

SPECIAL ARTICLE

CLINICAL PREDICTION RULES

Applications and Methodological Standards

JOHN H. WASSON, M.D., HAROLD C. SOX, M.D., RAYMOND K. NEFF, Sc.D., AND LEE GOLDMAN, M.D.

Abstract The objective of clinical prediction rules is to reduce the uncertainty inherent in medical practice by defining how to use clinical findings to make predictions. Clinical prediction rules are derived from systematic clinical observations. They can help physicians identify patients who require diagnostic tests, treatment, or hospitalization.

Before adopting a prediction rule, clinicians must evaluate its applicability to their patients. We describe methodological standards that can be used to decide whether a prediction rule is suitable for adoption in a clinician's prac-

tice. We applied these standards to 33 reports of prediction rules; 42 per cent of the reports contained an adequate description of the prediction rules, the patients, and the clinical setting. The misclassification rate of the rule was measured in only 34 per cent of reports, and the effects of the rule on patient care were described in only 6 per cent of reports.

If the objectives of clinical prediction rules are to be fully achieved, authors and readers need to pay close attention to basic principles of study design. (N Engl J Med 1985; 313:793-9.)

CLINICAL prediction rules are intended to help physicians interpret clinical information. In earlier eras, prediction rules were based on the experience of respected senior clinicians and took the form of clinical aphorisms. These distillations of experience are memorable but may oversimplify complex issues and are subject to many forms of bias. More recently, clinical prediction rules have been derived from studies involving many hundreds of patients and sophisticated mathematical analysis. Clinical prediction rules estimate the probability of a diagnostic outcome, such as the probability that a patient with chest pain is having an acute myocardial infarction, or link clinical characteristics to the choice of therapy, such as deciding how to treat a patient with pharyngitis. In either case, a physician may use a clinical prediction rule to classify patients according to the risk of disease or the potential benefit from therapy. In addition, since prediction rules are typically based on extensive prior experience, they can help a physician know what clinical data are important to obtain. By making objective the art of diagnosis, prediction rules may help clinicians cope with the inevitable uncertainties of clinical practice. An example of a clinical prediction rule is shown in Table 1.

There have been several instances in which a prediction rule did not classify patients as accurately as had been expected from the original report.¹⁻³ These reports appropriately raise concerns about the use of clinical prediction rules in patient care. Like any new technology, a prediction rule should be evaluated carefully before it is used. To help the practicing phy-

sician carry out this task, we describe methodological standards for creating and validating clinical prediction rules. We then use the standards to evaluate recently published prediction rules.

METHODS

In our review, we have included clinical prediction rules published during the period from 1981 through 1984 in the *New England Journal of Medicine*, the *Journal of the American Medical Association*, the *Annals of Internal Medicine*, and the *British Medical Journal*. We have included only published reports that used three or more clinical findings to categorize patients and that were based on clinical data collected by the authors. Thirty-three reports fulfilled these selection criteria.²⁻³⁴ We deleted three reports of our own work.³⁵⁻³⁷

We established seven methodological standards to guide our evaluation of the 33 reports. Two of us independently evaluated the adherence of each report to these seven methodological standards. Disagreement between the evaluators was infrequent and was resolved by a third review. We were concerned mainly with factors that might reduce the utility of a prediction rule if it were to be widely used in clinical practice. We focused on the definition of clinical outcomes and predictive findings. Four standards that affect the applicability of clinical prediction rules are also described. Finally, we reviewed each report for a description of the mathematical techniques used to derive clinical prediction rules. The results of this analysis are shown in the Appendix and are summarized in Table 2.

RESULTS

Bias in Clinical Prediction Rules

Three deficiencies in study design may affect internal validity: poorly defined outcome events, poorly defined predictive findings, and failure to "blind" those who define the outcome or the predictive findings.

Definition of the Outcome

A clinical prediction rule must clearly define the event to be predicted. A definition should be free of ascertainment bias.^{38,39} For example, if the end point is acute myocardial infarction, the criteria for the diagnosis must be stated clearly, and all patients should be equally available for the measurements that define the outcome. Thus, the follow-up of patients who are not admitted to the hospital must be sufficiently thorough

From the Departments of Medicine and Community and Family Medicine, Dartmouth Medical School, Hanover, N.H., and the White River Junction Veterans Administration Medical Center, White River Junction, Vt.; the Division of General Internal Medicine, Stanford University School of Medicine and the Palo Alto Veterans Administration Medical Center; the Division of Biostatistics, School of Public Health, and the Office of the Chancellor, University of California, Berkeley; and the Division of General Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston. Address reprint requests to Dr. Wasson at Dartmouth Medical School, Hanover, NH 03756.

Dr. Goldman is a Henry J. Kaiser Family Foundation Faculty Scholar in General Internal Medicine.

Table 1. Prototype of a Clinical Rule for Predicting Death.

	SCORE*
Predictive findings	
Age >75 yr	6
Severe pain	10
Emergency	5
Total points	0-21
Interpretation of the score	
High risk: >10 points (30% deaths)	
Low risk: ≤10 points (3% deaths)	

*Score shown is assigned if specified predictive finding is present; if it is absent, a score of 0 is assigned.

to identify any patients who have had a myocardial infarction.

If possible, the predicted outcome should be biologic rather than sociological or behavioral. Consider, for example, a rule that is designed to predict the decision to hospitalize a patient. If the decision is based in part on social and behavioral factors that are specific to one institution or patient population, the prediction rule may not accurately predict outcomes in another institution. By comparison, a rule that predicts an objective biologic diagnosis is more likely to be robust when applied in another institution.

As shown in Table 2, most of the reports (28 of 33) contained adequate definitions of outcome.

Definition of Predictive Findings

A report must include precise definitions of the predictive findings. However, the need for clear definitions does not preclude the use of "soft" clinical data, as long as these data can be defined precisely enough to have a similar meaning to everyone who may use them.⁴⁰

A report met the standard for an adequate definition of predictive findings if the authors described a method, usually a protocol, and a standard reporting form for recording the findings. Almost all the reports (32 of 33) met this standard.

Blinded Assessment of Outcome and Clinical Prediction

Some outcomes, such as a diagnosis of myocardial infarction, are defined in part by clinical findings that may be useful in a prediction rule. If findings are used both as predictors and as diagnostic criteria, the investigator may erroneously conclude that they are powerful predictors of a biologic event. This circular reasoning may be avoided simply by being certain that the list of clinical predictors does not include any diagnostic criteria. In addition, the investigator who assigns a patient's diagnosis should be blinded to any findings that are to be used as predictors. The analogous error in a retrospective study is awareness of the outcome when searching the patient's record for clinical findings that are to be used as predictors.

Studies met the criteria for avoiding circular reasoning if the authors stated that study diagnoses were assigned without awareness of clinical findings that were used as predictors. Of the 12 reports that used outcome definitions based in part on clinical predic-

tors, only 3 stated that the outcome assignment was performed without awareness of the clinical predictors. Three of the 11 retrospective studies met this standard.

Applicability of Clinical Prediction Rules

The predicted outcome should be pertinent to medical practice. For example, if the rule predicts complications of a disease, have the authors concentrated on the clinically important complications? The predictors should be feasible and relevant. A clinician derives little benefit from a prediction rule that requires the results of a test that is not readily available. The reader must also ask whether the investigator has reported the truly important predictive finding or has been misled by a surrogate. For example, if a report states that the uric acid level is a strong predictor of outcome in survivors of myocardial infarction, the reader should ask whether this variable was found to be important principally because the severity of congestive heart failure and the aggressive use of diuretics were not adequately analyzed.

The clinician should evaluate the investigators' patient population for features that might reduce the general applicability of the clinical prediction rule. For example, the prognosis of patients with cancer in a tertiary referral center may be quite different from that of similar patients in a community hospital. The study population should include a wide spectrum of patients, in terms of age, sex, and clinical characteristics. The type of practice should be described, as should the size of the catchment area for the study population and the proportion of self-referred patients. If study patients are not representative of the clinical practice in which the prediction rule is to be used, the clinician should be cautious about applying the rule.

We did not assess the relevance of predictive findings or outcome measures in evaluating these reports. A report met the standard for describing the characteristics of the patient population if the age and sex distributions of the patients were noted; 25 of the 33 reports included this information (Table 2). In 31 of the reports, we were able to ascertain the setting in which the study data were obtained. We did not analyze the proportion of self-referred patients, since this information was seldom stated.

Table 2. Methodological Standards and Published Clinical Prediction Rules.*

METHODOLOGICAL STANDARD	NO. OF REPORTS MEETING THE STANDARD (%)
1. Definition of outcome	28/33 (85)
Blind assessment when appropriate	3/12 (25)
2. Definition of predictive finding	32/33 (97)
Blind assessment when a retrospective study	3/11 (27)
Subtotal: reports meeting standards 1 and 2	19/33 (58)
3. Patient age and sex stated	25/33 (76)
4. Study site described	31/33 (94)
Subtotal: reports meeting standards 1-4	14/33 (42)
5. Test of misclassification rate	11/32 (34)
6. Effects of clinical use prospectively measured	2/33 (6)
7. Mathematical technique described	23/28 (82)

*Refer to the Appendix for the complete analysis.

Accuracy of the Prediction Rule

By placing patients into groups in which the likelihood of a diagnosis or outcome has been determined empirically, a clinical prediction rule can help the clinician make decisions that direct resources to patients who are at high risk. However, if the clinical prediction rule places a patient in the wrong risk group, the patient may receive unnecessary diagnostic tests or medical care or may suffer because the physician fails to intervene appropriately. It is therefore important for the clinician to know the expected error rate when using a prediction rule in clinical practice.

The error rate, which is the proportion of patients who are misclassified (Table 3), will almost always be higher when a clinical prediction rule is used prospectively in a new group of patients (the “test set”) than when it is used in the group from which it was derived (the “training set”). The discrepancy in the error rates for training and test sets will be minimized if the training set includes a large enough number of patients in the smallest outcome category. One empirical study⁴¹ and several theoretical analyses^{42,43} have shown that the smallest prognostic or diagnostic category in the training set should contain at least five patients for each predictive finding in the rule. Thus, if our prototypic rule (Table 1) used three findings to predict death, there should have been at least 15 patients in the training set who died. This rule of thumb should be used to estimate the number of patients needed for a study, but investigators should test the misclassification rate of the rule, as described below.

There are several ways to estimate what the misclassification rate would be if a clinical prediction rule were to be applied in a new clinical setting. In one class of methods, the rule is tested on patients in the study population from which the rule was derived (cross-validation methods). The other approach is to apply the prediction rule to a group of patients who were studied after the rule had been completed. These new patients can be from the clinical setting in which the rule was derived or from a different clinical environment.

Statistical cross-validation techniques, including the “jackknife” and “bootstrap” methods,⁴⁴⁻⁴⁷ can be employed when it is necessary to use the original study patients to validate a clinical prediction rule. In the jackknife method, sometimes called the “one-left-out” method, one patient is removed and the rule is re-derived and used to classify the excluded patient. The patient’s predicted state is compared with the true state. This process is repeated many times to determine the frequency with which the excluded patient is misclassified. Thus, if there are 100 patients, 1 patient is selected to be excluded. A rule is derived from the remaining 99 patients and applied to the 100th patient. This process is repeated systematically for all 100 patients. The misclassification rate is the fraction of the 100 test patients who were incorrectly classified.

The bootstrap method is similar in principle to the jackknife.^{44,47} Two error rates are measured. One is obtained by applying the original rule to a population

Table 3. Misclassification Rate of a Prediction Rule for Awakening after Cardiac Arrest.*

ACTUAL OUTCOME	PREDICTED OUTCOME		TOTAL
	AWAKENING	NOT AWAKENING	
Awakening	173	16	189
Not awakening	44	82	129
Total	217	98	315
Sensitivity of the rule	173/217 = 80%		
Specificity of the rule	82/98 = 84%		
Total error rate = $\frac{44+16}{315}$	= 19%		

*Modified from Longstreth et al.¹⁰

that was drawn randomly (with replacement) from the original population. The second error rate is obtained by using the new population to calculate a second rule, which is then applied directly to that new population. This process is repeated many times. Since each randomly drawn population is different, the rules all differ from the original rule and from each other. The difference between the two error rates in each population is averaged over all populations. The size of this average difference is an estimate of the stability of the rule as it might be applied to a new population. Both the jackknife and bootstrap techniques use the variability in the original data set to simulate the performance of the clinical prediction rule in a new population. However, the effects of biases in selecting patients or collecting data will not be eliminated, since all testing is carried out on patients from the training-set population.

The split-sample technique may also be used to test a prediction rule. The study population is divided into a training set and a test set. The training set may contain any proportion of the original sample; the test set contains the remaining patients. The rule, which was derived from the training set, is applied to each member of the test set. Like the jackknife or bootstrap technique, this cross-validation method will not eliminate the effects of biases in patient selection or data collection.

A preferable way to measure the error rate is to embark on a second independent study in the same location. Although seemingly similar to dividing the original sample for immediate cross-validation, a second study is a much more stringent test. The study design and data-gathering methods must be reproducible, and the classification accuracy of the prediction rule must remain stable over time.

The best way to test a prediction rule is to measure the misclassification rate in a new clinical setting. This form of prospective validation requires reproducible clinical methods and definitions. If the accuracy of a rule depends on unusual, practice-specific relationships between clinical predictors and outcome, the misclassification rate at other sites should differ from the rate at the training site. Prospective validation studies in a wide variety of settings are costly and time-consuming but essential.

The misclassification rate of a clinical prediction rule was tested in 11 of 32 reports. In 5 of the 11

reports, a previously published decision rule was tested.^{2,3,13,15,34} One of the six remaining studies used the split-sample method,¹⁰ and five^{17,22,23,27,28} described prospective testing in the clinical environment of the training set.

The Effect of a Clinical Prediction Rule on Patient Care

The ultimate measure of a clinical prediction rule is its effect on patient care. Even when methodological standards have been met, a prediction rule may have little clinical utility.

A clinical prediction rule may misclassify few patients, but the errors may have very serious consequences. Instead of trying to minimize the number of misclassified patients, one should try to minimize the chance of serious error in patient care. For example, it may be far worse to fail to admit a patient with a myocardial infarction than it is to admit a patient who does not have a myocardial infarction. Furthermore, even if the clinical prediction rule is accurate and seemingly appropriate for the clinical situation, sociological or behavioral factors may nullify its use. One prediction rule²⁷ is illustrative. The use of x-ray films for injured extremities was reduced by application of a prediction rule to identify patients in whom a fracture was unlikely. However, patients often demanded and received x-ray examinations that had not been indicated by the rule. Use of a clinical prediction rule may reassure the physician about the patient's status, but the patient may need additional reassurance.³⁶

The effect of a clinical prediction rule on patient care was described in two reports.^{24,27} In one of these reports, however, the misclassification rate of the prediction rule was not stated.²⁴

Mathematical Techniques for Developing Prediction Rules

The mathematical methods used to derive a prediction rule are often too complex to merit a detailed description in a clinical journal. Nonetheless, the method should be identified, and a detailed description and discussion should be cited in the reference list.

A variety of mathematical techniques have been used to derive clinical prediction rules. The simplest method is cross-tabulation of the potential predictive finding and the outcome by means of a two-by-two contingency table. In most cases, however, the number of potential predictors is so large and their relationships are so complex that multivariate statistical analyses are required.

Multivariate techniques include linear- or logistic-regression analysis and linear or quadratic discriminant analysis. Although each of these multivariate techniques is based on specific statistical assumptions that should guide its application,⁴⁸ all produce equations in which the likelihood of an outcome can be computed by summing the weights or values that the statistical technique calculates and assigns to the potential predictive factors. To use such a multivariate rule, the clinician should seek information from the patient on each potential predictive factor. The

weights corresponding to each of the factors that are present are then summed to give a score. In general, very high or very low scores are associated with either a very high or very low likelihood of the outcome that the rule is designed to predict. Intermediate values of the score are associated with an intermediate likelihood of the outcome. The investigator may suggest using a specific threshold value of the score as a guide to therapy.

Recursive partitioning analysis,⁴⁹ which builds an empirical tree diagram by repetitively splitting the patient population into smaller and smaller categories, can be considered a hybrid between cross-tabulation and discriminant analysis. With this method, the best predictor of disease is identified, and the entire population of study patients is divided into two groups: those who have the best predictor (and have a high likelihood of disease) and those who do not (and have a relatively low likelihood of disease). With the remaining predictors used to divide each group into subgroups, this process is repeated sequentially until the number of patients in each subgroup is small or until additional predictors cannot be identified. Each path along the branches of the tree corresponds to a sequence of clinical findings and defines a patient subgroup. Typically, the likelihood of the predicted outcome varies from one subgroup to another. The tree diagram may show the probability of a clinical outcome in each subgroup, and the authors may refer to the tree in suggesting how to treat certain subgroups of patients.

When the predicted outcome is the time until an event, such as death or disease recurrence, and this event is a censored observation, Kaplan-Meier survival curves,⁵⁰ Cox regression,⁵¹ and derivations of these methods¹² are usually employed to create a prediction rule.

Twenty-three (82 per cent) of the reports described the mathematical method used to create the clinical prediction rules. The methods used are listed in the Appendix.

DISCUSSION

Inadequacies are often revealed when methodological standards based on accepted statistical and epidemiologic principles are used to assess the medical literature.^{39,52-54} Few reports of clinical prediction rules met most of our methodological criteria. Adequate definition of outcome and clinical predictors, assessed in blinded fashion when appropriate, was provided in 58 per cent of the reports. Only 14 of the 33 reports adequately described the patients, the study site, the outcomes, and the predictive findings. The misclassification rate of the rule was estimated in 34 per cent of the reports, and the effects of the rule on patient care were evaluated in only two reports. Our results may not apply to reports in other medical journals or in the same journals during a different period. In some cases, there may have been excellent reasons for deviating from the principles underlying our criteria. Our purpose is to suggest standards for developing and reporting clinical prediction rules,

not to denigrate the contributions made by the cited authors.

Table 2 lists standards for developing and reporting clinical prediction rules. One cannot simply use the number of methodological criteria that were met to estimate the validity of a clinical prediction rule. In some instances, violation of one criterion may essentially invalidate a study. On the other hand, studies that meet few of the criteria may provide clinically useful, though somewhat preliminary, information. The standards are not all-inclusive, and additional criteria or modifications may be proposed as a result of future studies of the performance of clinical prediction rules. Such studies are critically important if clinical prediction rules are to be applied confidently to local practice, but certain precautions must be heeded. First, investigators should accurately follow the protocol of the original researchers. Often, the original study used rigorous diagnostic and follow-up methods. The subsequent investigator, who may wish to perform an inexpensive version of the original study, must resist the temptation to eliminate the necessary rigor. Second, the investigator may misunderstand the original report, either because of inexperience or because the published paper did not contain all the necessary definitions and details. In many areas of science, investigators review their proposed study methods with the original authors before beginning a validation experiment and consult with them after the study has been completed to see whether the original investigators can explain any apparent discrepancies. Such a policy is often prudent when investigators attempt to validate published clinical prediction rules. In many instances, cooperation between the original and subsequent investigators can lead to a multicenter study for the validation or revision of a clinical prediction rule. A coordinated, prospective evaluation is preferable to numerous, poorly standardized attempts at validation and eliminates any "publication bias" that might otherwise favor the publication of studies that are noteworthy only because they seem to invalidate the original clinical prediction rule. Prediction rules are a form of health care technology, and they should be studied with the same care as new drugs or devices.

Clinical prediction rules are explicit empirical statements that are formulated to improve the efficiency and accuracy of physicians' judgments. Our review of reports in four selected medical journals suggests that clinical prediction rules will more readily achieve this objective if basic principles of study design are adopted. Both authors and journal editors may decide that a prediction rule deserves to be disseminated before the ideal validation study — a multicenter testing of its effects on patient care — has been performed. In such cases, both the author and the editor should assure themselves that the prediction rule is sufficiently free of bias; that the patients, the clinical setting, and the mathematical techniques have been adequately described; and that the misclassification rate has been estimated by a cross-validation technique, such as the split-sample or bootstrap method.

We present standards by which investigators, editors, and readers may judge clinical prediction rules. Unless such standards are met, clinicians must be wary of basing actual practice on the recommendations of published clinical prediction rules.

REFERENCES

1. DeSmet AA, Fryback DG, Thornbury JR. A second look at the utility of radiographic skull examination for trauma. *AJR* 1979; 132:95-9.
2. 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; 310:573-7.
3. Centor RM, Yarbrough B, Wood JP. Inability to predict relapse in acute asthma. *N Engl J Med* 1984; 310:577-80.
4. Eisenberg M, Hallstrom A, Bergner L. The ACLS score: predicting survival from out-of-hospital cardiac arrest. *JAMA* 1981; 246:50-2.
5. Keene JS, Anderson CA. Hip fractures in the elderly: discharge predictions with a functional rating scale. *JAMA* 1982; 248:564-7.
6. Levy DE, Bates D, Caronna JJ, et al. Prognosis in nontraumatic coma. *Ann Intern Med* 1981; 94:293-301.
7. Thurston JH, Thurston DL, Hixon BB, Keller AJ. Prognosis in childhood epilepsy: additional follow-up of 148 children 15 to 23 years after withdrawal of anticonvulsant therapy. *N Engl J Med* 1982; 306:831-6.
8. Hutchinson TA, Thomas DC, MacGibbon B. Predicting survival in adults with end-stage renal disease: an age equivalence index. *Ann Intern Med* 1982; 96:417-23.
9. Fischl MA, Pitchenik A, Gardner LB. An index predicting relapse and need for hospitalization in patients with acute bronchial asthma. *N Engl J Med* 1981; 305:783-9.
10. Longstreth WT Jr, Diehr P, Inui TS. Prediction of awakening after out-of-hospital cardiac arrest. *N Engl J Med* 1983; 308:1378-82.
11. Stanton BA, Jenkins CD, Denlinger P, Savageau JA, Weintraub RM, Goldstein RL. Predictors of employment status after cardiac surgery. *JAMA* 1983; 249:907-11.
12. Multicenter Postinfarction Research Group. Risk stratification and survival after myocardial infarction. *N Engl J Med* 1983; 309:331-6.
13. Merrilees MA, Scott PJ, Norris RM. Prognosis after myocardial infarction: results of 15 year follow up. *Br Med J* 1984; 288:356-9.
14. Nichols RL, Smith JW, Klein DB, et al. Risk of infection after penetrating abdominal trauma. *N Engl J Med* 1984; 311:1065-70.
15. Palmeri ST, Harrison DG, Cobb FR, et al. A QRS scoring system for assessing left ventricular function after myocardial infarction. *N Engl J Med* 1982; 306:4-9.
16. McCarron MM, Schulze BW, Walberg CB, Thompson GA, Ansari A. Short-acting barbiturate overdosage: correlation of intoxication score with serum barbiturate concentration. *JAMA* 1982; 248:55-61.
17. Ryback RS, Eckardt MJ, Felsher B, Rawlings RR. Biochemical and hematologic correlates of alcoholism and liver disease. *JAMA* 1982; 248:2261-5.
18. Aronson MD, Komaroff AL, Pass TM, Ervin CT, Branch WT. Heterophil antibody in adults with sore throat: frequency and clinical presentation. *Ann Intern Med* 1982; 96:505-8.
19. Fowler AA, Hamman RF, Good JT, et al. Adult respiratory distress syndrome: risk with common predispositions. *Ann Intern Med* 1983; 98:593-7.
20. Mann J, Holdstock G, Harman M, Machin D, Loehry CA. Scoring system to improve cost effectiveness of open access endoscopy. *Br Med J* 1983; 287:937-40.
21. Gershel JC, Goldman HS, Stein REK, Shelov SP, Ziprkowski M. The usefulness of chest radiographs in first asthma attacks. *N Engl J Med* 1983; 309:336-9.
22. Moore RD, Smith CR, Lipsky JJ, Mellits ED, Lietman PS. Risk factors for nephrotoxicity in patients treated with aminoglycosides. *Ann Intern Med* 1984; 100:352-7.
23. Slap GB, Brooks JSJ, Schwartz JS. When to perform biopsies of enlarged peripheral lymph nodes in young patients. *JAMA* 1984; 252:1321-6.
24. Pozen MW, D'Agostino RB, Selker HP, Sytkowski PA, Hood WB Jr. 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; 310:1274-8.
25. Skinner HA, Holt S, Schuller R, Roy J, Israel Y. Identification of alcohol abuse using laboratory tests and a history of trauma. *Ann Intern Med* 1984; 101:847-51.
26. Hilton P, Stanton SL. Algorithmic method for assessing urinary incontinence in elderly women. *Br Med J* 1981; 282:940-2.
27. Brand DA, Frazier WH, Kohlhepp WC, et al. A protocol for selecting patients with injured extremities who need x-rays. *N Engl J Med* 1982; 306:333-9.
28. Ramsdale DR, Faragher EB, Bennett DH, Bray CL, Ward C, Beton DC. Preoperative prediction of significant coronary artery disease in patients with valvular heart disease. *Br Med J* 1982; 284:223-6.
29. Mendelow AD, Campbell DA, Jeffrey RR, et al. Admission after mild head injury: benefits and costs. *Br Med J* 1982; 285:1530-2.

30. Eisenberg RL, Heinekin P, Hedgcock MW, Federle M, Goldberg HI. Evaluation of plain abdominal radiographs in the diagnosis of abdominal pain. *Ann Intern Med* 1982; 97:257-61.
31. Wong ET, Freier EF. The differential diagnosis of hypercalcemia: an algorithm for more effective use of laboratory tests. *JAMA* 1982; 247: 75-80.
32. Degoulet P, Menard J, Vu H-A, et al. Factors predictive of attendance at clinical and blood pressure control in hypertensive patients. *Br Med J* 1983; 287:88-93.
33. Kraemer MJ, Richardson MA, Weiss NS, et al. Risk factors for persistent middle-ear effusions: otitis media, catarrh, cigarette smoke exposure, and atopy. *JAMA* 1983; 249:1022-5.
34. Berg AO, Heidrich FE, Fihn SD, et al. Establishing the cause of genitourinary symptoms in women in a family practice: comparison of clinical examination and comprehensive microbiology. *JAMA* 1984; 251:620-5.
35. Marton KI, Sox HC Jr, Krupp JR. Involuntary weight loss: a diagnostic and prognostic significance. *Ann Intern Med* 1981; 95:568-74.
36. Sox HC Jr, Margulies I, Sox CH. Psychologically mediated effects of diagnostic tests. *Ann Intern Med* 1981; 95:680-5.
37. Goldman L, Weinberg M, Weisberg M, et al. A computer-derived protocol to aid in the diagnosis of emergency room patients with acute chest pain. *N Engl J Med* 1982; 307:588-98.
38. Feinstein AR. *Clinical biostatistics*. St. Louis: CV Mosby, 1977.
39. Ransohoff DF, Feinstein AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *N Engl J Med* 1978; 299:926-30.
40. Feinstein AR. *Clinical biostatistics*. XLI. Hard science, soft data, and the challenges of choosing clinical variables in research. *Clin Pharmacol Ther* 1977; 22:485-96.
41. Rawlings RR, Rae DS, Graubard BI, Eckardt MJ, Ryback RS. A methodology for construction of a multivariate diagnostic instrument: an application to alcohol abuse screening. *Comput Biomed Res* 1982; 15:228-39.
42. Lachenbruch PA. Some misuses of discriminant analysis. *Methods Inf Med* 1977; 16:255-8.
43. Cornfield J. Statistical classification methods. In: *Computer diagnosis and diagnostic methods*. Jacques JA, ed. Springfield, Ill.: Charles C Thomas, 1972:108-30.
44. Diaconis P, Efron B. Computer-intensive methods in statistics. *Sci Am* 1983; 248(5):116-30.
45. Lachenbruch PA, Mickey RM. Estimation of error rates in discriminant analysis. *Technometrics* 1968; 10:1-11.
46. Fukonada K, Kessell DI. Estimation of classification error. *IEEE Trans Comput* 1971; 20:1521-7.
47. Efron B. Bootstrap methods: another look at the jackknife. *Ann Stat* 1979; 7:1-26.
48. Afifi AA, Azen SP. *Statistical analysis: a computer-oriented approach*. 2nd ed. New York: Academic Press, 1979.
49. Friedman JH. A recursive partitioning decision rule for nonparametric classification. *IEEE Trans Comput* 1977; 16:404-8.
50. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc* 1958; 53:457-81.
51. Cox DR. Regression models and life tables (with discussion). *J R Stat Soc (B)* 1972; 34:187-220.
52. Gifford RH, Feinstein AR. A critique of methodology in studies of anticoagulant therapy for acute myocardial infarction. *N Engl J Med* 1969; 280:351-7.
53. Sackett DL. Bias in analytic research. *J Chronic Dis* 1979; 32:51-63.
54. Sheps SB, Schechter MT. The assessment of diagnostic tests: a survey of current medical research. *JAMA* 1984; 252:2418-22.

Appendix. Analysis of Clinical Prediction Rules Published in Four Medical Journals from 1981 through 1984.*

CLINICAL PROBLEM	PREDICTION RULE COMPOSITION (REFERENCE NO.)	METHODOLOGICAL STANDARDS								EFFECTS OF CLINICAL USE	MATHEMATICAL TECHNIQUE DESCRIBED
		OUTCOME		PREDICTIVE FINDING		AGE AND SEX DEFINED	STUDY SITE DESCRIBED	TEST OF MISCLASSIFICATION RATE			
		definition	blind assessment if dependent on predictive findings	definition	blind assessment if retrospective						
Survival from out-of-hospital cardiac arrest	Four variables scored retrospectively (4)	+	NA	+	0	0	Ambulance	0	0	0	
Hospital discharge for elderly patients with hip fracture	Four variables scored retrospectively (5)	0	NA	+	0	0	Hospital	0	0	0	
Prognosis in nontraumatic coma	Algorithm: 12 variables (6)	+	0	+	NA	+	Hospital	0	0	CT	
Prognosis in childhood epilepsy	Four variables combined (7)	+	NA	+	NA	+	Clinic	0	0	R	
Survival in end-stage renal disease	Three variables scored retrospectively (8)	+	NA	+	+	+	Hospital	0	0	R	
Relapse and need for hospitalization in patients with acute asthma	Seven variables scored (9)	0	0	+	NA	+	Emergency dept.	0	0	D	
Validation of asthma rule (above)	Seven variables scored (2)	0	0	+	NA	+	Emergency dept.	PNa	0	a	
Validation of asthma rule (above)	Seven variables scored (3)	0	0	+	NA	0	Emergency dept.	PNa	0	a	
Awakening after out-of-hospital cardiac arrest	Four variables scored retrospectively (10)	+	NA	+	0	+	Hospital	SS	0	D	
Employment after cardiac surgery	Seven variables in multiple correlation (11)	+	NA	+	NA	+	Hospital	0	0	R	
Risk subgroups for death after myocardial infarction	Four variables combined (12)	+	NA	+	NA	+	Hospital	0	0	R	
Prognosis after myocardial infarction	Three variables combined (13)	+	NA	+	NA	+	Hospital	PSa	0	a	
Risk of infection after penetrating abdominal trauma	Four variables scored (14)	+	NA	+	NA	+	Hospital	0	0	R	

Appendix continues on page 799.

Appendix (cont.).

CLINICAL PROBLEM	PREDICTION RULE COMPOSITION (REFERENCE NO.)	METHODOLOGICAL STANDARDS								
		OUTCOME		PREDICTIVE FINDING		AGE AND SEX DEFINED	STUDY SITE DESCRIBED	TEST OF MISCLASSIFICATION RATE	EFFECTS OF CLINICAL USE	MATHEMATICAL TECHNIQUE DESCRIBED
		definition	blind assessment if dependent on predictive findings	definition	blind assessment if retrospective					
Left ventricular function after infarction	Ten variables scored and correlated (15)	+	NA	+	NA	0	Hospital	PNa	0	a
Serum barbiturate levels after overdose	Eight variables scored and correlated, some retrospectively (16)	+	NA	+	0	+	Hospital	0	0	0
Discriminating alcoholic from nonalcoholic liver disease	Twenty-five variables combined retrospectively (17)	0	0	+	0	+	Hospital	PS	0	D
Adults with sore throat likely to have heterophil	Four variables combined (18)	+	NA	+	NA	+	Clinic	0	0	CT
Adult respiratory distress predispositions	Eight variables combined (19)	+	NA	+	NA	+	Hospital	0	0	CT
Patients likely to have disease at endoscopy	Six variables scored (20)	+	NA	+	NA	+	Clinic	?	0	D
Children with acute asthma likely to have abnormal chest radiographs	Three variables combined (21)	+	+	+	NA	0	Emergency dept.	0	0	CT
Nephrotoxicity from aminoglycosides	Variables scored (22)	+	NA	+	NA	+	Hospital	PS	0	D
Abnormal lymph-node biopsy in young patients	Three variables scored retrospectively (23)	+	+	+	+	+	Hospital	PS	0	D
Improving coronary care unit admissions	Seven variables scored (24)	+	+	+	NA	+	Emergency dept.	0	+a	R
Identification of alcoholism	Four variables scored (25)	+	NA	+	NA	+	0	0	0	R
Assessment of urinary incontinence in women	Algorithm (26)	+	0	+	NA	+	Hospital	0	0	0
Patients with injured extremities who need radiographs	Algorithm: six variables (27)	+	NA	+	NA	0	Emergency dept.	PS	+	RP
Coronary artery disease in patients with valvular heart disease	Seven variables combined retrospectively (28)	+	NA	+	+	+	Hospital	PS	0	R
Intracranial hematoma after head injury	Four clinical variables combined retrospectively (29)	+	NA	0	0	0	Hospital	0	0	0
Abnormal abdominal radiographs in patients	Three variables combined (30)	+	NA	+	NA	0	Emergency dept.	0	0	CT
Causes of hypercalcemia	Retrospective algorithm: four variables (31)	+	0	+	0	+	0	0	0	D
Clinical dropout and blood-pressure control in hypertensive patients	Twelve variables scored (32)	+	NA	+	NA	+	Clinic	0	0	R
Risk for persistent ear effusions	Three variables combined retrospectively (33)	+	0	+	0	+	Hospital	0	0	CT
Causes of genitourinary symptoms in women	Four variables combined (34)	+	0	+	NA	+	Clinic	PNa	0	a
TOTAL POSITIVES		28/33	3/12	32/33	3/11	25/33	31/33	11/32	2/33	23/28

*A plus sign denotes methodological standard employed, 0 standard not employed, SS split-sample method, PS prospective study at the same institution, PN prospective study at a new institution, a that the study attempted to validate a previously published predictor rule, ? standard employed but not described, NA standard not applicable, CT cross-tabulation, R multivariate-regression techniques, D discriminant techniques, and RP recursive partitioning.