

Editorials

Application of Likelihood Ratios to Clinical Decision Rules: Defining the Limits of Clinical Expertise

See related article, p. 589.

[Gallagher EJ: Application of likelihood ratios to clinical decision rules: Defining the limits of clinical expertise. *Ann Emerg Med* November 1999;34:664-667.]

In this issue of *Annals*, Buckley et al¹ present an independent, prospective validation of their previously derived clinical decision rule targeted at predicting the presence or absence of ectopic pregnancy (EP).² These authors found that, among hemodynamically stable women presenting with abdominal pain or vaginal bleeding relatively early in pregnancy (<13 weeks), clinical features could be used to classify patients into high-, intermediate-, or low-risk strata for EP.

Buckley et al used recursive partitioning to develop their clinical decision rule.³ This statistical strategy, in contrast to the more conventional alternatives of logistic modeling⁴ or discriminant function analysis,⁵ was presumably used because recursive partitioning develops a *directional* clinical decision rule, aimed at either high sensitivity or specificity through optimal separation of dichotomous outcomes (ie, EP present versus absent).³

The analysis that follows uses likelihood ratios (LRs) to assess the performance of Buckley et al's clinical decision rule.² Although sensitivity/specificity and predictive values are also discussed, LRs were chosen in preference to these more traditional performance measures for the several reasons offered below. Unless otherwise indicated, all data are drawn from the authors' validation cohort.^{2,6}

SENSITIVITY AND SPECIFICITY

Although widely used in clinical practice, sensitivity and specificity fail to tell us what we really wish to know. In this instance, the clinician needs information on the stratum-specific probabilities of EP among patients categorized by the clinical decision rule as low, intermediate, or high risk for an EP. Instead, sensitivity

tells us how likely a patient is to be assigned to a particular risk stratum, *given that an EP is in fact present*.

Similarly, specificity tells us how likely a patient is to be assigned to a particular risk stratum, *given that an EP is in fact absent*. This represents a confusing inversion of customary clinical logic,⁷ since knowledge of whether the patient had an EP would presumably obviate the need for a clinical decision rule to explore the question in the first place.

In practice, one relies on a low rate of false-negative results (relative to true-positive results) to convert a clinical decision rule of high sensitivity into a “rule-out” maneuver.⁸ Conversely, a low rate of false-positive results (relative to true-negative results) produces a clinical decision rule of high specificity that can be used to “rule in” a target disorder.

For example, Buckley et al's low-risk criteria applied to their validation cohort reveal a sensitivity of 100% (95% confidence interval [CI] 89% to 100%) for this stratum relative to the intermediate- or high-risk strata.² Therefore, a patient with a low pretest probability of EP who was then classified by the clinical decision rule as low risk, would be likely to drop below the “testing threshold,”⁹ essentially eliminating the diagnosis of EP from further consideration.

In contrast, an examination of the high-risk stratum relative to those categorized as low risk within the same cohort² reveals a specificity of only about 83% (95% CI 77% to 90%) for high-risk patients, thus making intuitive translation of this information more difficult. Unlike the low-risk patients, whose 100% sensitivity meant no (or, considering the relative imprecision of the estimate, very few) false-negative misclassifications, high-risk patients' specificity reveals a false-positive misclassification rate of about 17%. The difficulty of translating the sensitivity and specificity of a decision rule into clinically useful terms only worsens as these performance measures drift further from perfect values of 100%. This in fact represents an intrinsic limitation of sensitivity and specificity, to which LRs offer a solution.

PREDICTIVE VALUES

Predictive values constitute the other traditional measure of a clinical decision rule's performance. In contrast to sensitivity and specificity, predictive values *do* tell us what we wish to know clinically by answering the following question: Given a patient classified as low, intermediate, or high risk for an EP, what is the stratum-specific probability that this individual *does*

(positive predictive value) or *does not* (negative predictive value) have an EP? Unfortunately, predictive values are vulnerable to variation in disease prevalence, making them too numerically unstable to transport from one patient population to another. This feature markedly limits their usefulness. As disease prevalence increases in a population, the positive predictive value of a clinical decision rule rises, and reciprocally, its negative predictive value falls. Similarly, as prevalence decreases, the reverse occurs, all without any change in the rule itself.

One means of adjusting for this prevalence-dependent variation in predictive values is to attempt to standardize them by calculating predictive increments and decrements (ie, the difference between pretest probability [prevalence] and posttest probability for each risk stratum). Using the same data from Buckley et al's validation cohort,² the overall prevalence of EP is approximately 7% (95% CI 5% to 10%), which is virtually identical to the posttest probability or prevalence of EP among women in the intermediate-risk category (also about 7%; 95% CI 5% to 11%), for a predictive increment of 0% among women assigned to the intermediate stratum.² Similar arithmetic reveals a posttest probability for EP of approximately 32% (95% CI 17% to 51%) among the high-risk group for a predictive increment of $32\% - 7\% = 25\%$ in women assigned to this stratum. Because the posttest probability of EP is about 0% (95% CI 0% to 3%) among those at low risk, the predictive decrement associated with this stratum is $0\% - 7\% = -7\%$.

Unfortunately, unless one is seeing a population in which the prevalence of EP is known to be about 7%, there is no firm assurance that predictive increments and decrements of comparable magnitude will be found among one's own patients, even when stratified according to the same risk profile. Although this rule was derived and validated on independent data sets as recommended,⁶ because the prevalence of EP in both the derivation and validation cohorts was virtually identical, methodologically appropriate independent validation will not erase the problem of application of this clinical decision rule to other populations with a higher or lower—or, as is typically the case, an unknown—prevalence of EP.

LIKELIHOOD RATIOS

LRs, in contrast, combine the prevalence-independent stability of sensitivity and specificity with the utility of

predictive values to provide a performance index for each stratum of a clinical decision rule. This is far more clinically useful and broadly applicable than either of the more traditional measures of test performance.¹⁰

LRs are defined as the likelihood that a particular test result would be found in a patient *with* the target disorder, relative to the likelihood of that same test result occurring in a patient *without* the target disorder. Application of Bayes' theorem to the LRs generated by Buckley et al's decision rule² produced the following:

$$\text{Clinically estimated pretest odds of EP} \times \\ \text{Stratum-specific LR derived from clinical} \\ \text{decision rule} = \text{Posttest odds of EP}$$

This simple equation succinctly reflects a logical convergence between the mathematical representation and conceptual strategy underlying clinical diagnosis (ie, that the principal goal of a clinical decision rule, as with any diagnostic test, is revision of disease probability).¹⁰ Because the LR is multiplied by the pretest odds to calculate posttest odds, from this it follows that LRs increase their power to alter the odds (or probability) of a target disorder in proportion to their divergence from unity (LR=1). An LR of 1 associated with any decision rule stratification is therefore not clinically helpful because disease probability is unaltered by this classification, leaving the patient's location on the diagnostic continuum unchanged.

Finally, LRs are particularly well suited to analysis of clinical decision rules containing more than 2 strata. The utility of this property has been well illustrated in a reanalysis of the Prospective Investigation Of Pulmonary Embolism Diagnosis (PIOPED) data.¹¹ This reanalysis showed convincingly that only those ventilation/perfusion (V/Q) scans at the extremes of the diagnostic spectrum provided clinically useful information.¹² Normal scans *decreased* the odds of pulmonary embolism by about 10-fold (LR=0.1); high probability scans *increased* the odds by nearly 20-fold (LR=17). However, the majority of scans were nondiagnostic (low/intermediate probability), with LRs in the neighborhood of 1.¹¹

The intermediate stratum of Buckley et al's decision rule had a similar LR=1 (95% CI 0.7 to 1.2),² indicating that, analogous to a nondiagnostic V/Q scan, further testing will be necessary to confirm or exclude the target diagnosis of EP. Women assigned to the intermediate stratum by the decision rule comprised 70% of these authors' combined derivation and validation cohorts.^{1,2}

SUMMARY FINDINGS

Clinicians will not be surprised that women presenting with pain or bleeding relatively early in pregnancy are at a roughly sixfold higher risk of harboring an EP if they have pelvic peritonitis or "definite" cervical motion tenderness, in contrast to otherwise comparable individuals with neither of these signs (LR=6; 95% CI 3 to 11) for high-risk stratum. Nor will they be surprised that women with an audible fetal heartbeat, tissue at the cervical os, or mild "menstrual-like" midline cramping without tenderness have at least a fivefold lower risk of EP, in contrast to otherwise comparable individuals in whom all 3 features are absent (LR≤0.2; 95% CI 0.0 to 0.4) for low-risk stratum.*

One of the most significant features of Buckley et al's meticulous work may lie in the demonstration of our limits of clinical expertise in the diagnosis of EP. Indeed, among the more than two thirds of patients (70%) assigned to the intermediate-risk group, the history and physical examination, as reflected by the stratum-specific LR=1, appear unable to alter the probability of EP to any greater extent than a nondiagnostic V/Q scan alters the probability of pulmonary embolus.¹² In both instances, further testing will be necessary to confirm or exclude the target diagnosis.

Although, as Buckley et al² correctly point out, a substantial proportion of patients with suspected EPs have human chorionic gonadotropin levels below the current discriminatory zone of transvaginal sonography,^{13,14} technology and examiner skills are rapidly improving. Serial determinations of human chorionic gonadotropin levels accompanied by serial transvaginal ultrasound—analogue to serial ultrasonographic examination of the lower extremities in suspected thromboembolic disease¹⁵⁻¹⁷—have been shown to provide excellent results in the diagnosis of EP, with LRs for ectopic detection in excess of 1,000 (95% CI 187 to >6000), and LRs for exclusion of EP less than 0.01 (95% CI 0.00 to 0.02).¹⁸

CLINICAL BOTTOM LINE

These data² lend strong support to the need for 24-hour, 7-day-a-week availability of timely pelvic ultrasonography, performed and interpreted by skilled operators, as an essential component of good and safe

*Estimate of LR (negative) of 0.2 derived by using midpoint of 95% CI for 100% sensitivity.

clinical care for pregnant women presenting to the emergency department with pain or bleeding.

1. Buckley RG, King KJ, Disney JD, et al: History and physical examination to estimate the risk of ectopic pregnancy: Validation of a clinical prediction model. *Ann Emerg Med* 1999;34:589-594.
 2. Buckley RG, King KJ, Disney JD, et al: Derivation of a clinical prediction model for the emergency department diagnosis of ectopic pregnancy. *Acad Emerg Med* 1998;5:951-960.
 3. Feinstein AR: *Multivariable Analysis: An Introduction*. New Haven, CT: Yale University Press, 1996:529-558.
 4. SAS Institute Inc, *SAS/STAT User's Guide*, version 6, ed 4, vol 2. Cary, NC: SAS Institute Inc, 1989:1071-1126.
 5. SAS Institute Inc, *SAS/STAT User's Guide*, version 6, ed , vol 1, Cary, NC: SAS Institute Inc, 1989:677-771.
 6. Laupacis A, Sekar N, Stiell IG: Clinical prediction rules. A review and suggested modifications of methodologic standards. *JAMA* 1997;277:488-494.
 7. Feinstein AR: *Clinical Epidemiology. The Architecture of Clinical Research*. Philadelphia: WB Saunders, 1985:419.
 8. Stiell IG, Greenberg GH, McKnight RD, et al: Decision rules for the use of radiography in acute ankle injuries: Refinement and prospective validation. *JAMA* 1993;269:1127-1132.
 9. Pauker SG, Kassirer JP: The threshold approach to clinical decision-making. *N Engl J Med* 1980;302:1109-1117.
 10. Gallagher EJ: Clinical utility of likelihood ratios. *Ann Emerg Med* 1998;31:391-397.
 11. The PIOPED Investigators: Value of the ventilation/perfusion scan in acute pulmonary embolism. Results of the prospective investigation of pulmonary embolism diagnosis (PIOPED). *JAMA* 1990;263:2753-2759.
 12. Jaeschke R, Guyatt GH, Sackett DL, et al: Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? *JAMA* 1994;271:703-707.
 13. Kaplan BC, Dart RG, Moskos M, et al: Ectopic pregnancy: Prospective study with improved diagnostic accuracy. *Ann Emerg Med* 1996;28:10-17.
 14. Lindblom B, Hahlin M, Sjoblom P: Serial human chorionic gonadotropin determinations fluoroimmunoassay for differentiation between intrauterine and ectopic gestation. *Am J Obstet Gynecol* 1989;161:397-400.
 15. Heijboer H, Buller HR, Lensing AW: A comparison of realtime compression ultrasonography with impedance plethysmography for the diagnosis of deep vein thrombosis in symptomatic outpatients. *N Engl J Med* 1993;329:1365-1369.
 16. Ginsberg JS: Management of venous thromboembolism. *N Engl J Med* 1996;335:1816-1828.
 17. Hull RD, Raskob GE, Ginsberg JS, et al: A noninvasive strategy for the treatment of patients with suspected pulmonary embolism. *Arch Intern Med* 1994;154:289-297.
 18. Barnhart K, Mennuti MT, Benjamin I, et al: Prompt diagnosis of ectopic pregnancy in an emergency department setting. *Obstet Gynecol* 1994;84:1010-1015.
-