Clinical prediction rules: What are they and what do they tell us?

Paul Beattie¹ and Roger Nelson²

¹University of South Carolina  ²Lebanon Valley College

USA

Introduction

Accurate clinical decision-making is a central component of patient care. Historically, clinicians have relied upon expert opinion, experience, and intuition to determine which evaluative procedures to perform, how to interpret the results of these procedures, and what interventions to choose. As the volume of scientifically sound research continues to grow, clinicians are increasingly incorporating these research findings with their clinical intuition and the unique circumstances of the individual patient, to make the most informed clinical decisions (Sackett 1998, Wang et al 2003).

A helpful way of incorporating research into clinical decision-making is through the use of clinical prediction rules. Randolph et al (1998) have described a clinical prediction rule as the process by which combinations of clinical findings that have been statistically demonstrated to be meaningful predictors of a condition or outcome of interest are used to categorise a heterogeneous group of patients into subgroups based on a shared likelihood of the presence of that condition or outcome. Simply, clinical prediction rules may be thought of as the combining of relevant clinical findings to calculate a numeric probability of the presence of a specific disorder or likelihood of an outcome, ie, they act as adjuncts to the evaluative process. The number and variety of predictors required for a meaningful clinical prediction rule may vary greatly as a function of the complexity of the outcome of interest for the population studied. For example, recently-described clinical prediction rules contain predictors based solely on physical examination findings (Bachmann et al 2003, Bachmann et al 2004), patient history (Buchsbaum et al 1991), a combination of patient history and physical examination (Hawker et al 2002, Wells et al 1997), or a combination of patient history, physical examination, and patient beliefs (Childs et al 2004, Flynn et al 2002).

The appropriate use of clinical prediction rules has the potential to assist in the process of clinical prediction. For example, clinical prediction rules can be valuable for patient screening by helping clinicians to decide when further evaluation is likely to yield meaningful findings (Bachmann et al 2003, Mauck et al 2005). This can improve the likelihood of detection of a condition when present. Conversely, clinical prediction rules can help clinicians to decide that further testing is not likely to yield meaningful findings, thereby reducing cost and patient anxiety that may be associated with over-evaluation when a condition is not present (Caragee and Hannibal 2004). Clinical prediction rules may provide likelihoods of a favourable or unfavourable outcome for patients with specific clinical findings and can be of great help in determining patient classification for treatment (Childs et al 2004, Kuipers et al 2006, Solomon et al 2004). In addition to impacting at the level of the individual patient, the use of validated clinical prediction rules across clinics can facilitate research by improving the homogeneity of patient classification which can, in turn, assist in the use of evidence-based guidelines.

In recent years there have been several publications describing the development of clinical prediction rules for use in physiotherapy practice (Table 1). While these studies represent a promising advance in evidence-based practice, it is important to note that all clinical prediction rules should be used with caution. For example, the term ‘clinical decision
rule’ is often used synonymously with clinical prediction rule (Laupacis et al 1997). Reilly and Evans (2006), however, argue that clinical decision-making is a separate domain from clinical prediction, ie, clinical prediction rules provide probabilities of a given diagnosis or prognosis but do not necessarily recommend decisions. Other authors agree and submit that clinical prediction rules have the potential to make important contributions to decision-making but they should not be used in isolation to direct it (Barry and McNamara 2005, Brehaut et al 2005, Cameron and Naylor 1999, Lang 2005). To ensure their appropriate utilisation, it is important that clinicians and researchers understand the strengths and weaknesses of clinical prediction rules and the conditions under which they may be optimally used. The purpose of this paper is to describe the potential role of clinical prediction rules in physiotherapy practice and to suggest strategies by which clinicians can determine the appropriateness of using clinical prediction rules for a given practice setting.

When are clinical prediction rules most needed?

As with other types of research evidence (Herbert et al 2001), clinical prediction rules are most needed in important areas of clinical uncertainty. For example, clinical prediction rules would be valuable in clinical scenarios where mis-inference may result in an increased risk of an adverse event and/or the unnecessary cost of care that may occur with utilisation of tests or interventions that fail to make meaningful contributions to patient outcome (McGinn et al 2000, Wasson et al 1985). The need for clinical prediction rules that assist in patient screening is enhanced when there is a concern that clinicians are failing to identify relevant but under-diagnosed conditions. For example, Riddle et al (2004) recently suggested that physiotherapists may underestimate the likelihood of proximal deep vein thrombosis, which is a potentially serious condition. The authors propose that the use of a previously-validated clinical prediction rule (Wells et al 1997) for screening at-risk patients may improve the likelihood of detection and result in improved outcomes. Along the same lines, Haggman et al (2004) submit that depression is commonly under-diagnosed and have identified a sensitive 2-question measure that helps identify at-risk patients.

Clinical prediction rules that assist in determining prognosis following intervention are valuable in classifying patients for intervention when there are conflicting reports regarding treatment effectiveness and/or a lack of meaningful patient classification systems to guide treatment (Dellito et al 1995, Flynn et al 2002, Hicks et al 2005, Kuijpers et al 2006). This is a common concern when treating individuals who have multifactorial conditions, such as low back pain, for whom establishing a prognosis and determining the optimal treatment is often difficult (DeBeard et al 2003, Riipinen et al 2005, Waddell 2004). Childs et al (2004) have attempted to address this concern by developing a clinical prediction rule that describes the likelihood of outcome following treatment with spinal manipulation. When used under similar conditions to those under which it was validated, this clinical prediction rule may provide valuable assistance for determining when to use spinal manipulation (Table 2).

<table>
<thead>
<tr>
<th>Study</th>
<th>Intended use</th>
<th>Sample</th>
<th>Level of validation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Childs et al (2004)</td>
<td>Identify patients with low back pain most likely to benefit from spinal manipulation</td>
<td>131 patients with low back pain</td>
<td>Level 2</td>
</tr>
<tr>
<td>Hicks et al (2005)</td>
<td>Identify patients with low back pain most likely to benefit from a lumbar stabilisation program</td>
<td>54 patients with low back pain</td>
<td>Derivation</td>
</tr>
<tr>
<td>Wainner et al (2005)</td>
<td>Identify the presence of carpal tunnel syndrome</td>
<td>82 patients with suspected cervical radiculopathy or carpal tunnel syndrome</td>
<td>Derivation</td>
</tr>
<tr>
<td>Tseng et al (2006)</td>
<td>Identify patients with neck pain most likely to benefit immediately from manipulation of the cervical spine</td>
<td>100 patients with neck pain</td>
<td>Derivation</td>
</tr>
<tr>
<td>Kuijpers et al (2006)</td>
<td>Identify the absolute risk of persistent shoulder problems</td>
<td>587 patients with shoulder problems</td>
<td>Derivation</td>
</tr>
</tbody>
</table>

Table 2. Predictors of favourable outcome following spinal manipulation (Flynn et al 2002, Childs et al 2004).

- Duration of symptoms < 16 days
- At least one hip with internal rotation of > 35 degrees
- Lumbar hypomobility
- No symptoms distal to the knee
- FABQW score of < 19

FABQW = Fear Avoidance Beliefs Questionnaire, Work Subscale
Table 3. The relation between the type of validity and method of testing the clinical prediction rule and the proposed use of the clinical prediction rule.

<table>
<thead>
<tr>
<th>Proposed use</th>
<th>Type of validity</th>
<th>Method of testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Screening: predicting the likelihood of the presence of a specific medical or</td>
<td>Criterion-referenced</td>
<td>Prospective, cross-sectional comparison of findings from clinical prediction rule to a ‘gold standard’ that indicates the presence or absence of the condition</td>
</tr>
<tr>
<td>psychobehavioural condition</td>
<td>Predictive</td>
<td>Prospective, longitudinal, comparison of findings from clinical prediction rule to measures of change in patient status over time</td>
</tr>
<tr>
<td>Prognosis: predicting the likelihood of a specific outcome</td>
<td>Prescriptive</td>
<td>Prospective, longitudinal, randomised, controlled design that compares outcomes following different interventions on subjects with the same score on the clinical prediction rule</td>
</tr>
<tr>
<td>Classification into treatment-based groups: predicting the likelihood of</td>
<td></td>
<td></td>
</tr>
<tr>
<td>outcome when a specific intervention is administered</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Although it would be desirable to have clinical prediction rules to assist most clinical decisions it is not practical to expect them to be developed for all potential clinical predictions (Upsur 2005). In many cases the amount of variation between patients in the target population is so high that it is not practical to expect a small number of variables to make a consistent, meaningful rule. This is illustrated by the differences in aetiologic and prognostic risk factors between patients with acute low back pain compared to those with chronic low back pain. Several clinical prediction rules have been described for patients with acute low back pain (Childs et al 2004, Flynn et al 2002, Hicks et al 2005). However, we were unable to locate any studies that have described clinical prediction rules for patients with chronic low back pain, which typically has much more variation in its manifestation than acute low back pain (Waddell 2004).

**What factors should be considered before using a clinical prediction rule?**

Several factors should be considered before using a clinical prediction rule in the clinical setting. These include: the method used to derive, validate and determine the impact of the clinical prediction rule (internal validity); the degree of accuracy in classifying patients (sensitivity and specificity; likelihood ratios); and the population for whom the use of the clinical prediction rule is appropriate (external validity).

**What is the method used to derive and validate the clinical prediction rule?**

To develop a clinical prediction rule, researchers start by defining an outcome of interest that is important and can be reliably measured, such as the presence or absence of a condition, or the likelihood of an outcome following classification. An exhaustive list of variables that are likely to be present at some level in most individuals, and that may predict the condition or outcome of interest, is identified by literature review, expert opinion and, if possible, by focus groups with patients who have the condition of interest (Laupacis et al 1997, Randolph et al 1998). Often a very large initial intake instrument is necessary to address this challenge (Dionne et al 2005). A concern is that the number of variables chosen dictates the sample size, ie, studies that attempt to derive clinical prediction rules in heterogeneous populations will require large sample sizes.

To assess the relationship between the outcome of interest and the predictor variables, a relevant sample of subjects and clinicians is identified and defined carefully. Prospective or retrospective data are collected that describe these relationships. It is important to note that different research methods are needed in the development phase of a clinical prediction rule based upon its proposed use. Clinical prediction rules being developed for use as screening tools require evidence of criterion-referenced validity in which the score of the clinical prediction rule has a strong relationship to a ‘gold standard’ that identifies the presence or absence of the condition. This gold standard is often derived from information obtained from imaging, biopsy, or surgery. Positive and negative responses on the clinical prediction rule should be highly predictive of positive and negative findings on the gold standard. Clinical prediction rules that are to be used for the establishment of prognosis must have evidence of predictive validity in which the initial score is strongly associated with change in patient status over time. Those clinical prediction rules that assist in treatment selection need evidence of prescriptive validity in which patients with similar scores are randomised to different treatment groups and one treatment group has clinically superior outcomes compared to others (Table 3).

As with any sound research design, the data collection must be as free from bias as possible. Subjects must be chosen in an unbiased fashion and the examiners who obtain the predictor variables must be masked to outcome, while those assessing outcome must be masked from the predictors. An adequate sample size should be addressed by power analysis (Simel et al 1991). Great care should be taken to maximise follow-up. Although intention-to-treat strategies have been described to account for subjects lost to follow-up (Childs et al 2004), subject attrition reduces the strength of the conclusions relating to clinical prediction rules (Laupacis et al 1997, Randolph et al 1998).

Following data collection, statistical analysis is performed to determine which combination of variables provides the best degree of diagnostic accuracy (Guyatt et al 1995, Rudy et al 1992). These findings provide a framework for further evaluation but do not supply evidence of validity for clinical use (Laupacis et al 1997, McGinn et al 2000, Randolph et al 1998). If the combination of variables has face validity and,
in the opinion of the investigators, has a reasonable degree of diagnostic accuracy, the clinical prediction rule is robust enough to continue to the validation phase.

Clinical prediction rules are structured to fit datasets from which they have been investigated. When used on other populations, they usually have reduced performance (Kocher et al. 2004). Considering this, the validation process consists of a series of studies on broader populations of clinicians and patients (Laupacis et al. 1997, McGinn et al. 2000, Reilly and Evans 2006). Initially, this is done under similar conditions, e.g., in a patient population and setting similar to that of the derivation study. The overall goals are to determine the degree to which the clinical prediction rule is reproducible on different populations and the degree to which its usage impacts upon clinical practice.

McGinn et al. (2000) have proposed a hierarchy of validation for clinical prediction rules (Table 4). The initial stages of validation – Levels 3 and 4 – help to insure that the initial derivation was not a random finding. Level 4 validation provides important preliminary information about the potential of the clinical prediction rule to classify subjects accurately under very restricted conditions while Level 3 clinical prediction rules have undergone validation in a narrow, prospective sample that is usually similar to the initial sample. Level 3 validation provides evidence that supports use of the clinical prediction rule, with caution, on samples of patients similar to those in the derivation process. If the results are favourable, the validation process extends to other populations and settings. This is referred to as Level 2 validation and acts to determine the generalisability or external validity of the clinical prediction rule to other patients and clinicians. Level 2 clinical prediction rules have undergone more extensive validation demonstrating consistent accuracy in one prospective study that includes a large variety of patients and clinicians or in numerous smaller studies with differing sample characteristics. Level 2 clinical prediction rules may be used for a wide variety of patients by a variety of clinicians in an array of settings consistent with the studies used for validation. The highest level of validation is Level 1. Level 1 clinical prediction rules have at least the validation characteristics of Level 2 clinical prediction rules and have undergone an impact analysis that demonstrates a change in clinician behaviour and favourable improvement of cost-effectiveness of patient care as determined by improved outcomes, reduced cost of care, and/or reduction of adverse events. Clinical prediction rules with Level 1 validation may therefore be used in a wide array of settings with confidence that their use will be beneficial (McGinn 2000). This level of validity is the most difficult to achieve but is necessary prior to determining if a clinical prediction rule is beneficial or harmful (Reilly and Evans 2006). Despite the large number of clinical prediction rules that have been reported in the recent literature most have yet to achieve Level 1 validation. In some instances attempts at Level 1 validation have revealed that the rules have failed to make a measurable impact upon clinical practice. For example, the impact of one of the most highly-validated clinical prediction rules, the Ottawa Ankle

<table>
<thead>
<tr>
<th>Stage of development</th>
<th>Rationale</th>
<th>Research strategy</th>
<th>Clinical utility</th>
</tr>
</thead>
<tbody>
<tr>
<td>Need</td>
<td>Concern over inadequate early detection of condition; use of ineffective treatments; excessive cost of care; poor outcomes</td>
<td>Clinical observations, review of literature, patient focus groups, expert clinician panel.</td>
<td>Proposed model only</td>
</tr>
<tr>
<td>Initial development</td>
<td>Identify all relevant predictors and relevant outcome measures</td>
<td>Determine reliability of measures</td>
<td>Needs further validation before clinical usage</td>
</tr>
<tr>
<td>Derivation</td>
<td>Determine variables that are the most powerful predictors</td>
<td>Sampling strategy, obtain measures, ensure complete follow-up</td>
<td>May be used for similar patients</td>
</tr>
<tr>
<td>Level 4 validation</td>
<td>Provide preliminary information regarding the stability of the proposed clinical prediction rule for limited, well-defined population</td>
<td>Not validated or validated with split-half of original data set, or retrospective data</td>
<td>May be used in a variety of settings consistent with patients and clinicians that were investigated</td>
</tr>
<tr>
<td>Level 3 validation</td>
<td>Determine if the proposed model is stable for different but similar sample</td>
<td>Prospective, similar sample and examiners</td>
<td>May be used for similar patients</td>
</tr>
<tr>
<td>Level 2 validation</td>
<td>Determine if the proposed model yields similar results for a variety of patients</td>
<td>Prospective with a variety of patients and clinicians. One large study or several small studies</td>
<td>May be used in a variety of settings consistent with patients and clinicians that were investigated</td>
</tr>
<tr>
<td>Level 1 validation</td>
<td>Determine if the proposed model improves overall clinical practice and changes clinical behaviour. Determine if the use of the rule improves patient outcomes</td>
<td>Prospective studies with a wide variety of subjects and clinicians, at least 1 impact study that describes improvement in clinical practice</td>
<td>May be used in a wide variety of settings with confidence that it can improve outcomes</td>
</tr>
</tbody>
</table>
Rules (Bachmann et al 2003), has recently been questioned. Brehaut et al (2005) reported that while most physicians use the rule, it is not considered a major determinant in decision making. Thus, although a rule may be validated for use on a wide population, its overall impact may be negligible.

**How accurate is the clinical prediction rule for its intended use?**

The main advantage of a clinical prediction rule is that it provides a numeric index of the post-test probability of a given condition or outcome. Estimates of sensitivity (the true positive rate) and specificity (the true negative rate) are typically derived. From these estimates positive or negative likelihood ratios can be calculated. Likelihood ratios describe how much a test will raise or lower the pre-test likelihood of a condition or outcome being present (Jaeschke et al 1994, Riddle and Stratford 1999, Fritz and Wainner 2001). Likelihood ratios less than 1.0 reduce the likelihood, while those greater than 1.0 increase it. Jaeschke et al (1994) propose that likelihood ratios < 0.1 and > 10 suggest large and conclusive changes in post-test likelihood; likelihood ratios of 0.2–0.1 and 5–10 suggest moderate changes; likelihood ratios of 0.5–0.2 and 2–5 suggest small, but possibly important changes while likelihood ratios between 0.5–2.0 are small and rarely important.

The interpretation of statistical indices is improved by reporting confidence intervals that describe the precision of the likelihood ratio (Childs et al 2005). Confidence intervals with smaller ranges indicate a high degree of precision of the estimate. There are no agreed values for diagnostic accuracy that determine if a test is adequate. Clinicians must determine the degree of error or misclassification that is tolerable. For an in-depth discussion of the interpretation of the accuracy of a clinical prediction rule, readers are referred to work of Justice et al (1999).

**Is a given clinical prediction rule appropriate for your clinical practice?**

Clinicians should ensure that the patients and clinicians for whom the clinical prediction rule is to be used share similar traits to those used for validation. Variations in such factors as the age of subjects, the a priori likelihood of a condition (Barry and McNamera 2005, Mauck et al 2005), or social factors (Lang 2005) can result in a large variation in the sensitivity and specificity of the clinical prediction rule to detect the likelihood of a condition and makes the cross-population use of a measure problematic. The presence of worker’s compensation may have a profound effect on prognosis and treatment response and must be clearly identified in a study (DeBeard et al 2003). The traits of the clinicians must be well described in terms of unique training and skills (Laupacis et al 1997, Randolph et al 1998). For example, a clinical prediction rule that incorporates assessment of lumbar spine mobility or spinal manipulation is not appropriate for a clinician who is not trained in these techniques. This concern is illustrated by Childs et al (2004) who concluded that a clinical prediction rule used to classify subjects with low back pain for treatment with spinal manipulation had Level 2 validation and could be used on a broad spectrum of patients. It should be noted, however, that this study was predominately performed using specially-trained physical therapists in the United States Military who were treating patients in military facilities. Although several clinics were involved in this study the overall heterogeneity of the population and the applicability of these results to other practitioners and patients are unknown.

**Limitations and misuse of clinical prediction rules**

Inappropriate use of existing clinical prediction rules, or using measures as clinical prediction rules that have not been developed for that purpose, can result in mis-inference and lead to inappropriate patient classification or treatment. For example, Waddell et al (1980) suggested that individuals who had three or more positive ‘Waddell signs’ may be at risk of having psychosocial factors that influence their clinical picture. The authors proposed this for use as a general screening tool based on the literature and on expert opinion. However, despite the vigorous opposition of Main and Waddell (1998) and others (Fishbain et al 2004), the Waddell signs have been used as a predictor of ‘malingering’ (Kiester and Duke 1999). Inappropriately classifying a patient of ‘malingering’ may cost an individual his or her job and result in long-term negative consequences.

In some instances, clinical prediction rules may be incorporated into clinical practice without adequate validation and fail to be validated with subsequent testing. Functional capacity evaluations have been used for many years to assist clinical decision making for injured workers with low back pain. However, there has been little evidence that describes the predictive validity of judgments based on these tests. In an attempt to address this, Gross et al (2004) and Gross and Battie (2004) tested the validity of a ‘functional capacity evaluation’ as an indicator of return to work and as a predictor of risk of recurrence for subjects with work-related back pain. Their findings disputed previous claims and indicated that the results of functional capacity evaluations were only weakly predictive of return to work and, surprisingly, were inversely related to sustained recovery, i.e., higher scores on the functional capacity evaluation were associated with a higher likelihood of recurrence. These findings are thought-provoking and illustrate the need for thorough validation of predictive measures. In addition to being a relatively expensive procedure, the potential for mis-inference from functional capacity evaluations could have a profound effect on the patient, the employer, the payer, and society.

**Conclusion**

We believe that expert opinion is still an important cornerstone of clinical decision making. However, opinions may be influenced by bias. The rising popularity of evidence-based health care has led to decision-making formats that attempt to minimise bias in judgments (Herbert et al 2001). Within this model, decisions regarding the need for further testing or classification of patients for treatment selection following patient examination are based on a combination of the clinician’s experience and intuition, the patient’s values, and the best research evidence (Sackett 1998). Clinical prediction rules use quantitative methods to build upon the body of literature and expert opinion and can provide quick and inexpensive estimates of probability. In our opinion, the development of valid clinical prediction rules should be an important goal of physical therapy research. Specific areas of consideration may include deriving and validating clinical prediction rules to screen patients for potentially serious conditions for which current tests lack adequate diagnostic accuracy or have unacceptable cost and risk, and to assist in classification of patients for treatments that are likely to result in substantially different outcomes in heterogeneous groups of patients.
Clinical prediction rules can be of great value to assist clinical decision making but they should not be used indiscriminately (Riddle et al. 2004). Clinical prediction rules are not a replacement for clinical judgment and should complement not supplant clinical opinion and intuition.

Correspondence Paul Beattie, Program in Physical Therapy, School of Public Health, University of South Carolina, USA. Email: pbeattie@gwm.sc.edu

References


Riddle DL and Stratford PW (1999) Interpreting validity indexes


