Clinical Prediction Rules

A Review and Suggested Modifications of Methodological Standards

Andreas Laupacis, MD; Nandita Sekar, MD; Ian G. Stiell, MD

Background.—Clinical prediction rules are decision-making tools for clinicians, containing variables from the history, physical examination, or simple diagnostic tests.

Objective.—To review the quality of recently published clinical prediction rules and to suggest methodological standards for their development and evaluation.

Data Sources.—Four general medical journals were manually searched for clinical prediction rules published from 1981 through 1994.

Study Selection.—Four hundred sixty potentially eligible reports were identified, of which 30 were clinical prediction rules eligible for study. Most methodological standards could only be evaluated in 29 studies.

Data Abstraction.—Two investigators independently evaluated the quality of each report using a standard data sheet. Disagreements were resolved by consensus.

Data Synthesis.—The mathematical technique was used to develop the rule, and the results of the rule were described in 100% (29/29) of the reports. All the rules but 1 (97% [28/29]) were felt to be clinically sensible. The outcomes and predictive variables were clearly defined in 83% (24/29) and 59% (17/29) of the reports, respectively. Blind assessment of outcomes and predictive variables occurred in 41% (12/29) and 79% (23/29) of the reports, respectively, and the rules were prospectively validated in 79% (11/14). Reproducibility of predictive variables was assessed in only 3% (1/29) of the reports, and the effect of the rule on clinical use was prospectively measured in only 3% (1/30). Forty-one percent (12/29) of the rules were felt to be easy to use.

Conclusions.—Although clinical prediction rules comply with some methodological criteria, for other criteria, better compliance is needed.

JAMA. 1997;277:488-494

CLINICAL PREDICTION rules (sometimes also called clinical decision rules) are tools designed to assist medical decision making and are intended for use by clinicians when caring for patients. They are usually created by multivariable analysis and either provide a probability of disease or outcome, or suggest a diagnostic or therapeutic course of action. An example of the former is a rule that provides clinicians with the likelihood that a patient with suspected coronary disease will die during the next 4 years, while an example of the latter is a rule that indicates which patients with ankle injuries require radiography.

Because these rules are used to make decisions about patient care, it is important that they be well developed and validated. In 1985, Wasson et al published methodological criteria for the evaluation of clinical prediction rules that have been useful to investigators creating new rules. The purpose of the present study was to review the quality of recently published clinical prediction rules using the criteria of Wasson et al and to suggest modifications to their criteria based on our experience developing prediction rules.

METHODS

For the purpose of this article, we defined a clinical prediction rule as a decision-making tool for clinicians that included 3 or more variables obtained from the history, physical examination, or simple diagnostic tests and that either provided the probability of an outcome or suggested a diagnostic or therapeutic course of action. Four general medical journals (Annals of Internal Medicine, British Medical Journal, JAMA, and New England Journal of Medicine) were manually searched for reports that might contain a clinical prediction rule published from January 1, 1991, through December 31, 1994. These are the same 4 journals that Wasson et al reviewed for 1981 through 1984.

Each report was read independently by 2 of the investigators to determine if it met our definition of a prediction rule. The methodological quality of each eligible report was evaluated indepen-
boli are postoperative pulmonary symptoms that are important. Any disagreements were resolved by consensus among the 3 authors.

Four hundred sixty potentially eligible reports were initially identified. Thirty-two reports fulfilled the inclusion criteria (Annals of Internal Medicine, n = 129,011; British Medical Journal, n = 530,218; JAMA, n = 85,033; and New England Journal of Medicine, n = 129,033). Two reports written by 1 of the investigators were excluded to reproduce the method of Wasson and colleagues. Of the remaining 30 reports, 15 derived a rule, 4 validated a previously derived rule, 10 derived and validated a rule, and 1 described a method of implementing a rule. The types of clinical problems and outcomes described in the reports are shown in Table 1.

The methodological standards described by Wasson et al are shown in Table 2. In the present report, we review the rationale for the standards and describe the degree to which the studies complied with the standards. One study described the impact of implementing a rule but did not provide much information about the earlier derivation or validation of the rule. Therefore, this study was included only when the effects of clinical use prospectively measured standard was assessed, and most standards were assessed in 29 studies. We suggest additional standards that were not included in the article by Wasson et al and describe the degree of compliance with these standards. A summary of how well each study complied with the standards is available from the authors on request.

RESULTS

Standards of Wasson et al

Outcome.—Definition and Clinical Importance.—The outcome being predicted by a rule should be clearly defined and clinically important. If a surrogate outcome is used, it must have a definite relationship with a clinically important outcome. For example, symptomatic postoperative pulmonary emboli are definitely clinically important. Asymptomatic proximal deep venous thrombi detected on routine surveillance venography are surrogate outcomes that many clinicians would consider important because of their propensity to embolize, even though their presence alone is not clinically important. On the other hand, asymptomatic distal deep venous thrombi might be considered inappropriate outcomes for a decision rule because their clinical importance is not clear, especially in the era of routine prophylaxis against thromboembolism.

In addition to being clinically important, the outcome measure must be defined in sufficient detail to allow readers to understand the definition and be able to replicate it in their own settings. The outcomes in the 30 studies are shown in Table 1 and were classified as definitely clinically important (28 studies), possibly clinically important (2 studies), and not clinically important (no studies). The 2 possibly clinically important outcomes were asymptomatic postoperative ST segment changes and the time to healing of leg ulcers. The outcome was well defined in 24 (83%) of 29 studies.

Blind Assessment.—The presence or absence of an outcome should ideally be determined without knowledge of the status of the predictor variables. The importance of blinding varies from extremely important when the assessment of the outcome measure is subject to...
interpretation (eg, venography) to less important when the outcome is death.

If a study did not comment on whether the outcome was categorized without knowledge of the predictor variables, it was assumed that assessment was not blinded. Twelve (41%) of the 29 studies reported that the outcome was assessed without knowledge of the predictor variables. The outcomes in the 17 studies that were not blinded were death (6 studies) and myocardial infarction, colon cancer, delirium, cerebral hemorrhage on computed tomographic scan, sinusitis, seizure, rate of healing of leg ulcer, high-grade carotid stenosis, traffic accidents, suicidal ideation, and death or a cardiac event (1 study each).

**Predictive Variables.**—**Identification and Definition.**—When developing a prediction rule, investigators identify a number of variables that they believe may predict the occurrence of the outcome. Many of these variables do not become part of the rule. This can occur because the variable has no predictive value, because it has predictive value on its own but does not add predictive value to the rule that is not provided by other variables, or because assessment of the variable is too unreliable to justify inclusion in the rule.

Readers of a report should look for a clear, clinically sensible, and reproducible definition of the variables in a prediction rule. Without this, they will not be able to confidently apply the rule in their own practice. Readers may also wish to know which variables were not included in the rule, to be assured that the investigators evaluated all predictive variables that were potentially important. Addition of an omitted variable might have made the rule even more predictive. However, the fact that a rule does not include a variable does not alter the value of the rule as developed.

The predictive variables should be collected prospectively using a data sheet specifically developed for the study. The assessors should be trained to assess patients and collect data in a standardized format. Retrospective collection of data from a chart or other document is not as complete or accurate as prospective data collection. Clinical prediction rules are sometimes developed using data from a study whose primary goal is not the development of a prediction rule. For example, the Stroke Prevention in Atrial Fibrillation Investigators performed a randomized trial evaluating antithrombotic regimens for the prevention of stroke in patients with atrial fibrillation. They then used these data to derive 2 prediction rules to stratify patients’ risk of stroke.15,16 Because the data were collected for a study, they were more complete and accurate than if the data had been derived from other records. However, since the primary purpose of the study was not the development of a prediction rule, it is unlikely that the predictive variables were collected with the same degree of standardization as the outcome variables. Thus, data collected as part of another study are usually superior to data from a retrospective review of records but usually not as good as data derived from a study whose primary purpose is the development of a prediction rule.

Seventeen (59%) of the 29 studies clearly identified and defined the variables in the prediction rule. In 10 of the 25 studies that derived a rule, it was not possible to determine all of the potentially predictive variables that were evaluated for possible inclusion in the rule.

**Blind Assessment of Predictor Variables.**—The rationale for assessing the predictor variables without knowledge of the patients’ outcome is similar to the rationale for blinding the assessment of the outcome measure. If a study was prospective and the predictive variables were clearly collected prior to the outcome event, then assessment was considered to have been blind. If the study was retrospective and the authors did not mention blinding, then assessment was considered not to have been blind. Twenty-three (79%) of the 29 studies met this standard.

**Patient Age and Sex Stated.**—When deciding whether a prediction rule is likely to be useful in your own patient population, it is important to compare your patients with those in the study. Wasson et al suggested that a description of the age and sex of the study patients was a minimum standard. We would expand this to include any patient characteristics that might make the predictive value of the rule different in the 2 patient populations. This does not mean that the characteristics of your patients must be identical to the characteristics of the patients for whom the rule was developed. For example, it seems unreasonable not to use a decision rule in the United States because it was developed using patients from Sweden. The patient characteristics that are important are those that are known to or are likely to affect the performance of the rule. For instance, a prediction rule for the identification of cervical spine injury in patients with neck trauma that was developed in fully alert patients is unlikely to be as reliable in patients with decreased levels of consciousness. Thus, the authors who derive such a rule should describe whether they included patients whose consciousness was impaired. Twenty-three studies (79%) adequately described the age, sex, and other important characteristics of the patients.

**Study Site Described.**—The nature of the site where the study was done should be described, since this will affect the type of patients as well as the training and experience of the clinicians deriving and validating the rule. Usually, the type of institution (primary, secondary, or tertiary), the setting (office, clinic, emergency department, or hospital ward), and whether the site was teaching or nonteaching should be described. This information should be explicitly provided because the mailing address of the principal investigator is sometimes not at the site where the study was actually done. A description of how patients were referred to the investigators (sometimes called the referral filter) should be provided. A qualitative description is sufficient. It is not necessary to provide a detailed, numerical description of the referral filter. Nineteen (66%) of the 29 studies met this standard.

**Test of Misclassification Rate.**—When discussing tests of the misclassification rate, Wasson and colleagues described methods of presenting the results of the rule (eg, sensitivity, specificity, accuracy, error rate) as well as methods of validating the rule. We have elected to divide these concepts into 2 new standards, describing the results of a clinical prediction rule and prospective validation, which will be described below.

**Effects of Clinical Use Prospectively Measured.**—Even a well-validated rule may not be used in routine clinical practice. There are a number of reasons for this. Physicians may feel that their patients and/or practice setting are different from those at the site where the rule was derived and validated and may thus feel uncomfortable using the rule. Because of the severe clinical consequences of missing some diagnoses (eg, sending a patient who had a myocardial infarction home from the emergency department), the false-negative rate associated with a rule may be too high. Physicians not involved in a rule’s development or evaluation may be concerned that they will not apply the rule correctly. If a rule is perceived as being complicated or significantly extends the time of the usual clinical encounter, it is unlikely that physicians will use it. Some decision rules may affect the way physicians and other health care workers are paid, which might make change difficult. Concerned patients may still ask for a test even though the rule may say the test is unnecessary. Finally, fear of legal action may inhibit the use of a prediction rule, although a successful suit is quite unlikely with a well-developed rule.
To assess a rule's effects in clinical use, it should be evaluated in a patient population other than the one in which it was developed and validated, and its effects on process and outcome should be documented. This requires a separate study in which the "gold standard" is no longer administered to all patients. For example, during the initial prospective validation of the Ottawa Ankle Rules, all patients had radiography to determine the accuracy of the rule. In contrast, to study the effects of the rule on clinical practice, patients for whom the rule suggested that radiography was not indicated were sent home without radiography. The latter study design allowed us to determine if physicians and patients were actually willing to use the rule and how its use affected clinical outcomes. Only 1 of the 30 studies evaluated the effects of the rule on clinical practice.

Mathematical Techniques Described.—Many mathematical techniques are available for the derivation of a clinical prediction rule. The simplest approach is a 2×2 cross-tabulation of each predictor variable with the outcome. This is called a univariate analysis, because only 1 predictive variable at a time is being compared with the outcome. It has the advantage of simplicity and transparency and allows one to determine which predictor variables are associated with the outcome. However, a univariate analysis does not allow the analyst to explore the relationship of the predictive variables with each other and with the outcome. To do this, multivariate statistical methods are needed, such as logistic regression, discriminant function analysis, and recursive partitioning.

With multivariate analyses one can determine which predictor variables provide independent information about the likelihood of an outcome. For example, in 1 study both tenderness and swelling over the medial malleolus were associated with an ankle fracture on univariate analysis. However, when multivariate analyses were used to develop the prediction rule, only tenderness was included in the rule. This is because the presence of tenderness and the presence of swelling over the medial malleolus are highly correlated, and swelling provides little information independent of that provided by tenderness.

In logistic regression, the outcome variable is a binary event (eg, fracture vs no fracture). The risk of developing the outcome is given as the logit or log-ratiohmitic odds, and this allows the calculation of the simple odds ratio. In discriminant function analysis, the outcome variable is categorical, meaning that the patient is classified as belonging to a category or group. Optimal performance of discriminant function analysis requires that many of the predictor variables be reported as continuous data. Recursive partitioning analysis successively divides the patients into subpopulations and results in 1 or more strata that include only patients with a particular outcome. This approach is particularly well suited when the objective is to completely separate one outcome from another, such as when trying to develop a rule with extremely high sensitivity. For example, with the use of recursive partitioning, we were able to determine that all patients with ankle injuries who had no tenderness over the posterior edge or tip of either malleolus and who were able to bear weight immediately after the injury and in the emergency department did not have a fracture. However, the other 2 techniques tend to derive prediction rules with the highest overall accuracy, which may mean that the sensitivity is lower than the investigators wish.

Any report of a clinical prediction rule should adequately describe and justify the mathematical technique used to derive the rule. Furthermore, the report should describe methods used to avoid several potential problems that might be encountered in the application of multivariate analyses. Did the study avoid the problem of overfitting the data with too few outcome events per predictor variable? A general rule is that there should be at least 10 outcome events per independent variable in the prediction rule. Did the authors specify how the predictor variables were selected? Was there a preliminary screen based on univariate association and reliability? There are many other factors that should be considered when performing multivariate analyses, but they are beyond the scope of this article. Neural networks have been used to develop clinical prediction rules but will not be discussed because their use is still inefrquent.

All of the studies described the mathematical technique used to develop the rule, and all used multivariate techniques. Twenty-eight studies used regression analyses, and 1 used recursive partitioning. In 8 of the studies it was not possible to determine the number of predictor variables assessed. Among the other 17 studies, the ratio of outcome events to predictor variables was less than 10:1 in 5 studies.

Additional Standards Not Suggested by Wasson et al

We have introduced 4 new standards.

Describing the results of a clinical prediction rule and prospective validation replace the test of misclassification rate standard of Wasson et al. Reproducibility and sensitivity are new standards.

Describing the Results of a Clinical Prediction Rule.—There are many methods of describing the results of a prediction rule. These include sensitivity, specificity, likelihood ratios, positive and negative predictive values, the proportion of patients with an outcome at a particular point in time, and survival curves.

To calculate sensitivity and specificity, the outcome must be dichotomous (eg, normal or abnormal, high risk or low risk). Sensitivity refers to the proportion of patients with the outcome in whom the results of the prediction rule are abnormal. Specificity refers to the proportion of patients who do not have the outcome of interest in whom the results are normal. Sensitivity and specificity are characteristics of the prediction rule itself. Clinicians are usually more interested in the likelihood that their patient has the abnormality of interest (eg, an ankle fracture in someone with an acute ankle injury). The characteristics that determine this are the positive predictive value (the probability of a fracture if the result of the rule is positive) and the negative predictive value (the probability of no fracture if the result of the rule is negative). The importance of high sensitivity or specificity depends on the clinical circumstances. In general, decision rules with high sensitivity will tend to rule out a disorder if the result is negative, while rules with high specificity will tend to rule it in if the result is positive.

In most situations, clinicians are concerned about not missing a disorder (eg, ensuring that a patient with acute chest pain does not have a myocardial infarction). In this case, the rule should have high sensitivity and thus a low false-negative rate. Unfortunately, as one increases the sensitivity of a rule, its specificity tends to decrease (and vice versa). Even a small decrease in sensitivity will mean that the frequency of false-negative results will increase. This might lead some physicians not to use the prediction rule at all. Thus, there is a conflict between achieving a maximum reduction in radiography (which would theoretically be achieved by lowering the sensitivity) and ensuring that physicians actually use the rule (most likely when the sensitivity is very high).

For example, the Ottawa Ankle Rules were designed to have a sensitivity of 100% because of concern that clinicians would not use the rules if they thought there was even a small risk of a missed fracture. This meant that the rule had a relatively low specificity of 40%, and use of the rule would have decreased...
the rate of radiography by only 36%. Another rule could have been developed with a sensitivity of 96% and a specificity of 58%, and the proportion of patients undergoing radiography could theoretically have been reduced by 50%. However, Steidl and colleagues chose to publish the rule with the sensitivity of 100% because they believed that many physicians would be reluctant to use the rule if there were any false-negative results. Additional empirical research is needed to determine if this belief is correct and to establish the proportion of false-negative results patients, clinicians, and society are willing to accept in different clinical circumstances.

The posttest probabilities associated with a positive or negative result using a prediction rule can easily be calculated from a 2x2 table. Another way of calculating predictive values is with the use of likelihood ratios. A likelihood ratio of 1.0 means that the likelihood of the outcome is the same as the pretest probability. A likelihood ratio greater than 1.0 increases the probability of disease, and a likelihood ratio less than 1.0 decreases it. Other sources should be consulted for more detail about the calculation of likelihood ratios.39 Likelihood ratios are particularly useful for tests or prediction rules with more than 2 response categories (eg, lung scans done because of suspected pulmonary embolus can be categorized as normal or low, intermediate, or high probability46), allowing the retention of the maximum amount of information from a test or decision rule.

Receiver operator characteristic curves are curves that result from plotting sensitivity vs 1-specificity, using different cutpoints in the data. Although they are useful for visually and statistically assessing the characteristics of 1 or more prediction rules,50,51 in our opinion they are of limited use to clinicians because they do not provide an estimate of the posttest probability of the disorder.

The probability of disease can be presented in a prediction rule as the percentage of patients with a disorder at a particular point in time or with survival curves. Compared with a probability at 1 time, survival curves have the advantage of providing visual information about the frequency of the disorder over time. Some disorders may have similar long-term survival rates yet very different time patterns of survival. For example, the mortality rate immediately after a ruptured abdominal aortic aneurysm is very high, yet patients who survive the initial event tend to do quite well. On the other hand, some cancers have 5-year survival rates similar to those associated with ruptured aortic aneurysms, yet their mortality rate is fairly constant during the 5 years.52

It is important that a measure of the amount of uncertainty or variability associated with any of the methods of describing the results be provided. This is most commonly expressed as the 95% confidence interval. For example, the 95% confidence interval around the sensitivity of 1.0 found in the validation study of the Ottawa Ankle Rules was 0.95 to 1.0.2 This means that one can be 95% sure that the sensitivity of the rule is between 0.95 and 1.0, although the best estimate of its sensitivity is 1.0. The wider the 95% confidence interval, the less confidence one can have in the accuracy of the results. The width of the confidence interval is negatively correlated with the number of subjects studied.

Twenty-three studies reported the probability of the outcome of interest at a particular point in time. Six of the 29 studies provided the sensitivity and specificity of the rule, and these values could be calculated from data provided in the report for another 14 studies. Five studies reported likelihood ratios, 8 provided receiver operator characteristic curves, and 5 provided survival curves. Some measure of variability was provided for 12 studies.

**Prospective Validation.—**Wasson et al53 described statistical techniques that can be used to validate a prediction rule using the patients in whom the rule was developed. However, these methods can never determine how well the rule will function in other patients. Therefore, in our opinion, it is essential to prospectively validate the rule in a group of patients different from the group in which it was derived, preferably with different clinicians. We call this standard prospective validation. It is important that individuals evaluating the rule use it correctly. It may be necessary to contact the developers of the rule, who can clarify uncertainties and perhaps provide materials (such as printed algorithms or guides) to ensure that the rule is used correctly.50,51 Prospective validation differs from the previously described standard, effects of clinical use prospectively measured, which determines how clinicians actually use the rule in practice and the effect of the rule on patient care and outcome. In Table 2, we have placed prospective validation before effects of clinical use prospectively measured because that is the order in which they occur when evaluating a prediction rule.

Among the 14 studies that validated a rule, this was done prospectively in 11 (79%), retrospectively in 2 (14%), and using statistical “validation” in the same population in which the rule was derived in 1 (7%).

**Reproducibility.—**Some signs and symptoms that are used in usual practice are not very reliable. This is true if the same physician repeats the maneuver (intraobserver reliability) or if different physicians perform the maneuver independently on the same patient (interobserver reliability). Since prediction rules will be used by a variety of physicians, we believe that interobserver variability is more important, and we suggest that predictive variables not be included in a rule if they are unreliable. We consider a k score or correlation coefficient less than 0.6 unreliable.4 For example, we evaluated the reliability of variables from the history and physical examination when developing the Ottawa Ankle Rules and found that the k score for agreement was less than 0.6 for 23 of 32 commonly assessed variables.4 Different clinicians will sometimes obtain different results when using the same rule in identical patients. Thus, it is also important to assess the interobserver reliability of the rule itself as well as of the individual predictor variables.5

Measuring reliability adds slightly to the complexity of the development process, since 2 examiners must be available at the same time. This is often difficult in a busy clinical setting (eg, the emergency department). However, it is not necessary to evaluate reliability in all patients. A representative subset that is large enough to give reasonably narrow confidence limits around the estimate of reliability will suffice. We assessed reliability in 100 of the 750 patients used to develop the Ottawa Ankle Rules.4 Only 1 (3%) of the 29 studies described the reproducibility of the predictor variables and none described the reproducibility of the rules themselves.

**Sensitivity.—**The likelihood that a prediction rule will be used is increased if it makes clinical sense, if it is easy to use, and if it suggests a course of action. The evaluation of sensitivity relies on judgment rather than statistical methods.5

The rule should demonstrate content validity.54 In other words, most clinicians should think that the items in the rule seem clinically sensible, that no obvious items are missing, that the method of aggregating component variables is reasonable, and that the items seem appropriate for the purpose of the rule. Ease of use includes factors such as time needed to apply the rule and simplicity of use. Rules that require extensive calculations or use of a calculator are less likely to be used than rules with simpler scoring schemes.
It is our impression, although we have no firm evidence for this, that rules are more likely to be used if they suggest a course of action rather than provide a probability of disease. This is likely to be particularly true in situations where a decision must be made quickly. For example, the Ottawa Ankle Rules suggest a course of action, namely, radiography or no radiography. We believe that if clinicians are provided with a probability of fracture instead, they might be uncertain about whether to order radiography. For example, if the rule indicates that the likelihood of a fracture is 1%, clinicians might have difficulty deciding whether this is high enough to warrant radiography and might abandon the rule. Further research is needed to determine whether rules that suggest a course of action are more likely to be used than rules that provide a probability of disease.

Twenty-eight rules were felt to be clinically sensible, while in 1 study we felt the rule was not clinically sensible. Rules were considered easy to use in 12 studies (41%), moderate in 7 studies (24%), and complicated in 10 studies (34%). All of the rules provided a probability of disease.

COMMENT

We defined a clinical prediction rule as a decision-making tool for clinicians that contained 3 or more variables obtained from the history, physical examination, and simple diagnostic tests. Clinical prediction rules are derived from data collected directly from patients. Although the results of formal decision analyses can be used at the bedside, the data are more likely to be derived from the literature, and their main purpose is usually to help formulate health policy. Therefore, we have not considered decision analyses in this study.

The methodological standards (they are really guides, but we have chosen to retain the term used by the authors) described by Wasson and colleagues in 1985 have provided excellent guidance for both the users and developers of clinical prediction rules. However, our experience with developing, validating, and implementing the Ottawa Ankle Rules and a decision rule for acute knee injuries has led us to suggest modifications to some standards of Wasson et al and the addition of other standards. The main changes have been increased emphasis on prospective validation, on the reliability of predictive variables and of the rules themselves, on the sensibility of the rules, and on the possible importance of deriving rules that suggest a course of action rather than provide a probability.

It is not possible to compare the methodological quality of the prediction rules evaluated by Wasson et al and the ones we evaluated, because they were not the same rules. However, the methodological quality of the 2 groups of studies seemed similar (Table 2). Most studies described the patients and study sites well and clearly defined most of the variables in the prediction rules. Blinding was not often used in the assessment of the outcomes. Eleven (79%) of the 14 studies that validated a rule did so prospectively in a different population from that in which the rule was developed, which is encouraging. However, the effect of the use of the rule on clinical outcomes was prospectively assessed in only 1 study, which is discouraging. Although the variables that were included in all the rules except 1 were felt to be clinically sensible, it was our impression that a number of the rules would not affect clinical practice, largely because of issues mentioned in the section on sensibility. Since affecting clinical practice is the raison d'être of clinical prediction rules, it is important that more studies evaluate this criterion.

Some groups have suggested standards for the reporting of clinical trials. Their hope is that this will encourage investigators both to accurately report what they have done and to comply more fully with methodological standards. A similar approach might be useful for clinical prediction rules, using the standards presented in Table 2. We also feel that more liberal description of some of the details of the derivation of a rule would be helpful. We frequently found it impossible to tell which potentially predictive variables were initially studied for possible inclusion in the rule. Knowing this is important for 2 reasons: (1) Clinicians who are surprised that a variable is not in a prediction rule can check whether it was even considered for inclusion. (2) Readers can determine whether the number of patients in the derivation study was appropriate for the number of predictive variables studied. The description need not be long—all that is required is a list of the variables considered.

In summary, properly developed and validated prediction rules can influence clinical practice. We reviewed 30 recently published prediction rules and found that their methodological quality was similar to that of rules evaluated 10 years previously. We suggested some modifications to the initial standards of Wasson et al for the development and evaluation of clinical prediction rules; we hope these modifications will be useful to those developing and using clinical prediction rules. We also hope a similar review in another 10 years will reveal an improvement in the methodological quality of prediction rules, especially more assessments of their impact on real practice.

Drs Laupacis and Stiell are scientists of the Medical Research Council of Canada. This work was funded in part by Medical Research Council group grant GR-15804 (University of Ottawa Group Grant in Decision Support Technology for Clinicians and Patients). We thank Elizabeth Drake, Ian Graham, PhD, Graham Nichol, MD, Annette O'Connor, PhD, and George Wells, PhD, for helpful comments on the manuscript and Karen Weeks for expert secretarial assistance.

References


25. Celani RW, Jamrozik D, Thomp-