Methodologic Standards for the Development of Clinical Decision Rules in Emergency Medicine

Ian G Stiell, MD, MSc, FRCPC
George A Wells, MSc, PhD

From the Division of Emergency Medicine, Departments of Medicine and Epidemiology and Community Medicine, and Ottawa Hospital Loeb Health Research Institute, University of Ottawa, Ottawa, Ontario, Canada.

Dr Stiell is a career scientist of the Medical Research Council of Canada.

Received for publication January 19, 1998. Revisions received July 6 and July 29, 1998. Accepted for publication September 11, 1998.

Address for reprints: *Ian G Stiell, MD, MSc, FRCPC, Clinical Epidemiology Unit, Ottawa Hospital Loeb Health Research Institute, 1053 Carling Avenue, Ottawa, Ontario, Canada, K1Y 4E9.*

0196-0644/99/$8.00 + 0
47/1/94993

The purpose of this review is to present a guide to help readers critically appraise the methodologic quality of an article or articles describing a clinical decision rule. This guide will also be useful to clinical researchers who wish to answer 1 or more questions detailed in this article. We consider the 6 major stages in the development and testing of a new clinical decision rule and discuss a number of standards within each stage. We use examples from emergency medicine and, in particular, examples from our own research on clinical decisions rules for radiography in trauma.

[Stiell IG, Wells GA: Methodologic standards for the development of clinical decision rules in emergency medicine. *Ann Emerg Med* April 1999;33:437-447.]

INTRODUCTION

Reports of clinical decision rules are becoming increasingly common throughout the medical literature and particularly within emergency medicine journals. Clinical decision rules (prediction rules) are designed to help physicians with diagnostic and therapeutic decisions at the bedside. We define a clinical decision rule as a decisionmaking tool that is derived from original research (as opposed to a consensus-based clinical practice guideline) and incorporates 3 or more variables from the history, physical examination, or simple tests.[1] These tools help clinicians cope with the uncertainty of medical decisionmaking and help clinicians improve their efficiency, an important issue as health care systems demand more cost-effective medical practice. A recently published example of a decision rule that helps emergency physicians cope with uncertainty is a guideline about which patients with community-acquired pneumonia require hospitalization.[2] A typical example of a decision rule to improve efficiency are the Ottawa Ankle Rules for the use of radiography in acute ankle injuries.[3-8]

Methodologic standards for the development of clinical decision rules have been described, originally by Wasson et al[9]

and Feinstein[10] and more recently by our own research group.[1,11] We consider the following to be the 6 important stages in the development and testing of a fully mature decision rule. First, is there a need for a decision rule? Has current practice been shown to be inefficient or highly variable? Second, was the rule derived according to rigorous methodologic standards? Third, has the rule been properly validated prospectively in a new patient group? Fourth, has the rule been successfully implemented into clinical practice and been shown to change behavior? Fifth, would use of the rule be cost-effective according to a formal health economic analysis? Sixth, how will the rule be disseminated to ensure widespread adoption in the emergency medicine community?

The purpose of this review is to present a guide to help readers critically appraise the methodologic quality of an article or articles describing a clinical decision rule. This guide will also be useful to clinical researchers who wish to answer 1 or more of the questions listed below. We will consider the 6 major stages in the development and testing of a new clinical decision rule and will discuss a number of standards within each stage (Figure 1). We will use examples from emergency medicine and, in particular, examples from our own research on clinical decision rules for radiography in trauma.

Figure 1.
Checklist of standards for 6 stages in the development of a clinical decision rule.

1. Is there a need for the decision rule?
Prevalence of the clinical condition
Current use of the diagnostic test
Variation in practice
Attitudes of physicians
Clinical accuracy of physicians
2. Was the rule derived according to methodologic standards?
Definition of outcome
Definition of predictor variables
Reliability of predictor variables
Selection of subjects
Sample size
Mathematical techniques
Sensibility of the decision rule
Accuracy
3. Has the rule been prospectively validated and refined?
Prospective validation
Selection of subjects
Application of the rule
Outcomes
Accuracy of the rule
Reliability of the rule
Physicians' interpretation
Refinement
Potential effect
4. Has the rule been successfully implemented into clinical practice?
Clinical trial
Effect on use
Accuracy of the rule
Acceptability
5. Would use of the rule be cost-effective?
6. How will the rule be disseminated and implemented?

1. IS THERE A NEED FOR THE DECISION RULE?

Clinicians should ask themselves whether there really appears to be a need for a particular decision rule, or whether the rule appears only to represent the analysis of a convenient set of data. Is there a demonstrated inefficiency or variation in current medical practice, and does there appear to be the potential for improved efficiency through guidelines or a decision rule?

Prevalence of the clinical condition

Does the proposed decision rule deal with a clinical problem commonly seen in emergency departments? A rule for an uncommon clinical entity is unlikely to be easily adopted by physicians and is unlikely to contribute significantly to the overall efficiency of emergency medicine practice. For example, because ankle injuries are one of the most common problems seen in US EDs, one might expect a decision rule for radiography to have an effect on the management of hundreds of thousands of patients annually.[12,13] We have estimated that $500 million is spent annually on ankle radiographs in North America and that even a modest reduction in the use of ankle radiography could lead to large health care savings. On the other hand, because injuries to the mandible are not common, a decision rule for this problem is unlikely to have a significant effect.

Current use of the diagnostic test

If the decision rule proposes to guide the ordering of a diagnostic test, are there data clearly demonstrating that current use of the test is inefficient? For example, we have previously shown that 87.2% of ankle and foot radiographs and 92.4% of knee radiographs ordered in EDs are negative for fracture.[14,15] Although we have no exact data, we suspect that a majority of hip radiographs ordered are, in fact, positive for fracture and that, consequently, hip radiography would not be a productive area for a decision rule.

Variation in practice

Is there significant variability in clinical practice among similar physicians or similar institutions? The low yield of

a diagnostic test for positive results may not necessarily imply inefficiency, especially if the cost of a "missed" diagnosis is high. One would expect the yield of cervical spine radiography to be low in that emergency physicians naturally tend to be cautious when dealing with the possibility of a broken neck. Hence, our recent finding that results of 98.5% of cervical spine radiographs ordered in Canadian EDs are negative would not surprise many physicians. What does suggest a need for a decision rule, however, is our finding that use of cervical spine radiography varies twofold among equivalent busy hospitals and sixfold among attending staff physicians in these EDs.[16] We have also demonstrated great variation in the use of computed tomography (CT) for patients with minor head injury.[17]

Attitudes of physicians

The need for a clinical decision rule is further supported if there is evidence that emergency physicians believe that many of the diagnostic tests they order are unnecessary. For example, we demonstrated in studies involving more than 1,700 patients that experienced attending physicians expected the likelihood of fracture to be 10% or less in 57.8% of ankle injury cases and 75.6% of knee injury cases.[15] Despite ordering radiography for most patients, these same physicians indicated that they would have been theoretically comfortable with no radiography for 45.9% of patients with ankle injuries and 55.5% of patients with knee injuries. In a mail survey, we gathered compelling evidence that Canadian emergency physicians strongly support the development of decision rules for cervical spine radiography as well as CT for minor head injury.[18] These data have led to 2 large national research projects to derive and validate rules for radiography in head and neck trauma.

Clinical accuracy of physicians

There is little likelihood that a decision rule will ultimately diminish reliance on diagnostic tests if physicians are unable to accurately predict a patient's outcome on the basis of clinical findings alone. We were convinced of the potential for improved efficiency in the use of ankle and knee radiography when we demonstrated that, based on history and physical examination alone, experienced emergency physicians could very accurately discriminate between fracture and nonfracture cases (respective areas under the receiver operating characteristic curves .88 and .87).[14,15]

2. WAS THE RULE DERIVED ACCORDING TO METHODOLOGIC STANDARDS?

Research methodology standards for the derivation of a clinical decision rule were first reviewed in 1985 in a landmark paper by Wasson et al.[9] Feinstein,[10] one of the fathers of clinical epidemiology, later added to the literature of evidence-based patient assessment in his book *Clinimetrics*. More recently, our research group at the University of Ottawa Clinical Epidemiology Unit assessed clinical decision rule articles in 4 major medical journals and proposed modifications to Wasson et al's original methodologic standards.[1] In the following discussion, numbers in parentheses indicate the percentage of published studies meeting these criteria (in *Journal of the American Medical Association, New England Journal of Medicine, British Medical Journal,* and *Annals of Internal Medicine*).

Definition of outcome

The outcome being identified by the clinical decision rule should be clinically important (100% of reviewed studies met this criterion)[1] and clearly defined (83%). Survival (eg, death), radiologic results (eg, fracture), and laboratory results (eg, elevated WBC count) are all biologic outcomes that can be clearly defined and reproduced in other settings. Although survival and fractures are clinically important outcomes, most emergency physicians would not consider a laboratory marker such as WBC count to be any more than a weak surrogate outcome. A behavioral outcome such as hospital admission may be dependent on local factors and difficult to replicate, and as a result not be of high clinical relevance. Similarly, the term "positive computed tomography" in a head injury study cannot be considered clearly defined unless the investigators explicitly indicate that the outcome refers only to "acute brain findings attributable to trauma" rather than to chronic or soft tissue findings.[19,20]

To avoid the danger of observation bias, the outcome measure should be assessed blindly (41%), that is, without knowledge of the status of the predictor variables. This standard is more important when evaluation of the outcome is "soft" or subject to interpretation (eg, days off work), and less important for a "hard" outcome such as death.

Definition of predictor variables

Potential predictor variables for a decision rule should be clearly defined (59%) and ideally collected in a prospective standardized fashion. Investigators should ensure that the physicians in their study have been adequately trained to

evaluate the patients and collect data according to well-standardized assessment techniques. Clinical data tend to be most reliable when collected prospectively and recorded on a data collection form designed specifically for a decision rule study. Less satisfactory are data collected prospectively as part of a clinical trial and then subjected to a post hoc secondary data analysis. Large administrative databases may also be used as a source of predictor and outcome variables but they frequently lack key clinical variables. Data collected from review of clinical records lack precision and are missing information and are generally unacceptable other than for assessing feasibility.

Again, to avoid observation bias, the assessment of predictor variables should be done without knowledge of the outcome (79%). Knowing, for example, that a patient has a knee fracture, might very well influence how a physician interprets and records the physical examination on a data form.

Reliability of predictor variables

Decision rules are highly dependent on findings from the clinical examination, so unless these findings are reliable the resultant rule will not be dependable. Reliability refers to the consistency or reproducibility of the findings by the same clinician (intraobserver reliability) or by different clinicians (interobserver reliability).[21-23] In assessing reliability, the coefficient used for agreement depends on the level of measurement, namely, the simple κ for dichotomous or nominal data,[23,24] the weighted κ for ordinal data,[23] and the intraclass correlation coefficient for interval data.[25] Several articles have been published regarding the determination of sample size for reliability studies from the perspective both of the precision of estimation and of comparison of coefficients.[26-29]

We believe that the reliability of the predictor variables (3%) should be explicitly assessed and that only those with good agreement beyond that expected by chance alone should be considered for a decision rule. For example, during the derivation of the Ottawa Ankle Rules we assessed 32 variables for interobserver agreement by having patients assessed independently by pairs of physicians.[30] Twenty-three of these findings were discarded when they proved to have κ values of less than .6. What constitutes acceptable agreement depends somewhat on the coefficient to be used and on the nature of the clinical problem.

Selection of subjects

Readers must be able to understand the generalizability of the findings of a decision rule study, as well as its applicability to their own patients. Hence, the study subjects should be well described in terms of inclusion criteria, method of selection, clinical and demographic characteristics (79%), and the study setting (66%). Explicit inclusion criteria allow readers to understand clearly what types of patients are being studied and, therefore, to which patients the derived rule may be applicable. For example, the Ottawa Ankle Rules were developed only with patients aged 18 years and older and should not be considered appropriate for pediatric patients. Similarly, if a decision rule for cervical spine radiography were derived in patients who were alert and stable then clinicians could not apply the rule to patients who were obtunded or in unstable condition. Some studies fail to specifically define their study population and enroll patients chosen for a particular diagnostic test at the discretion of the treating physician. Because different physicians may use different criteria for ordering tests, the reader of these studies has a difficult time deciding if the study patients were similar to his or her own.

Ideally, the method of patient selection is free of bias so that study subjects encompass a wide clinical and demographic spectrum and are representative of all patients seen at the site with the designated condition. The investigators should report all pertinent characteristics of patients included in the study as well as of those eligible but not included. For example, a study of knee radiography that systematically excluded more severely injured patients because the physicians were too busy to complete the data form may result in a decision rule with limited applicability.

The study site should be described in sufficient detail to permit the reader to make a comparison to his own emergency department. Was the study actually conducted in an ED or, rather, in a clinic or office setting? Was the hospital a primary, secondary, or tertiary care facility? Was this a teaching institution? What was the "referral filter" (ie, what proportion of patients were self-referred as opposed to being sent in by a physician?) As an example, the status of patients seen in a neurologist's clinic for migraine might be expected to differ considerably from that of patients presenting to an ED with severe intractable migraine.

Sample size

The authors should justify the number of subjects enrolled in the study. Of particular importance is that the sample size be appropriate for the type of multivariate analysis chosen. There may be problems with overfitting the data if there are too few outcome events per predictor variable. A commonly used "rule of thumb" is that there should be at least 10 outcome events per independent variable in the prediction rule.[9] Another important consideration in

choosing the study sample size is the degree of precision in the confidence interval (CI) around the measure of accuracy (eg, the sensitivity of the rule). For example, a study to develop a prediction rule for rib fractures would require at least 40 subjects with positive radiographic results if the rule encompassed 4 predictor variables. If all rib fractures were identified in such a study, the rule could be said to have a sensitivity of 1.0; however, the CI around the sensitivity would be .91 to 1.0.

Ideally, sample size should be determined on the basis of the mathematical modeling technique used and the estimation and relationship of predictors used in model development. In most instances, such detailed information will be unknown. However, if pilot data were available that could be used to demonstrate the feasibility of a particular statistical model and provide initial estimates of the parameters and their covariances, then a sample size that suitably bounds the CI for estimating the response could be used as the basis for sample size determination. As an example, if the logistic regression model is appropriate and pilot data are available, sample size calculations using this approach have been developed.[31,32]

Mathematical techniques

The mathematical techniques used to derive a decision rule should be adequately described and justified (100%). Many techniques are available, from a simple 2×2 cross-tabulation of each predictor variable with the outcome to sophisticated multivariate analyses. Although univariate analyses are easy to perform, they do not allow the exploration of the relationship of predictor variables with each other and with the outcome. To achieve this end, multivariate statistical approaches such as logistic regression and recursive partitioning are commonly used. The variables to be included in a multivariate analysis are usually "screened" by assessment of the univariate association with the outcome, as well as an assessment of the reproducibility of the variable.

χ^2 Recursive partitioning analysis progressively divides the patients into a subpopulation that includes only patients with a particular outcome.[33-35] This approach is quite appropriate when the objective of the study is to develop a decision rule with a very high sensitivity. For example, recursive partitioning was used to develop the Ottawa Ankle Rules, which identify a subgroup of patients with zero probability of having a significant fracture (Figure 2).[3,36]

Although logistic regression also predicts the likelihood of a binary outcome (eg, myocardial infarction or no myocardial infarction), this probability is given as the "logit" or log odds, which can be used to calculate a simple odds ratio.[37] Logistic regression analyses tend to lead to decision rules with higher overall accuracy (ie, better overall classification of all patients) but possibly less than optimal sensitivity (ie, <100% classification of abnormal patients).[38]

Sensibility of the decision rule

Feinstein[10] coined the term "sensibility" to describe whether a decision rule is clinically reasonable (97%), easy to use (41%), and provides a course of action (0%). Assessment of sensibility depends on judgment rather than on statistical methods. Ideally, a decision rule should demonstrate content validity, which means that most clinicians would consider the items in the rule to be clinically sensible, that no obvious items are missing, that the method of grouping the individual variables is reasonable, and that the items seem appropriate for the purpose of the rule. Ease of use depends on factors such as the length of time needed to apply the rule and simplicity of interpretation. In the ED, it is unlikely that physicians would embrace a rule that requires extensive calculations or use of a calculator. Similarly, a rule will not be useful if it depends on laboratory results that are not available in a timely fashion.

We believe that decision rules are more likely to be used if they suggest a course of action rather than if they merely provide a probability of outcome. For example, a rule to predict the need for a cranial CT for nontrauma patients proposes a "CT/no CT" course of action rather than giving a percentage probability of abnormality on CT.[39] In the case of the Ottawa Ankle Rules, clinicians are also advised whether to order radiography. We believe that merely giving the probability of fracture (eg, 2%) may lead to uncertainty in the clinician's decision to order a radiograph. An estimate of an outcome probability may, however, be useful for both the clinician and the patient in situations where the relative costs and benefits of a treatment option are less clear. A recently developed clinical model for deep vein thrombosis stratifies ED patients into low-, medium-, and high-probability groups.[40-42]

Accuracy

Authors should make an effort to present the accuracy or classification performance (100%) based on the population from which the decision rule was derived. If the outcome is dichotomous (eg, fracture or no fracture), then a 2×2 table should be presented along with calculations of sensitivity, specificity, negative predictive value, and positive predictive value, with the respective 95% CIs.[43,44] Sensitivity and specificity are characteristics of the decision rule itself and refer to the classification accuracy for patients with an

abnormal outcome and a normal outcome, respectively. Negative and positive predictive values are greatly influenced by the prevalence of the outcome in the population being studied.

In many emergency medicine scenarios, high sensitivity, or the ability to "rule out" an outcome if the result is negative, is important. Alternately, a rule with a high specificity will "rule in" a condition or outcome if the result is positive. For conditions where the cost of missing the outcome is serious (eg, myocardial infarction,[45,46] fracture), then a very high sensitivity is required. Unfortunately, a very high sensitivity tends to be associated with a decreased specificity and, consequently, overall decreased efficiency of the decision rule. For example, although the Ottawa Ankle Rules are designed to have a sensitivity of 100%, they have a relatively low specificity of 40%. Merely dropping the sensitivity to 96% would increase the specificity to 58% and greatly increase the potential for reducing radiography. However, we believe that many North American clinicians would be very uncomfortable using a rule with a sensitivity of only 96%.[47-49]

The posttest probability of an outcome associated with the prediction of a decision rule can be calculated by the use of likelihood ratios.[43,50,51] A likelihood ratio greater than 1.0 increases the probability of abnormality and a likelihood ratio less than 1.0 decreases it. Likelihood ratios are particularly useful for rules with more than 2 response categories. Receiver operator operating characteristic curves result from the plotting of sensitivity versus (1–specificity) using different cut points in the data.[50,52] This technique allows a comparison of several rules and gives an overall estimate of

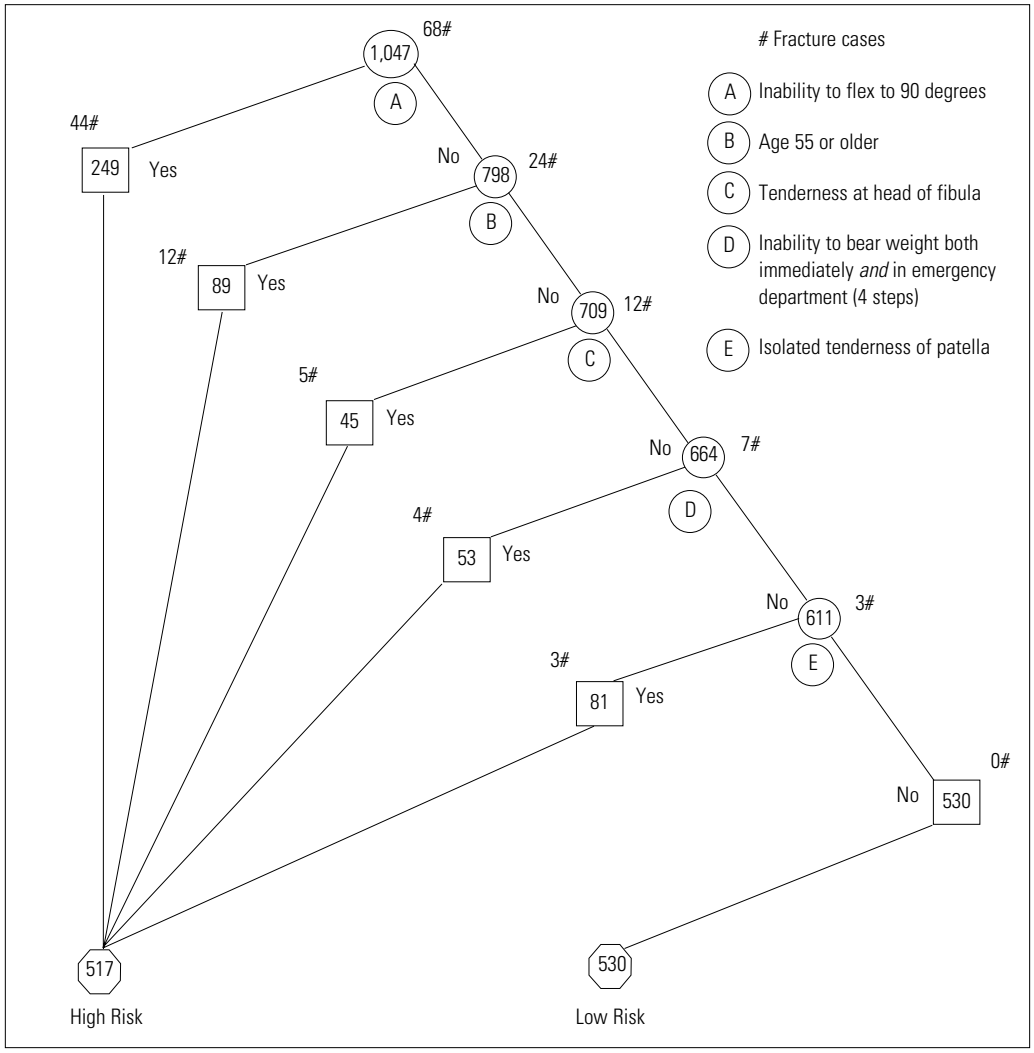

Figure 2.
Example of χ^2 recursive partitioning analysis used to derive a clinical decision rule.

accuracy, but is of limited clinical use because posttest probability is not provided.

Another important performance index of decision rules is an estimate of the potential impact of use. This estimation is usually overlooked and, although often optimistic, gives the reader a reasonable idea of the potential savings associated with implementation of the rule. For example, a rule for knee radiography would be of questionable value if it proved to be 100% sensitive for important fractures (which were normally never missed) but had such a low specificity that use of radiography would actually increase. Alternately, a decision rule for the use of abdominal CT in blunt trauma that led to increased use of CT might be valuable if subsequent laparotomy rates could be decreased.

3. HAS THE RULE BEEN PROSPECTIVELY VALIDATED AND REFINED?

Unfortunately, many clinical decision rules are not prospectively assessed to determine, in a new patient population, their accuracy, reproducibility, acceptance to clinicians, or potential effect on practice. This validation process is very important because many statistically derived rules or guidelines fail to perform well when tested in a new population.[53-55] The reason for this poor performance may be statistical—overfitting or instability in the original derived model[56]—or may be related to differences in prevalence of disease or differences in how the decision rule is applied.[57,58] The ideal characteristics of a validation study are described below.

Prospective validation

The clinical decision rule should be applied prospectively to a completely new patient population. Under ideal circumstances this would be performed in a new clinical setting by different clinicians from those involved in the derivation study.

Selection of subjects

The patients should be chosen in an unbiased fashion, preferably in a complete population-based sample for that setting, for example, all patients seen in a 12-month period in a particular ED. Patients who are not included in this validation phase need to be described so that their characteristics can be compared with those subjects who are included in the validation study. Readers should be reassured that the patients included represent a wide spectrum of ages and severity for the clinical condition under question.

Application of the rule

Investigators conducting a validation study have an obligation to ensure that they fully understand how the decision rule is to be applied. If necessary, the investigator should discuss the accurate application of the decision rule with the original researcher and should be prepared to accurately present the decision rule to the study physicians. This must involve a brief but adequate training session (eg, 15 minutes during a rounds presentation) and training tools such as posters, pocket cards, and audiovisual aids.[59-63] The emergency medicine literature includes clear examples where decision rules have faired suboptimally in validation studies when the treating physicians were not fully taught how to use the rule or were not in fact aware that they were assessing decision rules.[64-67] We believe that the purpose of validation studies is to rigorously assess the efficacy or accuracy of the rule itself. If the rule proves to be accurate, then subsequent implementation studies can serve to assess the effectiveness of the rule in "the real world."

Outcomes

Ideally, all patients would be subjected to the gold or criterion standard to determine their true outcome compared with that predicted by the decision rule. For example, all patients undergoing a prospective validation of the Ottawa Ankle Rule underwent ankle radiography to determine the presence of fracture.[4] In many cases, however, the criterion standard is not normally applied to all potential patients. In this instance, a suitable and reasonable proxy outcome may be substituted. For example, in the case of the Ottawa Knee Rule, knee radiography was not uniformly used for all knee injury patients in the study hospitals. Consequently, the investigators incorporated a proxy outcome for the 30% of patients who normally would not undergo radiography.[68] In this instance, the proxy outcome included a structured telephone interview consisting of 5 explicit telephone questions (relating to pain, ability to walk, return to work, and need for medical care) applied by a registered nurse. Patients who could not satisfy all criteria were asked to return for radiographic examination.

Accuracy of rule

As described above, accuracy is best presented by a 2×2 table with calculation of classification performance in terms of sensitivity, specificity, negative predictive value,

and positive predictive value, all with 95% CIs. The sample size of a validation study primarily depends on 2 factors. The first factor, as described above, is the width of the CI around the target sensitivity. The second factor is the need to have an adequate number of outcomes to allow a satisfactory multivariate analysis in the refinement process.

Reliability of rule

Rarely do investigators determine the interobserver agreement between physicians for assessing the actual rule itself. We believe that the reliability of interpretation of the decision rule (0%) should be explicitly measured in a validation study (eg, radiography indicated—yes/no).[4,68] This is very important in that a rule that cannot be reliably assessed may lead to the misclassification of some patients and potentially to the misdiagnosis of serious outcomes.

Physicians' interpretation

An important aspect of validation that is often overlooked is the ease with which clinicians can apply this rule. The investigator should make efforts to determine the interobserver agreement for interpreting the decision rule between pairs of physicians. Furthermore, the investigator should determine the accuracy with which the physicians interpret the decision rule, as well as their comfort with its use (the latter may be determined by a simple survey question).

Refinement

A validation study provides a unique opportunity for the investigator to review the value of each of the component variables in the decision rule and possibly to improve on the rule. Toward this end, the investigator might well consider reevaluating several variables that proved to be valuable but not essential in the original derivation study. By reassessing the reproducibility and accuracy of a limited number of variables, it is quite possible that the study may lead to a more accurate rule or a simplified rule. In the case of the Ottawa Ankle Rules, the validation study led to a more streamlined ankle rule without the requirement for the age criterion. Furthermore, the refinement process led to a more accurate foot rule with slightly different criteria.[4] Once refined, of course, the rule must then be validated again in a new patient set.

Potential effect

Having assessed the classification performance of the decision rule as well as the baseline ordering rate for the test in question, the investigator is now in a good position to estimate the potential effect of the decision rule. In other words, the investigator can now calculate the potential savings that might be realized if the decision rule were accurately and completely implemented into practice. This determination is very important in predicting the potential usefulness of the rule, as well as in determining the need or rationale for an implementation study.

4. HAS THE RULE BEEN SUCCESSFULLY IMPLEMENTED INTO CLINICAL PRACTICE?

Very few clinical decision rules have been implemented and shown to alter clinical practice in what has been termed "the next painful step" for evaluating decision aids.[69] A decision rule that has been shown to be valid and reliable still faces significant barriers to implementation in terms of both patient and physician acceptance.[48,49,70] Emergency physicians are concerned about the medicolegal consequences of missing a diagnosis. Patients must place their trust in a physician whom they have never met before and whom they will not likely see again in follow-up. Hence, many emergency physicians have the perception that their patients will only be satisfied with some form of diagnostic investigation.[47] Consequently, widespread acceptance of a clinical decision rule must be preceded by an implementation study that clearly shows that clinical behavior can be altered.

Clinical trial

Although the ideal implementation trial would be a randomized controlled trial, such a study is often not feasible when the intervention under study is a cognitive guideline that is learned by the individual physician. Several implementation trials have adopted before-after comparisons that also incorporate concurrent control groups from other hospitals.[6,71] To enhance generalizability, an implementation trial should be conducted in as many different types of settings as possible and involve as many types of physician as possible.[8,72] The physicians must be carefully trained to accurately interpret and apply the decision rule and must have adequate reminders and cues so that they learn to routinely apply the rule. For example, we have used pocket cards and posters as training tools, as well as memos and follow-up by study nurses, to encourage the accurate use of the Ottawa Ankle and Ottawa Knee Rules.

Effect on use

The primary outcome measure of an implementation trial is generally the effect on use of the resource in question (eg, ordering of radiography). Follow-up of patients is important to ensure that those denied a procedure or test at the study hospital do not go elsewhere to obtain the same test. Ideally, physician practice would also be followed at a later date to assess the long-term effect of the decision rule.[73]

Accuracy of the rule

Careful follow-up is important to ensure that there are few missed diagnoses and that the sensitivity and specificity of the decision rule can be recalculated.

Acceptability

Physicians should be surveyed both formally as part of the study and informally to determine their comfort with applying the rule, as well as their impression regarding its ease of use. Finally, the patients should be surveyed to determine their attitudes to the process, and especially their satisfaction with care that may not include a diagnostic test.

5. WOULD USE OF THE RULE BE COST-EFFECTIVE?

If an implementation trial does, in fact, show that the decision rule alters clinical behavior, then a formal economic evaluation conducted by a health economist might be conducted.[74,75] The objective of such a study would be to clearly demonstrate the health care savings that might be associated with widespread use of the decision rule. Economic assessment is concerned with choosing between alternative uses of resources. Resources are limited and choices must be made.

Three basic concepts are involved in any economic evaluation, namely: the type of analysis that is performed, the type of costs and benefits included, and the point of view from which the analysis is taken. There are 3 types of analysis: cost identification, cost-effectiveness, and cost-benefit. The choice is determined by whether benefits are included in the analysis and the manner in which benefits are valued. Three types of costs and benefits can be considered—direct, indirect, and intangible. Several points of view can be considered including those of the patient, health care provider, health care payer, and society. An economic analysis must include reasonable assumptions regarding the accuracy and effectiveness of the rule, as well as of the costs involved. Sensitivity analyses allow one to assess the robustness of the effect of the rule under various conditions.

6. HOW WILL THE RULE BE DISSEMINATED AND IMPLEMENTED?

For a clinical guideline to have widespread effect on health care delivery, there must be an active plan for dissemination and implementation. We are well aware that the simple passive diffusion of original study results (through publication in medical journals or presentation at scientific meetings) is unlikely to significantly alter clinical practice in emergency medicine or in any other discipline.[76,77] Strategies to ensure the dissemination and implementation of clinical research are currently the subject of intensive health services research.[78-80] Dissemination is a more active process that involves targeting modified information for a specific audience.[81,82] Examples include secondary sources such as metaanalyses, reviews, practice guidelines, and consensus statements that are distributed by journal publication, the lay press, targeted mailings, or a campaign of visiting speakers. Implementation is the most active process and uses organizational and behavioral tools applied locally and persistently to overcome barriers to the use of the new information by practitioners.

Adoption of innovations (eg, decision rules) is affected by a number of factors, including the attributes of the innovation itself, the characteristics of the physician, the practice setting, legal and financial issues, regulation, as well as patient factors.[83-85] The attributes of the innovation that facilitate adoption are relative advantage (new practice is demonstrably superior to the old), compatibility (similar to prior experience or practice), complexity (ease of incorporation into practice), "trialability" (practitioner can "try it out"), and "observability" (can observe practice of other physicians).[83,86]

Previous research has shown that some interventional strategies are much more effective than others in changing physician behavior.[83,84,87] Relatively weak approaches are traditional lecture-based conferences and seminars and unsolicited, mailed information. More effective is audit and feedback, which is best when given concurrently rather than later, when directed at specific individuals, and when delivered by opinion leaders (respected local clinicians). The strongest implementation strategies are considered to be concurrent reminder systems (posters, pocket cards, sheets, computer-embedded prompts), academic detailing (face-to-face education in

the physician's setting), and the use of multiple interventions concurrently.[79]

How should the effect of dissemination or implementation be assessed? Generally, studies evaluate either the process of care or patient outcome.[88] Examples of process of care evaluations might include rates of cervical spine radiography use, timely thrombolytic administration, or appropriate hospital admission for pneumonia. Analogous examples of important patient outcomes would be morbidity from missed cervical spine fractures, mortality from myocardial infarction, or morbidity and mortality from undertreated pneumonia. Ideally, investigators should discuss appropriate strategies for dissemination and implementation once their clinical decision rule has been proved to be valid and effective.

In summary, very few clinical decision rules have been successfully derived, validated, and adopted into clinical practice. This is because development of an effective decision rule is a long, rigorous, and expensive process. Emergency medicine has many clinical situations that would benefit from a decision rule; this provides much research opportunity for our specialty. This article has outlined important methodologic issues for both the critical reader and the emergency medicine researcher to consider.

REFERENCES

1. Laupacis A, Sekar N, Stiell IG: Clinical prediction rules: A review and suggested modifications of methodological standards. *JAMA* 1997;277:488-494.

2. Fine MJ, Auble TE, Yealy DM, et al: A prediction rule to identify low-risk patients with community-acquired pneumonia. *N Engl J Med* 1997;336:243-250.

3. Stiell IG, Greenberg GH, McKnight RD, et al: A study to develop clinical decision rules for the use of radiography in acute ankle injuries. *Ann Emerg Med* 1992;21:384-390.

4. Stiell IG, Greenberg GH, McKnight RD, et al: Decision rules for the use of radiography in acute ankle injuries: Refinement and prospective validation. *JAMA* 1993;269:1127-1132.

5. McDonald CJ: Guidelines you can follow and can trust: An ideal and an example. *JAMA* 1994;271:872-873.

6. Stiell IG, McKnight RD, Greenberg GH, et al: Implementation of the Ottawa ankle rules. *JAMA* 1994;271:827-832.

7. Wasson JH, Sox CH: Clinical predition rules: Have they come of age? *JAMA* 1996;275:641-642.

8. Stiell I, Wells G, Laupacis A, et al: A multicentre trial to introduce clinical decision rules for the use of radiography in acute ankle injuries. *BMJ* 1995;311:594-597.

9. Wasson JH, Sox HC, Neff RK, et al: Clinical prediction rules: Application and methodological standards. *N Engl J Med* 1985;313:793-799.

10. Feinstein AR: *Clinimetrics.* New Haven, CT: Yale University Press, 1987.

11. Stiell IG: Cranial computed tomography for nontrauma patients: Critical appraisal of a decision rule. *Acad Emerg Med* 1997;4:650-652.

12. McCaig LF: National Hospital Ambulatory Medical Care Survey: 1992 emergency department summary. *Advance Data* 1994;245:1-12.

13. National Center for Health Statistics: *National Hospital Ambulatory Medical Care Survey 1992.* Hyattsville, MD: National Center for Health Statistics, 1994.

14. Stiell IG, McDowell I, Nair RC, et al: Use of radiography in acute ankle injuries: Physicians' attitudes and practice. *Can Med Assoc J* 1992;147:1671-1678.

15. Stiell IG, Wells GA, McDowell I, et al: Use of radiography in acute knee injures: Need for clinical decision rules. *Acad Emerg Med* 1995;2:966-973.

16. Stiell IG, Wells GA, Vandemheen K, et al: Variation in emergency department use of cervical spine radiography for alert, stable trauma patients. *Can Med Assoc J* 1997;156:1537-1544.

17. Stiell IG, Wells GA, Vandemheen K, et al: Variation in ED use of computed tomography for patients with minor head injury. *Ann Emerg Med* 1997;30:14-22.

18. Graham ID, Stiell IG, Laupacis A, et al: Emergency physicians' attitudes toward the use of clinical decison rules for radiography. *Acad Emerg Med* 1998;5:134-140.

19. Madden C, Witzke DB, Sanders AB, et al: High-yield selection criteria for cranial computed tomography after acute trauma. *Acad Emerg Med* 1995;2:248-253.

20. Borczuk P: Predictors of intracranial injury in patients with mild head trauma. *Ann Emerg Med* 1995;25:731-736.

21. Kramer MS, Feinstein AR: Clinical biostatistics: LIV. The biostatistics of concordance. *Clin Pharmacol Ther* 1982;29:111-123.

22. Landis JR, Koch GG: The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-174.

23. Fleiss JL: *Statistical Methods for Rates and Proportions*, ed 2. New York: John Wiley & Sons, 1981.

24. Roberts C, McNamee R: A matrix of kappa-type coefficients to assess the reliability of nominal scales. *Stat Med* 1998;17:471-488.

25. Fleiss JL: The design and analysis of chemical experiments, in Barnett V, Bradley RA, Hunter JS (eds): *Reliability of Measurement.* New York: John Wiley & Sons, 1986:1-31.

26. Flack VF, Afifi AA, Lachenbruch PA, et al: Sample size determinations for the two rater kappa statistics. *Psychometrika* 1988;53:321-325.

27. Walter SD, Eliasziw M, Donner A: Sample size and optimal designs for reliability studies. *Stat Med* 1998;17:101-110.

28. Donner A: Sample size requirements for the comparison of two or more coefficients of interobserver agreement. *Stat Med* 1998;17:1157-1168.

29. Donner A, Eliasziw M, Klar N: Testing the homogeneity of kappa statistics. *Biometrics* 1996;52:176-183.

30. Stiell IG, McKnight RD, Greenberg GH, et al: Interobserver agreement in the examination of acute ankle injury patients. *Am J Emerg Med* 1992;10:14-17.

31. Flack VF, Eudey TL: Sample size determinations using logistic regression with pilot data. *Stat Med* 1993;12:1079-1084.

32. Hsieh FY: Sample size tables for logistic regression. *Stat Med* 1989;8:795-802.

33. Friedman JHL A recursive partitioning decision rule for nonparametric classification. *IEEE Trans Comput* 1977;16:404-408.

34. Ciampi A, Thiffault J, Nakache J, et al: Stratification by stepwise regression, correspondence analysis and recursive partition: A comparison of three methods of analysis for survival data with covariates. *Comput Stat Data Anal* 1986;4:185-204.

35. Ciampi A, Hogg SA, McKinney S, et al: RECPAM: A computer program for recursive partition and amalgamation for censored survival data and other situations frequently occurring in biostatistics. I. Methods and program features. *Comput Meth Prog Biomed* 1988;26:239-256.

36. Stiell IG: A study to develop clinical decision rules for the emergency department use of radiography in acute ankle injuries [thesis]. Ottawa: University of Ottawa, 1991.

37. Hosmer DW, Lemeshow S: *Applied Logistic Regression.* New York: John Wiley & Sons, 1989.

38. Heckerling PS, Tape TG, Wigston RS: Clinical prediction rule for pulmonary infiltrates. *Ann Intern Med* 1990;113:664-670.

39. Rothrock SG, Buchanan C, Green SM, et al: Cranial computed tomography in the emergency evaluation of adult patients without a recent history of head trauma: A prospective analysis. *Acad Emerg Med* 1997;4:654-661.

40. Wells PS, Hirsh J, Anderson DR, et al: Accuracy of clinical assessment of deep-vein thrombosis. *Lancet* 1995;345:1326-1330.

41. Wells PS, Anderson DR, Bormanis J, et al: Value of assessment of pretest probability of deep-vein thrombosis in clinical management. *Lancet* 1997;350:1795-1798.

42. Wells PS, Anderson DR, Stiell IG, et al: The use of a clinical model to manage patients with suspected deep venous thrombosis [abstract]. *Acad Emerg Med* 1998;5:500.

43. Sackett DL, Haynes RB, Guyatt GH, et al: *Clinical Epidemiology: A Basic Science for Clinical Medicine,* ed 2. Toronto: Little, Brown, 1991.

44. Diamond GA: Limited assurances. *Am J Cardiol* 1989;63:99-100.

45. Pozen MW, D'agostino RB, Selker HP, et al: A predictive instrument to improve coronary-care-unit admission practices in acute ischemic heart disease: A prospective multicenter clinical trial. *N Engl J Med* 1984;310:1273-1278.

46. Goldman L, Cook EF, Brand DA: A computer protocol to predict myocardial infarction in emergency department patients with chest pain. *N Engl J Med* 1988;318:797-803.

47. Feinstein AR: The "chagrin factor" and qualitative decision analysis. *Arch Intern Med* 1985;145:1257-1259.

48. Long AE: Radiographic decision-making by the emergency physician. *Emerg Med Clin North Am* 1985;3:437-446.

49. Svenson J: Need for radiographs in the acutely injured ankle. *Lancet* 1988;1:244-245.

50. Sox HC, Blatt MA, Higgins MC, et al: *Medical Decision Making.* Boston: Butterworths, 1988.

51. Koopman PAR: Confidence intervals for the ratio of two binomial proportions. *Biometrics* 1984;40:513-517.

52. Hanley JA, McNeil BJ: The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;143:29-36.

53. Fischl MA, Pitchenik A, Gardner LB: An index predicting relapse and need for hospitalization in patients with actue bronchial asthma. *N Engl J Med* 1981;305:783-799.

54. Rose CC, Murphy JG, Schwartz JS: Performance of an index predicting the response of patients with acute bronchial asthma to intensive emergency department treatment. *N Engl J Med* 1984;310:573-577.

55. Centor RM, Yarbrough B, Wood JP: Inability to predict relapse in acute asthma. *N Engl J Med* 1984;310:577-580.

56. Charlson ME, Ales KL, Simon R, et al: Why predictive indexes perform less well in validation studies. *Arch Intern Med* 1987;147:2155-2161.

57. Poses RM, Cebul RD, Collins M, et al: The importance of disease prevalence in transporting clinical prediction rules: The case of streptococcal pharyngitis. *Ann Intern Med* 1986;105:586-591.

58. Wigton RS, Connor JL, Centor RM: Transportability of a decision rule for the diagnosis of streptococcal pharyngitis. *Arch Intern Med* 1986;146:81-83.

59. Pigman EC, Klug RK, Sanford S, et al: Evaluation of the Ottawa clinical decision rules for the use of radiography in acute ankle and midfoot injuries in the emergency department: An independent site assessment. *Ann Emerg Med* 1994;24:41-45.

60. McBride KL: Validation of the Ottawa ankle rules: Experience at a community hospital. *Can Fam Physician* 1997;43:459-465.

61. Solomito AL, Singal BM, Radack M: Ankle radiography in the emergency department: A prospective validation of Ottawa ankle rules [abstract]. *Acad Emerg Med* 1994;1:A64.

62. Chande VT: Decision rules for roentgenography of children with acute ankle injuries. *Arch Pediatr Adolesc Med* 1995;149:255-258.

63. Diercks DB, Hall KN, Hamilton CA: Validation of the Ottawa knee rules in an American urban teaching emergency department [abstract]. *Acad Emerg Med* 1997;4:408-409.

64. Kerr L, Kelly AM, Grant J, et al: Failed validation of a clinical decision rule for the use of radiography in acute ankle injury. *N Z Med J* 1994;107:294-295.

65. Stiell IG, Greenberg GH, McKnight RD, et al: Ottawa ankle rules for radiography of acute injuries. *N Z Med J* 1995;108:11.

66. Lucchesi GM, Jackson RE, Peacock WF, et al: Sensitivity of the Ottawa Rules. *Ann Emerg Med* 1995;26:1-5.

67. Stiell IG, Greenberg GH, McKnight RD, et al: The "real" Ottawa ankle rules. *Ann Emerg Med* 1996;27:103-104.

68. Stiell IG, Greenberg GH, Wells GA, et al: Prospective validation of a decision rule for the use of radiography in acute knee injuries. *JAMA* 1996;275:611-615.

69. Lee TH: Evaluating decision aids: the next painful step. *J Gen Intern Med* 1990;5:528-529.

70. Matthews MG: Guidelines for selective radiological assessment of inversion ankle injuries. *BMJ* 1986;293:959.

71. Stiell IG, Wells GA, Hoag RA, et al: Implementation of the Ottawa knee rule for the use of radiography in acute knee injuries. *JAMA* 1997;278:2075-2078.

72. Auleley G-R, Ravaud P, Giraudeau B, et al: Implementation of the Ottawa Ankle Rules in France: A multicenter randomized controlled trial. *JAMA* 1997;277:1935-1939.

73. Verbeek RP, Stiell IG, Hebert G, et al: Ankle radiograph utilization after learning a decision rule: A 12-month follow-up. *Acad Emerg Med* 1997;4:776-779.

74. Anis AH, Stiell IG, Stewart DG, et al: Cost-effectiveness analysis of the Ottawa ankle rules. *Ann Emerg Med* 1995;26:422-428.

75. Nichol G, Stiell IG, Wells GA, et al: Cost-benefit analysis of implementation of the Ottawa knee rule [abstract]. *Acad Emerg Med* 1997;4:433.

76. Lee TH, Cooper HL: Translating good advice into better practice. *JAMA* 1997;278:2108-2109.

77. Greer AL: The state of the art versus the state of the science: The diffusion of new medical technologies into practice. *Int J Technol Assess Health Care* 1988;4:5-26.

78. Greco PJ, Eisenberg JM: Changing physicians' practices. *N Engl J Med* 1993;329:1271-1274.

79. Anderson G: Implementing practice guidelines. *Can Med Assoc J* 1993;148:753-755.

80. Hayward RSA, Laupacis A: Initiating, conducting and maintaining guidelines development programs. *Can Med Assoc J* 1993;148:507-512.

81. Lomas J: Diffusion, dissemination, and implementation: Who should do what? *Ann N Y Acad Sci* 1993;703:226-235.

82. Lomas J: Teaching old (and not so old) docs new tricks: Effective ways to implement research findings, in Dunn EV, Norton PG, Stewart M, et al (eds): *Disseminating Research/Changing Practice.* Thousand Oaks, CA: Sage Publications, 1994:1-18.

83. Davis DA, Taylor-Vaisey A: A systematic review of theoretic concepts, practical experience and research evidence in the adoption of clinical practice guidelines. *Can Med Assoc J* 1997;157:408-416.

84. Davis DA, Thomson MA, Oxman AD, et al: A systematic review of the effect of continuing medical education strategies. *JAMA* 1995;274:700-705.

85. Basinski ASH, Naylor CD, Cohen MM, et al: Standards, guidelines, and clinical policies. *Can Med Assoc J* 1992;146:833-837.

86. Grilli R, Lomas J: Evaluating the message: The relationship between compliance rate and the subject of a practice guideline. *Med Care* 1994;32:202-213.

87. Oxman AD, Thomson MA, Davis DA, et al: No magic bullets: A systematic review of 102 trials of interventions to improve professional practice. *Can Med Assoc J* 1995;153:1423-1431.

88. Grimshaw JM, Russell IT: Effect of clinical guidelines on medical practice: A systematic review of rigorous evaluations. *Lancet* 1993;342:1317-1322.