Clinical Prediction Rules

Andrew Jull, Alba DiCenso, and Gordon Guyatt

The following Editorial Board members also made substantive contributions to this chapter: Rien de Vos and Mary van Soeren.

We gratefully acknowledge the work of Thomas McGinn, Peter Wyer, David Naylor, and Ian Stiell on the original chapter that appears in the Users’ Guides to the Medical Literature, edited by Guyatt and Rennie.

In This Chapter

Clinical Prediction Rules

Developing a Clinical Prediction Rule

Validation

Interpreting the Results

Testing the Rule’s Impact
This chapter outlines the criteria for deciding on the strength of inferences you can make about the accuracy and impact of clinical prediction rules. Examples of clinical prediction rules of interest to nurses include an antenatal index to predict postpartum depression, a risk assessment tool (STRATIFY) to predict which elderly inpatients will fall, a prediction rule to identify patients in whom venous leg ulcers will heal with limb compression bandages, and a risk index to predict, at the point of initial triage, mortality at 30 days in patients with ST-elevation myocardial infarction.

### Clinical Scenario

**Can a Clinical Prediction Rule Reduce Unnecessary Ankle Radiographs?**

You are a nurse educator responsible for the continuing nursing education program for nurse practitioners in outpost settings. You have been conducting a series on how to best use research evidence to guide practice. The nurse practitioners are interested in strategies that minimize the need to send patients out for consultations and diagnostic tests because of the associated inconvenience to patients and costs to the health care system. Some of the nurses are using clinical prediction guides to assist with this process, but they are not confident in their ability to identify high-quality clinical prediction guides. In planning your education session, you decide to identify a clinical prediction guide that is relevant to nurse practitioners and has been shown to be useful. From your own clinical practice, you are familiar with the Ottawa ankle rules (Figure 22-1), and you recently came across a systematic review on the accuracy of these rules in *BMJ*. You have a careful look at the review and are confident that it is methodologically sound. The review concludes that this clinical prediction rule is accurate for excluding fractures of the ankle and midfoot. The clinical prediction rule has a sensitivity of almost 100% and modest specificity. The authors estimated that use of the Ottawa ankle rules would reduce the number of unnecessary radiographs by 30% to 40%. You decide to search for the original studies on the development and validation of the Ottawa ankle rules to help illustrate the criteria for evaluating clinical prediction rules. In the systematic review, you find references to the original articles that described the development of the Ottawa ankle rules in 1992 and their validation.

This chapter outlines the criteria for deciding on the strength of inferences you can make about the accuracy and impact of clinical prediction rules. Examples of clinical prediction rules of interest to nurses include an antenatal index to predict postpartum depression, a risk assessment tool (STRATIFY) to predict which elderly inpatients will fall, a prediction rule to identify patients in whom venous leg ulcers will heal with limb compression bandages, and a risk index to predict, at the point of initial triage, mortality at 30 days in patients with ST-elevation myocardial infarction.

### Clinical Prediction Rules

Establishing a patient’s diagnosis and determining the prognosis are closely linked activities. Diagnoses and assessments of patients’ prognoses often determine the recommendations we make to patients. Clinical experience provides us with an intuitive sense of which findings on history, physical examination, and laboratory or radiologic investigation are critical in making an accurate diagnosis or assessment of a patient’s prognosis. Although intuition can often be extraordinarily accurate, it can also be misleading. Clinical prediction rules attempt to test, simplify, and increase the accuracy of clinicians’ diagnostic and prognostic assessments.

A **clinical prediction rule** can be defined as a clinical tool that quantifies the individual contributions of various components of the history, physical examination, and basic laboratory results toward the diagnosis, prognosis, or likely response to treatment of
individual patients. This definition also applies to what have been called clinical prediction guides and clinical decision rules.

*Prediction* implies helping clinicians to determine a future clinical event more accurately. *Decision* implies directing clinicians to a specific course of action. As you will see, application of clinical prediction rules sometimes results in a decision and other times in a prediction. It can also result in a probability or likelihood ratio (LR) that a clinician can apply to a current diagnostic problem. In this last application, the term *clinical diagnosis rule* or *clinical diagnosis guide* may be more accurate. We use the term *clinical prediction rule* regardless of whether the output of the rule is a suggested clinical course of action, the probability of a future event, or an increase or decrease in the likelihood of a particular diagnosis.

Whatever the clinical prediction rule generates—a decision, a prediction, or a change in diagnostic probability—it is most likely to be useful when decision making is complex, the clinical stakes are high, or opportunities exist to achieve cost savings without compromising patient care.

Developing and testing a clinical prediction rule involves three steps: (1) creation or derivation of the rule, (2) testing or validation of the rule, and (3) assessment of the impact of the rule on clinical behavior—the *impact analysis*. The validation process may require several studies to test the accuracy of the rule fully at different clinical sites.
(Figure 22-2). Each step in the development of a clinical prediction rule may be published separately by different authors, or all three steps may be included in one article. Table 22-1 presents a hierarchy of evidence that can guide clinicians in assessing the full range of evidence supporting the use of a clinical prediction rule in their practice. We now review the steps in the development and testing of a clinical prediction rule and relate each stage of the process to the hierarchy of evidence presented in Table 22-1.

**DEVELOPING A CLINICAL PREDICTION RULE**

The development of the Ottawa ankle rules was described in an article by Stiell and colleagues that was published in 1992. Developers of clinical prediction rules begin by constructing a list of potential predictors of the outcome of interest—in this case, ankle fractures demonstrated on ankle radiograph. The list typically includes items from the history, physical examination, and basic laboratory tests. The investigators then examine a group of patients and determine (1) whether the candidate clinical predictors are present and (2) each patient’s status on the outcome of interest—in this case, the result of the ankle radiograph. Statistical analysis reveals which predictors are most powerful and which predictors can be omitted from the rule without loss of predictive power. Typically, the statistical techniques used in this process are based on logistic regression (see Chapter 31, Regression and Correlation). Other techniques include discriminant analysis, which produces equations similar to regression analysis; recursive partitioning analysis, which divides the patient population into smaller and smaller groups based on discriminating risk factors; and artificial neural networks, which apply non-linear statistics to pattern recognition problems.

Clinical prediction rules that have been derived but not validated should not be considered ready for clinical application (see Table 22-1). Although these rules have predicted the outcome in one sample, it does not necessarily mean that they will predict the outcome in another sample or in the same sample when applied by clinicians in

---

**Step 1: Derivation**

**Step 2: Validation**

**Step 3: Impact Analysis**

**Evidence of reproducible accuracy**

**Evidence that rule changes clinician behavior and improves patient outcomes and/or reduces costs**

**Level of Evidence**

| IV | III | II | I |

*Figure 22-2.* Development and testing of a clinical prediction rule.
practice. However, investigators interested in validating a clinical prediction rule first need to determine whether the derivation process has been well done and whether the rule meets certain criteria. Table 22-2 summarizes criteria for determining whether the derivation process was well done. Interested readers can find a complete discussion of the derivation process and criteria for assessing the process in an article by Laupacis and colleagues.8

VALIDATION

There are three reasons that even rigorously derived clinical prediction rules are not ready for application in clinical practice without further validation. First, prediction rules derived from one set of patients may reflect associations between given predictors and outcomes that occur primarily because of the play of chance. If that is so, a different set of predictors will emerge in a different group of patients, even if these patients come from the same setting. Second, predictors may be idiosyncratic to the population, the clinicians using the rule, or other aspects of study design. If that is so, the rule may fail in different settings. Finally, because of problems with the feasibility of rule application in clinical settings, clinicians may fail to implement a rule comprehensively or accurately. The result would be that a rule succeeds in theory but fails in practice.
Statistical methods can deal with the first of these problems. For instance, investigators can split a population into two groups and use one to develop the rule and the other to test it. Alternatively, investigators can use more sophisticated statistical methods built on the same logic. Conceptually, these approaches involve removing one patient from the sample, generating the rule using the remainder of patients, and testing the rule on the patient who was removed from the sample. One repeats this procedure, sometimes referred to as a *bootstrap technique*, in sequence for every patient under study.

Although statistical validations within the same setting or group of patients reduce the likelihood that the rule reflects the play of chance rather than true associations, they fail to address the other two threats to validity. The success of a clinical prediction rule may be peculiar to the particular populations of patients and clinicians involved in the derivation study. Even if this is not so, clinicians may have difficulties using the rule in practice, compromising its predictive power. Thus, to ascend from level IV in our hierarchy of evidence, studies must test the use of the rule by clinicians in clinical practice.

A clinical prediction rule developed to predict serious outcomes (e.g., heart failure or ventricular arrhythmia) in patients with syncope highlights the importance of validation. Investigators derived the rule using data from 252 patients who presented to an emergency department; subsequently, the investigators attempted to validate it prospectively in a sample of 374 patients. The prediction rule assigned patients a score from 0 to 4, depending on the number of clinical predictors present. The probability of poor outcomes corresponding to almost every score in the derivation set was approximately twice that of the validation set. For example, in the derivation set, the risk of a poor outcome in a patient with a score of 3 was estimated to be 52%. By contrast, a patient with the same score in the validation set had a probability of a poor outcome of only 27%. This variation in results may have been caused by differences in the severity of syncope cases in the two studies or differences in criteria for generating a score of 3. Because there is a risk that a clinical prediction rule will provide misleading information when applied in real-world clinical settings, rules that have been developed but not validated are designated as level IV in our hierarchy (see Table 22-1).

### Table 22-2  Methodological Standards for Derivation of a Clinical Prediction Rule

- Were all important predictors included in the derivation process?
- Were all important predictors present in a significant proportion of the study population?
- Were the outcome event and predictors clearly defined?
- Were those assessing the outcome event blinded to the presence of the predictors, and were those assessing the presence of predictors blinded to the outcome event?
- Was the sample size adequate (including an adequate number of outcome events)?
- Does the rule make clinical sense?
Despite this major limitation, clinicians can still extract clinically relevant messages from an article describing the development of a clinical prediction rule. Clinicians can identify the most important predictors and consider them more carefully in their own practice. They can also consider giving less importance to variables that failed to show predictive power. For instance, in developing a clinical prediction rule to predict mortality from pneumonia, investigators found that the white blood cell count had no bearing on subsequent mortality. Hence, clinicians could put less weight on the white blood cell count when they make decisions about admitting patients with pneumonia to hospital.

To move up the hierarchy, clinical prediction rules must provide additional evidence of validity. After developing the Ottawa ankle rules, Stiell and colleagues refined and prospectively validated them. Validation of a clinical prediction rule involves demonstrating that its repeated application as part of the process of clinical care leads to the same results. Ideally, validation entails application of the rule prospectively in a new population with a prevalence and spectrum of disease that differ from those of patients in the derivation set. It is important to be sure that a clinical prediction rule performs similarly in different populations and with different clinicians who work in numerous institutions. It is also important to be sure that it works well when clinicians consciously apply it as a rule, rather than as a statistical derivation from a large number of potential predictors.

If the setting in which a prediction rule was originally developed was limited and its validation was confined to this setting, application by clinicians in other settings will be less secure. Validation in a similar setting can take several forms. Most simply, after developing the prediction rule, the investigators return to their population, identify a new sample of patients, and test the rule’s performance. Thus, we classify rules that have been validated in the same—or very similar—limited or narrow population as was used in the development phase as level III on our hierarchy, and we recommend that clinicians use the result cautiously (see Table 22-1). If patients in the derivation set are from a sufficiently heterogeneous population across various institutions, testing the rule in the same population provides strong validation. Validation in a new population provides clinicians with strong inferences about the usefulness of the rule, corresponding to level II in our hierarchy (see Table 22-1). The more numerous and diverse the settings in which the rule is tested and found to be accurate, the more likely it is that it will generalize to untested settings.

The Ottawa ankle rules were derived in two large, university-based emergency departments in Ottawa, Canada and then prospectively validated in a large sample of patients from the same emergency departments. At this stage, the rules would be classified as level II in our hierarchy because of the large number and diversity of patients and physicians involved in the study. Since that initial validation, the rules have been validated in many different clinical settings, with relatively consistent results. This evidence further strengthens our inference about its predictive power.

Many clinical prediction rules are derived and then validated in a small, narrowly selected group of patients (level III). One such rule was derived to predict preserved left ventricular function after myocardial infarction. The initial derivation and validation were performed on 314 patients who had been admitted to a tertiary care center. The prediction rule was derived using 162 patients and then validated using 152 patients in the same setting. When the rule predicted that left ventricular function had been
preserved, this was, in fact, true in 99% of patients. At this stage in the rule development, the rule would be considered to be level III, to be used only in settings similar to those of the validation study (i.e., in similar cardiac care units). The rule was further validated in two larger trials, one involving 213 patients\textsuperscript{16} from one site and a larger trial involving 1891 patients from several different institutions.\textsuperscript{17} In both settings, of patients predicted to have preserved left ventricular function, 11% actually had abnormal left ventricular function. This drop in accuracy changes the potential use and implications of the rule in clinical practice. At this point in development, the rule would be designated as level II, meaning that the rule could be used in clinical settings with a high degree of confidence, but with adjusted results. This example highlights the importance of validating a clinical prediction rule on a diverse patient population before applying it broadly.

Regardless of whether investigators validated a rule in a similar, narrow (level III) population or a broad, heterogeneous, or different (level II) population, the results allow stronger inferences if the investigators adhered to certain methodological standards (Table 22-3). Interested readers can find a complete discussion of the validation process and criteria for assessing this process in an article by Laupacis and colleagues.\textsuperscript{8}

If investigators evaluating predictor status of study patients are aware of the outcome, or if those assessing the outcome are aware of patients’ status with respect to predictors, the assessments may be biased. For instance, in a clinical prediction rule developed to predict the presence of pneumonia in patients presenting with cough,\textsuperscript{18} the authors did not mention blinding during either the derivation process or the validation process. Knowledge of history or physical examination findings could have influenced the judgments of the unblinded radiologists.

The investigators testing the Ottawa ankle rules enrolled consecutive patients, obtained radiographs for all of them, and ensured that clinicians assessing the clinical predictors were unaware of the radiologic results and that radiologists assessing ankle fractures had no knowledge of the clinical data.

**INTERPRETING THE RESULTS**

Regardless of the level of evidence associated with a clinical prediction rule, its usefulness depends on its predictive power. Investigators may report their results in various ways. First, the results may dictate a specific course of action. The ankle component of the

<table>
<thead>
<tr>
<th>Table 22-3  Methodological Standards for Validation of a Clinical Prediction Rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Were the patients chosen in an unbiased fashion, and do they represent a wide spectrum of severity of disease?</td>
</tr>
<tr>
<td>• Was there a blinded assessment of the criterion standard for all patients?</td>
</tr>
<tr>
<td>• Was there an explicit and accurate interpretation of the predictor variables and the actual rule without knowledge of the outcome?</td>
</tr>
<tr>
<td>• Was there 100% follow-up of those enrolled?</td>
</tr>
</tbody>
</table>
Ottawa ankle rules states that an ankle series of radiographs is indicated only for patients with pain near the malleoli plus either localized bone tenderness at the posterior edge or tip of either malleolus or inability to bear weight (see Figure 22-1). Underlying this decision are the sensitivity and specificity of the rule as a diagnostic test (see Chapter 6, Diagnosis). In the development process, all patients with fractures had a positive result (sensitivity of 100%), but only 40% of those without fractures had a negative result (specificity of 40%). These results suggest that if clinicians order radiographs only for patients with a positive result, they will not miss any fractures and will avoid the test in 40% of patients without a fracture.

The validation study confirmed these results; in particular, the test maintained a sensitivity of 100%. This is particularly reassuring because the sample size was sufficiently large to result in a narrow confidence interval (CI) around the estimate of sensitivity (95% CI, 93% to 100%). Thus, clinicians adopting the rule would miss very few, if any, fractures.

Another way of reporting clinical prediction rule results is in terms of probability of the target condition being present given a particular result. For example, a prediction rule for pulmonary embolus derived and validated by Wells and colleagues accurately placed patients into low (3.4%; 95% CI, 2.2% to 5%), intermediate (28%; 95% CI, 23.4% to 32.2%), or high (78%; 95% CI, 69.2% to 89.6%) probability categories. When investigators report prediction rule results in this fashion, they are implicitly incorporating all clinical information. In doing so, they remove the need for clinicians to consider independent information when deciding about the likelihood of a diagnosis or about a patient’s prognosis.

Finally, prediction rules may report their results as likelihood ratios (LRs) or as absolute or relative risks. For example, the CAGE (Cut down, Annoyed, Guilty, Eye-opener), a prediction rule for detecting alcoholism, has been reported as likelihood ratios (e.g., for a CAGE score of 0/4, LR = 0.14; for a score of 1/4, LR = 1.5; for a score of 2/4, LR = 4.3; for a score of 3/4, LR = 13; and for a score of 4/4, LR = 100). To interpret these data, remember that likelihood ratios greater than 1.0 increase the probability that the target condition (i.e., alcoholism) is present, and the greater is the likelihood ratio, the greater is the increase in probability. Likelihood ratios less than 1.0 decrease the probability that the target condition is present, and the smaller is the likelihood ratio, the greater is the decrease in probability. In this example, the probability of alcoholism depends on a combination of the prevalence of alcoholism in the community and score on the CAGE prediction rule. When investigators report their results as likelihood ratios, they are implicitly suggesting that clinicians should use other, independent information to generate a pretest (or prerule) probability. Clinicians can then use the likelihood ratios generated by the rule to establish a posttest probability. (For approaches to using likelihood ratios, see Chapter 6, Diagnosis.)

**TESTING THE RULE’S IMPACT**

Given that the Ottawa ankle rules were developed in 1992, numerous studies have been conducted to examine their impact. The systematic review identified at the beginning of this chapter included 32 studies that investigated the accuracy of the Ottawa ankle rules. Meta-analysis included 27 studies (reporting on 15,581 patients) that had prospective
data collection and blinded assessment of radiographs. The clinical prediction rules had a sensitivity of almost 100%, a modest specificity (median of 31.5%), and a pooled LR for a negative result of the ankle rule of 0.08 (95% CI, 0.03 to 0.18).

This review tells us about the accuracy of the rule. However, demonstration of accuracy is insufficient to warrant a confident recommendation for use of the rule in clinical practice. Use of clinical prediction rules involves remembering predictor variables and often entails making calculations to determine a patient’s probability of having the target outcome. Pocket cards and computer algorithms can facilitate the task of using complex clinical prediction rules. Nonetheless, they demand time and energy, and their use is warranted only if they change clinicians’ behavior and only if that behavior change results in improved patient outcomes or reduced costs while maintaining quality. If these conditions are not met, attempts to use a clinical prediction rule systematically will be a waste of time, regardless of its accuracy.

An accurate prediction rule may not change behavior or improve outcomes for several reasons. First, clinicians’ intuitive estimation of probabilities may be as good as, if not better than, the rule. If this is so, use of a clinical prediction rule will not improve their practice. Second, clinicians may not use the rule because the calculations involved are cumbersome. Even worse, they could make errors in the calculations. Third, practical barriers could impede acting on the results of the clinical prediction rule. For instance, in the case of the Ottawa ankle rules, clinicians may be sufficiently concerned about protecting themselves against litigation that they may order radiographs despite a prediction rule result suggesting a negligible probability of fracture. These considerations lead us to classify a clinical prediction rule with evidence of accuracy in diverse populations as level II and to insist on a positive result from a study of impact before the rule ascends to level I.

Ideally, an impact study would randomize patients or larger administrative units to apply or not apply the clinical prediction rule and would follow up patients for all relevant outcomes (including quality of life, morbidity, and resource utilization). Randomization of individual patients is unlikely to be appropriate because one would expect participating clinicians to incorporate the rule into the care of all patients. A suitable alternative is to randomize institutions or practice settings and conduct analyses appropriate to these larger units of randomization. Another potential design is to assess the outcomes of a single group before and after clinicians begin to use the clinical prediction rule, but the choice of a before/after study substantially reduces the strength of inference.

With respect to the Ottawa ankle rules, investigators conducted a randomized trial to examine outcomes associated with application of the decision rules. Six emergency departments were randomized to use or not use the decision rules.21 One center dropped out before the study began, leaving a total of five emergency departments: two in the intervention group and three in the usual-care group. The intervention consisted of the following: (1) introducing the prediction rules at a general meeting, (2) distributing pocket cards summarizing the rules, (3) posting the rules throughout the emergency department, and (4) applying preprinted data collection forms to each patient chart. In the control group, the only intervention was the introduction of preprinted data
collection forms without the Ottawa rules attached to each chart. The investigators entered a total of 1911 eligible patients into the study, 1005 in the control group and 906 in the intervention group. There were 691 radiographs requested for intervention group patients and 996 requested for control group patients. In an analysis focused on the ordering physician, the investigators found that the mean proportion of patients referred for radiography was 78.9% in the intervention group and 99.6% in the control group ($P = 0.03$). The investigators noted three missed fractures in the intervention group, none of which led to adverse outcomes. Thus, the investigators demonstrated a positive resource utilization impact of the Ottawa ankle rules (decreased test ordering) without an increase in adverse outcomes, thus moving the clinical prediction rule to level I in the hierarchy (see Table 22-1).

Clinical Resolution

You conduct the continuing nurse education program for outpost nurse practitioners, and they are enthusiastic about using the Ottawa ankle rules at their sites with the objective of sending fewer patients for unnecessary radiographs. In addition, the nurses note in their evaluations that the session has armed them with the knowledge they need to critique studies that describe other clinical prediction rules that may be applicable to their practice.

The next day, as you reflect on the session, you wonder whether the nurse practitioners will indeed use the Ottawa ankle rules. In preparing for the session, you came across a study by Cameron and Naylor that reported an initiative in which clinicians who were expert in the use of the Ottawa ankle rules trained 16 other workers to teach the use of the rules. These persons used slides, overhead projections, and a 13-minute instructional video to train their colleagues locally and regionally in the use of the rules. Surprisingly, this program led to no change in the use of ankle radiography. You are also aware that little research has been done on evaluating practice change in nursing. You decide to work more closely with this group of nurse practitioners to identify and implement strategies to reinforce their learning and evaluate changes in their clinical practice.

Clinical prediction rules inform our clinical judgment and have the potential to change clinical behavior and reduce unnecessary costs while maintaining quality of care and patient satisfaction. The challenge for clinicians is to evaluate the strength of a rule and its likely impact, as well as to find ways of incorporating level I rules efficiently into their daily practice. Clinicians can access a summary of clinical prediction rules and associated levels of evidence on the Internet (http://www.mssm.edu/medicine/general-medicine/ebm).

REFERENCES


