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

ARTICLE INFO

ABSTRACT

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. Citation tracking and hand-searching of relevant journals were used as supplemental strategies.
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 (Riihimaki, 1996; Andersson, 1998; Walker, 1999) that has been 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 the 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 appropriate, 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 the identification of prescriptive rules that facilitate treatment decision-making by identifying variables that moderate the magnitude of the treatment-effect. These reviews have specifically excluded studies concerning diagnosis and prognosis, thereby preventing a complete assessment of the available CPRs a physiotherapist may consider in their clinical management of LBP. Further, the quality appraisal systems used in these reviews have not been reflective of the consensus of the common methodological standards for CPR development.

As no universally-accepted standardised tool currently exists for the methodological appraisal of studies of CPRs (Fritz, 2009), previous systematic reviews have used a variety of means to evaluate the quality of included studies. Some reviews have utilised standardised tools that were developed to appraise prognostic (Beneciuk et al., 2009) and diagnostic studies (Bachmann et al., 2004; Hess et al., 2008). Criticism in this approach has 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., 2009) 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 X2.0.1[1]) 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; Stiell 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; Stiell 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 Luecken, 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[2].

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

---

[1] EndNote version X2.0.1, Thomson Reuters, 2141 Palomar Airport Road, Suite 350, Carlsbad, CA 92011, USA.

[2] Stata version 11.0, StataCorp LP, 4905 Lakeway Drive, College Station, Texas 77845, USA.
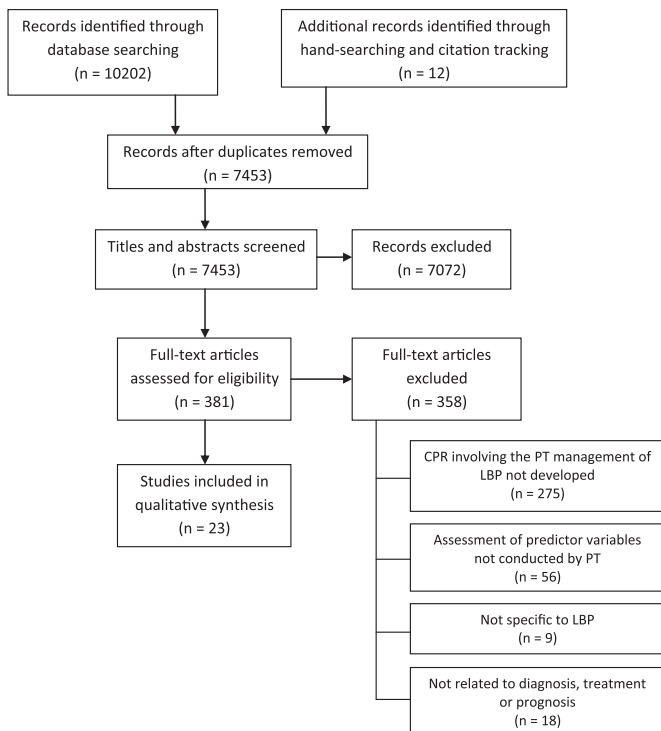
Fig. 1. Flow chart of search strategy and study selection.

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 sub-groups 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.

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

[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.

Table 2
Validation study quality appraisal.

	Prescriptive							
	Childs et al., 2003	Childs et al., 2004	Childs et al., 2006	Cleland et al., 2006	Cleland et al., 2009	Fritz et al., 2006	Hancock et al., 2008b	Hallegraeff et al., 2009
Prospective validation in new patient population	No[a]	✔	✔	✔	✔	No	✔	✔
Different clinical setting to derivation study	No[a]	No	No	✔	✔	✔	✔	✔
Different clinicians to derivation study	No[a]	No[a]	No[a]	✔	✔	✔	✔	✔
Representative sample	No	✔	✔	No	✔	No	✔	No
The rule is applied accurately	No	✔	✔	✔	✔	✔	✔	No
Complete follow-up	✔	✔	✔	✔	✔	✔	✔	✔
Accuracy of the rule in the validation study sample described	No	✔	No[b]	No	No	No	✔	No
Assessment of the inter-observer reliability of the rule	No	No	No	No	No	No	No[a]	No
Specifically uses the term "Clinical Prediction Rule"	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes

[a] Unclear. Insufficient information.
[b] Absence of target outcome in sub-group preventing appropriate statistical analysis.

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; Cook et al., 2010) and in two Physical Therapy podcasts (Fritz et al., 2009a, 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 valida-tion 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., 2009a).

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 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 QUADCPR) 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.

Acknowledgements

Appendix 1. Database search strategies

MEDLINE, EMBASE and the Cochrane Database of Systematic Reviews.

1	Validat$.mp. or Predict$.ti. or Rule$.mp.
2	(Predict$ and (Outcome$ or risk$ or model$)).mp.
3	((History or Variable$ or Criteria or Scor$ or Characteristic$ or Finding$ or Factor$) and (Predict$ or Model$ or Decision$.mp. or Identifi$ or Prognos$)).mp.
4	Decision$.mp. and ((Model$ or Clinical$).mp. or Logistic Models/)
5	(Prognostic and (History or Variable$ or Criteria or Scor$ or Characteristic$ or Finding$ or Factor$ or Model$)).mp.
6	4 or 1 or 3 or 2 or 5
7	dorsalgia.ti.ab.
8	exp Back Pain/
9	backache.ti.ab.
10	exp Low Back Pain/
11	(lumbar adj pain).ti.ab.
12	coccyx.ti.ab.
13	coccydynia.ti.ab.
14	sciatica.ti.ab.
15	sciatica/
16	spondylosis.ti.ab.
17	lumbago.ti.ab.
18	11 or 7 or 9 or 17 or 12 or 15 or 14 or 8 or 16 or 10 or 13
19	6 and 18
20	("1990" or "1991" or "1992" or "1993" or "1994" or "1995" or "1996" or "1997" or "1998" or "1999" or "2000" or "2001" or "2002" or "2003" or "2004" or "2005" or "2006" or "2007" or "2008" or "2009").yr.
21	19 and 20

CINAHL, AMED.

S22	S13 and S21
S21	S12 and S20
S20	S15 or S16 or S17 or S18 or S19
S19	prognostic and (history or variable* or criteria or scor* or characteristic* or finding* or factor* or model*)
S18	decision* and (model* or clinical* or mh Logistic Models)
S17	(history or variable* or criteria or scor* or characteristic* or finding* or factor*) and (predict* or model* or decision* or identif* or prognos*)
S16	predict* and (outcome* or risk* or model*)
S15	Validat* or ti Predict* or Rule*
S14	S12 and S13
S13	yr 1990 or yr 1991 or yr 1992 or yr 1993 or yr 1994 or yr 1995 or yr 1996 or yr 1997 or yr 1998 or yr 1999 or yr 2000 or yr 2001 or yr 2002 or yr 2003 or yr 2004 or yr 2005 or yr 2006 or yr 2007 or yr 2008 or yr 2009
S12	S1 or S2 or S3 or S4 or S5 or S6 or S7 or S8 or S9 or S10 or S11
S11	ti lumbago or ab lumbago
S10	ti spondylosis or ab spondylosis
S9	mh sciatica
S8	ti sciatica or ab sciatica
S7	ti coccydynia or ab coccydynia
S6	ti coccyx or ab coccyx
S5	ti (lumbar n0 pain) or ab (lumbar n0 pain)
S4	mh Low Back Pain+
S3	ti backache or ab backache
S2	mh Back Pain+
S1	ti dorsalgia or ab dorsalgia

Appendix 2a. Diagnostic clinical prediction rules included in qualitative synthesis.

CPR	Variables	Publication	Stage of rule development	Sample	Results/outcome
Radiographic instability	Lumbar Flexion > 53°, lack of hypomobility with intervertebral motion testing (2 variables)	Fritz et al., 2005b	Derivation	n = 49, LBP +/− leg pain, referred for imaging on suspicion of instability, mean 39.2 years old, 57% female, median 78 days of symptoms, 57% prevalence of target condition.	If 2 variables positive, +LR[a] = 12.8 (95% CI 0.79–211.6). If 1 variable positive, +LR = 4.3 (95% CI 1.8–10.6).
Diskogenic pain CPR1	CP[b], PPE[c], VABLE[d], Ext Loss[e] (4 variables)	Laslett et al., 2006a	Derivation	n = 216, LBP +/− leg pain, referred to specialist diagnostic centre, mean 44.2 years old, 43% female, mean 158 weeks of symptoms, 35% prevalence of target condition. Only 107 patients received reference standard.	If 1 or more variables positive, then +LR = 1.9 (95% CI 1.1–3.2) and −LR[f] = 0.37 (95% CI 0.21–0.65). If 2 variables positive, then +LR = 6.7 (95% CI 0.95–50) and −LR = 0.73(0.61–0.97).
Diskogenic pain CPR2	No CP, PPE, VABLE, Ext Loss (4 variables)	Laslett et al., 2006a	Derivation	n = 216, LBP +/− leg pain, referred to specialist diagnostic centre, mean 44.2 years old, 43% female, mean 158 weeks of symptoms, 35% prevalence of target condition. Only 107 patients received reference standard.	If 2 variables positive, then sensitivity = 37% (95% CI 24–50) and specificity = 100% (95% CI 82–100). LR's not calculated due to 100% specificity.
Diskogenic pain CPR3	PPE, VABLE, Ext Loss (3 variables)	Laslett et al., 2006a	Derivation	n = 216, LBP +/− leg pain, referred to specialist diagnostic centre, mean 44.2 years old, 43% female, mean 158 weeks of symptoms, 35% prevalence of target condition. Only 107 patients received reference standard.	If 2 variables positive, then +LR = 6.5 (95% CI 0.9–46.3) and −LR = 0.77 (95% CI 0.66–0.9).
SIJ mediated pain CPR1	Distraction, Compression, Thigh thrust, Gaenslen's (right), Gaenslen's (left), Sacral Thrust (6 variables)	Laslett et al., 2005	Derivation	n = 48, buttock pain +/− LBP +/− leg pain, referred to specialist diagnostic centre with suspicion of SIJ pain, mean 42.1 years old, 67% female, mean 32 months of symptoms, 33% prevalence of target condition.	If 3 or more variables positive, then +LR = 4.29 (95% CI 2.34–8.58) and −LR = 0.8 (95% CI 0.14–0.37)
SIJ mediated pain CPR2	Distraction, Thigh Thrust, Compression, Sacral Thrust (4 variables)	Laslett et al., 2005	Derivation	n = 48, buttock pain +/− LBP +/− leg pain, referred to specialist diagnostic centre with suspicion of SIJ pain, mean 42.1 years old, 67% female, mean 32 months of symptoms, 33% prevalence of target condition.	If 2 positives, then +LR = 4 (95% CI 2.13–8.08) and −LR = 0.16 (95% CI 0.04–0.47)

SIJ 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 +/− LBP +/− leg pain, referred to specialist diagnostic centre with suspicion of SIJ pain, insufficient data to report precise demographic details, 26% prevalence of target condition.	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).
SIJ 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 +/− LBP +/− leg pain, referred to specialist diagnostic centre with suspicion of SIJ pain, insufficient data to report precise demographic details, 32% prevalence of target condition.	If no CP/periphalisation 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)
SIJ 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 for invasive procedures, mean 51 years old, 78% female, mean 98 months of symptoms, 45% prevalence of target condition.	If 3 or more positives, then +LR = 4.02 (95% CI 2.04−7.89) and −LR 0.19 (95% CI 0.07−0.47)
Z-jt mediated pain CPR1	Age ≥ 50, symptoms best walking, symptoms best sitting, onset pain is paraspinal, MSPQ[g] > 13, ext/rot test[h], 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.	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 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.	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 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.	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 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.	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 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.	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).
Vertebral fracture	Female sex, age > 70, significant trauma, prolonged use of corticosteroids (4 variables)[i]	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.	If 2 or more positives, then +LR = 15.5 (95% CI 7.2−24.6). If 3 or more positives, then +LR = 218.3(95% CI 45.6−953.8).

[a] +LR = positive likelihood ratio.
[b] CP = centralization phenomenon.
[c] PPE = persistent low back pain between episodes of acute low back pain.
[d] VABLE = subjective report of 'vulnerability' when in the semi-stooped position or when performing twisting actions.
[e] Ext Loss = visual estimation of moderate or major loss of lumbar extension range of movement.
[f] −LR − negative likelihood ratio.
[g] MSPQ = Modified somatic perception questionnaire.
[h] Ext/Rot test = Extension/Rotation test.
[i] 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.

CPR	Variables	Publication	Stage of rule development	Sample	Results/outcome	Methodological notes
Spinal manipulation	Duration of symptoms < 16 days, FABQ-W[a] < 19, at least 1 hip with >35° IR ROM[b], hypomobility with lumbar spring testing, no symptoms distal to knee (5 variables)	Flynn et al., 2002	Derivation	$n = 71$, LBP +/− leg pain, baseline ODQ[c] 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)	Single-arm design. Therefore unable to identify treatment-effect modifiers.
		Childs et al., 2003	Validation	$n = 2$ (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.	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.
		Childs et al., 2004	Validation	$n = 131$ (RCT), LBP +/− leg pain, baseline ODQ score ≥ 30%, referred to physiotherapy, mean 33.9 years old, 42% female, median 27 days of symptoms, 29% prevalence of target outcome at 1/52 and 50% at 4/52.	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. For patients receiving manipulation, CPR positive status had +LR = 13.2 (95% CI 3.4−52.1). For patients CPR positive the NNT with manipulation = 1.3 (95% CI 1.1−1.9)	RCT. Therefore treatment-effect modifiers able to be identified.
		Childs et al., 2006	Validation	$n = 131$ (RCT), LBP +/− leg pain, baseline ODQ score ≥ 30%, referred to physiotherapy, mean 33.9 years old, 42% female, median 27 days of symptoms.	Aimed 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.	Secondary analysis of 2004 RCT. Therefore treatment-effect modifiers able to be identified.
		Cleland et al., 2006	Validation	$n = 12$ (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.	Aimed 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 intervention.	All patients CPR positive, therefore unable to determine rule performance. Research design prevents identification of treatment-effect modifiers.
		Hancock et al., 2008b	Validation	$n = 239$ (RCT), LBP < 6/52 duration, presenting to general practitioner, mean 40.7 years old, 44% female, mean 9 days of symptoms.	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.	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.
		Cleland et al., 2009	Validation	$n = 112$ (RCT), LBP +/− leg pain, attending an outpatient physiotherapy clinic, modified ODQ baseline score >25%, all CPR positive (≥4/5 = positive), mean 40.3 years old, 52% female, median 45 days of symptoms.	Aimed 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.

Spinal manipulation – pragmatic rule	Duration of symptoms < 16 days, no symptoms distal to knee (2 variables)	Fritz et al., 2005a	Derivation	$n = 141$ (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.	If both criteria positive, then +LR = 7.2 (95% CI 3.2–16.1).	Single-arm design. Therefore unable to identify treatment-effect modifiers.
		Fritz et al., 2006	Validation	$n = 215$ (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.	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.
		Hallegraeff et al., 2009	Validation	$n = 64$ (RCT), acute LBP, all CPR positive (2/2 = positive), mean 39 years old, 45% female, 31% had symptoms less than 1/52.	Significant interaction for disability at 2.5 weeks between CPR status (including the additional criterion of age > 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 >35 years.
Lumbar traction	FABQ-W < 21, no neurological deficit, age > 30, non-manual work job status (4 variables)	Cai et al., 2009	Derivation	$n = 129$, diagnosis related to the lumbosacral 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.	If 3 or more positives, then +LR = 3.04 (95% CI 2.04–4.53). If all 4 positive, then +LR = 9.36 (95% CI 3.13–28.0).	Single-arm design. Therefore unable to identify treatment-effect modifiers.
Stabilisation exercise – success	Age < 40 years, average SLR[e] > 91°, aberrant movement present, positive prone instability test (4 variables)	Hicks et al., 2005	Derivation	$n = 54$, 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).	If 3 or more positives, then +LR = 4.0 (95% CI 1.6–10.0). If 2 or more positives, then +LR = 1.9 (95% CI 1.2–2.9).	Single-arm design. Therefore unable to identify treatment-effect modifiers.
Stabilisation exercise – failure	Prone instability test, aberrant movement, hypermobility, FABQ physical activity subscale > 8 (4 variables)	Hicks et al., 2005	Derivation	$n = 54$, LBP +/− leg pain, referred to outpatient physiotherapy clinics, mean 42.4 years old, 57% female, mean 41 days of symptoms, 72% prevalence of target outcome (not failure).	In the absence of 2 or more positives (ie. 1 or 0 positives), then −LR[f] = 0.18 (95% CI 0.08–0.38).	Single-arm design. Therefore unable to identify treatment-effect modifiers.
McKenzie approach (MDT[g])	<12/52 duration, centralization or abolition of symptoms with MDT loading strategies (2 variables)	May et al., 2008	Derivation	$n = 102$ (secondary analysis of single-arm of RCT), back and neck pain patients referred by GP's to Physiotherapy, study sample demographics not provided.	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.
Specific exercise program for Ankylosing Spondylitis	SF-36 Physical Role > 37, SF-36 Bodily Pain > 27, BASDAI[h] > 31 (3 variables)	Alonso-Blanco et al., 2009	Derivation	$n = 35$, 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.	If 2 or more positives, then +LR = 11.2 (95% CI 1.7–76.0). If 3 or more positives, then +LR = 2.6 (95% CI 1.6–4.0).	Single-arm design. Therefore unable to identify treatment-effect modifiers.

[a] FABQ-W = Fear avoidance beliefs questionnaire work subscale.
[b] IR ROM = internal rotation range of movement.
[c] ODQ = Oswestry disability questionnaire.
[d] +LR = positive likelihood ratio.
[e] SLR = straight leg raise.
[f] −LR = negative likelihood ratio.
[g] MDT = mechanical diagnosis and therapy.
[h] BASDAI = Bath ankylosing spondylitis disease activity index.

Appendix 2c. Prognostic clinical prediction rules included in qualitative synthesis.

CPR	Variables	Publication	Stage of rule development	Sample	Results/outcome	Notes
6 month pain outcome for acute/subacute LBP	Baseline pain intensity (0–10 NRS[a]), CP[b] (present = 1, absent = 0) (2 variables)	George et al., 2005	Derivation	$n = 28$ (secondary analysis of sub-group in earlier clinical trial), LBP < 60 days duration, aged 18–55 years, demographic details of this sub-group not reported.	6 month pain intensity (0–10 NRS) = 0.97 + 0.27(Pain 0–10 NRS) − 1.6 (CP).	Analysis limited to only those patients that were classified for 'specific exercise'.
6 month disability outcome for acute/subacute LBP	Baseline disability (ODQ[c]), FABQ-W[d], CP (present = 1, absent = 0) (3 variables)	George et al., 2005	Derivation	$n = 28$ (secondary analysis of sub-group in earlier clinical trial), LBP < 60 days duration, aged 18–55 years, demographic details of this sub-group not reported.	6 month disability (ODQ) = 4.4 + 0.24 (ODQ) + 0.34(FABQ-W) − 10 (CP).	Analysis limited to only those patients that were classified for 'specific exercise'.
Time to recovery from acute LBP	Baseline pain ≤ 7/10, duration of current episode ≤5 days, and ≤1 previous episodes (3 variables)	Hancock et al., 2009b	Derivation	$n = 239$ (RCT), LBP +/− leg pain <6/52, presenting to GPs, mean age 40.7 years, 44% female, mean 9 days of symptoms.	If 3 variables positive, then median days to recovery (from baseline assessment) = 6 (95% CI 4–8). If no variables are positive, then median days to recovery = 22 (95% CI 11–33).	All arms of study included in analysis.

[a] NRS = numerical rating scale.
[b] CP = centralisation phenomenon.
[c] ODQ = Oswesty Disability Questionnaire.
[d] FABQ-W = Fear Avoidance Beliefs Questionnaire Work Subscale.

References

Allison SC. On "A guide to interpretation of studies". Hancock M, et al. Physical Therapy. 2009;89:698–704. Physical Therapy 2009;89(10):1098–9.

Alonso-Blanco C, Fernandez-de-las-Penas C, Cleland JA. Preliminary clinical prediction rule for identifying patients with ankylosing spondylitis who are likely to respond to an exercise program: a pilot study. American Journal of Physical Medicine & Rehabilitation 2009;88(6):445–54.

Andersson GB. Epidemiology of low back pain. Acta Orthopaedica Scandinavica Supplementum 1998;281:28–31.

Assmann SF, Pocock SJ, Enos LE, Kasten LE. Subgroup analysis and other (mis)uses of baseline data in clinical trials. Lancet 2000;355(9209):1064–9.

Bachmann LM, Haberzeth S, Steurer J, ter Riet G. The accuracy of the Ottawa knee rule to rule out knee fractures: a systematic review. Annals of Internal Medicine 2004;140(2):121–4.

Beattie P, Nelson R. Clinical prediction rules: what are they and what do they tell us? Australian Journal of Physiotherapy 2006;52(3):157–63.

Beneciuk JM, Bishop MD, George SZ. Clinical prediction rules for physical therapy interventions: a systematic review. Physical Therapy 2009;89(2):114–24.

Brennan GP, Fritz JM, Hunter SJ, Thackeray A, Delitto A, Erhard RE. Identifying subgroups of patients with acute/subacute "nonspecific" low back pain: results of a randomized clinical trial. Spine 2006;31(6):623–31.

Brookes ST, Whitely E, Egger M, Smith GD, Mulheran PA, Peters TJ. Subgroup analyses in randomized trials: risks of subgroup-specific analyses; power and sample size for the interaction test. Journal of Clinical Epidemiology 2004;57(3):229–36.

Cai C, Pua YH, Lim KC. A clinical prediction rule for classifying patients with low back pain who demonstrate short-term improvement with mechanical lumbar traction. European Spine Journal 2009;18(4):554–61.

Childs J, Fritz J, Piva S, Erhard R. Clinical decision making in the identification of patients likely to benefit from spinal manipulation: a traditional versus an evidence-based approach. Journal of Orthopaedic and Sports Physical Therapy 2003;33(5):259–72.

Childs JD, Cleland JA. Development and application of clinical prediction rules to improve decision making in physical therapist practice. Physical Therapy 2006;86(1):122–31.

Childs JD, Flynn TW, Fritz JM. A perspective for considering the risks and benefits of spinal manipulation in patients with low back pain. Manual Therapy 2006;11(4):316–20.

Childs JD, Fritz JM, Flynn TW, Irrgang JJ, Johnson KK, Majkowski GR, et al. A clinical prediction rule to identify patients with low back pain most likely to benefit from spinal manipulation: a validation study. Annals of Internal Medicine 2004;141(12):920–8.

Cleland J, Fritz J, Kulig K, Davenport T, Eberhart S, Magel J, et al. Comparison of the effectiveness of three manual physical therapy techniques in a subgroup of patients with low back pain who satisfy a clinical prediction rule: a randomized clinical trial. Spine 2009;34(25):2720–9.

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. Journal of Orthopaedic & Sports Physical Therapy 2006;36(4):209–14.

Cochrane Back Group. Search strategies for back group specialized registry, http://www.mrw.interscience.wiley.com/cochrane/clabout/articles/BACK/frame.html; 2009 [accessed January 2009].

Concato J, Feinstein AR, Holford TR. The risk of determining risk with multivariable models. Annals of Internal Medicine 1993;118(3):201–10.

Cook C. Potential pitfalls of clinical prediction rules. Journal of Manual and Manipulative Therapy 2008;16(2):69–71.

Cook C, Brismee JM, Pietrobon R, Sizer Jr P, Hegedus E, Riddle DL. Development of a quality checklist using Delphi methods for prescriptive clinical prediction rules: the QUADCPR. Journal of Manipulative and Physiological Therapeutics 2010;33(1):29–41.

Dahri K, Loewen P. The risk of bleeding with warfarin: a systematic review and performance analysis of clinical prediction rules. Thrombosis & Haemostasis 2007;98(5):980–7.

Eagles D, Stiell IG, Clement CM, Brehaut J, Kelly AM, Mason S, et al. International survey of emergency physicians' priorities for clinical decision rules. Academic Emergency Medicine 2008;15(2):177–82.

Flynn T, Fritz J, Whitman J, Wainner R, Magel J, Rendeiro D, et al. A clinical prediction rule for classifying patients with low back pain who demonstrate short-term improvement with spinal manipulation. Spine 2002;27(24):2835–43.

Fritz J, Hancock M, Herbert R, Riddle D. July 2009 debate: Clinical prediction rules – part 1, http://www.scienceaudio.net/ptj/200907/ptj_200907_debate_1.mp3; 2009a [accessed 27.02.10].

Fritz J, Hancock M, Herbert R, Riddle D. July 2009 debate: Clinical prediction rules – part 2, http://www.scienceaudio.net/ptj/200907/ptj_200907_debate_2.mp3; 2009b [accessed 27.02.10].

Fritz JM. Clinical prediction rules in physical therapy: coming of age? Journal of Orthopaedic & Sports Physical Therapy 2009;39(3):159–61.

Fritz JM, Brennan GP, Leaman H. Does the evidence for spinal manipulation translate into better outcomes in routine clinical care for patients with occupational low back pain? A case-control study. Spine Journal: Official Journal of the North American Spine Society 2006;6(3):289–95.

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 Family Practice 2005a;6(29).

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(13):1363–71.

Fritz JM, Lindsay W, Matheson JW, Brennan GP, Hunter SJ, Moffit SD, 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(26):E793–800.

Fritz JM, Piva SR, Childs JD. Accuracy of the clinical examination to predict radiographic instability of the lumbar spine. European Spine Journal 2005b;14(8): 743–50.

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. Physical Therapy 2004;84(2):173–90.

George SZ, Bialosky JE, Donald DA. The centralization phenomenon and fear-avoidance beliefs as prognostic factors for acute low back pain: a preliminary investigation involving patients classified for specific exercise. Journal of Orthopaedic & Sports Physical Therapy 2005;35(9):580–8.

Hallegraeff JM, de Greef M, Winters JC, Lucas C. Manipulative therapy and clinical prediction criteria in treatment of acute nonspecific low back pain. Perceptual & Motor Skills 2009;108(1):196–208.

Hancock M, Maher RD, Maher CG. A guide to interpretation of studies investigating subgroups of responders to physical therapy interventions. Physical Therapy 2009a;89(7):698–704.

Hancock MJ, Maher CG, Herbert RD. Answer to the letter to the editor of J. Hebert et al. concerning "Hancock MJ, Maher CG, Latimer J, Herbert RD, McAuley JH (2008) Independent evaluation of a clinical prediction rule for spinal manipulative therapy: a randomised controlled trial. European Spine Journal 2008a; 17(10):1403–4.

Hancock MJ, Maher CG, Latimer J, Herbert RD, McAuley JH. Independent evaluation of a clinical prediction rule for spinal manipulative therapy: a randomised controlled trial. European Spine Journal 2008b;17(7):936–43.

Hancock MJ, Maher CG, Latimer J, Herbert RD, McAuley JH. Can rate of recovery be predicted in patients with acute low back pain? Development of a clinical prediction rule. European Journal of Pain 2009b;13(1):51–5.

Hebert JJ, Perle SM. Letter to the editor concerning "Independent evaluation of a clinical prediction rule for spinal manipulative therapy: a randomised controlled trial" (M. Hancock et al.). European Spine Journal 2008;17(10):1401–2.

Henschke N, Maher CG, Refshauge KM, Herbert RD, Cumming RG, Bleasel J, et al. Prevalence of and screening for serious spinal pathology in patients presenting to primary care settings with acute low back pain. Arthritis & Rheumatism 2009;60(10):3072–80.

Hess EP, Thirganasambandamoorthy V, Wells GA, Erwin P, Jaffe AS, Hollander JE, et al. Diagnostic accuracy of clinical prediction rules to exclude acute coronary syndrome in the emergency department setting: a systematic review. CJEM Canadian Journal of Emergency Medical Care 2008;10(4):373–82.

Hicks GE, Fritz JM, Delitto A, McGill SM. Preliminary development of a clinical prediction rule for determining which patients with low back pain will respond to a stabilization exercise program. Archives of Physical Medicine and Rehabilitation 2005;86(9):1753–62.

Ingui BJ, Rogers MA. Searching for clinical prediction rules in MEDLINE. Journal of the American Medical Informatics Association 2001;8(4):391–7.

Katz MH. Multivariable analysis. A practical guide for clinicians. 2nd ed. New York: Cambridge University Press; 2006.

Kent P, Keating JL. Classification in nonspecific low back pain: what methods do primary care clinicians currently use? Spine 2005;30(12):1433–40.

Kent PM, Keating JL, Taylor NF. Primary care clinicians use variable methods to assess acute nonspecific low back pain and usually focus on impairments. Manual Therapy 2009;14(1):88–100.

Kraemer HC, Frank E, Kupfer DJ. Moderators of treatment outcomes: clinical, research, and policy importance. Journal of the American Medical Association 2006;296(10):1286–9.

Kraemer HC, Gibbons RD. Why does the randomized clinical trial methodology so often mislead clinical decision making? Focus on moderators and mediators of treatment. Psychiatric Annals 2009;39(7):736–45.

Kraemer HC, Wilson GT, Fairburn CG, Agras WS. Mediators and moderators of treatment effects in randomized clinical trials. Archives of General Psychiatry 2002;59(10):877–83.

Laslett M, Aprill CN, McDonald B, Oberg B. Clinical predictors of lumbar provocation discography: a study of clinical predictors of lumbar provocation discography. European Spine Journal 2006a;15(10):1473–84.

Laslett M, Aprill CN, McDonald B, Young SB. Diagnosis of sacroiliac joint pain: validity of individual provocation tests and composites of tests. Manual Therapy 2005;10(3):207–18.

Laslett M, McDonald B, Aprill CN, Tropp H, Oberg B. Clinical predictors of screening lumbar zygapophyseal joint blocks: development of clinical prediction rules. Spine Journal 2006b;6(4):370–9.

Laslett M, Young SB, Aprill CN, McDonald B. Diagnosing painful sacroiliac joints: a validity study of a McKenzie evaluation and sacroiliac provocation tests. Australian Journal of Physiotherapy 2003;49(2):89–97.

Laupacis A, Sekar N, Stiell IG. Clinical prediction rules: a review and suggested modifications of methodological standards. Journal of the American Medical Association 1997;277(6):488–94.

Lewis S. Regression analysis. Practical Neurology 2007;7(4):259–64.

Long A, Donelson R, Fung T. Does it matter which exercise? A randomized control trial of exercise for low back pain. Spine 2004;29(23):2593–602.

MacKinnon DP, Luecken LJ. How and for whom? Mediation and moderation in health psychology. Health Psychology 2008;27(Suppl. 2):S99–100.

May S, Gardiner E, Young S, Klaber-Moffett J. Predictor variables for a positive long-term functional outcome in patients with acute and chronic neck and back pain treated with a McKenzie approach: a secondary analysis. Journal of Manual and Manipulative Therapy 2008;16(3):155–60.

May S, Rosedale R. Prescriptive clinical prediction rules in back pain research: a systematic review. Journal of Manual and Manipulative Therapy 2009;17(1): 36–45.

McGinn T, Wyer P, Wisnivesky J, Devereaux PJ, Stiell I, Richardson S, et al. Advanced topics in diagnosis: clinical prediction rules. In: Guyatt G, Rennie D, Meade MO, Cook DJ, editors. Users' guides to the medical literature. A manual for evidence-based clinical practice. 2nd ed. USA: McGraw Hill Medical; 2008. p. 491–505.

McGinn TG, Guyatt GH, Wyer PC, Naylor CD, Stiell IG, Richardson WS. Users' guides to the medical literature: XXII: how to use articles about clinical decision rules. Evidence-based medicine working group. JAMA 2000;284(1):79–84.

Randolph AG, Guyatt GH, Calvin JE, Doig G, Richardson WS, Cook DJ. Understanding articles describing clinical prediction tools. Critical Care Medicine 1998;26(9): 1603–12.

Raney NH, Petersen EJ, Smith TA, Cowan JE, Rendeiro DG, Deyle GD, et al. Development of a clinical prediction rule to identify patients with neck pain likely to benefit from cervical traction and exercise. European Spine Journal 2009;18(3): 382–91.

Reilly BM, Evans AT. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. Annals of Internal Medicine 2006; 144(3):201–9.

Riihimaki H. Epidemiology and pathogenesis of non-specific low back pain: what does the epidemiology tell us? Bulletin of the Hospital for Joint Diseases 1996; 55(4):197–8.

Sackett D, Hayes R, Guyatt G, Tugwell P. Clinical epidemiology: a basic science for clinical medicine. 2nd ed. Boston, USA: Brown & Co; 1991.

Stanton TR, Hancock MJ, Maher CG, Koes BW. Critical appraisal of clinical prediction rules that aim to optimize treatment selection for musculoskeletal conditions. Physical Therapy 2010;90(6):843–54.

Stanton TR, Maher CG, Hancock M. On clinical prediction rules for physical therapy interventions. Physical Therapy 2009;89(4):394.

Stiell IG, Wells GA. Methodologic standards for the development of clinical decision rules in emergency medicine. Annals of Emergency Medicine 1999;33(4): 437–47.

Szadek KM, van der Wurff P, van Tulder MW, Zuurmond WW, Perez RS. Diagnostic validity of criteria for sacroiliac joint pain: a systematic review. Journal of Pain 2009;10(4):354–68.

Tamariz LJ, Eng J, Segal JB, Krishnan JA, Bolger DT, Streiff MB, et al. Usefulness of clinical prediction rules for the diagnosis of venous thromboembolism: a systematic review. American Journal of Medicine 2004;117(9):676–84.

Teyhen DS, Flynn TW, Childs JD, Abraham LD. Arthrokinematics in a subgroup of patients likely to benefit from a lumbar stabilization exercise program. Physical Therapy 2007;87(3):313–25.

Toll DB, Janssen KJM, Vergouwe Y, Moons KGM. Validation, updating and impact of clinical prediction rules: a review. Journal of Clinical Epidemiology 2008; 61(11):1085–94.

Turner JA, Holtzman S, Mancl L. Mediators, moderators, and predictors of therapeutic change in cognitive-behavioral therapy for chronic pain. Pain 2007; 127(3):276–86.

van der Wurff P, Buijs EJ, Groen GJ. A multitest regimen of pain provocation tests as an aid to reduce unnecessary minimally invasive sacroiliac joint procedures. Archives of Physical Medicine & Rehabilitation 2006;87(1):10–4.

Walker BF. The prevalence of low back pain in Australian adults. A systematic review of the literature from 1966–1998. Asia-Pacific Journal of Public Health 1999;11(1):45–51.

Wasson JH, Sox HC, Neff RK, Goldman L. Clinical prediction rules. Applications and methodological standards. New England Journal of Medicine 1985;313(13): 793–9.

Werner EL, Indahl A, Werner EL, Indahl A. Knowledge, practice and attitudes to back pain among doctors, physiotherapists and chiropractors. Tidsskrift for Den Norske Laegeforening 2005;125(13):1794–7.

Wisnivesky JP, Serebrisky D, Moore C, Sacks HS, Iannuzzi MC, McGinn T. Validity of clinical prediction rules for isolating inpatients with suspected tuberculosis. A systematic review. Journal of General Internal Medicine 2005;20(10):947–52.