

CHAPTER
1**INTRODUCTION**

| Chapter Outline | Page |
|---|-------------|
| Definition | 2 |
| CPR Development | 3 |
| Derivation | 3 |
| Validation | 5 |
| Impact Analysis | 6 |
| Summary | 7 |
| Common Methodological Shortcomings of CPR Derivation Studies | 7 |
| Study Design | 7 |
| Sample | 8 |
| Variable Selection | 8 |
| Blinding | 9 |
| Outcome Measurement | 9 |
| Statistical Reporting | 10 |
| Reliability of Predictor Tests and Measures | 10 |
| CPR Quality Assessment | 11 |

The field of physical therapy (PT) has been undergoing a paradigm shift recently as the importance of evidence in practice is realized. Evidence-based medicine has been defined as the conscientious, explicit, and judicious use of the best available evidence.¹ Effective use of the current evidence requires the clinician to draw on clinical experience and assess patient values as well as to collect, analyze, and implement the available high-quality research. The process of incorporating evidence

into physical therapy practice can be quite challenging as the literature continues to grow and evolve at a seemingly insurmountable pace. To further complicate decision making, the interventions used in high-quality research are often controlled to such an extent that their results may not be realistically applied to individual patients in everyday practice. Given these challenges, the physical therapist would benefit from tools specifically aimed at facilitating evidence-based diagnostic, prognostic, and interventional decision making. One such tool has existed for decades within the medical literature but has only recently begun to appear in the PT research; it is called the clinical prediction rule.

Definition

Clinical prediction rules (CPRs) are algorithmic decision tools designed to aid clinicians in determining a diagnosis, prognosis, or likely response to an intervention. They use a parsimonious set of clinical findings from the history, physical examination, and diagnostic test results that have been analyzed and found to be statistically meaningful predictors of a condition or outcome of interest.^{2–10} One example is a CPR by Flynn et al.¹¹ to identify patients with acute low back pain (LBP) who will benefit from lumbar manipulation. The authors found five variables that were associated with success from manipulation: duration of symptoms less than 16 days, symptoms do not extend below the knee, hypomobility in at least one lumbar spine segment, either hip with internal rotation greater than 35°, and a Fear–Avoidance Beliefs Questionnaire Work subscale score of less than 19. The authors found that patients who possess four or more of the above variables are likely to benefit from lumbar manipulation.

CPRs use statistical models to identify the complex interaction of predictive variables in clinical practice. They are in turn able to help guide the clinician and provide him or her with a more efficient way to accurately subgroup patients while also reducing potential biases.^{4–6,9} CPRs are therefore highly useful in situations where decision making is complex due to heterogeneous conditions (e.g., LBP), the clinical stakes are high (e.g., deep vein thrombosis), or there is an opportunity for cost savings without compromising the quality of care (e.g., ordering of radiographs).^{7,12}

There are three distinct types of CPRs: diagnostic, prognostic, and interventional. Diagnostic CPRs (DCPRs) help clinicians determine the probability that a patient has a particular condition. Prognostic CPRs (PCPRs) provide information about the likely outcome of patients with a specific condition, and interventional CPRs (ICPRs), also called prescriptive CPRs, aid clinicians in determining which patients are likely to respond favorably to an intervention or set of interventions.^{2–10}

CPR Development

Before initiating the development of a CPR the need for such a tool should be identified. The greatest need is believed to be in highly prevalent conditions characterized by diagnostic uncertainty, diagnostic heterogeneity, or high practice variability.¹²⁻¹⁴ The process of preplanning for CPR development aids the researcher in identifying areas where an evidence base for clinical decision making is lacking, thus potentially improving the clinical utility of the tool.

CPRs exist along a continuum of quality and validity for clinical use. According to McGinn et al,⁷ progress along this continuum involves a three-step process: derivation, validation, and impact analysis. CPRs can also be graded on a scale from I to IV depending on where they are on the developmental continuum with Level I indicating the highest stage of development and Level IV indicating the lowest. Those that have been derived but have yet to undergo validation or impact analysis are considered Level IV.^{3,4,7,9} The progression from a Level IV to a Level III CPR includes validation of the original variable set in a “narrow” population (i.e., similar to the derivation study). This often involves examining the predictive value of the CPR in a separate population within the original derivation study. Due to the similar nature of the two samples, these CPRs still demonstrate limited generalizability into PT practice. CPRs that have been validated in “broad” populations (i.e., larger, more diverse patient samples) are considered Level II prediction rules. These rules can be applied confidently in clinical care as they have demonstrated stability in their predictive ability across diverse populations. Lastly, CPRs that have been implemented on a large scale and have demonstrated the ability to affect the quality and/or economy of care are considered Level I.^{3,4,7,9} To help the clinician better understand the development process a brief description of each step is included.

Derivation

Although many research designs may be utilized to derive a CPR (e.g., randomized clinical trials, retrospective analyses, cross-sectional analyses), currently the prospective cohort design is most frequently used in the PT literature.⁹ In such a design, a group of patients are selected based upon specific inclusion and exclusion criteria. All patients undergo a standardized examination, receive a reference standard test (DCPRs), specific treatment (ICPRs), or wait a specified period of time (PCPRs) and finally undergo an outcomes assessment. With this study design there is no comparison or control group.⁹

The derivation process begins by selecting the target condition through the preplanning process described earlier (e.g., conditions lacking evidence, treatment variability).^{4,7,9} Once the target condition has been chosen, the outcome of interest must

be clearly defined so the clinician may accurately frame his or her expected outcome and the most appropriate measures can be selected.^{3–10} For DCPRs this means using the most reliable and valid reference standard available to ensure that the condition is properly identified. In the case of ICPRs and PCPRs, it is important to use outcome measures with strong psychometric properties. The outcome tools should possess a recognized acceptable level of measurement error to help confirm statistically meaningful change; this is known as the minimal detectable change (MDC).¹⁵ They should also possess a difference score determined to be clinically meaningful to the patient, which is referred to as the minimal clinically important difference or (MCID).¹⁶ The use of such statistically sound measures helps to improve the likelihood that recorded changes are significant, meaningful, and are not due to chance.

Once the reference standard and/or outcome measures have been selected, the set of potential predictor variables must be established. These are all of the items that will be examined during the study for their relationship to the desired outcome. Potential predictor variables may include items from the history, physical exam, and self-report instruments (including psychosocial factors). The researchers must find a balance between choosing a select list of potential predictor variables, since the data will likely be collected under the time constraints typical of a busy clinical setting, while still including all relevant variables that might improve the predictive ability of the CPR.^{5,7,9} For this reason reliable and valid variables are often identified through prior research studies. However, if such research is lacking, a larger set of predictor variables may be chosen based upon clinical experience and expert opinion. This comprehensive approach helps to minimize the likelihood that a potential predictor variable is overlooked.

After undergoing a standardized historical and physical examination, the patients receive either the reference standard test (DCPRs), the treatment under investigation (ICPRs), or they wait a predetermined amount of time (PCPRs). The outcome(s) for each subject is (are) then determined. In most CPRs the outcome is dichotomized into two distinct possibilities. For DCPRs, the reference standard test is used to determine the presence or absence of the target condition. In the case of both ICPRs and PCPRs the outcome measure is used to determine whether the patient is either “successful” or “nonsuccessful” or whether symptoms are persistent or nonpersistent.

To determine successful from nonsuccessful, the investigators will establish the magnitude of change on the measure that they believe represents a true improvement, or a true lack of improvement, in the patient’s status. Frequently, the MCID for the outcome measure is used to distinguish between groups.¹⁵ An example of this would be the CPR to identify patients with acute LBP who respond favorably to manipulation. The investigators chose a 50% improvement on the modified Oswestry Disability Index to indicate a significant change in the patient’s disability

after manipulation.¹¹ This represents a fourfold increase over the commonly accepted MCID of 6 points or 12% for this scale.¹⁷

Once patients have been dichotomized, the categories (e.g., success vs. nonsuccess) are used as the reference standard to compare the individual predictor variables to determine univariate (individual) associations.^{5,9} Predictors with statistically significant univariate associations are retained for further analysis. The variables are then entered into a multivariate analysis to determine their contribution to a group of significant predictors. This process is important as occasionally variables with strong individual significance do not contribute to the greater accuracy of a group of predictors. An example of this can be found in the CPR by Flynn et al.¹¹ In this study, 11 individual variables met the proposed threshold for significance of $p < 0.15$. Of the 11 variables only five were retained in the final model after multivariate analysis. The final five variables were not necessarily the most significant predictors as indicated by the initial univariate analysis; however, as a group of variables they produced the most significant level of prediction. Thus, the final output of the multivariate analysis is a minimal set of predictor variables that contribute maximal predictive value for the outcome of interest. It is this final set of select predictor variables that comprise the derivation-level CPR.

Although the majority of PT CPRs are derived through prospective cohort studies, there are inherent limitations to this approach with the most important one being the lack of a control group. Without a control group the true treatment effect cannot be established; therefore, although responders are identified, the clinician still does not know if the intervention(s) applied is (are) the most effective for that subgroup. For this reason researchers have recommended the randomized clinical trial (RCT) as the approach of choice due to its controlled manner, ability to reduce bias, and ability to identify treatment effect modifying variables.¹⁸ Regardless of the study design, all agree that the derived CPR should undergo a validation process to confirm the predictive ability of the final variable set.³⁻¹⁰

Validation

As indicated by McGinn et al,⁸ the next step in the development of a CPR involves validation of the predictor variables in a narrow and/or broad patient sample. The primary purpose of this step is to confirm that the original predictor variables are neither due to chance or study design, nor are they specific to the patients or setting utilized in the derivation study.³⁻⁹ To accomplish these goals, the validation studies use a new cohort of therapists; a new patient sample derived using the original inclusion/exclusion criteria; and a different treatment setting. Although validation studies utilize various designs, they are most commonly prospective cohort studies or randomized clinical trials. Medical prediction rules

such as those that produce risk scores (e.g., deep vein thrombosis, pulmonary embolism) frequently validate the original variables through prospective cohort designs with a different patient population or various combinations of additional diagnostic tests in an attempt to further increase the diagnostic accuracy.^{7,8} Validation using a randomized clinical trial is common in the PT literature where interventions are tested and compared either to no intervention or other competing interventions.⁹ In this scenario the patient's status on the CPR (i.e., met or not met) is determined prior to randomization. Patients are then randomized using a block design to receive treatment that either matches or does not match the treatment that would be recommended by the results of the CPR. A block design is used to ensure that equal amounts of patients who are positive and negative on the CPR are in both groups to allow for a valid comparison. The subjects then undergo the treatment they were assigned to and their outcome is collected. Once the outcome has been determined, the results are analyzed to determine whether individuals who received a treatment that matched their status on the CPR had superior results compared to those who received unmatched treatment.⁹ Randomization also allows for implementation of a competing intervention and in so doing allows the researchers to assess the interaction of the CPR with patients receiving an alternative treatment. Regardless of the approach to validation, this is an extremely important step in the evolutionary process as it helps to improve the clinician's confidence that the results are reproducible and applicable to a larger more diverse patient population.

Impact Analysis

The third and final step in the CPR development process is the determination of the impact the rule has on clinical practice.³⁻⁹ Reilly and colleagues have suggested the ideal study design for such an analysis would be a randomized clinical trial where the randomization occurs between study sites rather than between treating clinicians.³ Site randomization enhances the prospects of the rule becoming part of the site's standard operating procedure thus increasing the likelihood of implementation. Upon analysis the researchers should consider the impact of the rule on patient care, its accuracy with or without real-life modifications, and the safety and efficiency of its use. By assessing the multifaceted impact of the rule, the researchers will then be able to determine whether the CPR will truly affect clinical decision making. If the rule is found to affect decision making, it is considered a Level I CPR and may then be referred to as a "clinical decision rule." Practice change is a multifaceted process and can occur for varied reasons thus making the identification of the rule's true impact difficult. Implementation can have many barriers including economic, a lack of resources, rule complexity and its associated cognitive burden,

medicolegal fear, and clinician preferences. For this reason, very few prediction rules achieve a Level I status.^{3,7,9}

Summary

Ideally, clinical prediction rules that are routinely used in practice will have undergone full development from derivation to broad validation with a subsequent impact analysis. Some have suggested that rules that have been derived but lack validation (Level IV) are not appropriate for clinical use due to the potential for chance variables and inaccurate findings.⁸ In the case of medical prediction rules, the importance of diagnostic accuracy is vital, as the risk of missing a condition may have dire consequences. For this reason, this textbook only includes medical screening CPRs that have undergone broad-based validation (Levels I and II). With regard to PT rules, the risk–benefit ratio is such that if the CPR is followed properly and inclusion/exclusion criteria are applied, the physical therapist is typically dealing with a low-risk intervention and high levels of potential benefit.⁹ For this reason, Level IV PT CPRs are included within this text. Although these CPRs may have limited generalizability, sound methodology can allow for the judicious use of their findings as a component of best evidence to guide decision making, particularly in areas where little research exists. It is therefore recommended that clinicians critically analyze derivation-level CPRs to confirm quality and individual patient applicability before implementing them into clinical care.

Common Methodological Shortcomings of CPR Derivation Studies

Study Design

The ideal study design to derive DCPRs is the cross-sectional design as it analyzes the effect of variables at one point in time. Longitudinal cohort studies have been proposed as the optimal means to determine factors that may influence the prognosis of an individual with a particular condition over a period of time (PCPR). The goal of an interventional CPR is to identify factors that may influence the treatment effect of an intervention on a subgroup of patients. For this reason, an experimental design such as an RCT has been suggested as the ideal approach to derive these rules. Researchers have indicated that the proper identification of predictor variables for a subgroup of patients can only be achieved through the simultaneous investigation in both an experimental group and a control group.¹⁸ Due to the high cost of RCTs and the large number of subjects required to achieve statistically meaningful results when examining several potential predictor variables, researchers frequently employ cohort-based designs to derive ICPRs. Although this study

design does not possess the methodological rigor of the RCT, it begins to identify potential predictor variables and assists in hypothesis generation, making future RCT validation studies more manageable.¹⁸

Sample

The astute clinician should begin by analyzing the patient sample to determine whether it is representative of the typical population that would receive the diagnostic test or therapeutic intervention.^{8,10} The size of the sample should be adequate to consistently identify predictor variables while also considering the risk of missing an outcome.⁹ For this reason researchers have suggested that at least 10–15 patients per potential predictor variable be included in the study.¹⁹ In practice, the determination of the sample size frequently depends on the potential for harm should the patient be misclassified. In situations where the consequences of a missed outcome may be detrimental, the sample size should be large. This is well demonstrated in the medical literature as many CPRs have sample sizes in the thousands.^{9,20–25}

Unlike medical diagnostic studies, PT studies generally involve very low risk, thus a lesser degree of precision is acceptable and fewer subjects need to be recruited. Commonly 10–15 subjects per predictor variable in the final CPR model is followed as a recruitment standard.^{9,19} Using this suggestion, the recommended recruitment for the Flynn et al.⁵ CPR would be 50 subjects as the final CPR contained five variables. Recently a systematic review of PT CPRs has recommended increasing the number of subjects recruited from 10 to 15 per predictor variable in the final model to 10 to 15 subjects per variable entered into multivariate analysis.²⁶ In this instance Flynn et al.⁵ would have a recommended recruitment of 110 subjects as 11 variables were entered into the multivariate analysis. Given that CPRs derived by physical therapists are commonly underpowered, we believe that this recommendation could strengthen the methodological quality of future derivation studies while still maintaining a manageable level of subject recruitment.

Variable Selection

Beyond sampling, another methodological shortcoming that should be considered includes deriving a prediction rule from a small number of potential predictor variables. It is important to include all the potential predictor variables that may influence the derived rule.^{5,9} Variables should be included from the history and physical examination, as well as self-report measures and diagnostic tests. With this being said, it is not always realistic to include a large number of potential predictors as the time required for clinical examination may limit subject participation. For this reason, many initial predictor sets are derived from prior studies, which have identified

a prognostic, interventional, or diagnostic link between the variable and the condition or intervention of interest. It is therefore recommended that CPR derivation studies use available evidence to guide and justify the size of the initial set of potential predictor variables.

Blinding

The importance of blinding is well recognized in other forms of research and the same is true for CPRs. Ideally, the examination will be performed prior to testing on the reference criterion or outcome measures.^{3-6,8,10} If the outcome results are not yet known, the examining clinician is inherently blinded; however, if the outcome is known before the exam, it is necessary for the researchers to blind the clinicians from the results to ensure they are not biased. In cases where the outcome is collected after the examination, it is also important to blind the clinician who collects the outcome from the results of the examination to eliminate bias. For this reason, separate clinicians should perform the examination and the intervention. Often in PT research it is very difficult to conceal the intervention from the patient; thus patients are generally aware of the intervention they are receiving, making a double-blinded design rare.

Outcome Measurement

As mentioned earlier, the outcome measures used to identify predictor variables must possess sufficient validity, reliability, and responsiveness.³⁻¹⁰ Since the results of the outcome measure are used to determine the “true” outcome for the patient, the validity of the measure is especially important. One concern along these lines is that the determination of change in many of the interventional CPRs is based upon retrospective self-report measures such as Global Rating of Change (GROC) scales. The most commonly utilized GROC scale in the PT literature has been described by Jaeschke et al.¹⁶ It is comprised of a 15-point rating scale ranging from 0 (“about the same”) to +7 (“a very great deal better”) or -7 (“a very great deal worse”). Other scales do exist, but they have not been frequently utilized in the derivation of PT CPRs.^{27,28} Regardless of which global rating scale is used, it should be recognized that they all rely on the patient to determine his or her level of change over a set period of time. Schmitt et al.²⁹ have demonstrated the potential for recall bias with these measures, as patients may have difficulty recalling their initial status over greater lengths of time. The results, in turn, become a representation of the patient’s current status rather than his or her change status. Schmitt et al. therefore suggest caution in utilizing the results of GROC scales as long-term (≥ 4 weeks) outcome measures.²⁹ They also recommend utilizing multi-item questionnaires to assess functioning across multiple tasks and constructs thus avoiding the potential influence of

one difficult task on the patient's perceived level of improvement. Examples of such questionnaires in the PT literature include but are not limited to the Neck Disability Index (NDI),³⁰ modified Oswestry Disability Index (ODI),¹⁷ Disabilities of the Arm, Shoulder, and Hand (DASH),³¹ Lower Extremity Functional Scale (LEFS),³² and the Patient Specific Functional Status (PSFS).³³

Statistical Reporting

Insufficient statistical reporting is another potential flaw found in many CPRs. The most commonly used statistical analysis to derive CPRs is logistic regression. Logistic regressions compare many tests and measures to a dichotomous outcome and determine a parsimonious group of variables that best predict the outcome.⁶⁻⁹ However, regression analyses can provide other information that is frequently not reported in CPR studies. The statistical significance (expressed as the p-value) of the model derived from the regression analysis can provide readers with the probability that the selected group of variables came about by chance alone. The omission of this statistic precludes the reader from making an accurate judgment on how much he or she should trust the results. Another statistic generated is the R^2 , or how much variance in the outcome measure the predictor variables account for. This is important as a small R^2 value (i.e., they predict a small portion of the variance in the outcome measure) suggests that there may be other variables that could also contribute to prediction of the outcome. Both of these statistics, probability (p) and R^2 , should be included to allow the physical therapist to more fully and accurately assess the strength of the CPR being considered.¹⁰

Reliability of Predictor Tests and Measures

Lastly, accurate determination of change requires not only a blinded evaluator and a valid and responsive outcome measure/reference criterion but also an acceptable level of inter-rater reliability among the tests and measures.¹⁰ Similar to the determination of sample size, the level of acceptable reliability may vary depending upon the CPR and the condition being studied. For example, the reliability of the tests and measures required to diagnose a potentially fatal pulmonary embolism must be very high. On the contrary, the reliability of a measure to determine who will benefit from lumbar stabilization can be lower as the potential for harmful consequences is much less. Reliability therefore lies on a continuum; however, it has been suggested that an acceptable kappa value is ≥ 0.60 , and ≥ 0.70 is considered an acceptable intra-class correlation coefficient.¹⁰ Reliability statistics should be reported within the body of the CPR. If prior reliability studies do not exist, researchers should perform an internal reliability study to confirm adequate levels among the clinicians involved. A review of the reliability statistics is available in Chapter 3.

CPR Quality Assessment

The aforementioned list of common methodological shortcomings comprises the overall quality of the CPR study. In determining whether a specific CPR is appropriate and relevant to a particular patient, the clinician must consider the quality of the study's methodology. One might expect an increased level of clinical confidence in the findings of studies that demonstrate high methodological quality and thus consider such CPRs appropriate for implementation into clinical practice. For this reason, the authors have chosen to analyze the quality of derivation-level CPRs to assist the readers in their decision as to whether the rule is appropriate for practice prior to validation. A flow diagram outlining the overall decision process has also been provided in Appendix E.

The first assessment measure that may be used to retrospectively analyze the quality of derivation-level *prognostic* studies was developed by Kuijpers et al.³⁴ It consists of 18 criteria assessing quality in seven categories: study population, response information, follow-up, treatment, outcome, prognostic factors, and data presentation. Each item is scored as positive, negative, or unclear, with positive scores receiving 1 point, and negative or unclear scores receiving 0 points. The positive scores are added, divided by the total possible points (18), and multiplied by 100 to yield a percentage. Kuijpers et al. recommend an arbitrary cut-off score of at least 60% to indicate a “high-quality” study, and a score of less than 60% to represent a “low-quality” study³⁴ (**Figure 1.1**).

Recently the tool utilized by Kuijpers et al.³⁴ has been modified to accommodate analysis of derivation-level *interventional* CPRs. Beneciuk et al.²⁶ have removed the question regarding response rate and have added an eighth category concerning masking of the outcome assessors and the treating clinicians. The authors noted an intraclass correlation coefficient (ICC) of .73 (95% CI .27–.91), indicating moderate to good inter-rater reliability with this tool.²⁶ The authors again recommend an arbitrary cut-off score of 60% to indicate a “high-” vs. “low-” quality study (**Figure 1.2**).

Currently, a quality assessment tool for derivation-level *diagnostic* CPRs does not exist, and therefore we have not assigned a quality score to these rules. To help the reader assess the quality of Level IV DCPRs, we have formulated a list of items adapted from prior quality assessment tools, which we believe address many of the elements necessary for a well-designed diagnostic derivation study. The list is intended merely as an assessment guide as it has not undergone the rigors of peer review or a Delphi process for its construction. It is recommended that the clinician critically analyze these CPRs using his or her knowledge of evidence-based practice as well as our table as a guide to determine whether the application of the DCPR is appropriate to his or her specific patient (**Table 1.1**).

Figure 1.1

Criteria for assessing the methodological quality of prognostic studies.

- A. Positive if patients were identified at a uniform point (inception cohort) in the course of their disease (first episode, with restriction to duration of symptoms, of shoulder pain in lifetime, or first treated episode of shoulder pain).
- B. Positive if criteria were formulated for at least: age, duration of symptoms, relevant comorbidity (i.e., cervical radiculopathy, luxation)/systemic diseases.
- C. Positive if it was described in what setting the patients were recruited (i.e., general practice, hospital, occupational setting).
- D. Positive if the response rate was $\geq 75\%$.
- E. Positive if information was presented about patient/disease characteristics of responders and non-responders or if there was no selective response.
- F. Positive if a prospective design was used, also positive in case of an historical cohort in which the determinants had been measured before outcome was determined.
- G. Positive if the follow-up period was at least 6 months.
- H. Positive if the total number of participants was $\geq 80\%$ on the last moment of follow-up compared to the number of participants at baseline.
- I. Positive if demographic/clinical information (patient/disease characteristics such as age, sex, and other potential prognostic predictors) was presented for completers and those lost to follow-up/dropouts at the main moment of outcome measurement, or no selective dropouts/lost to follow-up, or no dropouts/lost to follow-up.
- J. Positive if treatment subsequent to inclusion in cohort is fully described or standardized. Also positive in case of no treatment given.
- K. Positive if standardized questionnaires of objective measurements of at least 1 of the following 5 outcome measures were used for each follow-up measurement: pain, general improvement, functional status, general health status, or lost days of work.
- L. Positive if standardized questionnaires or objective measurements were used at baseline for at least 4 of the following 8 potential prognostic factors: age, sex, pain, functional status, duration of complaints, neck complaints, physical workload, or dominant shoulder affected.
- M. Positive if standardized questionnaires or objective measurements were used at baseline of at least 1 of the following 6 potential prognostic factors: depression, somatisation, distress, fear and avoidance, coping strategies, or psychosocial work-related factors (i.e., social support, psychological demands, job decision latitude).
- N. Positive if frequency, percentage or mean, median (Inter Quartile Range), and standard deviation/CI were reported for the most important outcome measures.
- O. Positive if frequency, percentage or mean, median (Inter Quartile Range), and standard deviation/CI were reported for the most important prognostic factors.
- P. Positive if univariate crude estimates were provided for the association of a prognostic factor with outcome.
- Q. Attempt is made to determine a set of prognostic factors with the highest prognostic value.
- R. Positive if the number of cases in the multivariate analysis was at least 10 times the number of independent variables in the analysis (Altman, 1991).

Source: Reprinted from Kuijpers T, van der Windt D, van der Heijden G, Bouter LM. Systematic review of prognostic cohort studies on shoulder disorders. *Pain*. 2005;109:429–430. Reprinted with permission from the International Association for the Study of Pain.

- A. Positive if subjects were identified at an early uniform point (inception cohort) in the course of the condition (first episode, with restriction to duration of symptoms mentioned, of their respective complaint or first physical therapy-related intervention episode for their respective complaint).
- B. Positive if criteria were formulated for at least age, duration of symptoms, and relevant comorbidities.
- C. Positive if setting in which subjects were treated was described.
- D. Positive if information was presented about subject or condition characteristics of responders or nonresponders or if there was no selective process.
- E. Positive if a prospective design was used (immediate or same-day follow-up was not considered prospective).
- F. Positive if the follow-up period was ≥ 6 months.
- G. Positive if the total number of subjects was $\geq 80\%$ at the last moment of the final follow-up compared with the number of subjects at baseline.
- H. Positive if demographic or clinical information (subject or condition characteristics, such as age, sex, and other potential prognostic predictors) was presented for subjects completing the study and those lost to follow-up/dropouts at the main moment of baseline outcome measurement, or no selective dropouts/lost to follow-up, or no dropouts/lost to follow-up.
- I. Positive if the intervention subsequent to inclusion in a cohort was fully described or standardized (treating clinicians had to adhere to a strict protocol and were not permitted to adjust the intervention on the basis of their independent decision-making processes).
- J. Positive if standardized questionnaires or quantitative measurements of at least 1 of the following 5 outcome measures were used for each follow-up measurement: pain, general improvement, functional status, general health status, or lost days of work.
- K. Positive if masking of the outcome assessor and treating clinician was achieved. In studies in which self-administered questionnaires were used, masking of the outcome assessor portion of this criterion would be considered acceptable but would have no bearing on the treating clinician status.
- L. Positive if standardized questionnaires or objective measurements were used at baseline for at least 4 of the following 6 potential prognostic factors: age, sex, pain, functional status, duration of complaints, or physical work load.
- M. Positive if standardized questionnaires or objective measurements were used at baseline for at least 1 of the following 7 potential prognostic factors: depression, somatization, distress, fear-avoidance, coping strategies, anxiety, or psychosocial work-related factors (social support, psychological demands, and job decision latitude).
- N. Positive if frequency, percentage, or mean, media, and standard deviation or confidence interval were reported for the most important outcome factors.
- O. Positive if frequency, percentage, or mean, media, and standard deviation or confidence interval were reported for the most important prognostic factors.
- P. Positive if univariate crude estimates were provided for the association of a prognostic factor with outcome.
- Q. Positive if an attempt was made to determine a set of prognostic factors with the highest prognostic value.
- R. Positive if the number of cases in the multivariate analysis was at least 10 times the number of independent variables in the multivariate analysis (on the basis of the final clinical prediction rule model, not the initial prospective variables).

Figure 1.2

Criteria for assessing the methodological quality of interventional studies.

Source: Reprinted from Beneciuk JM, Bishop MD, George SZ. Clinical prediction rules for physical therapy interventions: a systematic review. *Phys Ther.* 2009;89:10–11. Reprinted with permission of the American Physical Therapy Association. This material is copyrighted, and any further reproduction or distribution is prohibited.

Table 1.1 CPR Quality Assessment of Diagnostic Studies

| Item | Description | Yes | No |
|------|--|-----|----|
| 1 | Inception cohort | | |
| 2 | Prospective and consecutive subjective enrollment | | |
| 3 | Description of setting | | |
| 4 | Description of subject's baseline demographics | | |
| 5 | Clear inclusion/exclusion criteria | | |
| 6 | Recognized valid/reliable reference standard | | |
| 7 | Explanation for predictor variable selection | | |
| 8 | Reliable predictor variables (ICC \geq 0.70; Kappa \geq 0.60) | | |
| 9 | Prospective application of reference standard within a reasonable time frame after the examination | | |
| 10 | Detailed description of positive/negative on reference standard | | |
| 11 | Blinded examiner | | |
| 12 | Blinded interpretation of reference standard | | |
| 13 | Diagnostic accuracy of significant individual predictor variables reported | | |
| 14 | Variables exceeding set cut score for univariate significance entered into regression model | | |
| 15 | Results of regression analysis reported with 95% CIs | | |
| 16 | Statistical significance of the model reported | | |
| 17 | Full description of retained predictor variables presented | | |
| 18 | 10–15 subjects per variable presented in the final clinical prediction rule | | |
| 19 | Were study withdrawals/dropouts < 10% | | |
| | Total | | |

Validation of a CPR indicates a higher level of development, and this step is recommended prior to routine implementation of the rule into clinical care. Despite this step the clinician must apply the same critical analysis to these studies as one would to others under consideration for incorporation into patient care, as the mere presence of a validation study does not automatically make the CPR appropriate for use. May and Rosedale¹⁴ have recently adapted a set of proposed quality criteria to retrospectively evaluate *validation* studies of interventional CPRs. The quality measure consists of 10 criteria analyzing the methodological standards of the study. Each item is scored as met (1 point) or not met (0 points). The positive scores are added, divided by the total possible points (10), and multiplied by 100 to yield a percentage. The authors once again recommend an arbitrary cut-off score of 60% or greater to indicate a “high-quality” study, and a score of less than 60% to represent a “low-quality” study. This tool has not currently been validated, and, in our opinion, it requires further expansion to encompass the full quality of a PT-based validation study. However, we have chosen to utilize it as a component of this text in an

effort to provide the reader with a means of analyzing the existing PT literature on this topic (Figure 1.3).

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

Figure 1.3

Criteria for assessing the methodological quality of interventional validation studies.

Source: Adapted from May S, Rosedale R. Prescriptive clinical prediction rules in back pain research: a systematic review. *JMMT*. 2009;17:38. Courtesy of the Journal of Manual and Manipulative Therapy.

References

1. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *BMJ*. 1996;312:71–72.
2. 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;10:948–956.
3. Reilly BM, Evans AT. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. *Ann Int Med*. 2006;144:201–210.
4. Beattie P, Nelson R. Clinical prediction rules: what are they and what do they tell us? *Aust J Physiother*. 2006;52:157–173.
5. Randolph AG, Guyatt GH, Calvin JE, Doig G, Richardson WS. Understanding articles describing clinical prediction rules. Evidence based medicine in critical care group. *Crit Care Med*. 1998;26:1603–1612.
6. Wasson JH, Sox HC, Neff RK, Goldman L. Clinical prediction rules: application and methodological standards. *NEJM*. 1985;313:793–799.
7. McGinn TG, Guyatt GH, Wyer PC, Naylor C, Stiell IG, Richardson W. Users guide to medical literature: XXII. How to use articles about clinical decision rules. *JAMA*. 2000;284:79–84.
8. Laupacis A, Sekar N, Stiell IG. Clinical prediction rules and suggested modifications of methodological standards. *JAMA*. 1997;277:488–494.
9. Childs JD, Cleland JA. Development and application of clinical prediction rules to improve decision making in physical therapy practice. *Phys Ther*. 2006;86:122–131.
10. Cook CE. Potential pitfall of clinical prediction rules. *JMMT*. 2008;16:69–71.

11. Flynn T, Fritz J, Whitman J, et al. A clinical prediction rule for classifying patients with low back pain who demonstrate short-term improvement with spinal manipulation. *Spine*. 2002;27:2835–2843.
12. Fritz JM. Clinical prediction rules in physical therapy: coming of age? *J Orthop Sports Phys Ther*. 2009;39:159–161.
13. Stiell IG, Wells GA. Methodological standards for the development of clinical decision rules in emergency medicine. *Ann Emerg Med*. 1999;33:437–447.
14. May S, Rosedale R. Prescriptive clinical prediction rules in back pain research: a systematic review. *JMMT*. 2009;17:36–45.
15. Resnik L, Dobrzykowski E. Guide to outcome measurement for patients with low back syndrome. *J Orthop Sports Phys Ther*. 2003;33:307–318.
16. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials*. 1989;10:407–415.
17. Fritz JM, Irrgang JJ. A comparison of a modified Oswestry low back pain disability questionnaire and the Quebec back pain disability scale. *Phys Ther*. 2001;81:776–788.
18. Hancock M, Herbert RD, Maher CG. A guide to interpretation of studies investigating subgroups of responders to physical therapy interventions. *Phys Ther*. 2009;89:698–704.
19. Concato J, Feinstein AR, Holford TR. The risk of determining risk with multivariate methods. *Ann Int Med*. 1993;118:201–210.
20. Stiell IG, Greenberg GH, McKnight RD, Nair RC, McDowell I, Worthington JR. A study to develop clinical decision rules for the use of radiography in acute ankle injuries. *Ann Emerg Med*. 1992;21:384–390.
21. Stiell IG, Greenberg GH, Wells GA, et al. Derivation of a decision rule for the use of radiography in acute knee injuries. *Ann Emerg Med*. 1995;26:405–413.
22. Stiell IG, Wells GA, Vandemheen KL, et al. The Canadian C-spine rule for radiography in alert and stable trauma patients. *JAMA*. 2001;286:1841–1848.
23. Stiell IG, Wells GA, Vandemheen KL, et al. The Canadian CT head rule for patients with minor head injury. *Lancet*. 2001;357:1391–1396.
24. Wells PS, Anderson DR, Rodger M, et al. Derivation of a simple clinical model to categorize patient's probability of pulmonary embolism: increasing the models utility with the SimpliRed D-dimer. *Thromb Haemost*. 2000;83:416–420.
25. Mower WR, Hoffman JR, Herbert M, Wolfson AB, Pollack CV, Zucker MI. Developing a decision instrument to guide computed tomographic imaging of blunt head injury patients. *J Trauma*. 2005;59:954–959.
26. Beneciuk JM, Bishop MD, George SZ. Clinical prediction rules for physical therapy interventions: a systematic review. *Phys Ther*. 2009;89:114–124.
27. Dworkin RH, Turk DC, Farrar JT, et al. Core outcome measures for chronic pain clinical trials. IMMPACT recommendations. *Pain*. 2005;113:9–19.
28. Stewart M, Maher CG, Refshauge KM, Bogduk N, Nicholas M. Responsiveness of pain and disability measures for chronic whiplash. *Spine*. 2007;32:580–585.
29. Schmitt J, Di Fabio RP. The validity of prospective and retrospective global change criterion measures. *Arch Phys Med Rehabil*. 2005;86:2270–2276.
30. Vernon H, Mior S. The neck disability index: a study of reliability and validity. *J Manipulative Physiol Ther*. 1991;14:409–415.

31. Hudak PL, Amadio PC, Bombardier C. The Upper Extremity Collaborative Group (UECG). Development of an upper extremity outcome measure: the DASH (disabilities of the arm, shoulder and hand). *Am J Ind Med.* 1996;29:602–608.
32. Binkley JM, Stratford PW, Lott SA, Riddle DL. The Lower Extremity Functional Scale (LEFS): scale development, measurement properties, and clinical application. North American Orthopaedic Rehabilitation Research Network. *Phys Ther.* 1999;79:371–383.
33. Stratford P, Gill C, Westaway M, Binkley J. Assessing disability and change on individual patients: a report of a patient specific measure. *Physiotherapy Canada.* 1995;47:258–263.
34. Kuijpers T, van der Windt D, van der Heijden G, Bouter LM. Systematic review of prognostic cohort studies on shoulder disorders. *Pain.* 2004;109:420–431.

