EDITORIAL

Potential Pitfalls of Clinical Prediction Rules

What Are Clinical Prediction Rules?

A clinical prediction rule (CPR) is a combination of clinical findings that have statistically demonstrated meaningful predictability in determining a selected condition or prognosis of a patient who has been provided with a specific treatment[1,2]. CPRs are created using multivariate statistical methods, are designed to examine the predictive ability of selected groupings of clinical variables[3,4], and are intended to help clinicians make quick decisions that may normally be subject to underlying biases[5]. The rules are algorithmic in nature and involve condensed information that identifies the smallest number of indicators that are statistically diagnostic to the targeted condition[6]. The number of derived or validated CPRs is increasing[6], specifically in rehabilitation medicine where prescriptive studies have been developed for musculoskeletal interventions for low back pain[7,8], cervical pain[9,10], and knee dysfunction[11,12].

Clinical prediction rules may best be classified into three distinct groups: 1) diagnostic, 2) prognostic, and 3) prescriptive[1,13]. Studies that focus on predictive factors related to a specific diagnosis are known as *diagnostic* CPRs. Clinical prediction rules that are designed to predict an outcome such as success or failure are considered *prognostic*. Clinical prediction rules designed to target the most effective interventions are identified as *prescriptive*, and these require prospective, longitudinal, randomized controlled trials that compare outcomes after selected interventions for subjects who meet a similar score on the CPR[1].

Clinical prediction rules are generally developed using a 3-step method[14]. First, CPRs are derived prospectively using multivariate statistical methods to examine the predictive ability of selected groupings of clinical variables[3]. The second step involves validating the CPR in a randomized controlled trial to reduce the risk that the predictive factors developed during the derivation phase were selected by chance[14]. The third step involves conducting an impact analysis to determine the extent that the CPR improves care, reduces costs, and accurately defines the targeted objective[14].

Although there is little debate that carefully constructed CPRs can improve clinical practice, to my knowledge, there are no guidelines that specify methodological requirements for CPRs for infusion into *all* clinical practice environments. Guidelines are created to improve the rigor of study design and reporting. The following editorial outlines potential methodological pitfalls in CPRs that may significantly weaken the transferability of the algorithm. Within the field of rehabilitation, most CPRs have been prescriptive; thus, my comments here are reflective of prescriptive CPRs.

Methodological Pitfalls

CPRs are designed to specify a homogenous set of characteristics from a heterogeneous population of prospectively selected consecutive patients[5,15]. Typically, the resulting applicable population is a small subset of a larger sample and may only represent a small percentage of the clinician's actual daily caseload. The setting and location of the larger sample should be generalizable[15,16], and subsequent validity studies require assessment of the CPR in different patient groups, in different environments, and with a typical patient group seen by most clinicians[16]. Because many CPRs are developed based on a very distinct group, that may or may not be reflective of a typical population of patients, the spectrum transportability[17] of many current CPR algorithms may be limited.

Clinical prediction rules use outcome measures to determine the effectiveness of the intervention. Outcome measures must have a single operational definition[5] and require enough responsiveness to truly capture appropriate change in the condition[14]; in addition, these measures should have a well constructed cut-off score[16,18] and be collected by a blinded administrator[15]. The selection of an appropriate anchor score for measurement of actual change is currently debated[19-20]. Most outcome measures use a patient recall-based questionnaire such as a global rating of change score (GRoC), which is appropriate when used in the short term but suffers from recall bias when used in long-term analyses[19-21]. Other studies may use minimally detectable change scores that were originally validated using the GROC and also may be affected by both recall bias and differences in sample severity or pathology. Lastly, outcome measures that use scores that are influenced by administrative factors (discharge date, length of stay, patient charges), socio-demographic factors, or internal behavioral characteristics (changes in fear avoidance or attitude) are not consistent among populations[5].

A potential drawback for CPRs is the failure to maintain the quality of the tests and measures used as predictors in the algorithm. The prospective test and measures should be independent of one another during modeling[16]; each should be performed in a meaningful, acceptable manner[4]; and clinicians or data administrators should be blinded to the patient's outcomes measures and condition[22]. Furthermore, the tests should demonstrate acceptable reliability ($\geq$ 0.60)[15] and require administration within an acceptable timeframe of the outcome measure[22]; equivocal or indeterminable results necessitate reporting[22]. Recognizing the likelihood of a true positive finding in the absence of any information will avoid the representative heuristic pitfall that may compel us toward identifying a clinical test as positive simply because the result fits the pattern of other findings[23]. CPRs that use tests and measures with reliability or agreement below 0.60 may result in variable findings depending on the clinician who performs the examination and depending on the findings of other tests and measures.

It is my impression that the most frequent current pitfall of CPRs is associated with the failure to meet statistical assumptions during regression modeling. CPRs are typically underpowered falling below the suggested requirements of 10 to 15 subjects for each prospective predictor variable[24]. Validation cohorts require sampling sizes of 100 or greater with use of logistic regression (used as a standard for CPR assessment)[25]. Rarely is the statistical significance of the model reported in the rehabilitation-based CPRs, nor is the R^2 or R^2-equivalent of the model identified[5]. An R^2 or R^2-equivalent outlines the strength of association of the predictor variables (both independently and as a group) in explaining the variance of the outcome measure. Low R^2 or R^2-equivalents may suggest that other variables more accurately predict the outcome of the study[5] and generally suggest a low effect size of the independent variables identified and retained in the analyses[26]. Most CPRs do report confidence intervals, and when reported, wide confidence intervals imply poor precision or too small of a sample size[15].

Once a CPR is developed, it is important to recognize the true benefit of the tool. It has been suggested that for true impact on clinical practice, CPRs should provide a LR+ of 5 or greater[27]. CPR derivations performed on high-risk groups, where failure to provide the appropriate intervention is highly undesirable, should have sensitivity values that are greater than specificity values[28]. This indicates that the final algorithm will accurately provide all of the best treatment(s) possible versus assuring that only those specific to the problem are used[28].

CPRs should have clinical sensibility. Clinical sensibility implies that the tool makes inherent clinical sense, that it's easy to use, that the tests and measures are truly related to the outcome, and that clinician perception does not overly alter the findings of the tool[15]. Consequently, tests and measures that vary in clinical interpretation (e.g., spring tests of the spine) or that are potentially explained by factors beyond the original scope of the examination (e.g., hip osteoarthritis when addressing hip procedures that affect the knee) may not be as useful as factors that are more explicit during clinical assessment.

Lastly, most rehabilitation-related CPRs are derivation studies, which are the initial steps in the development of clinical decision rules. Derivation studies lack validation and require follow-up studies in diverse centers with different populations of patients and different clinicians. Whether the findings from a derivation study stand up to the scrutiny of further assessment is unknown[15]. In essence, adoption of a derivation-only CPR runs the risk of improper treatment. Careful attention should be made before blindly adopting derivation studies or basing treatment pathways on these tools.

Summary

Is this editorial an attack on clinical prediction rules? Actually, it's quite the contrary. Prescriptive CPRs are useful tools for a select and discrete population of patients. As manually oriented clinicians, we have long realized that sub-sets of the population benefit from manual therapy more so than others. CPRs allow us to isolate a sub-set of desired patient characteristics and to define which techniques are most useful for that population. The current rehabilitation-based CPRs have opened the door for additional research to improve our accuracy as clinicians. Unfortunately, many of the present rehabilitation-based CPRs may have methodological weaknesses that may allow questioning of the utility of the instrument. Although there is no such thing as a "perfect" study, better and more rigorous designs should provide additional, profound and clinically applicable findings. As a clinician and a researcher, I am an advocate of CPRs.

Chad E. Cook PT, PhD, MBA,
OCS, FAAOMPT

REFERENCES

1. Beattie P, Nelson R. Clinical prediction rules: What are they and what do they tell us? *Aust J Physiother* 2006;52:157–163.
2. Randolph A, Guyatt H, Calvin JE, Doig G, Richardson WS. Understanding articles describing clinical prediction tools. *Crit Care Med* 1998;26:1603–1612.
3. Hier DB, Edlestein G. Deriving clinical prediction rules from stroke outcome research. *Stroke* 1991;22:1431–1436.
4. Kuijpers T, van der Heijden GJMG, Vergouwe Y, et al. Good generalizability of a prediction rule for prediction of persistent shoulder pain in the short term. *J Clin Epidemiol* 2007;60:947–953.
5. Wasson JH, Sox HC, Neff RK, Goldman L. Clinical prediction rules: Applications and methodological standards. *New Engl J Med* 1985;313:793–799.
6. Brehaut JC, Stiell IG, Visentin L, Graham ID. Clinical decision rules "in the real world": How a widely disseminated rule is used in everyday practice. *Acad Emerg Med* 2005; 12:948–956.
7. Childs JD, Fritz JM, Flynn TW, et al. A clinical prediction rule to identify patients with low back pain most likely to benefit from spinal manipulation: A validation study. *Ann Intern Med* 2004;141:920–928.
8. Hicks GE, Fritz JM, Delitto A, McGill SM. Preliminary development of a clinical prediction rule for determining which patients with low back pain will respond to a stabilization exercise program. *Arch Phys Med Rehabil* 2005;86:1753–1762.
9. Cleland JA, Childs JD, Fritz JM, Whitman JM, Eberhart SL. Development of a clinical prediction rule for guiding treatment of a subgroup of patients with neck pain: Use of thoracic spine manipulation, exercise, and patient education. *Phys Ther* 2007;87:9–23.
10. Tseng YL, Wang WT, Chen WY, Hou TJ, Chen TC, Lieu FK. Predictors for the im-

mediate responders to cervical manipulation in patients with neck pain. *Man Ther* 2006;11:306–315.
11. Lesher JD, Sutlive TG, Miller GA, Chine NJ, Garber MB, Wainner RS. Development of a clinical prediction rule for classifying patients with patellofemoral pain syndrome who respond to patellar taping. *J Orthop Sports Phys Ther* 2006;36:854–866.
12. Currier LL, Froechlich PJ, Carow SD, et al. Development of a clinical prediction rule to identify patients with knee pain and clinical evidence of knee osteoarthritis who demonstrate a favorable short-term response to hip mobilization. *Phys Ther* 2007;87:1106–1119.
13. Reilly BM, Evans AT. Translating clinical research into clinical practice: Impact of using prediction rules to make decisions. *Ann Intern Med* 2006;144:201–209.
14. Childs JD, Cleland JA. Development and application of clinical prediction rules to improve decision-making in physical therapist practice. *Phys Ther* 2006;86:122–131.
15. Laupacis A, Sekar M, Stiell IG. Clinical prediction rules: A review and suggested modifications of methodological standards. *JAMA* 1997;277:488–494.
16. Knottnerus JA. Diagnostic prediction rules: Principles, requirements, and pitfalls. *Prim Care* 1995;22:341–363.
17. Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. *Ann Intern Med* 1999;130:515–524.
18. McConnochie KM, Roghmann KJ, Pasternack J. Developing clinical prediction rules and evaluating observational patterns using categorical clinical markers. *Med Decis Making* 1993;13:30–42.
19. Norman GR, Stratford P, Regehr G. Methodological problems in the retrospective computation of responsiveness to change: The lesson of Cronbach. *J Clin Epidemiol* 1997;50:869–879.
20. Schmitt JC, Di Fabio RP. Reliable change and minimum important difference (MID) proportions facilitated group responsiveness comparisons using individual threshold criteria. *J Clin Epidemiol* 2004;57:1008–1018.
21. Schmitt JC, Di Fabio RP. The validity of prospective and retrospective global change criterion measures. *Arch Phys Med Rehabil* 2005;86:2270–2276.
22. Whiting P, Rutjes AV, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: A tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 2003;10(3):25.
23. Klein JG. Five pitfalls in decisions about diagnosis and prescribing. *BMJ* 2005;330:781–783.
24. Concato J, Feinstein AR, Holford TR. The risk of determining risk with multivariate methods. *Ann Intern Med* 1993;118:201–210.
25. Vergouwe Y, Steyerberg EW, Eijkemans MS, Habbema J. Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. *J Clin Epidemiol* 2005;58:475–483.
26. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, NJ: Erlbaum, 1988.
27. Jaeschke R, Guyatt GH, Sackett DL. Users guide to the medical literature. III. How to use an article about a diagnostic test. What are the results and will they help me? *JAMA* 1994;271:703–707.
28. McGinn TG, Guyatt GH, Wyer PC, et al. Evidence-based medicine working group. User's guides to the medical literature. XXII. How to use articles about clinical decision rules. *JAMA* 2000;284:79–84.