevant features that modify the effect of a given intervention, but instead reflect non-specific prognostic factors (Hancock et al., 2009a).

It has been argued that the biologic plausibility of predictor variables be carefully considered throughout the derivation of a CPR to minimise the likelihood of including factors that reflect chance associations with the target outcome (Childs and Cleland, 2006; Fritz et al., 2009b; Raney et al., 2009). However, the primary function of a CPR is to accurately predict a target outcome and not to identify the determinants of that outcome. The composite of factors that together accurately predict a given outcome are of most value, regardless of whether this relationship is confounded by other variables (Katz, 2006). To illustrate this point, consider that although carrying a cigarette lighter will not cause lung cancer, it may accurately predict a greater likelihood of developing the disease (Katz, 2006). Excluding predictive variables that are not believed at the time to be causally related to the target outcome may result in the development of CPRs with inferior predictive accuracy. Consequently, the process of rigorous validation of derived CPRs is the most suitable method to identify and...
exclude those variables that previously reflected chance associations with the target outcome (McGinn et al., 2008).

4.2. Methodological quality

Substantial variation was observed in the methodological quality of the fifteen included derivation studies. In addition to the previously mentioned research-design limitations of many prescriptive CPR studies, other common methodological shortcomings included the omission of important predictor variables, not providing a justification for the sample size and not including an appropriate number of outcome events per independent predictor when performing multivariable regression analysis.

Including the most probable predictor variables in the investigation aims to ensure that important relevant factors are not omitted (Laupacis et al., 1997). However, this needs to be balanced with restricting the analysis to a pre-determined small number of variables, ideally for only one outcome, to reduce the likelihood of eliciting findings that are due to chance and random error (Assmann et al., 2000). Researchers should consider examining the results of secondary-analyses of randomised controlled trials and the findings of single-arm treatment studies to help guide the selection of variables (Fritz et al., 2005).

Only one of the included derivation studies explicitly justified the size of the studied population. Larger sample sizes enable more precise estimates of a rule’s predictive power, which in turn enhances confidence in its clinical application (Childs and Cleland, 2006; McGinn et al., 2008). A further consideration is that the investigation of treatment-effect modifiers in prescriptive CPRs requires much larger sample sizes in comparison to identifying main effects between treatment groups. Simulation studies have demonstrated that a study with an 80% power of detecting a given overall effect would require four times the number of subjects to maintain this power in detecting an interaction effect of the same magnitude (Brookes et al., 2004).

Researchers developing CPRs need to carefully consider the prevalence of the target outcome or condition when determining the sample size to ensure that there is a sufficient number of outcome events to satisfy the assumptions implicit to the statistical analysis. Seventy percent of the included derivation studies that used multivariable regression analysis did not have an adequate number of outcome events per independent variable in the model. Guidelines for the development of multivariable logistic regression and Cox proportional hazard models advocate a minimum of ten outcome events per independent variable to reduce the likelihood of identifying erroneous associations and to improve the precision of the findings (Concato et al., 1993). For multiple linear regression, it is recommended that there should be at least ten patients for every variable selected (Lewis, 2007).

Similar to the variance observed in the derivation studies, the methodological quality for the eight included validation studies varied substantially. No validation study included in this review investigated the inter-observer reliability of the CPR. Guidelines on the validation of CPRs have recommended that researchers examine the inter-observer reliability of the rule, at least within a subset of the study population, to ensure consistency in the interpretation of a patient’s status on the rule (Laupacis et al., 1997; Stiell and Wells, 1999).

4.3. Study limitations

The search strategy employed in this review has been demonstrated to have high sensitivity for the detection of CPR studies (Ingui and Rogers, 2001) and has been used in other systematic reviews (Dahri and Loewen, 2007; Beneciuk et al., 2009; May and Rosedale, 2009). However, due to inconsistent nomenclature used to describe these clinical tools, it is plausible that not all potentially eligible studies were identified.

The primary aim of this review was the identification and appraisal of CPRs in the physiotherapy management of LBP. Due to substantial between-discipline practice differences in the assessment of LBP (Kent et al., 2009), it was determined a priori that for a study to be included, the assessment of potential predictor variables was required to be performed by a physiotherapist. This eligibility criterion resulted in the exclusion of studies that had developed CPRs using other LBP treatment providers for the assessment of predictor variables. While outside the scope of the present review, the value and validity of such CPRs for physiotherapy practice arguably merits investigation.

The sensitive operational definition of a CPR used in this review enabled the inclusion of studies that may not have explicitly used the term “clinical prediction rule”. Consequently, the methodological standards that would be considered by researchers explicitly aiming to develop a CPR may not have been considered in the design of these other studies. As the quality appraisal tool used in this review reflects these well-cited standards for CPR development, it is perhaps not surprising that a large variation of quality was observed between those studies that did and did not explicitly use the term “clinical prediction rule”.

The methodological appraisal tool used in this review was developed via a systematic process that aimed to minimise bias in the selection of appropriate quality criteria. While we believe this approach represents an improvement upon that used in previous systematic reviews of CPRs, our checklist has not been formally validated, and consequently the results need to be interpreted with caution. The degree of between-rater agreement was high for the majority of the quality criteria, however, it is clear that some variables particularly those relating to the appraisal of validation studies would benefit from measures to further improve rater concordance. An important consideration is that the quality criteria used in this review reflects the well-cited methodological standards that are common to diagnostic, prescriptive and prognostic forms of CPRs. Although this approach appropriately reflects the primary aim of this review and enables a qualitative comparison of the included studies, it is acknowledged that the omission of appraisal criteria that are specific to the development of each particular form of CPR may represent a potential limitation of the present study. Recently, a quality checklist for prescriptive derivation-based CPRs (the QUADCP) has been developed using Delphi methods (Cook et al., 2010). While this checklist will require further investigation of its reliability and validity, and is not advocated for the retrospective appraisal of CPR studies, it constitutes an important contribution in providing clear methodological guidelines for developing future studies aiming to derive prescriptive rules.

5. Conclusions

This review is the first to systematically locate, appraise and determine the clinical readiness of diagnostic, prescriptive and prognostic CPRs involving the physiotherapy management of LBP in all phases of their development. Twenty-five unique rules were identified across fifteen derivation and eight validation studies. No impact studies were located. The current body of evidence does not enable confident direct clinical application of any of the identified CPRs. Further validation studies utilizing appropriate research designs and rigorous methodology are required to determine the performance and generalizability of the derived CPRs to other patient populations, clinicians and clinical settings.
### Table 1: Database search strategies

<table>
<thead>
<tr>
<th>Search Strategy</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Appendix 1</strong></td>
<td>Database search strategies</td>
</tr>
<tr>
<td><strong>Appendix 2</strong></td>
<td>CPR Variables Publication Stage of rule</td>
</tr>
</tbody>
</table>

### Acknowledgements

We thank Dane P. Fehlberg for his assistance with screening and MEDLINE, EMBASE and the Cochrane Database of Systematic Reviews.
### SJJ mediated pain CPR3
- **Distraction, Thigh Thrust, Gaenslen’s test, Compression, Sacral Thrust** (5 variables)
- **Laslett et al., 2003** Derivation
- \( n = 43 \) (subset of patients from Laslett et al., 2005 using different reference standard), buttock pain + leg pain, referred to specialist diagnostic centre with suspicion of SJJ pain, insufficient data to report precise demographic details, 26% prevalence of target condition.

#### Derivation

If 3 or more positives, then +LR = 4.16 (95% CI 2.16–8.39) and −LR = 0.12 (95% CI 0.02–0.49).

### SJJ mediated pain CPR4
- **No CP/peripheralisation, Distraction, Thigh Thrust, Gaenslen’s test, Compression, Sacral Thrust** (6 variables)
- **Laslett et al., 2003** Derivation
- \( n = 34 \) (subset of patients from Laslett et al., 2005 using different reference standard), buttock pain + leg pain, referred to specialist diagnostic centre with suspicion of SJJ pain, insufficient data to report precise demographic details, 32% prevalence of target condition.

#### Derivation

If no CP/peripheralisation and if 3 or more positives of remaining variables, then +LR = 6.97 (95% CI 2.7–20.27) and −LR = 0.11 (95% CI 0.02–0.44).

### SJJ mediated pain CPR5
- **Distraction, Compression, Thigh Thrust, Patrick sign, Gaenslen’s test** (5 variables)
- **van der Wurff et al., 2006** Derivation
- \( n = 60 \), buttock pain + leg pain, referred to specialist diagnostic centre with suspicion of SJJ pain, insufficient data to report precise demographic details, 26% prevalence of target condition.

#### Derivation

If 4 or more positives, then +LR = 7.6 (95% CI 4.5–13.7) and −LR = 0.0 (95% CI 0.0–0.35).

### Z-jt mediated pain CPR1
- **Age ≥ 50, symptoms best walking, symptoms best sitting, onset pain is paraspinal, MSPQ > 13, ext/rot test**, no CP (7 variables)
- **Laslett et al., 2006b** Derivation
- \( n = 120 \), LBP + leg pain, referred to specialist diagnostic centre with suspicion of z-jt pain, mean 43 years old, 46% female, mean 158 weeks of symptoms, 11% prevalence of target condition.

#### Derivation

If 2 or more positives, then +LR = 1.6 (95% CI 1.5–1.8) and −LR = 0.0 (95% CI 0.0–0.69).

### Z-jt mediated pain CPR2
- **Age ≥ 50, symptoms best walking, symptoms best sitting, onset pain is paraspinal, MSPQ > 13, ext/rot test** (6 variables)
- **Laslett et al., 2006b** Derivation
- \( n = 120 \), LBP + leg pain, referred to specialist diagnostic centre with suspicion of z-jt pain, mean 43 years old, 46% female, mean 158 weeks of symptoms, 11% prevalence of target condition.

#### Derivation

If 1 or more positives, then +LR = 1.4 (95% CI 1.3–1.5) and −LR = 0.0 (95% CI 0.0–0.95).

### Z-jt mediated pain CPR3
- **Age ≥ 50, symptoms best walking, symptoms best sitting, onset pain is paraspinal, MSPQ > 13** (5 variables)
- **Laslett et al., 2006b** Derivation
- \( n = 120 \), LBP + leg pain, referred to specialist diagnostic centre with suspicion of z-jt pain, mean 43 years old, 46% female, mean 158 weeks of symptoms, 11% prevalence of target condition.

#### Derivation

If 2 or more positives, then +LR = 2.0 (95% CI 1.8–2.5) and −LR = 0.0 (95% CI 0.0–0.49).

### Z-jt mediated pain CPR4
- **Age ≥ 50, symptoms best walking, symptoms best sitting, onset pain is paraspinal, ext/rot test** (5 variables)
- **Laslett et al., 2006b** Derivation
- \( n = 120 \), LBP + leg pain, referred to specialist diagnostic centre with suspicion of z-jt pain, mean 43 years old, 46% female, mean 158 weeks of symptoms, 11% prevalence of target condition.

#### Derivation

If 3 or more positives, then +LR = 9.7 (95% CI 5.0–18.8) and −LR = 0.17 (95% CI 0.05–0.6).

### Z-jt mediated pain CPR5
- **Age ≥ 50, symptoms best walking, symptoms best sitting, onset pain is paraspinal, ext/rot test** (5 variables)
- **Laslett et al., 2006b** Derivation
- \( n = 120 \), LBP + leg pain, referred to specialist diagnostic centre with suspicion of z-jt pain, mean 43 years old, 46% female, mean 158 weeks of symptoms, 11% prevalence of target condition.

#### Derivation

If 2 or more positives, then +LR = 15.5 (95% CI 7.2–24.6).

### Vertebral fracture
- **Female sex, age > 70, significant trauma, prolonged use of corticosteroids** (4 variables)
- **Henschke et al., 2009** Derivation
- \( n = 1172 \), acute LBP + leg pain patients presenting to a primary care provider, mean 44 years old, 47% female, 59% had duration of less than one week, 0.7% prevalence of target condition.

#### Derivation

If 3 or more positives, then +LR = 218.3 (95% CI 45.6–953.8).

---

**Notes:**
- +LR = positive likelihood ratio.
- −LR = negative likelihood ratio.
- CP = centralization phenomenon.
- PPE = persistent low back pain between episodes of acute low back pain.
- VABLE = subjective report of ‘vulnerability’ when in the semi-stooped position or when performing twisting actions.
- Ext Loss = visual estimation of moderate or major loss of lumbar extension range of movement.
- Mod = Modified somatic perception questionnaire.
- Ext/Rot test = Extension/Rotation test.
- Predictor variables not exclusively assessed by physiotherapists. Physiotherapists = 72.6%, general practitioners = 22.8%, chiropractors = 4.6%.
## Appendix 2b. Prescriptive clinical prediction rules included in qualitative synthesis.

<table>
<thead>
<tr>
<th>CPR</th>
<th>Variables</th>
<th>Publication</th>
<th>Stage of rule development</th>
<th>Sample</th>
<th>Results/outcome</th>
<th>Methodological notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spinal manipulation</td>
<td>Duration of symptoms &lt; 16 days, FABQ-W &lt; 19, at least 1 hip with &gt;35° IR ROM, hypomobility with lumbar spring testing, no symptoms distal to knee (5 variables)</td>
<td>Flynn et al., 2002</td>
<td>Derivation</td>
<td><em>n</em> = 71, LBP +/− leg pain, baseline ODQ score ≥ 30%, referred to physiotherapy, mean 37.6 years old, 41% female, mean 42 days of symptoms, 45% prevalence of target outcome. If 4 or more positives, then +LRd = 24.38 (95% CI 4.63−139.41)</td>
<td>Only the patient with all 5 criteria positive experienced dramatic improvement in pain and disability following manipulation. Research design prevents identification of treatment-effect modifiers.</td>
<td>Single-arm design. Therefore unable to identify treatment-effect modifiers.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Childs et al., 2003</td>
<td>Validation</td>
<td><em>n</em> = 2 (case reports), 54 and 26 year old males, LBP and buttock pain respectively. One patient met 5 CPR criteria, the other patient met just 1 (or 2) criteria.</td>
<td></td>
<td>RCT. Therefore treatment-effect modifiers able to be identified.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Childs et al., 2004</td>
<td>Validation</td>
<td><em>n</em> = 131 (RCT), LBP +/− leg pain, baseline ODQ score ≥ 30%, referred to physiotherapy, mean 33.9 years old, 42% female, median 27 days of symptoms, 20% prevalence of target outcome at 1/52 and 50% at 4/52.</td>
<td>Significant 3 way-interaction between CPR status (≥4/5 − positive), Rx-group and time for pain and disability. For dichotomized outcome (success/failure) the interaction between CPR status and Rx-group strongly predicted success.</td>
<td>RCT. Therefore treatment-effect modifiers able to be identified.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Childs et al., 2006</td>
<td>Validation</td>
<td><em>n</em> = 131 (RCT), LBP +/− leg pain, baseline ODQ score ≥ 30%, referred to physiotherapy, mean 33.9 years old, 42% female, median 27 days of symptoms.</td>
<td>Aimsed to investigate if CPR status is predictive of a worsening in disability. No patient that was CPR positive and received manipulation worsened, preventing appropriate statistical analysis.</td>
<td>Secondary analysis of 2004 RCT. Therefore treatment-effect modifiers able to be identified.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cleland et al., 2006</td>
<td>Validation</td>
<td><em>n</em> = 12 (case series), LBP, ODQ score ≥ 30%, referred to physiotherapy, all CPR positive (≥4/5 − positive), mean 39 years old, 42% female, median 19 days of symptoms.</td>
<td>Aimsed to investigate generalizability of CPR status to another high-velocity thrust manipulation procedure. 11 out of 12 patients (92%) achieved the target outcome of 'success' at 1/52 following manipulation. All patients CPR positive, therefore unable to determine rule performance. Research design prevents identification of treatment-effect modifiers.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hancock et al., 2008b</td>
<td>Validation</td>
<td><em>n</em> = 239 (RCT), LBP &lt; 6/52 duration, presenting to general practitioner, mean 40.7 years old, 44% female, mean 9 days of symptoms.</td>
<td>Non-significant 3-way interaction between Rx-group, CPR status (≥4/5 − positive) and time for pain (p = 0.805) and disability (p = 0.6). Patients that were CPR positive had better pain and disability outcomes independent of treatment group.</td>
<td>RCT. Therefore treatment-effect modifiers able to be identified. Spinal manipulative technique differed to derivation study. Only 5% of sample received high-velocity thrust manipulation.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cleland et al., 2009</td>
<td>Validation</td>
<td><em>n</em> = 112 (RCT), LBP +/− leg pain, attending an outpatient physiotherapy clinic, modified ODQ baseline score &lt;25%, all CPR positive (≥4/5 − positive), mean 40.3 years old, 52% female, median 45 days of symptoms.</td>
<td>Aimsed to investigate the generalizability of CPR to another high-velocity thrust manipulation procedure and a non-thrust manipulative technique. No difference between the 2 high-velocity thrust procedures in pain and disability at any time point. Outcomes poorer in the non-thrust group. All patients CPR positive, therefore unable to determine rule performance. RCT. Therefore treatment-effect modifiers able to be identified.</td>
<td></td>
</tr>
<tr>
<td>Diagnosis/Decision</td>
<td>Predictor Variables</td>
<td>Study</td>
<td>Study Details</td>
<td>Result</td>
<td>Notes</td>
<td></td>
</tr>
<tr>
<td>--------------------</td>
<td>---------------------</td>
<td>-------</td>
<td>---------------</td>
<td>--------</td>
<td>-------</td>
<td></td>
</tr>
<tr>
<td>Spinal manipulation – pragmatic rule</td>
<td>Duration of symptoms &lt; 16 days, no symptoms distal to knee (2 variables)</td>
<td>Fritz et al., 2005a</td>
<td>Derivation</td>
<td>n = 141 (data from 2 previous studies (Flynn et al., 2002; Childs et al., 2004)), LBP +/− leg pain, baseline ODQ score ≥ 30%, referred to physiotherapy, mean 35.5 years old, 49% female, median 22 days of symptoms, 45% prevalence of target outcome.</td>
<td>If both criteria positive, then $\text{LR} + = 7.2$. Single-arm design. Therefore unable to identify treatment-effect modifiers.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fritz et al., 2006</td>
<td>Validation</td>
<td>n = 215 (retrospective review of clinical database), occupational LBP, receiving Rx in outpatient physiotherapy clinic, all CPR positive (2/2 − positive), mean 35.9 years old, 32% female, mean 5.3 days of symptoms.</td>
<td>66.5% received manipulation (49.8% thrust and 16.7% non-thrust). Patients receiving manipulation experienced greater reductions in pain and disability with treatment, compared to those not receiving manipulation. Research design prevents the identification of treatment-effect modifiers. All patients CPR positive, therefore unable to determine rule performance.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hallegaard et al., 2009</td>
<td>Validation</td>
<td>n = 64 (RCT), acute LBP, all CPR positive (2/2 − positive), mean 39 years old, 45% female, 66.5% received manipulation (49.8% thrust and 16.7% non-thrust). Patients receiving manipulation experienced greater reductions in pain and disability with treatment, compared to those not receiving manipulation.</td>
<td>Significant interaction for disability at 2.5 weeks between CPR status (including the additional criterion of age &gt; 35 years) and Rx-group. No significant interactions for pain or lumbar spinal mobility. All patients CPR positive (by derivation study criteria), therefore unable to determine rule performance. RCT. Therefore treatment-effect modifiers able to be identified. Analysis performed with the additional CPR criterion of age &gt; 35 years. Single-arm design. Therefore unable to identify treatment-effect modifiers.</td>
<td></td>
</tr>
<tr>
<td>Lumbar traction</td>
<td>FABQ-W &lt; 21, no neurological deficit, age &gt; 30, non-manual work job status (4 variables)</td>
<td>Cai et al., 2009</td>
<td>Derivation</td>
<td>n = 129, diagnosis related to the lumbar spine +/− leg pain, referred from orthopaedics to physiotherapy, mean 30.9 years old, 16% female, mean 40 weeks of symptoms, 19% prevalence of target outcome.</td>
<td>If 3 or more positives, then $\text{LR} + = 3.04$ (95% CI 2.04–4.53). If all 4 positive, then $\text{LR} + = 9.36$ (95% CI 3.2–16.1). Single-arm design. Therefore unable to identify treatment-effect modifiers.</td>
<td></td>
</tr>
<tr>
<td>Stabilisation exercise – success</td>
<td>Age &lt; 40 years, average SLR$^c &gt; 91^c$, aberrant movement present, positive prone instability test (4 variables)</td>
<td>Hicks et al., 2005</td>
<td>Derivation</td>
<td>n = 54, LBP +/− leg pain, referred to outpatient physiotherapy clinics, mean 42.4 years old, 57% female, mean 41 days of symptoms, 13% prevalence of target outcome (success).</td>
<td>If 3 or more positives, then $\text{LR} + = 4.0$ (95% CI 1.6–10.0). If 2 or more positives, then $\text{LR} + = 1.9$ (95% CI 1.2–2.9). Single-arm design. Therefore unable to identify treatment-effect modifiers.</td>
<td></td>
</tr>
<tr>
<td>Stabilisation exercise – failure</td>
<td>Prone instability test, aberrant movement, hypermobility, FABQ physical activity subscale &gt; 8 (4 variables)</td>
<td>Hicks et al., 2005</td>
<td>Derivation</td>
<td>n = 54, LBP +/− leg pain, referred to outpatient physiotherapy clinics, mean 42.4 years old, 57% female, mean 41 days of symptoms, 33% prevalence of target outcome (success).</td>
<td>In the absence of 2 or more positives (ie. 1 or 0 positives), then $\text{LR} -$ = 0.18 (95% CI 0.08–0.38). Single-arm design. Therefore unable to identify treatment-effect modifiers.</td>
<td></td>
</tr>
<tr>
<td>McKenzie approach (MDT$^a$)</td>
<td>&lt;12/52 duration, centralization or abolition of symptoms with MDT loading strategies (2 variables)</td>
<td>May et al., 2008</td>
<td>Derivation</td>
<td>n = 102 (secondary analysis of single-arm of RCT), back and neck pain patients referred by GP’s to Physiotherapy, study sample demographics not provided.</td>
<td>For those patients with back pain, the presence of both predictor variables gave a probability of success (‘liberal’ definition provided in study) of 68.9%. The absence of both variables gave a probability of success of 10%. Single-arm design. Therefore unable to identify treatment-effect modifiers.</td>
<td></td>
</tr>
<tr>
<td>Specific exercise program for Ankylosing Spondylitis</td>
<td>SF-36 Physical Role &gt; 37, SF-36 Bodily Pain &gt; 27, BASDAI$^b$ &gt; 31 (3 variables)</td>
<td>Alonso-Blanco et al., 2009</td>
<td>Derivation</td>
<td>n = 35, patients with AS referred to physiotherapy clinic, mean 45.7 years old, 20% female, mean 9.7 years of symptoms, 46% prevalence of target outcome.</td>
<td>For those patients with back pain, the presence of both predictor variables gave a probability of success (‘liberal’ definition provided in study) of 68.9%. The absence of both variables gave a probability of success of 10%. Single-arm design. Therefore unable to identify treatment-effect modifiers.</td>
<td></td>
</tr>
</tbody>
</table>

$^a$ FABQ-W – Fear avoidance beliefs questionnaire work subscale.  
$^b$ IR ROM – internal rotation range of movement.  
$^c$ ODQ – Oswestry disability questionnaire.  
$^d$ $+\text{LR}$ – positive likelihood ratio.  
$^e$ $-\text{LR}$ – negative likelihood ratio.  
$^f$ $\text{LR}$ – negative likelihood ratio.  
$^g$ MDT – mechanical diagnosis and therapy.  
$^h$ BASDAI – Bath ankylosing spondylitis disease activity index.
Appendix 2c. CPR Variables Publication Stage of rule

Baseline pain intensity (0–10 NRS)
6 month pain intensity (0–10 NRS)

• Analysed limited to only those patients that were unavailable to do specific exercise.
• Analysed limited to only those patients that were unavailable to do specific exercise.

All arms of study included in analysis.

6 month disability (ODQ)
Baseline disability (ODQ)

• Analysed limited to only those patients that were unavailable to do specific exercise.
• Analysed limited to only those patients that were unavailable to do specific exercise.

Duration, aged 18–55 years.

Outcome for acute/subacute LBP (3 variables)

Fear Avoidance Beliefs Questionnaire Work Subscale.


References


Haskins R. Haskins et al. / Manual Therapy 17 (2012) 9–21


