Annals of Internal Medicine | Academia and Clinic

Translating Clinical Research into Clinical Practice: Impact of Using Prediction Rules To Make Decisions

Brendan M. Reilly, MD, and Arthur T. Evans, MD, MPH

Clinical prediction rules, sometimes called clinical decision rules, have proliferated in recent years. However, very few have undergone formal impact analysis, the standard of evidence to assess their impact on patient care. Without impact analysis, clinicians cannot know whether using a prediction rule will be beneficial or harmful. This paper reviews standards of evidence for developing and evaluating prediction rules; important differences between prediction rules and decision rules; how to assess the potential clinical impact of a prediction rule before translating it into a decision rule; methodologic issues critical to successful impact analysis, including defining outcome measures and estimating sample size; the importance of close collaboration between clinical investigators and practicing clinicians before, during, and after impact analysis; and the need to measure both efficacy and effectiveness when analyzing a decision rule's clinical impact. These considerations should inform future development, evaluation, and use of all clinical prediction or decision rules.

Ann Intern Med. 2006;144:201-209. www.annals.org
For author affiliations, see end of text.

Clinical prediction rules use clinical findings (history, physical examination, and test results) to make a diagnosis or predict an outcome (1–5). They quantify the relative importance of particular findings when evaluating an individual patient. How frequently these and other prediction rules are used in clinical practice is unknown, but they have become popular teaching aids in many training programs. This is not surprising, given that prediction rules were originally intended "to help physicians interpret clinical information . . . [and] know what clinical data are important to obtain" (1). Recently, however, experts have equated prediction rules with decision rules, a subtle but important change. Laupacis and colleagues (2) proposed that the purpose of prediction rules is to "suggest a diagnostic or therapeutic course of action." More ambitiously, the Evidence-Based Medicine Working Group (6) posited that prediction rules can "change clinical behavior and reduce unnecessary costs while maintaining quality of care and patient satisfaction." This shift in purpose from predicting to decision making is notable for 2 reasons.

First, very few prediction rules recommend decisions. Instead, most prediction rules provide diagnostic or prognostic probabilities, typically using a score (**Table 1**) or risk-stratification algorithm (**Figure 1**). Proponents of such rules—which intend to assist clinicians without telling them what to do—assume that accurate predictions will improve clinical decisions. This assumption is questionable. For example, how the distinctions between low and moderate probabilities predicted by the rule in **Figure 1** will (or should) affect clinicians' decisions is not obvious. Indeed, several studies suggest that access to accurate clinical predictions has an unpredictable effect on clinicians' decisions (7–10). Second, unlike those depicted in **Figures 1** and **2** and **Table 1**, very few prediction rules have undergone formal impact analysis to determine whether they improve outcomes when used in clinical practice (6). In fact, few published prediction rules have described any clinical effects of their use (1, 2, 11). Thus, when using a prediction rule, clinicians usually do not know whether it will improve (or worsen) patient care.

Table 2 summarizes standards of evidence for developing and evaluating prediction rules, including the 4 levels of evidence proposed by the Evidence-Based Medicine Working Group (6, 12–47). We propose a fifth level of evidence because we believe that broad verification of a decision rule's clinical impact is no less important than that of the prediction rule on which it is based (48–51). These progressive evidentiary standards emphasize that a prediction rule rises to the level of a decision rule only if clinicians use its predictions to help make decisions for patients. This distinction is important because analyzing a decision rule's impact requires different methods and outcome measures than those needed to validate a prediction rule (52). For example, in advancing from a level 3 rule for predicting venous thrombosis (**Table 1**) to a level 5 decision rule, developers dichotomized the rule's predictions into 2 rather than 3 probability levels, further stratified these probabilities after bedside D-dimer testing, and then analyzed the impact of the rule's use on reducing unnecessary diagnostic imaging (outcome measure) (3, 48). These substantial differences between prediction rules and decision rules may partly explain why so few impact analyses have been performed (**Table 2**).

Accordingly, in our paper, we describe and illustrate the steps necessary to conduct an impact analysis of the prediction rule depicted in **Figure 1** (52). The decision it

See also:

Print
Editorial comment. 210
Related articles . . . . . 157 and 165

Web-Only
CME quiz
Conversion of figures and tables into slides

Table 1. **Clinical Prediction Rule for Deep Venous Thrombosis***

Variable	Score	Predicted Risk (95% CI), %
Active cancer	1	
Paralysis, paresis, or immobilization of the leg	1	
Bedridden > 3 d or major surgery within 4 wk	1	
Tenderness along deep venous system	1	
Entire leg swollen	1	
Pitting edema greater than that in the other leg	1	
Collateral superficial veins (nonvaricose)	1	
Alternative diagnosis at least as likely as deep venous thrombosis	–2	
Total score		
<1		3 (2–6)
1–2		17 (12–23)
>2		75 (63–84)

* Data are from reference 3.

addresses—the triage of patients in the emergency department with suspected acute cardiac ischemia—exemplifies occasions when decision rules are most likely to be useful: when "decision making is complex, when the clinical stakes are high, or when there are opportunities to achieve cost savings without compromising patient care" (6).

We should note the differences between decision rules and other types of decision aids that are not discussed in our paper. *Decision analysis* quantifies the value of specified outcomes and uses data from the published literature to formulate health care policy (although it can be applied at the bedside as well). Most *decision-support tools* are designed to prevent errors when implementing decisions that are already made, whereas decision rules are designed to help clinicians make decisions. *Practice guidelines* address several issues in caring for patients with a particular syndrome, whereas decision rules address 1 discrete decision at 1 point in the continuum of care. Furthermore, when developed properly (**Table 2**), decision rules are exclusively evidence-based, their predictions are empirically validated, and their benefits are proven in clinical trials. In contrast, most practice guidelines reflect a consensus of expert opinion that is only sometimes supported by strong scientific evidence. Decision rules should not replace these other types of decision aids; rather, they can complement and strengthen them.

Picking a Prediction Rule with Potential Impact

In designing an impact analysis, the first step is to select the clinical prediction rule to use. For some clinical problems, such as triaging patients presenting to the emergency department with chest pain, several competing clinical prediction rules must be considered. Among the 5 previously published prediction rules relevant to this problem (4, 8, 53–55), 1 rule used only electrocardiographic predictors, which we did not find clinically sensible (54). We rejected 3 other rules because they predicted disease (the *probability* of acute myocardial infarction or unstable angina) rather than a full complement of health outcomes important to patients and physicians (including cardiac arrest, respiratory failure, and death) (8, 53, 55). Only Goldman and colleagues' rule (4) to predict *complications* of acute ischemic heart disease (**Figure 1**) aligned with our objective to reduce unnecessary admissions to inpatient monitored beds without increasing complications in patients triaged to less intensive care settings. Therefore, we investigated this prediction rule's validity, sensibility, and impact potential in our setting.

Consider Validity

The prediction rule (**Figure 1**) was derived in more than 10 000 patients from 6 hospitals and was subsequently validated in nearly 5000 additional patients in 1 of those hospitals. These studies met level 2 evidentiary standards (**Table 2**) but also raised questions about the rule's predictive accuracy in the patient population served by large public hospitals. To address these questions, we used the prediction rule to risk-stratify consecutive cohorts of patients in our institution and found that both their distribution of risk strata and their risk-stratified complication rates were similar to those of Goldman and colleagues' validation cohort (4, 56, 57). These level 3 studies increased confidence in the predictive validity of the rule and also identified opportunities for improvement by showing that many patients admitted to our inpatient telemetry unit were patients at very low risk (according to the prediction rule) who had no complications (57).

Consider Sensibility

The evaluation of sensibility requires judgment, not statistical methods. Laupacis and colleagues (2) proposed that sensible prediction rules are those with logic that is clinically sensible *and* predictors that are both comprehensive (no potential predictors omitted from consideration) and appropriate for the purpose of the rule. Thus, input from clinicians who might use the prediction rule should be obtained when assessing its sensibility.

Although our physicians were impressed by the rigor of the rule's development, they insisted that certain predictors were lacking: age, risk factors for atherosclerosis, results of previous diagnostic testing (coronary angiography), previous treatment (coronary revascularization), and additional electrocardiographic criteria for myocardial ischemia or infarction (new left bundle-branch block).

Consider Impact Potential

In their validation study of the prediction rule, Goldman and colleagues (4) demonstrated that the prediction rule's specificity was far superior to that of physicians' actual (unaided) decisions (57% vs. 39%; $P < 0.001$), meaning that the prediction rule outperformed the physicians in correctly identifying patients who will not have complications. However, the rule's sensitivity was slightly inferior to that of physicians' unaided decisions (91% vs. 96%; $P = 0.04$). This suggested that the prediction rule could reduce

unnecessary use of inpatient monitored beds but with an important (and perhaps unacceptable) tradeoff: more patients having major cardiac complications in unmonitored locations.

To explore this issue further, we performed a simulated impact analysis of the prediction rule, using 20 written cases presented to 147 physicians at our institution (58). This exercise showed that the prediction rule's predictions were more sensitive and more specific than our physicians' (simulated) decisions. In addition, we documented marked variation in decision making by our physicians, none of whom equaled the prediction rule in both sensitivity and specificity. Thus, we hypothesized that our physicians, by using the prediction rule, might improve their specificity without reducing their sensitivity in correctly identifying patients who would have complications. If so, use of the prediction rule would achieve the impact that we sought.

Preparing for Impact Analysis

After identifying a valid, sensible prediction rule with potential impact, one must consider several issues before undertaking impact analysis. These issues concern not only the logic of the decision rule and the definition of its impact measures but also the importance of respecting clinicians' practical concerns and clinical judgment when designing and implementing the rule.

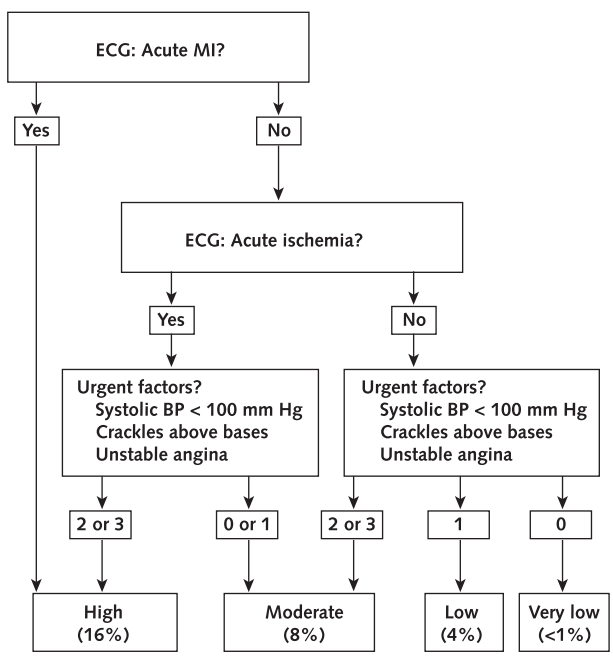
Translate Predictions into Decisions

The first priority is to determine whether the rule will be an assistive prediction rule (provide probabilities without recommending decisions, as in **Table 1** and **Figure 1**) or a directive decision rule (explicitly recommending decisions, as in **Figure 2**). This important generic question deserves further research in head-to-head comparisons: Does a directive or an assistive construction of the same prediction rule have greater clinical impact? On the basis of research in behavioral psychology and decision-support systems, we believe that directive rules will have greater impact (59–63). However, assistive rules may be less threatening to some physicians' autonomy and may seem more respectful of their clinical judgment. Thus, we strongly believe that physicians' use of a directive decision rule should be discretionary, not mandatory.

The second priority is to define the patient population to which the rule is applied. Goldman and colleagues (4) applied their rule to all patients presenting to the emergency department with chest pain (unexplained by trauma or an abnormal chest radiograph), but this does not mean that the same eligibility criteria must be used when applying the prediction rule as a decision rule. For example, why apply the rule to patients for whom clinicians can confidently make an alternative diagnosis and *exclude the possibility* of acute cardiac ischemia (64)? Predicting complications of acute cardiac ischemia in such patients does not make clinical sense, and applying the rule to this group will inflate the specificity of the predictions and lead to unnecessary care of patients who are inappropriately risk-stratified. Thus, in our impact analysis, we applied the rule only to patients in whom acute cardiac ischemia remained a diagnostic possibility after evaluation by emergency physicians (52). This had 3 important implications. First, it showed respect for physicians' judgments about when to use the rule. Second, the candidate triage decisions for recommendation by our decision rule decreased to 3 options (coronary care unit, inpatient telemetry unit, and observation unit) from the original 4 options (which also included discharge home). This revision in the rule's target population, however, required that we validate physicians' judgments when they assigned an alternative diagnosis and discharged the patient home without applying the rule.

Finally, we should note that no prediction rule can achieve consistent perfection (100% sensitivity) in ruling out its targeted clinical outcome. Developers of the Ottawa Ankle Rule (**Figure 2**) constructed it by using a cut-point with 100% sensitivity (36% specificity) in their validation studies because they surmised that clinicians would be less likely to use the rule if its negative predictive value was less than 100% (2, 5). (An alternative cut-point with 96% sensitivity and 58% specificity would have had greater impact on their primary outcome [unnecessary use of radiography].) Subsequent studies in other settings have found that the sensitivity of the rule is imperfect (96% to 99%) (65, 66). This is important because recommending deci-

Figure 1. **Goldman and colleagues' clinical prediction rule for major cardiac complications for patients with chest pain.**



BP = blood pressure; ECG = electrocardiography; MI = myocardial infarction. Adapted from reference 4.

Table 2. **Developing and Evaluating Clinical Prediction Rules**

Level of Evidence	Definitions and Standards of Evaluation	Implications for Clinicians	Systematic Reviews, *n**		
			Wasson et al., 1985 (1); 1981–1984 (*n* = 36)	Laupacis et al., 1997 (2); 1991–1994 (*n* = 32)	Present Study, 2006; 2000–2003 (*n* = 41)†
Level 1: Derivation of prediction rule	Identification of predictors using multivariate model; blinded assessment of outcomes	Needs validation and further evaluation before using clinically in actual patient care‡	20	15	10 (12–21)
Level 2: Narrow validation of prediction rule	Verification of predictors when tested prospectively in 1 setting; blinded assessment of outcomes	Needs validation in varied settings; may use predictions cautiously in patients similar to sample studied‡	10	4	10 (22–31)
Level 3: Broad validation of prediction rule	Verification of predictive model in varied settings with wide spectrum of patients and physicians	Needs impact analysis; may use predictions with confidence in their accuracy‡	4	11	16 (32–47)
Level 4: Narrow impact analysis of prediction rule used as decision rule	Prospective demonstration in 1 setting that use of prediction rule improves physicians' decisions (quality or cost-effectiveness of patient care)	May use cautiously to inform decisions in settings similar to that studied‡	2	2	1 (52)
Level 5: Broad impact analysis of prediction rule used as decision rule	Prospective demonstration in varied settings that use of prediction rule improves physicians' decisions for wide spectrum of patients	May use in varied settings with confidence that its use will benefit patient care quality or effectiveness	0	0	4 (48–51)

* The review by Wasson et al. (1) includes 33 studies from 4 general medical journals (*New England Journal of Medicine*, *Journal of the American Medical Association*, *British Medical Journal*, and *Annals of Internal Medicine*). We added 3 studies not included in their original report because those studies were performed by the authors themselves. The review by Laupaucis et al. (2) includes 30 studies from the same 4 general medical journals. We added 2 studies not included in their original report because those studies were performed by the authors themselves. Our review includes all prediction rules reported in the same 4 general medical journals from January 2000 through December 2003, including 1 study (52) written by ourselves.
† In this column, the numbers in parentheses are reference numbers.
‡ Adapted from the Evidence-Based Medicine Working Group (6).

sions based on imperfect (but highly accurate) prediction rules is not easy, partly because of medicolegal concerns about recommending and making wrong decisions (however rare) (67). Medicolegal implications of using decision rules are beyond the scope of our paper, but we believe that the imperfection of a decision rule should not prevent its use unless usual decisions (made without the rule) are demonstrably better. Impact analysis is the only way to address this question (6).

Get Clinicians' Input

In general, decision rules will improve physicians' specificity more than sensitivity. However, physicians will be more concerned about the rule's sensitivity than its specificity. Why? Because physicians appropriately ascribe greater value to true-positive decisions (correct decisions to provide care to patients who need it) than to true-negative decisions (correct decisions to withhold care from patients who do not need it). In fact, in some settings, physicians' sensitivity may be superior to that of the rule, as noted in Goldman and colleagues' validation study (4).

Our physicians sought to improve the sensitivity of the prediction rule: They insisted that a new left bundle-branch block be considered as an electrocardiographic predictor of acute ischemia. In addition, they argued that patients who are stratified as low risk by the prediction rule could inappropriately include patients presenting with acute pulmonary edema, ongoing ischemic pain despite maximal medical therapy, or unstable angina after recent coronary revascularization (52). They insisted that such emergent clinical presentations be recommended for coronary care unit admission, not telemetry unit admission.

We agreed to these changes for 2 reasons. First, Goldman and colleagues' validation study (4) did not disprove our physicians' concerns about the sensitivity of the prediction rule in these small subsets of patients. Second, our physicians agreed to provide the data needed to measure the accuracy of both the original and modified prediction rules to analyze the impact of the modifications.

Anticipate Potential Obstacles

In addition to clinicians' general disdain for "cookbook medicine" (decision aids and guidelines whose algorithmic simplicity seems to disrespect clinical complexity) (68), many more specific obstacles will impede the use of decision rules and an assessment of their impact. These will vary with the rule's content and purpose, but **Table 3** summarizes common obstacles, as well as strategies to overcome them. Most relate to physicians—after all, decision rules work by influencing physicians' attitudes and behavior (60, 63, 69–72)—but they also include barriers presented by patients (for example, comorbid conditions or preferences not addressed by the rule), the clinical setting

(for example, resource availability), and the decision rule instrument itself (for example, its ease of use).

Define Impact

Just as the impact of a decision rule differs from the accuracy of its prediction rule, the *actual* impact of a decision rule will differ from its *potential* impact. Clinicians will not always follow the rule's recommendations: They may not consult the rule at all, they may apply it inaccurately or unreliably, they may deliberately overrule its recommendations, or they may be unable to implement its recommendations for various reasons. Therefore, an impact analysis of a decision rule involves 4 assessments. First, did implementing a decision rule actually achieve its purported impact on patient care? Second, was actual impact greater or less than potential impact (measured by analyzing the rule's recommended decisions, regardless of implementation)? Third, was the accuracy of the original prediction rule preserved? Finally, did modifications to the prediction rule and its target population improve (or worsen) its accuracy?

These different objectives require different outcome measures, which previous reviews have not addressed. Sensitivity and specificity are accepted standards for measuring the accuracy of the original and modified prediction rule. We propose an analogous pair of measures—safety and efficiency—for assessing the actual and potential impact of a decision rule. Safety is defined as the proportion of all patients experiencing the predicted outcome (for example, cardiac complications) who receive the targeted diagnostic or therapeutic intervention (triage to intensive care). Efficiency is the proportion of all patients *not* experiencing the predicted outcome who do not receive the targeted intervention (that they do not need). Thus defined, safety and efficiency are not synonymous with the overall safety and efficiency of patient care, but they define the impact of a particular decision rule. Safety and efficiency are also not synonymous with sensitivity and specificity. A decision rule's safety and efficiency will only rarely replicate its prediction rule's sensitivity and specificity: when the decision rule's logic exactly follows the prediction rule's logic *and* when clinicians follow that logic exactly in all patients.

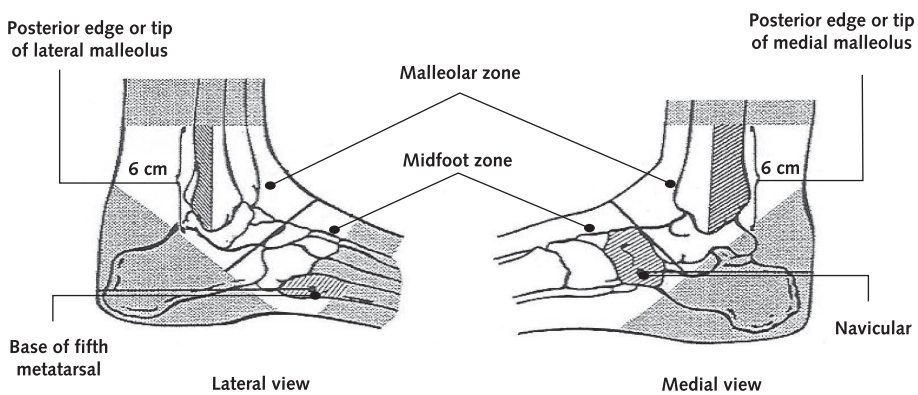
Performing Impact Analysis

We think that beneficial impact is more likely when decision aids provide specific recommendations, provide them automatically as part of clinicians' routine workflow, and use computer-based decision support (63). Although we lacked computer-based support, our impact analysis showed that the efficiency of physicians' decisions improved without compromising safety when the decision rule was the standard of care in our emergency department, compared with time periods before and after the study period (52). These results showed that use of the decision rule improved physicians' decisions (level 4 evidence in **Table 2**) and also provided further (level 3) validation of the original prediction rule's sensitivity and specificity, which were not improved by any modification we made to the prediction rule.

Use Appropriate Study Design

A randomized, controlled trial is the ideal design for impact analysis (1, 2). Alternatives to a randomized trial include a "before–after" impact analysis (measure outcomes before, during, and after using the decision rule)

Figure 2. **Ottawa Ankle Rule.**



Ankle radiography required if pain in the malleolar zone and one of the following: bone tenderness at posterior edge or tip of lateral or medial malleolus, or inability to bear weight immediately and in emergency department.

Foot radiography required if pain in the midfoot zone and one of the following: bone tenderness at base of fifth metatarsal or at navicular, or inability to bear weight immediately and in emergency department.

Adapted from reference 5.

Table 3. **Strategies To Overcome Barriers to Effective Use of Decision Rules**

Barrier	Strategic Approach
Before introducing decision rule	
Skepticism of guidelines and "cookbook medicine" or umbrage about diminished autonomy	Emphasize and enable discretionary use of decision rule
Conviction that clinical judgment is superior to the decision rule or that physicians' decision making is not root cause of problem	Perform simulated impact analysis to compare clinical judgment with decision rule and to measure the effect of physician decisions independent of all other factors
Distrust of accuracy of the rule's predictors or the "translation" of predictions into decisions	Review rule's derivation and validation and solicit input from physicians about logic of decision rule
Fear of medicolegal risk	Establish decision rule as standard of care
Disinterest in addressing system inefficiencies	Collect local data on prevalence and impact of problem and how decision rule could facilitate physicians' own tasks
During use of decision rule (impact analysis)	
Weak incentives for using rule consistently and accurately	Track physicians' (accurate) use of the rule and provide feedback about impact on patient outcomes
Conviction that overruling the decision rule is often justified	Track physicians' overruling of decision rule to assess whether clinical judgment improves on decision rule
Concern that important factors are not addressed by decision rule (e.g., patient comorbid condition or resource availability)	Review prediction rule's derivation and track whether any excluded factor affects predictions or outcomes
Concern that improving efficiency will threaten patient safety	Solicit local consensus about tradeoffs after reviewing anticipated tradeoffs based on simulated impact analysis
After impact analysis establishes benefit	
Decision rule "instrument" is not easy to use	Solicit physicians' input and redesign format
Absence of supportive infrastructure (available during impact analysis) to sustain decision rule use	Redesign procedures
Natural "regression to the mean" of previous physician behaviors	Institute continuous performance improvement measures
Fear of unintended consequences of decision rule use	Solicit concerns and measure outcomes

and an "on–off" impact analysis (measure outcomes in alternating time periods when the decision rule is or is not available). However, these designs are weaker, subject to temporal and "wash over" confounding.

Although a multi-institutional randomized study is the preferred trial design, the risk for contaminating intervention and control groups is high and the logistic and economic challenges of multicenter studies are formidable, especially without strong previous evidence of impact. Single-site (level 4) impact analysis is important because it measures the actual effects of using the rule in clinical practice, which is critical information when planning multisite (level 5) studies.

Consider Inclusion Criteria

Differences in inclusion criteria—denoting different target populations—between an impact analysis of a decision rule and the previous validation studies of its prediction rule must be stipulated. This helps to understand unintended consequences of its use and differences that may be observed in its predictive accuracy.

Consider Outcome Measures

Measuring the impact of a decision rule only by its predictive value or accuracy is tempting, but these and other measures may be misleading. Negative predictive value, for example, is a popular metric because deciding which patients will *not* need interventions (tests, treatment, or hospitalization) is usually a major outcome of interest (5, 48, 52). However, when the prevalence of targeted outcome events is very low (for example, 4% prevalence of major complications in both Goldman and colleagues' study [4] and our impact analysis [52]), a high negative predictive value reflects the low outcome prevalence as much as the predictive "value" of the rule itself.

Accuracy and other measures (sensitivity, specificity, and area under the receiver-operating characteristic curve) may be misleading because they assume equivalent societal value for true-positive and true-negative results. In addition, these measures can vary with the overall prevalence of outcomes, which is itself determined by the eligibility criteria that define the target population.

Thus, outcome measures for decision rules should include the traditional predictive values, as well as safety and efficiency, as previously defined. For physicians, negative predictive value and safety will seem most important because their primary concern is to minimize "missed" patients who have the targeted outcome. For payers, positive predictive value and efficiency will be important because their major concern is cost-effectiveness. Although differences in perspective between practicing physicians' and payers' perspectives are beyond the scope of our paper, both perspectives must recognize, and develop metrics to describe, the inevitable tradeoffs between them. A metric analogous to marginal cost-effectiveness (73, 74) can summarize these tradeoffs: the additional number of patients triaged more efficiently for every 1 extra patient triaged unsafely. This is expressed as $\Delta E(1 - P) \div \Delta S(P)$, where

ΔE is the increase in efficiency, ΔS is the decrease in safety, and P is the prevalence of targeted outcomes in the patient population. Further research about this metric and its impact on shared perspectives is needed.

Use Blinding

Assessment of outcome measures for decision rules must be blinded to patients' risk stratification and decisions recommended by the decision rule (1, 2). Ideally, this means that 1 group of clinicians uses the decision rule to make clinical decisions for patients and a different group, unaware of the rule's predictions or recommendations, assesses patients' clinical outcomes and impact measures. The potential for bias without blinding is especially great when any outcome event has a subjective component. For example, in Goldman and colleagues' prediction rule (**Figure 1**), recurrent ischemia requiring urgent coronary revascularization accounts for approximately one half of all major complications, yet assessment of "urgency" is subjective (4, 52). Thus, impact analysis of decision rules requires blinded outcomes assessment.

Estimate Sample Size

Most prediction or decision rules target clinically important outcomes that are uncommon but important. This means that an impact analysis designed to detect improvement or equivalency in safety or sensitivity (whose identical denominators are patients experiencing the targeted outcome) will often require thousands or tens of thousands of study participants. Measuring differences in efficiency and specificity (whose identical denominators are patients not experiencing the targeted outcome) typically requires many fewer participants because denominators and critical differences are larger. Before undertaking an impact analysis, therefore, one must calculate sample size for all major impact measures, especially when use of the decision rule may improve 1 impact measure at the expense of another.

Understanding Potential versus Actual Impact: Efficacy versus Effectiveness

Beneficial impact in a research study (efficacy) does not guarantee beneficial impact in "real-world" clinical practice (effectiveness). For example, after its translation into a decision rule incorporating D-dimer test results, the prediction rule in **Table 1** demonstrated a beneficial impact on unnecessary test use in a well-designed randomized, controlled trial involving several sites (48). This study, which meets level 5 evidentiary criteria in **Table 2**, analyzed the rule's impact (efficacy) by testing it in a research protocol that mandated its use and carefully monitored its correct application. Such a study has no opportunity to measure the decision rule's effectiveness when used (or not used or used incorrectly) by practicing clinicians.

In our impact analysis, actual efficiency was clearly inferior to potential efficiency (36% vs. 48%; $P < 0.001$), while actual safety was possibly superior to potential safety (94% vs. 89%; $P = 0.6$) (52). How could we understand these differences? They imply that the primary benefit of using the decision rule—improved efficiency—would increase further if physicians simply followed the rule's recommendations in every case. On the other hand, they suggest that by exercising clinical judgment and sometimes overruling the decision rule's recommendations, physicians might improve its safety.

Further analysis showed that the differences between actual and potential impact reflect complex interactions between clinicians and the decision rule. For example, our physicians did not consult the rule in 17% of study patients. In these patients, physicians' decisions were less efficient and less safe. Furthermore, when physicians consulted the rule, their actual decisions disagreed with the decision rule recommendations in 26% of patients. In these cases, physicians' decisions were less efficient than those that agreed with the decision rule. What was responsible for these disagreements? In most cases, physicians purposefully overruled the rule's recommendations, and efficiency decreased as a result. In the remainder of cases, however, physicians' preferred decisions were overruled by factors beyond their control (primarily, lack of available inpatient beds). In these cases, efficiency *increased* because most of these patients were downgraded to observation status against physicians' wishes. Remarkably, no disagreement affected safety because none of the 256 affected patients experienced a cardiac complication.

These differences between the decision rule's actual and potential impact provide little insight into how physicians' judgment might improve the rule's actual safety. However, they show that the rule's potential efficiency will be undermined when physicians overrule its recommendations or do not consult it at all.

Conclusion

Clinical prediction rules provide powerful tools to improve clinical decision making, but very few have undergone rigorous impact analysis to determine their actual effects on patient care. The potential impact of a prediction rule can be estimated by assessing its predictive validity and clinical sensibility and by measuring its potential to improve current decision making. However, the actual impact of a prediction rule also will depend on how its predictions are translated into decisions and how clinician input is effectively incorporated before, during, and after testing in actual practice. Measures of safety and efficiency are clinically relevant outcomes of its use, but the impact of clinicians' disagreements with, or misuse of, the rule when making actual decisions must be assessed. These considerations should inform the future development and evaluation of prediction rules so that more can be used confidently as decision rules to help clinicians improve the quality and cost-effectiveness of care.

From Cook County (Stroger) Hospital and Rush Medical College, Chicago, Illinois.

Potential Financial Conflicts of Interest: None disclosed.

Requests for Single Reprints: Brendan M. Reilly, MD, Cook County (Stroger) Hospital, Administration Building, Room 1528, 1900 West Polk Street, Chicago, IL 60612; e-mail, breilly@cchil.org.

Current author addresses are available at www.annals.org.

References

1. **Wasson JH, Sox HC, Neff RK, Goldman L.** Clinical prediction rules. Applications and methodological standards. N Engl J Med. 1985;313:793-9. [PMID: 3897864]

2. **Laupacis A, Sekar N, Stiell IG.** Clinical prediction rules. A review and suggested modifications of methodological standards. JAMA. 1997;277:488-94. [PMID: 9020274]

3. **Wells PS, Anderson DR, Bormanis J, Guy F, Mitchell M, Gray L, et al.** Value of assessment of pretest probability of deep venous thrombosis in clinical management. Lancet. 1997;350:1795-8. [PMID: 9428249]

4. **Goldman L, Cook EF, Johnson PA, Brand DA, Rouan GW, Lee TH.** Prediction of the need for intensive care in patients who come to the emergency departments with acute chest pain. N Engl J Med. 1996;334:1498-504. [PMID: 8618604]

5. **Stiell IG, McKnight RD, Greenberg GH, McDowell I, Nair RC, Wells GA, et al.** Implementation of the Ottawa ankle rules. JAMA. 1994;271:827-32. [PMID: 8114236]

6. **McGinn TG, Guyatt GH, Wyer PC, Naylor CD, Stiell IG, Richardson WS.** Users' guides to the medical literature: XXII: how to use articles about clinical decision rules. Evidence-Based Medicine Working Group. JAMA. 2000;284:79-84. [PMID: 10872017]

7. **Lee TH, Pearson SD, Johnson PA, Garcia TB, Weisberg MC, Guadagnoli E, et al.** Failure of information as an intervention to modify clinical management. A time-series trial in patients with acute chest pain. Ann Intern Med. 1995;122:434-7. [PMID: 7856992]

8. **Selker HP, Beshansky JR, Griffith JL, Aufderheide TP, Ballin DS, Bernard SA, et al.** Use of the acute cardiac ischemia time-insensitive predictive instrument (ACI-TIPI) to assist with triage of patients with chest pain or other symptoms suggestive of acute cardiac ischemia. A multicenter, controlled clinical trial. Ann Intern Med. 1998;129:845-55. [PMID: 9867725]

9. **Brand DA, Frazier WH, Kohlhepp WC, Shea KM, Hoefer AM, Ecker MD, et al.** A protocol for selecting patients with injured extremities who need x-rays. N Engl J Med. 1982;306:333-9. [PMID: 7054709]

10. **Cameron C, Naylor CD.** No impact from active dissemination of the Ottawa Ankle Rules: further evidence of the need for local implementation of practice guidelines. CMAJ. 1999;160:1165-8. [PMID: 10234347]

11. **Lee TH.** Evaluating decision aids: the next painful step [Editorial]. J Gen Intern Med. 1990;5:528-9. [PMID: 2266437]

12. **Wang TJ, Massaro JM, Levy D, Vasan RS, Wolf PA, D'Agostino RB, et al.** A risk score for predicting stroke or death in individuals with new-onset atrial fibrillation in the community: the Framingham Heart Study. JAMA. 2003;290:1049-56. [PMID: 12941677]

13. **Stiell IG, Wells GA, Vandemheen KL, Clement CM, Lesiuk H, De Maio VJ, et al.** The Canadian C-spine rule for radiography in alert and stable trauma patients. JAMA. 2001;286:1841-8. [PMID: 11597285]

14. **Mehr DR, Binder EF, Kruse RL, Zweig SC, Madsen R, Popejoy L, et al.** Predicting mortality in nursing home residents with lower respiratory tract infection: The Missouri LRI Study. JAMA. 2001;286:2427-36. [PMID: 11712938]

15. **Collins S, Myatt M.** Short-term prognosis in severe adult and adolescent malnutrition during famine: use of a simple prognostic model based on counting clinical signs. JAMA. 2000;284:621-6. [PMID: 10918712]

16. **Johnston SC, Gress DR, Browner WS, Sidney S.** Short-term prognosis after emergency department diagnosis of TIA. JAMA. 2000;284:2901-6. [PMID: 11147987]

17. **Boersma E, Poldermans D, Bax JJ, Steyerberg EW, Thomson IR, Banga JD, et al.** Predictors of cardiac events after major vascular surgery: role of clinical characteristics, dobutamine echocardiography, and beta-blocker therapy. JAMA. 2001;285:1865-73. [PMID: 11308400]

18. **Ferreira FL, Bota DP, Bross A, Mélot C, Vincent JL.** Serial evaluation of the SOFA score to predict outcome in critically ill patients. JAMA. 2001;286:1754-8. [PMID: 11594901]

19. **Nielsen OW, Hansen JF, Hilden J, Larsen CT, Svanegaard J.** Risk assessment of left ventricular systolic dysfunction in primary care: cross sectional study evaluating a range of diagnostic tests. BMJ. 2000;320:220-4. [PMID: 10642232]

20. **Hasbun R, Abrahams J, Jekel J, Quagliarello VJ.** Computed tomography of the head before lumbar puncture in adults with suspected meningitis. N Engl J Med. 2001;345:1727-33. [PMID: 11742046]

21. **Benavente O, Eliasziw M, Streifler JY, Fox AJ, Barnett HJ, Meldrum H, et al.** Prognosis after transient monocular blindness associated with carotid-artery stenosis. N Engl J Med. 2001;345:1084-90. [PMID: 11596587]

22. **Bailey WC, Gerald LB, Kimerling ME, Redden D, Brook N, Bruce F, et al.** Predictive model to identify positive tuberculosis skin test results during contact investigations. JAMA. 2002;287:996-1002. [PMID: 11866647]

23. **Walter LC, Brand RJ, Counsell SR, Palmer RM, Landefeld CS, Fortinsky RH, et al.** Development and validation of a prognostic index for 1-year mortality in older adults after hospitalization. JAMA. 2001;285:2987-94. [PMID: 11410097]

24. **van Walraven C, Forster AJ, Parish DC, Dane FC, Chandra KM, Durham MD, et al.** Validation of a clinical decision aid to discontinue in-hospital cardiac arrest resuscitations. JAMA. 2001;285:1602-6. [PMID: 11268268]

25. **Ozuah PO, Ozuah TP, Stein RE, Burton W, Mulvihill M.** Evaluation of a risk assessment questionnaire used to target tuberculin skin testing in children. JAMA. 2001;285:451-3. [PMID: 11242430]

26. **Berkley JA, Ross A, Mwangi I, Osier FH, Mohammed M, Shebbe M, et al.** Prognostic indicators of early and late death in children admitted to district hospital in Kenya: cohort study. BMJ. 2003;326:361. [PMID: 12586667]

27. **Imperiale TF, Wagner DR, Lin CY, Larkin GN, Rogge JD, Ransohoff DF.** Using risk for advanced proximal colonic neoplasia to tailor endoscopic screening for colorectal cancer. Ann Intern Med. 2003;139:959-65. [PMID: 14678915]

28. **Haydel MJ, Preston CA, Mills TJ, Luber S, Blaudeau E, DeBlieux PM.** Indications for computed tomography in patients with minor head injury. N Engl J Med. 2000;343:100-5. [PMID: 10891517]

29. **Casey BM, McIntire DD, Leveno KJ.** The continuing value of the Apgar score for the assessment of newborn infants. N Engl J Med. 2001;344:467-71. [PMID: 11172187]

30. **Miller ST, Sleeper LA, Pegelow CH, Enos LE, Wang WC, Weiner SJ, et al.** Prediction of adverse outcomes in children with sickle cell disease. N Engl J Med. 2000;342:83-9. [PMID: 10631276]

31. **Weber JE, Shofer FS, Larkin GL, Kalaria AS, Hollander JE.** Validation of a brief observation period for patients with cocaine-associated chest pain. N Engl J Med. 2003;348:510-7. [PMID: 12571258]

32. **Lee DS, Austin PC, Rouleau JL, Liu PP, Naimark D, Tu JV.** Predicting mortality among patients hospitalized for heart failure: derivation and validation of a clinical model. JAMA. 2003;290:2581-7. [PMID: 14625335]

33. **Morrow DA, Antman EM, Parsons L, de Lemos JA, Cannon CP, Giugliano RP, et al.** Application of the TIMI risk score for ST-elevation MI in the National Registry of Myocardial Infarction 3. JAMA. 2001;286:1356-9. [PMID: 11560541]

34. **Cadarette SM, Jaglal SB, Murray TM, McIsaac WJ, Joseph L, Brown JP, et al.** Evaluation of decision rules for referring women for bone densitometry by dual-energy x-ray absorptiometry. JAMA. 2001;286:57-63. [PMID: 11434827]

35. **Hasbun R, Vikram HR, Barakat LA, Buenconsejo J, Quagliarello VJ.** Complicated left-sided native valve endocarditis in adults: risk classification for mortality. JAMA. 2003;289:1933-40. [PMID: 12697795]

36. **D'Agostino RB Sr, Grundy S, Sullivan LM, Wilson P.** Validation of the Framingham coronary heart disease prediction scores: results of a multiple ethnic groups investigation. JAMA. 2001;286:180-7. [PMID: 11448281]

37. **Gage BF, Waterman AD, Shannon W, Boechler M, Rich MW, Radford MJ.** Validation of clinical classification schemes for predicting stroke: results from the National Registry of Atrial Fibrillation. JAMA. 2001;285:2864-70. [PMID: 11401607]

38. **Antman EM, Cohen M, Bernink PJ, McCabe CH, Horacek T, Papuchis G, et al.** The TIMI risk score for unstable angina/non-ST elevation MI: a method for prognostication and therapeutic decision making. JAMA. 2000;284:835-42. [PMID: 10938172]

39. **Brindle P, Emberson J, Lampe F, Walker M, Whincup P, Fahey T, et al.**

Predictive accuracy of the Framingham coronary risk score in British men: prospective cohort study. BMJ. 2003;327:1267. [PMID: 14644971]
40. **Daly K, Beale R, Chang RW.** Reduction in mortality after inappropriate early discharge from intensive care unit: logistic regression triage model. BMJ. 2001;322:1274-6. [PMID: 11375229]
41. **Schoonhoven L, Haalboom JR, Bousema MT, Algra A, Grobbee DE, Grypdonck MH, et al.** Prospective cohort study of routine use of risk assessment scales for prediction of pressure ulcers. BMJ. 2002;325:797. [PMID: 12376437]
42. **Ramachandran S, French JM, Vanderpump MP, Croft P, Neary RH.** Using the Framingham model to predict heart disease in the United Kingdom: retrospective study. BMJ. 2000;320:676-7. [PMID: 10710574]
43. **Arozullah AM, Khuri SF, Henderson WG, Daley J.** Development and validation of a multifactorial risk index for predicting postoperative pneumonia after major noncardiac surgery. Ann Intern Med. 2001;135:847-57. [PMID: 11712875]
44. **Gilbert K, Larocque BJ, Patrick LT.** Prospective evaluation of cardiac risk indices for patients undergoing noncardiac surgery. Ann Intern Med. 2000;133:356-9. [PMID: 10979880]
45. **Wells PS, Anderson DR, Rodger M, Stiell I, Dreyer JF, Barnes D, et al.** Excluding pulmonary embolism at the bedside without diagnostic imaging: management of patients with suspected pulmonary embolism presenting to the emergency department by using a simple clinical model and d-dimer. Ann Intern Med. 2001;135:98-107. [PMID: 11453709]
46. **Hoffman JR, Mower WR, Wolfson AB, Todd KH, Zucker MI.** Validity of a set of clinical criteria to rule out injury to the cervical spine in patients with blunt trauma. National Emergency X-Radiography Utilization Study Group. N Engl J Med. 2000;343:94-9. [PMID: 10891516]
47. **Stiell IG, Clement CM, McKnight RD, Brison R, Schull MJ, Rowe BH, et al.** The Canadian C-spine rule versus the NEXUS low-risk criteria in patients with trauma. N Engl J Med. 2003;349:2510-8. [PMID: 14695411]
48. **Wells PS, Anderson DR, Rodger M, Forgie M, Kearon C, Dreyer J, et al.** Evaluation of D-dimer in the diagnosis of suspected deep-vein thrombosis. N Engl J Med. 2003;349:1227-35. [PMID: 14507948]
49. **Marrie TJ, Lau CY, Wheeler SL, Wong CJ, Vandervoort MK, Feagan BG.** A controlled trial of a critical pathway for treatment of community-acquired pneumonia. CAPITAL Study Investigators. Community-Acquired Pneumonia Intervention Trial Assessing Levofloxacin. JAMA. 2000;283:749-55. [PMID: 10683053]
50. **English DR, Burton RC, del Mar CB, Donovan RJ, Ireland PD, Emery G.** Evaluation of aid to diagnosis of pigmented skin lesions in general practice: controlled trial randomised by practice. BMJ. 2003;327:375. [PMID: 12919990]
51. **Montgomery AA, Fahey T, Peters TJ, MacIntosh C, Sharp DJ.** Evaluation of computer based clinical decision support system and risk chart for management of hypertension in primary care: randomised controlled trial. BMJ. 2000;320:686-90. [PMID: 10710578]
52. **Reilly BM, Evans AT, Schaider JJ, Das K, Calvin JE, Moran LA, et al.** Impact of a clinical decision rule on hospital triage of patients with suspected acute cardiac ischemia in the emergency department. JAMA. 2002;288:342-50. [PMID: 12117399]
53. **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;310:1273-8. [PMID: 6371525]
54. **Brush JE Jr, Brand DA, Acampora D, Chalmer B, Wackers FJ.** Use of the initial electrocardiogram to predict in-hospital complications of acute myocardial infarction. N Engl J Med. 1985;312:1137-41. [PMID: 3920520]
55. **Goldman L, Cook EF, Brand DA, Lee TH, Rouan GW, Weisberg MC, et al.** A computer protocol to predict myocardial infarction in emergency department patients with chest pain. N Engl J Med. 1988;318:797-803. [PMID: 3280998]
56. **Reilly B, Durairaj L, Husain S, Acob C, Evans A, Hu TC, et al.** Performance and potential impact of a chest pain prediction rule in a large public hospital. Am J Med. 1999;106:285-91. [PMID: 10190376]
57. **Durairaj L, Reilly B, Das K, Smith C, Acob C, Husain S, et al.** Emergency department admissions to inpatient cardiac telemetry beds: a prospective cohort study of risk stratification and outcomes. Am J Med. 2001;110:7-11. [PMID: 11152858]
58. **Reilly BM, Evans AT, Schaider JJ, Wang Y.** Triage of patients with chest pain in the emergency department: a comparative study of physicians' decisions. Am J Med. 2002;112:95-103. [PMID: 11835946]
59. **Grol R, Dalhuijsen J, Thomas S, Veld C, Rutten G, Mokkink H.** Attributes of clinical guidelines that influence use of guidelines in general practice: observational study. BMJ. 1998;317:858-61. [PMID: 9748183]
60. **Grimshaw JM, Shirran L, Thomas R, Mowatt G, Fraser C, Bero L, et al.** Changing provider behavior: an overview of systematic reviews of interventions. Med Care. 2001;39:II2-45. [PMID: 11583120]
61. **Rutter D, Quine L.** Changing Health Behavior: Intervention and Research with Social Cognition Models. Buckingham, United Kingdom: Open University Pr; 2002.
62. **Michie S, Johnston M.** Changing clinical behaviour by making guidelines specific. BMJ. 2004;328:343-5. [PMID: 14764503]
63. **Kawamoto K, Houlihan CA, Balas EA, Lobach DF.** Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. BMJ. 2005;330:765. [PMID: 15767266]
64. **Goldman L, Kirtane AJ.** Triage of patients with acute chest pain and possible cardiac ischemia: the elusive search for diagnostic perfection. Ann Intern Med. 2003;139:987-95. [PMID: 14678918]
65. **Lucchesi GM, Jackson RE, Peacock WF, Cerasani C, Swor RA.** Sensitivity of the Ottawa rules. Ann Emerg Med. 1995;26:1-5. [PMID: 7793707]
66. **Verma S, Hamilton K, Hawkins HH, Kothari R, Singal B, Buncher R, et al.** Clinical application of the Ottawa ankle rules for the use of radiography in acute ankle injuries: an independent site assessment. AJR Am J Roentgenol. 1997;169:825-7. [PMID: 9275906]
67. **Hurwitz B.** How does evidence based guidance influence determinations of medical negligence? BMJ. 2004;329:1024-8. [PMID: 15514351]
68. **Genuis SJ, Genuis SK.** Resisting cookbook medicine. BMJ. 2004;329:179.
69. **Cabana MD, Rand CS, Powe NR, Wu AW, Wilson MH, Abboud PA, et al.** Why don't physicians follow clinical practice guidelines? A framework for improvement. JAMA. 1999;282:1458-65. [PMID: 10535437]
70. **Grol R.** Improving the quality of medical care: building bridges among professional pride, payer profit, and patient satisfaction. JAMA. 2001;286:2578-85. [PMID: 11722272]
71. **Casalino LP.** The unintended consequences of measuring quality on the quality of medical care. N Engl J Med. 1999;341:1147-50. [PMID: 10511617]
72. **James BC.** Making it easy to do it right [Editorial]. N Engl J Med. 2001;345:991-3. [PMID: 11575294]
73. **Weinstein MC, Stason WB.** Foundations of cost-effectiveness analysis for health and medical practices. N Engl J Med. 1977;296:716-21. [PMID: 402576]
74. **Eddy DM.** Clinical decision making: from theory to practice. Cost-effectiveness analysis. Is it up to the task? JAMA. 1992;267:3342-8. [PMID: 1597918]

Current Author Addresses: Dr. Reilly: Cook County (Stroger) Hospital, Administration Building, Room 1528, 1900 West Polk Street, Chicago, IL 60612.

Dr. Evans: Cook County (Stroger) Hospital, Administration Building, Room 1600, 1900 West Polk Street, Chicago, IL 60612.