# When Do Clinical Decision Rules Improve Patient Care?

Steven M. Green, MD

0196-0644/$-see front matter

http://dx.doi.org/10.1016/j.annemergmed.2013.02.006

A **podcast** for this article is available at www.annemergmed.com.

**SEE RELATED ARTICLES, P. 107, 117, AND 126.**

[Ann Emerg Med. 2013;62:132-135.]

## INTRODUCTION

Clinical decision rules are everywhere in medicine today. These impressive-looking decision algorithms and scoring systems are widely promoted as tools to either improve diagnosis or decrease expensive testing.

As clinicians, we find decision rules psychologically appealing because they would appear to create order out of disorder. Medicine is inherently subjective; however, decision rules strive to transform such murky complexity into structured and tangible tools. Certainly a scientific decision instrument statistically derived from 10,000 or more patients must be superior to the frailty and variability of clinician judgment, mustn't it? Many would assume so.

Despite their many strengths, decision rules can also have important limitations. Before adoption, I suggest that readers scrutinize each such rule on the following questions. Three high-quality articles in this issue of *Annals* will help illustrate the principles discussed—the first a derivation of a rule to decrease abdominal computed tomography (CT) in children with blunt torso trauma,[1] the second a contrast of the relative accuracy of decision rules for pulmonary embolism risk stratification when compared with unstructured clinical judgment,[2] and the third an attempted validation of the modified Alvarado score—a previously described decision rule for appendicitis.[3]

## DOES THE RULE ADDRESS A CLINICALLY RELEVANT QUESTION?

Decision rules that address trivial or uncommon questions are likely to be more trouble than they are worth. The 3 studies highlighted in this issue, however, address vital clinical concerns using optimal patient-oriented rather than surrogate outcomes.

## HAS THE RULE BEEN RIGOROUSLY DERIVED?

The best decision rules demonstrate rigorous standards for derivation, which have been detailed elsewhere.[4] *Annals* will only publish newly derived decision rules that, like that of Holmes et al,[1] are of substantial methodological rigor.

## HAS THE RULE BEEN EXTERNALLY VALIDATED?

Decision rules naturally perform their very best in the derivation sample because they are statistically modeled to depict that specific data set. When retested in a new patient sample, decision rules typically perform less well or at times fail altogether. In this issue, for example, Meltzer et al[3] were unable to validate the modified Alvarado score for appendicitis in adults—mirroring the known deficiencies of this rule in children.[5,6] Despite the best intentions of its creators, the modified Alvarado score should therefore be abandoned.

Until successfully validated, decision rules should generally not be applied clinically. The newly derived rule by Holmes et al,[1] for example, should not be used for patient care unless and until its accuracy has been replicated in another large patient sample. The Wells criteria and modified Geneva criteria discussed by Penaloza et al,[2] on the other hand, are examples of decision rules that have been validated in multiple diverse settings, thus supporting sufficient accuracy for clinical use.

## DOES THE RULE REQUIRE 1-WAY RATHER THAN 2-WAY APPLICATION?

Classic examples of effective 2-way decision rules are the Ottawa Ankle Rules and the cervical spine National Emergency X-ray Utilization Study (NEXUS). A patient should generally receive imaging if the Ottawa or NEXUS criteria are met and generally should not receive it when they are absent. A quality 2-way decision rule is maximally useful in that it tells you both when and when not to act.

A 1-way decision rule can be useful but has a dark side. The pulmonary embolism rule-out criteria (PERC), for example, are intended to be used in 1-way fashion. Low-risk patients lacking any of the 8 criteria need not be worked up for pulmonary embolism; however, the converse does not necessarily apply. Patients with positive PERC criteria should not automatically be worked up; indeed, most should not.

The pitfall of 1-way rules is the natural human propensity to apply them in a 2-way fashion, even when we know that we shouldn't. If we start working up everyone with any PERC criteria, then the net overall effect of this decision rule will be to increase rather than decrease testing.[7]

**Table.** Use of abdominal CT to identify intra-abdominal injury with intervention in Holmes et al.[1]

*A*, CT ordering by gestalt.

| | Intra-abdominal Injury With Intervention | | |
|---|---|---|---|
| | **Yes** | **No** | |
| CT ordered in ED | 191 | 5,189 | 5,380 |
| No CT ordered in ED | 12 | 6,652 | 6,664 |
| | 203 | 11,841 | 12,044 |

*ED*, Emergency department.
Raw sensitivity=191/203=94% (95% confidence interval [CI] 90% to 97%).
Sensitivity excluding 11 too unstable for CT and taken directly to laparotomy=191/192=99% (95% CI 97% to 100%).
Specificity=6,652/11,841=56% (95% CI 55% to 57%).
Likelihood ratio of positive gestalt=2.3 (95% CI 2.2 to 2.3).
Likelihood ratio of negative gestalt=0.01 (95% CI 0.001 to 0.07).
Percentage of all children receiving CT=45%.

*B*, CT ordering by 1-way application of the clinical decision rule.

| | Intra-abdominal Injury With Intervention | | |
|---|---|---|---|
| | **Yes** | **No** | |
| Both positive gestalt and positive rule | 196 | 3,930 | 4,126 |
| Negative gestalt or negative rule | 7 | 7,911 | 7,918 |
| | 203 | 11,841 | 12,044 |

Sensitivity=196/203=97% (95% confidence interval [CI] 93% to 99%).
Specificity=7,911/11,841=67% (95% CI 66% to 68%).
Likelihood ratio positive=2.9 (95% CI 2.8 to 3.0).
Likelihood ratio negative=0.05 (95% CI 0.02 to 0.11).
Percentage of all children receiving CT=34%.

*C*, CT ordering by 2-way application of the clinical decision rule.

| | Intra-abdominal Injury with Intervention | | |
|---|---|---|---|
| | **Yes** | **No** | |
| Rule positive | 197 | 6,813 | 7,010 |
| Rule negative | 6 | 5,028 | 5,034 |
| | 203 | 11,841 | 12,044 |

Sensitivity=197/203=97% (95% confidence interval [CI] 94% to 99%).
Specificity=5,028/11,841=42% (95% CI 42% to 43%).
Likelihood ratio of positive rule=1.7 (95% CI 1.6 to 1.7).
Likelihood ratio of negative rule=0.07 (95% CI 0.03 to 0.15).
Percentage of all children receiving CT=58%.

The decision rule newly presented by Holmes et al[1] was found to be sensitive (97%) but not specific (42%), and thus the authors propose that it be applied, like PERC, only in 1-way fashion. Children lacking any of their 7 derived criteria should generally not receive an abdominal CT scan; however, the converse does not apply. Instead, children with positive criteria should undergo imaging as guided by clinician judgment. If physicians correctly apply this rule in strictly 1-way fashion, the net effect in this "best-case scenario" derivation sample would be to decrease overall radiography by 11% (from 45% baseline to 34%), as shown in the Table. To illustrate the dark side, however, if clinicians get lazy with this rule and apply it strictly both ways, then abdominal CT scanning would increase by 13% (from 45% baseline to 58%). If they are lazy just half of the time, then the net rate of CT scanning would approximate baseline, ie, there would be no benefit from the rule. Thus, we must be very careful with 1-way decision rules. If we lack the fortitude to strictly apply them unilaterally, then they may adversely affect rather than improve our clinical behavior.

## DOES THE RULE IMPROVE ON PREEXISTING CLINICAL PRACTICE?

The fundamental purpose of a decision rule is to improve clinical care, not just to predict what we are already doing. Are the outcomes of importance superior when applying the rule compared with the outcomes that resulted from unstructured clinical judgment (ie, gestalt)? The Ottawa Ankle Rules, for example, improve on clinical judgment because they are just as sensitive but more specific. A rule that replicates but does not improve on gestalt, however, would not appear to have contributed anything.[7]

The decision rule by Holmes et al[1] provides the opportunity to discuss this principle because it can be contrasted with their clinicians' baseline judgment used to order CTs during the study. Study physicians didn't use the rule—which of course had not yet been derived—but instead used their standard discretion to obtain imaging. Their baseline gestalt performance was actually quite good. After excluding the 11 children who were too unstable for CT scanning and went directly to the operating room for laparotomy, CTs were ordered for all but 1 of children who had an intra-abdominal injury associated with intervention (99% sensitivity). Physicians of course ordered many CTs in children who didn't have this outcome in question, with a specificity of 56% (Table).

How does the new rule compare with baseline gestalt? When applied to its derivation data set, the sensitivity and specificity of the decision rule are both lower—97% and 42%, respectively (Table). It can be argued that the substantial specificity gap is irrelevant because the authors recommend only 1-way application of the rule. Even if we overlook the caveat about unilateral rules discussed in the last section, it still remains that the new rule is of roughly similar and perhaps lesser sensitivity. The decision rule missed 6 children with intra-abdominal injury, whereas the physicians themselves only missed 1. It would appear that this decision rule does not improve on gestalt, nor perhaps even simply replicate it.

The other 2 studies in this issue also highlight this same surprising but perhaps not uncommon outcome. Penaloza et al[2] contrasted the Wells criteria and the revised Geneva score with physicians' gestalt assessment of pulmonary embolism risk, and in both cases clinician judgment outperformed the rules. Meltzer et al[3] similarly found gestalt superior to the modified Alvarado score for detecting appendicitis.

What do we do with a decision rule that does not improve on baseline clinician gestalt? Critics would argue that such a rule

has failed and should not be adopted.[7] Such experiments are still worthwhile additions to the medical literature because documentation of such failure will guide future research. Not every clinical condition can be predicted by mathematical modeling, even when attempted with enormous, multicenter studies.

Holmes et al[1] maintain that their decision rule may be useful not because it would improve detection of intra-abdominal injury but rather because it might reduce overall CT use by 11% if physicians apply the rule only to patients for whom they would have ordered imaging (Table).[1] This positive effect of course assumes that physicians apply the rule only when their gestalt suggests imaging (1-way) and that they never apply it otherwise (2-way). We physicians are only human. Are we collectively capable of such a trick?

## DOES THE RULE APPLY TO YOUR PRACTICE SETTING AND YOUR TARGET POPULATION?

A decision rule may not apply to your practice if it was derived in a different setting, a different practice style, or in different patient types. The study by Holmes et al[1] provides an example of how decision rules don't always translate effectively from setting to setting.

The new decision rule has a surprising omission for a study of blunt abdominal trauma: it excludes focused abdominal sonography for trauma (FAST).[8-11] Most modern trauma centers assess the need for abdominal CT based upon to a primary survey and a rapid FAST scan.[10,11] FAST is no doubt more specific than any clinical variable studied by Holmes et al but was not factored into their decision rule because so few participating children's hospitals were using the technology.[1,9]

Obviously, it is imperative that pediatric trauma centers without FAST get with the program and attain the same quality care standard as general and adult trauma centers. The evidence supporting FAST is beyond reasonable dispute[8-11] and, indeed, all 6 children missed by the decision rule had hemoperitoneum that would likely have been quickly detected by FAST.[1] This decision rule newly derived by Holmes et al[1] is thus already obsolete in trauma centers using FAST. The modern clinical question instead is, when do children with blunt torso trauma and a negative FAST result need abdominal CT?

## IS THE RULE EASY TO REMEMBER AND APPLY?

Decision rules are famously ignored by clinicians, and the more complicated they are, the more likely they are to be ignored. The new rule by Holmes et al[1] has 7 points, and physicians will be challenged to reliably list them by memory. The Wells criteria, modified Geneva criteria, and modified Alvarado score are all similarly complex and likely beyond the typical recall capacity for most clinicians.

Electronic medical records and decision tools may permit automatic calculation of rule results; however, this will only be effective if clinicians remember to input all specific rule elements or somehow are forced to do so.

## SUMMARY

Holmes et al[1] remark that decision rules "aid and empower clinicians by providing evidence with regard to risk"; however, the devil is in the details of application. Their particular new decision rule, if successfully validated, may reduce abdominal CT scanning by up to 11% in settings wherein committed clinicians can selectively and consistently apply the rule in a 1-way fashion. Elsewhere, however, the rule may be perceived as too complex, insufficiently sensitive, less accurate than gestalt, and having the potential to increase rather than decrease CT usage. Penaloza et al[2] make a compelling case that skilled clinicians should likely abandon 2 common decision rules for pulmonary embolism risk stratification because they are inferior to gestalt. Meltzer et al[3] disprove a widely described scoring system for appendicitis in adults.

Decision rules are everywhere in medicine today but often add little or nothing to solid professional judgment. Before we adopt them, we should carefully scrutinize the strengths and limitations of each.

*Supervising editor:* Michael L. Callaham, MD

*Author affiliations:* From the Department of Emergency Medicine, Loma Linda University Medical Center & Children's Hospital, Loma Linda, CA.

*Funding and support:* By *Annals* policy, all authors are required to disclose any and all commercial, financial, and other relationships in any way related to the subject of this article as per ICMJE conflict of interest guidelines (see www.icmje.org). The author has stated that no such relationships exist.

Dr. Callaham was the supervising editor on this article. Dr. Green did not participate in the editorial review or decision to publish this article.

*Address for correspondence:* Steven M. Green, MD, E-mail steve@viridissimo.com.

### REFERENCES

1. Holmes JF, Lillis K, Monroe D, et al. Identifying children at very low risk of clinically important blunt abdominal injuries. *Ann Emerg Med*. 2013;62:107-116.
2. Penaloza A, Vershuren F, Meyer G, et al. Comparison of the unstructured clinician gestalt, the Wells score, and the revised Geneva score to estimate pretest probability for suspected pulmonary embolism. *Ann Emerg Med*. 2013; 62:117-124.
3. Meltzer A, Baumann BM, Chen EH, et al. Poor sensitivity of a modified Alvarado score in adults with suspected appendicitis. *Ann Emerg Med*. 2013;62:126-131.
4. Stiell IG, Wells GA. Methodologic standards for the development of clinical decision rules in emergency medicine. *Ann Emerg Med*. 1999;33:437-447.
5. Schneider C, Kharbanda A, Bachur R. Evaluating appendicitis scoring systems using a prospective pediatric cohort. *Ann Emerg Med*. 2007;49:778-784.

6. Kulik DM, Uleryk EM, Maguire JL. Does this child have appendicitis? a systematic review of clinical prediction rules for children with acute abdominal pain. *J Clin Epidemiol*. 2013;66:95-104.
7. Schriger DL, Newman DH. Medical decisionmaking: let's not forget the physician. *Ann Emerg Med*. 2012;59:219-220.
8. Noble VE, Blaivas M, Blankenship R, et al. Decision rule for imaging utilization in blunt abdominal trauma—where is ultrasound [letter]? *Ann Emerg Med*. 2010;55:487-489.
9. Fox JC, Boysen M, Gharahbaghian L, et al. Test characteristics of focused assessment of sonography for trauma for clinically significant abdominal free fluid in pediatric blunt abdominal trauma. *Acad Emerg Med*. 2011;18:477-482.
10. American College of Emergency Physicians. Policy statement: emergency ultrasound guidelines. Approved 2008. Available at: http://www.acep.org/workarea/downloadasset.aspx?id=32878. Accessed January 31, 2013.
11. Diercks DB, Mehrotra A, Nazarian DJ, et al. Clinical policy: critical issues in the evaluation of adult patients presenting to the emergency department with acute blunt abdominal trauma. *Ann Emerg Med*. 2011;57:387-404.

**The 2013 Council Resolutions, including any amendments to the ACEP Bylaws, can be found on the ACEP Web site at http://www.acep.org/council/.**